The Theory and Practice of Induction by Alignment

CJ McCartney

March 2018

Abstract

Induction is the discovery of models given samples. This paper demonstrates formally from first principles that there exists an optimally likely model for any sample, given certain general assumptions. Also, there exists a type of encoding, parameterised by the model, that compresses the sample. Further, if the model has certain entropy properties then it is insensitive to small changes. In this case, approximations to the model remain well-fitted to the sample. That is, accurate classification and prediction is practicable for some samples. Artificial neural networks are implementations of supervised machine learning. The paper explains why the least-squares gradient-descent optimisation of a neural net can be well-fitted in some cases, even without regularisation techniques. Then the paper derives directly from theory a practicable unsupervised machine learning algorithm that optimises the likelihood of the model by maximising the alignment of the model variables. Alignment is a statistic which measures the law-likeness or the degree of dependency between variables. It is similar to mutual entropy but is a better measure for small samples. If the sample variables are not independent then the resultant models are well-fitted. Furthermore, the models are structures that can be analysed because they consist of trees of context-contingent sub-models that are built layer by layer upwards from the substrate variables. In the top layers the variables tend to be diagonalised or equational. In this way, the model variables are meaningful in the problem domain.

1 Preface

Although this document is still in the format of a paper, it has grown to be the length of a book. In order to be more accessible there is an 'Overview' section at the beginning that covers the important points of the theory and some interesting parts of the practice. The overview also has a summary of the set-theoretic notation used throughout. The complete theory and various practical implementations are the following sections. The section 'Induction' also begins with a review of relevant parts of the earlier sections. The paper finishes with some appendices on various related issues, including an appendix 'Useful functions'.

Readers interested mainly in implementation should focus on sections 'Overview', 'Substrate structures', 'Shuffled history', 'Rolls', 'Computation time and representation space', 'Rolled alignment', 'Decomposition alignment', 'Computation of alignment', 'Tractable alignment-bounding' and 'Practicable alignment-bounding'.

Terms in italics have a mathematical definition to avoid ambiguity. So '*independent*' is a well defined property, whereas 'independent' has its dictionary definition.

For further discussion see `http://greenlake.co.uk`.

2 Overview

This section provides an overview of the main points of the paper. Detailed explanations are excluded for brevity. The overview is presented as a series of assertions of fact, but only some are proven and many are conjectured, especially statements regarding correlations. In some cases, however, there are multiple strands of evidence that corroborate a conjecture. This is particularly true for the conjectures regarding the general *induction* of *models* given *samples*. Given a set of *induction* assumptions these conjectures relate (i) the maximisation of the *likelihood* of a *sample*, and also the minimisation of the *likelihood's sensitivity* to *model* and *distribution*, to (ii) properties such as *encoding space*, *entropy* and *alignment*. The different sets of *induction* assumptions can be categorised in various complementary ways: (a) *classical induction* versus *aligned induction*, (b) *law-like* conditional *draws* of *samples* from *distributions* versus the *compression* of *encodings* of *samples* by *model*, (c) simple *transform models* versus *layered, contingent models*, and (d) intractable theoretical *induction* assumptions versus tractable and practicable *induction* assumptions. The existence of working implementations of practicable *induction* such as *artificial neural networks* and *alignment inducers* provides concrete support to the theory.

2.1 Notation

The notation is briefly summarised below. The appendices contain further details.

The notation used throughout this discussion is conventional set-theoretic with some additions. Sets are often defined using set-builder notation, for example $Z = \{\mathrm{f}(x) : x \in X,\ \mathrm{p}(x)\}$ where $\mathrm{f}(x)$ is a function and $\mathrm{p}(x)$ is a predicate. Tuples can be defined similarly where the order is not important, for example, $\sum(\mathrm{f}(x) : x \in X,\ \mathrm{p}(x))$.

The powerset function is defined as $\mathrm{P}(A) := \{X : X \subseteq A\}$.

The partition function B is the set of all partitions of an argument set. A partition is a set of non-empty disjoint subsets, called components, which union to equal the argument, $\forall P \in \mathrm{B}(A)\ \forall C \in P\ (C \neq \emptyset)$, $\forall P \in \mathrm{B}(A)\ \forall C, D \in P\ (C \neq D \implies C \cap D = \emptyset)$ and $\forall P \in \mathrm{B}(A)\ (\bigcup P = A)$.

A relation $A \in \mathrm{P}(\mathcal{X} \times \mathcal{Y})$ between the set $\mathcal{X}$ and the set $\mathcal{Y}$ is a set of pairs, $\forall (x, y) \in A\ (x \in \mathcal{X}\ \wedge\ y \in \mathcal{Y})$. The domain of a relation is $\mathrm{dom}(A) := \{x : (x, y) \in A\}$ and the range is $\mathrm{ran}(A) := \{y : (x, y) \in A\}$.

Functions are special cases of relations such that each element of the domain appears exactly once. Functions can be finite or infinite. For example, $\{(1, 2), (2, 4)\} \subset \{(x, 2x) : x \in \mathbf{R}\}$. The powerset of functional relations between sets is denoted $\to$. For example, $\{(x, 2x) : x \in \mathbf{R}\} \in \mathbf{R} \to \mathbf{R}$. The application of the function $F \in \mathcal{X} \to \mathcal{Y}$ to an argument $x \in \mathcal{X}$ is denoted by $F(x) \in \mathcal{Y}$ or $F_x \in \mathcal{Y}$. Functions $F \in \mathcal{X} \to \mathcal{Y}$ and $G \in \mathcal{Y} \to \mathcal{Z}$ can be composed $G \circ F \in \mathcal{X} \to \mathcal{Z}$. The inverse of a function, $\mathrm{inverse} \in (\mathcal{X} \to \mathcal{Y}) \to (\mathcal{Y} \to \mathrm{P}(\mathcal{X}))$, is defined $\mathrm{inverse}(F) := \{(y, \{x : (x, z) \in F,\ z = y\}) : y \in \mathrm{ran}(F)\}$, and is sometimes denoted F^{-1}. The range of the inverse is a partition of the domain, $\mathrm{ran}(F^{-1}) \in \mathrm{B}(\mathrm{dom}(F))$.

Functions may be recursive. Algorithms are represented as recursive functions.

The powerset of bijective relations, or one-to-one functions, is denoted $\leftrightarrow$. The cardinality of the domain of a bijective function equals the range, $F \in \mathrm{dom}(F) \leftrightarrow \mathrm{ran}(F) \implies |\mathrm{dom}(F)| = |\mathrm{ran}(F)|$.

Total functions are denoted with a colon. For example, the left total function $F \in X :\to Y$ requires that $\mathrm{dom}(F) = X$ but only that $\mathrm{ran}(F) \subseteq Y$.

An order D on some set X is a choice of the enumerations, $D \in X :\leftrightarrow: \{1 \ldots |X|\}$. Given the order, any subset $Y \subseteq X$ can be enumerated. Define $\mathrm{order}(D, Y) \in Y :\leftrightarrow: \{1 \ldots |Y|\}$ such that $\forall a, b \in Y \ (D_a \leq D_b \implies \mathrm{order}(D, Y)(a) \leq \mathrm{order}(D, Y)(b))$.

The set of natural numbers $\mathbf{N}$ is taken to include 0. The set $\mathbf{N}_{>0}$ excludes 0. The *space* of a non-zero natural number is the natural logarithm, $\mathrm{space}(n) := \ln n$. The set of rational numbers is denoted $\mathbf{Q}$. The set of log-rational numbers is denoted $\ln \mathbf{Q}_{>0} = \{\ln q : q \in \mathbf{Q}_{>0}\}$. The set of real numbers is denoted $\mathbf{R}$.

The factorial of a non-zero natural number $n \in \mathbf{N}_{>0}$ is written $n! = \prod\{1 \ldots n\}$.

The unit-translated gamma function is the real function that corresponds to the factorial function. It is defined $(\Gamma_!) \in \mathbf{R} \to \mathbf{R}$ as $\Gamma_! x = \Gamma(x+1)$ which is such that $\forall n \in \mathbf{N}_{>0} \ (\Gamma_! n = \Gamma(n+1) = n!)$.

Given a relation $A \subset \mathcal{X} \times \mathcal{Y}$ such that an order operator is defined on the range, $\mathcal{Y}$, the max function returns the maximum subset, $\max \in \mathrm{P}(\mathcal{X} \times \mathcal{Y}) \to (\mathcal{X} \to \mathcal{Y})$

$$\max(A) := \{(x, y) : (x, y) \in A, \ \forall (r, s) \in A \ (s \leq y)\}$$

For convenience define the functions $\mathrm{maxd}(A) := \mathrm{dom}(\max(A))$ and $\mathrm{maxr}(A) := m$, where $\{m\} = \mathrm{ran}(\max(A))$. The corresponding functions for minimum, min, mind and minr, are similarly defined.

Given a relation $A \subset \mathcal{X} \times \mathcal{Y}$ such that the arithmetic operators are defined on the range, $\mathcal{Y}$, the sum function is defined $\mathrm{sum}(A) := \sum(y : (x, y) \in A)$. The relation can be normalised, $\mathrm{normalise}(A) := \{(x, y/\mathrm{sum}(A)) : (x, y) \in A\}$. Define notation $\hat{A} := \mathrm{normalise}(A)$. A normalised relation is such that its sum is one, $\mathrm{sum}(\hat{A}) = 1$.

The set of *probability functions* $\mathcal{P}$ is the set of rational valued functions such that the values are bounded $[0, 1]$ and sum to 1, $\mathcal{P} \subset \mathcal{X} \to \mathbf{Q}_{[0,1]}$ and $\forall P \in \mathcal{P} \ (\mathrm{sum}(P) = 1)$. The normalisation of a positive rational valued function $F \in \mathcal{X} \to \mathbf{Q}_{\geq 0}$ is a *probability function*, $\hat{F} \in \mathcal{P}$.

The *entropy* of positive rational valued functions, $\text{entropy} \in (\mathcal{X} \to \mathbf{Q}_{\geq 0}) \to \mathbf{Q}_{\geq 0} \ln \mathbf{Q}_{>0}$, is defined as $\text{entropy}(N) := -\sum(\hat{N}_x \ln \hat{N}_x : x \in \text{dom}(N),\ N_x > 0)$. The *entropy* of a singleton is zero, $\text{entropy}(\{(\cdot, 1)\}) = 0$. *Entropy* is maximised in uniform functions as the cardinality tends to infinity, $\text{entropy}(X \times \{1/|X|\}) = \ln |X|$.

Given some finite function $F \in \mathcal{X} \to \mathcal{Y}$, where $0 < |F| < \infty$, a *probability function* may be constructed from its distribution, $\{(y, |X|) : (y, X) \in F^{-1}\}^\wedge \in (\mathcal{Y} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$. The *probability function* of an arbitrarily chosen finite function is likely to have high *entropy*.

A *probability function* $P(z) \in (X :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$, parameterised by some parameter $z \in Z = \text{dom}(P)$, has a corresponding *likelihood function* $L(x) \in Z :\to \mathbf{Q}_{\geq 0}$, parameterised by coordinate $x \in X$, such that $L(x)(z) = P(z)(x)$. The *maximum likelihood estimate* $\tilde{z}$ of the parameter, z, at coordinate $x \in X$ is the mode of the *likelihood function*,

$$
\begin{aligned}
\{\tilde{z}\} &= \text{maxd}(L(x)) \\
&= \text{maxd}(\{(z, P(z)(x)) : z \in Z\}) \\
&= \{z : z \in Z,\ \forall z' \in Z\ (P(z)(x) \geq P(z')(x))\}
\end{aligned}
$$

A list is a object valued function of the natural numbers $\mathcal{L}(\mathcal{X}) \subset \mathbf{N} \to \mathcal{X}$, such that $\forall L \in \mathcal{L}(\mathcal{X})\ (L \neq \emptyset \implies \text{dom}(L) = \{1 \ldots |L|\})$. Two lists $L, M \in \mathcal{L}(\mathcal{X})$ may be concatenated, $\text{concat}(L, M) := L \cup \{(|L| + i, x) : (i, x) \in M\}$.

A tree is recusively defined as a tree valued function of objects, $\text{trees}(\mathcal{X}) = \mathcal{X} \to \text{trees}(\mathcal{X})$. The nodes of the tree $T \in \text{trees}(\mathcal{X})$ are $\text{nodes}(T) := T \cup \bigcup\{\text{nodes}(R) : (x, R) \in T\}$, and the elements are $\text{elements}(T) := \text{dom}(\text{nodes}(T))$. The paths of a tree $\text{paths}(T) \subset \mathcal{L}(\mathcal{X})$ is a set of lists. Given a set of lists $Q \subset \mathcal{L}(\mathcal{X})$ a tree can be constructed $\text{tree}(Q) \in \text{trees}(\mathcal{X})$.

2.2 Occam's Razor

Let $X \subset \mathcal{X}$ be a finite set of micro-states, $0 < |X| < \infty$. Consider a system of n distinguishable particles, each in a micro-state. The set of states of the system is the set of micro-state functions of particle identifier, $\{1 \ldots n\} :\to X$. The cardinality of the set of states is $|X|^n$.

Each state implies a distribution of particles over micro-states,

$$I = \big\{(R, \{(x, |C|) : (x, C) \in R^{-1}\}) : R \in \{1 \ldots n\} :\to X\big\}$$

That is, a state $R \in \{1 \ldots n\} :\to X$ has a particle distribution $I(R) \in X \to \{1 \ldots n\}$ such that $\text{sum}(I(R)) = n$.

The cardinality of states for each particle distribution, $I(R)$, is the multinomial coefficient,

$$\begin{aligned} W &= \{(N, |D|) : (N, D) \in I^{-1}\} \\ &= \{(N, \frac{n!}{\prod_{(x,\cdot) \in N} N_x!}) : (N, \cdot) \in I^{-1}\} \end{aligned}$$

That is, there are $W(I(R))$ states that have the same particle distribution, $I(R)$, as state R. The normalisation of the state distribution over particle distributions is a *probability function*, $\hat{W} \in ((X \to \{1 \ldots n\}) \to \mathbf{Q}_{>0}) \cap \mathcal{P}$.

In the case where the number of particles is large, $n \gg \ln n$, the logarithm of the multinomial coefficient of a particle distribution $N \in X \to \{1 \ldots n\}$ approximates to the scaled *entropy*,

$$\ln \frac{n!}{\prod_{(x,\cdot) \in N} N_x!} \approx n \times \text{entropy}(N)$$

so the probability of the particle distribution varies with its *entropy*, $\hat{W}(N) \sim \text{entropy}(N)$.

The least probable particle distributions are singletons,

$$\text{mind}(W) = \big\{\{(x, n)\} : x \in X\big\}$$

because they have only one state, $\forall x \in X\ (W(\{(x, n)\}) = 1)$. The *entropy* of a singleton distribution is zero, $\text{entropy}(\{(x, n)\}) = 0$.

In the case where the number of particles per micro-state is integral, $n/|X| \in \mathbf{N}_{>0}$, the modal particle distribution is the uniform distribution,

$$\text{maxd}(W) = \big\{\{(x, n/|X|) : x \in X\}\big\}$$

The *entropy* of the uniform distribution is maximised, $\text{entropy}(\{(x, n/|X|) : x \in X\}) = \ln |X|$.

The normalisation of a particle distribution $N \in X \to \{1 \ldots n\}$ is a micro-state *probability function*, $\hat{N} \in (X \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$, which is independent of the number of particles, $\mathrm{sum}(\hat{N}) = 1$.

So in the case where a problem domain is parameterised by an *unknown* micro-state *probability function* otherwise arbitrarily chosen from a *known* subset $Q \subseteq (X \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$, where the number of particles is *known* to be large, the *maximum likelihood estimate* $\tilde{P} \in Q$ is the *probability function* with the greatest *entropy*, $\forall P \in Q\ (\mathrm{entropy}(\tilde{P}) \geq \mathrm{entropy}(P))$ or $\tilde{P} \in \mathrm{maxd}(\{(P, \mathrm{entropy}(P)) : P \in Q\})$.

2.3 Histograms

2.3.1 States, histories and histograms

The set of all *variables* is $\mathcal{V}$. The set of all *values* is $\mathcal{W}$. A *system* $U \in \mathcal{V} \to \mathrm{P}(\mathcal{W})$ is a functional mapping between *variables* and non-empty sets of *values*, $\forall (v, W) \in U\ (|W| > 0)$. The *variables* of a *system* is the domain, $\mathrm{vars}(U) := \mathrm{dom}(U)$.

In a *system* of finite *variables*, $\forall v \in \mathrm{vars}(U)\ (|U_v| < \infty)$, each *variable* has a set of discrete *values*. The *values* need not be ordered. The *valency* of a *variable* v is the cardinality of its *values*, $|U_v|$. The *volume* of a set of *variables* in a *system* $V \subseteq \mathrm{vars}(U)$ is the product of the *valencies*, $\prod_{v \in V} |U_v| \geq 1$.

The set of *states* is the set of *value* valued functions of *variable*, $\mathcal{S} = \mathcal{V} \to \mathcal{W}$. The *variables* of a *state* $S \in \mathcal{S}$ is the function domain, $\mathrm{vars}(S) := \mathrm{dom}(S)$.

The *state*, S, is in a *system* U if (i) the *variables* of the *state* are *variables* of the *system*, $\mathrm{vars}(S) \subseteq \mathrm{vars}(U)$, and (ii) the *value* of each *variable* in the *state* is in the *system*, $\forall v \in \mathrm{vars}(S)\ (S_v \in U_v)$.

Given a set of *variables* in a *system* $V \subseteq \mathrm{vars}(U)$, the *cartesian* set of all possible *states* is $\prod_{v \in V}(\{v\} \times U_v)$, which has cardinality equal to the *volume* $\prod_{v \in V} |U_v|$.

The *variables* $V = \mathrm{vars}(S)$ of a *state* S may be *reduced* to a given subset $K \subseteq V$ by taking the subset of the *variable-value* pairs,

$$S \,\%\, K := \{(v, u) : (v, u) \in S,\ v \in K\}$$

A set of *states* $Q \subset \mathcal{S}$ in the same *variables* $\forall S \in Q\ (\mathrm{vars}(S) = V)$ may be *split* into a subset of its *variables* $K \subseteq V$ and the complement $V \setminus K$,

$$\mathrm{split}(K, Q) \quad = \quad \{(S \ \%\ K,\ S \ \%\ (V \setminus K)) : S \in Q\}$$

Two *states* $S, T \in \mathcal{S}$ are said to *join* if their union is also a *state*, $S \cup T \in \mathcal{S}$. That is, a *join* is functional,

$$\begin{aligned} S \cup T \in \mathcal{S} \quad &\iff \quad |\mathrm{vars}(S) \cup \mathrm{vars}(T)| = |S \cup T| \\ &\iff \quad \forall v \in \mathrm{vars}(S) \cap \mathrm{vars}(T)\ (S_v = T_v) \end{aligned}$$

States in disjoint *variables* always *join*, $\forall S, T \in \mathcal{S}\ (\mathrm{vars}(S) \cap \mathrm{vars}(T) = \emptyset \implies S \cup T \in \mathcal{S})$. *States* in the same *variables* only *join* if they are equal, $\forall S, T \in \mathcal{S}\ (\mathrm{vars}(S) = \mathrm{vars}(T) \implies (S \cup T \in \mathcal{S} \iff S = T))$.

The set of *event identifiers* is the universal set $\mathcal{X}$. An *event* (x, S) is a pair of an *event identifier* and a *state*, $(x, S) \in \mathcal{X} \times \mathcal{S}$. A *history* H is a *state* valued function of *event identifiers*, $H \in \mathcal{X} \to \mathcal{S}$, such that all of the *states* of its *events* share the same set of *variables*, $\forall (x, S), (y, T) \in H\ (\mathrm{vars}(S) = \mathrm{vars}(T))$. The set of *histories* is denoted $\mathcal{H} \subset \mathcal{X} \to \mathcal{S}$.

The set of *variables* of a *history* is the set of the *variables* of any of the *events* of the *history*, $\mathrm{vars}(H) = \mathrm{vars}(S)$ where $(x, S) \in H$.

The *event identifiers* of a *history* need not be ordered, so a *history* is not necessarily sequential or chronological.

The inverse of a *history*, H^{-1}, is called the *classification*. So a *classification* is an *event identifier* set valued function of *state*, $H^{-1} \in \mathcal{S} \to \mathrm{P}(\mathcal{X})$. The *event identifier* components are non-empty, $\forall (S, X) \in H^{-1}\ (X \neq \emptyset)$.

The *reduction* of a *history* is the *reduction* of its *events*, $H\%V := \{(x, S\%V) : (x, S) \in H\}$.

The *addition* operation of *histories* is defined as the disjoint union of the *events* if both *histories* have the same *variables*,

$$H_1 + H_2 \quad := \quad \{((x, \cdot), S) : (x, S) \in H_1\} \ \cup\ \{((\cdot, y), T) : (y, T) \in H_2\}$$

where $\mathrm{vars}(H_1) = \mathrm{vars}(H_2)$. The *size* of the *sum* equals the sum of the *sizes*, $|H_1 + H_2| = |H_1| + |H_2|$.

The *multiplication* operation of *histories* is defined as the product of the *events* where the *states join*,

$$\begin{aligned} H_1 * H_2 \ := \ \{((x,y), S \cup T) : (x,S) \in H_1,\ (y,T) \in H_2, \\ \forall v \in \mathrm{vars}(S) \cap \mathrm{vars}(T)\ (S_v = T_v)\} \end{aligned}$$

The *size* of the *product* equals the product of the *sizes* if the *variables* are disjoint, $\mathrm{vars}(H_1) \cap \mathrm{vars}(H_2) = \emptyset \implies |H_1 * H_2| = |H_1| \times |H_2|$. The *variables* of the *product* is the union of the *variables* if the *size* is non-zero, $H_1 * H_2 \neq \emptyset \implies \mathrm{vars}(H_1 * H_2) = \mathrm{vars}(H_1) \cup \mathrm{vars}(H_2)$.

The set of all *histograms* $\mathcal{A}$ is a subset of the positive rational valued functions of *states*, $\mathcal{A} \subset \mathcal{S} \to \mathbf{Q}_{\geq 0}$, such that each *state* of each *histogram* has the same set of *variables*, $\forall A \in \mathcal{A}\ \forall S,T \in \mathrm{dom}(A)\ (\mathrm{vars}(S) = \mathrm{vars}(T))$.

The set of *variables* of a *histogram* $A \in \mathcal{A}$ is the set of the *variables* of any of the elements of the *histogram*, $\mathrm{vars}(A) = \mathrm{vars}(S)$ where $(S,q) \in A$. The *dimension* of a *histogram* is the cardinality of its *variables*, $|\mathrm{vars}(A)|$. The *counts* of a *histogram* is the range. The *states* of a *histogram* is the domain. Define the shorthand $A^{\mathrm{S}} := \mathrm{dom}(A)$. The *size* of a *histogram* is the sum of the *counts*, $\mathrm{size}(A) := \mathrm{sum}(A)$. The *size* is always positive, $\mathrm{size}(A) \geq 0$. If the *size* is non-zero the normalised *histogram* has a *size* of one, $\mathrm{size}(A) > 0 \implies \mathrm{size}(\hat{A}) = 1$. In this case the normalised *histogram* is a *probability function*, $\mathrm{size}(A) > 0 \implies \hat{A} \in \mathcal{P}$.

The *volume* of a *histogram* A of *variables* V in a *system* U is the *volume* of the *variables*, $\prod_{v \in V} |U_v|$.

A *histogram* with no *variables* is called a *scalar*. The *scalar* of *size* z is $\{(\emptyset, z)\}$. Define $\mathrm{scalar}(z) := \{(\emptyset, z)\}$. A *singleton* is a *histogram* with only one *state*, $\{(S,z)\}$. A *uniform histogram* A has unique non-zero *count*, $|\{c : (S,c) \in A,\ c > 0\}| = 1$.

The set of *integral histograms* is the subset of *histograms* which have integal *counts* $\mathcal{A}_{\mathrm{i}} = \mathcal{A} \cap (\mathcal{S} \to \mathbf{N})$. A *unit histogram* is a special case of an *integral histogram* in which all its *counts* equal one, $\mathrm{ran}(A) = \{1\}$. The *size* of a *unit histogram* equals its cardinality, $\mathrm{size}(A) = |A|$. A set of *states* $Q \subset \mathcal{S}$ in the same *variables* may be promoted to a *unit histogram*, $Q^{\mathrm{U}} := Q \times \{1\} \in \mathcal{A}_{\mathrm{i}}$.

The *unit effective histogram* of a *histogram* is the *unit histogram* of the *states* where the *count* is non-zero. Define the shorthand $A^{\mathrm{F}} := \{(S,1) : (S,c) \in$

$A,\ c > 0\} \in \mathcal{A}_{\mathrm{i}}$.

Given a *system* U define the *cartesian histogram* of the set of *variables* V as $V^{\mathrm{C}} := \big(\prod_{v \in V} (\{v\} \times U_v)\big) \times \{1\} \in \mathcal{A}_{\mathrm{i}}$. The *size* of the *cartesian histogram* equals its cardinality which is the *volume* of the *variables*, $\mathrm{size}(V^{\mathrm{C}}) = |V^{\mathrm{C}}| = \prod_{v \in V} |U_v|$. The *unit effective histogram* is a subset of the *cartesian histogram* of its *variables*, $A^{\mathrm{F}} \subseteq V^{\mathrm{C}}$, where $V = \mathrm{vars}(A)$. A *complete histogram* has the *cartesian* set of *states*, $A^{\mathrm{S}} = V^{\mathrm{CS}}$.

A *partition* P is a partition of the *cartesian states*, $P \in \mathrm{B}(V^{\mathrm{CS}})$. The *partition* is a set of disjoint *components*, $\forall C, D \in P\ (C \neq D \implies C \cap D = \emptyset)$, that union to equal the *cartesian states*, $\bigcup P = V^{\mathrm{CS}}$. The *unary partition* is $\{V^{\mathrm{CS}}\}$. The *self partition* is $V^{\mathrm{CS}\{\}} = \{\{S\} : S \in V^{\mathrm{CS}}\}$. A *partition variable* $P \in \mathrm{vars}(U)$ in a *system* U is such that its set of *values* equals its set of *components*, $U_P = P$. So the *valency* of a *partition variable* is the cardinality of the *components*, $|U_P| = |P|$.

A *regular histogram* A of *variables* V in *system* U has unique *valency* of its *variables*, $|\{|U_v| : v \in V\}| = 1$. The *volume* of a *regular histogram* is $d^n = |V^{\mathrm{C}}| = \prod_{v \in V} |U_v|$, where *valency* d is such that $\{d\} = \{|U_v| : v \in V\}$ and *dimension* $n = |V|$.

The *counts* of the *integral histogram* $A \in \mathcal{A}_{\mathrm{i}}$ of a *history* $H \in \mathcal{H}$ are the cardinalities of the *event identifier* components of its *classification*, $A = \mathrm{histogram}(H)$ where $\mathrm{histogram}(H) := \{(S, |X|) : (S, X) \in H^{-1}\}$. In this case the *histogram* is a distribution of *events* over *states*. If the *history* is bijective, $H \in \mathcal{X} \leftrightarrow \mathcal{S}$, then its *histogram* is a *unit histogram*, $A = \mathrm{ran}(H) \times \{1\}$.

A *sub-histogram* A of a *histogram* B is such that the *effective states* of A are a subset of the *effective states* of B and the *counts* of A are less than or equal to those of B, $A \leq B := A^{\mathrm{FS}} \subseteq B^{\mathrm{FS}}\ \wedge\ \forall S \in A^{\mathrm{FS}}\ (A_S \leq B_S)$. The *histogram* of a *sub-history* $G \subseteq H$ is a *sub-histogram*, $\mathrm{histogram}(G) \leq \mathrm{histogram}(H)$.

The *reduction* of a *histogram* is the *reduction* of its *states*, adding the *counts* where two different *states reduce* to the same *state*,

$$A\%V := \{(R, \sum(c : (T, c) \in A,\ T \supseteq R)) : R \in \{S\%V : S \in A^{\mathrm{S}}\}\}$$

Reduction leaves the *size* of a *histogram* unchanged, $\mathrm{size}(A\%V) = \mathrm{size}(A)$, but the number of *states* may be fewer, $|(A\%V)^{\mathrm{S}}| \leq |A^{\mathrm{S}}|$. The *reduction* to the empty set is a *scalar*, $A\%\emptyset = \{(\emptyset, z)\}$, where $z = \mathrm{size}(A)$. The *histogram* of a *reduction* of a *history* equals the *reduction* of the *histogram* of the *history*,

$$\text{histogram}(H \ \% \ V) = \text{histogram}(H) \ \% \ V$$

The *addition* of *histograms* A and B is defined,

$$\begin{aligned} & A + B := \\ & \quad \{(S, c) : (S, c) \in A, \ S \notin B^{\mathrm{S}}\} \cup \\ & \quad \{(S, c + d) : (S, c) \in A, \ (T, d) \in B, \ S = T\} \cup \\ & \quad \{(T, d) : (T, d) \in B, \ T \notin A^{\mathrm{S}}\} \end{aligned}$$

where $\text{vars}(A) = \text{vars}(B)$. The *sizes* add, $\text{size}(A + B) = \text{size}(A) + \text{size}(B)$. The *histogram* of an *addition* of *histories* equals the *addition* of the *histograms* of the *histories*,

$$\text{histogram}(H_1 + H_2) = \text{histogram}(H_1) + \text{histogram}(H_2)$$

The *multiplication* of *histograms* A and B is the product of the *counts* where the *states join*,

$$A * B := \{(S \cup T, cd) : (S, c) \in A, \ (T, d) \in B, \ \forall v \in \text{vars}(S) \cap \text{vars}(T) \ (S_v = T_v)\}$$

If the *variables* are disjoint, the *sizes* multiply, $\text{vars}(A) \cap \text{vars}(B) = \emptyset \implies \text{size}(A * B) = \text{size}(A) \times \text{size}(B)$. *Multiplication* by a *scalar* scales the *size*, $\text{size}(\text{scalar}(z) * A) = z \times \text{size}(A)$. The *histogram* of a *multiplication* of *histories* equals the *multiplication* of the *histograms* of the *histories*,

$$\text{histogram}(H_1 * H_2) = \text{histogram}(H_1) * \text{histogram}(H_2)$$

The *reciprocal* of a *histogram* is $1/A := \{(S, 1/c) : (S, c) \in A, \ c > 0\}$. Define *histogram division* as $B/A := B * (1/A)$.

A *histogram* A is *causal* in a subset of its *variables* $K \subset V$ if the *reduction* of the *effective states* to the subset, K, is functionally related to the *reduction* to the complement, $V \setminus K$,

$$\{(S \ \% \ K, \ S \ \% \ (V \setminus K)) : S \in A^{\mathrm{FS}}\} \ \in \ K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$$

or

$$\text{split}(K, A^{\mathrm{FS}}) \ \in \ K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$$

A *histogram* A is *diagonalised* if no pair of *effective states* shares any *value*, $\forall S, T \in A^{\mathrm{FS}} \ (S \neq T \implies S \cap T = \emptyset)$. A *diagonalised histogram* A is *fully* *diagonalised* if its *effective* cardinality equals the minimum *valency* of its *variables*, $|A^{\mathrm{F}}| = \mathrm{minr}(\{(v, |U_v|) : v \in V\})$. The cardinality of the *effective states* of a *fully diagonalised regular histogram* is the *valency*, $|A^{\mathrm{F}}| = d$, where $\{d\} = \{|U_v| : v \in V\}$. In a *diagonalised histogram* the *causality* is bijective or equational,

$$\forall u, w \in V\ (\{(S\%\{u\}, S\%\{w\}) : S \in A^{\mathrm{FS}}\}\ \in\ \{u\}^{\mathrm{CS}} \leftrightarrow \{w\}^{\mathrm{CS}})$$

Given some *slice state* $R \in K^{\mathrm{CS}}$, where $K \subset V$ and $V = \mathrm{vars}(A)$, the *slice histogram*, $A * \{R\}^{\mathrm{U}} \subset A$, is said to be *contingent* on the *incident slice state*. For example, if the *slice histogram* is *diagonalised*, $\mathrm{diagonal}(A * \{R\}^{\mathrm{U}}\ \%\ (V \setminus K))$, then the *histogram*, A, is said to be *contingently diagonalised*.

The *perimeters* of a *histogram* $A \in \mathcal{A}$ is the set of its *reductions* to each of its *variables*, $\{A\%\{w\} : w \in V\}$, where $V = \mathrm{vars}(A)$. The *independent* of a *histogram* is the product of the normalised *perimeters* scaled to the *size*,

$$A^{\mathrm{X}} := Z * \prod_{w \in V} \hat{A}\%\{w\}$$

where $z = \mathrm{size}(A)$ and $Z = \mathrm{scalar}(z) = A\%\emptyset$. The *independent* of a *histogram* is such that (i) the *states* are a superset, $A^{\mathrm{XS}} \supseteq A^{\mathrm{S}}$, (ii) the *size* is unchanged, $\mathrm{size}(A^{\mathrm{X}}) = \mathrm{size}(A)$, and (iii) the *variables* are unchanged, $\mathrm{vars}(A^{\mathrm{X}}) = \mathrm{vars}(A)$. A *histogram* is said to be *independent* if it equals its *independent*, $A = A^{\mathrm{X}}$. The *independent* of an *independent histogram* is the *independent*, $A^{\mathrm{XX}} = A^{\mathrm{X}}$. The scaled *product* of (i) any *reduction* of a normalised *independent histogram* to any subset of its *variables* $K \subseteq V$, and (ii) the *reduction* to the complement, $V \setminus K$, is the *independent*, $Z * (\hat{A}^{\mathrm{X}}\ \%\ K) * (\hat{A}^{\mathrm{X}}\ \%\ (V \setminus K)) = A^{\mathrm{X}}$.

Scalar histograms are *independent*, $\{(\emptyset, z)\} = \{(\emptyset, z)\}^{\mathrm{X}}$. *Singleton histograms*, $|A^{\mathrm{F}}| = 1$, are *independent*, $\{(S, z)\} = \{(S, z)\}^{\mathrm{X}}$. If the *histogram* is *mono-variate*, $|V| = 1$, then it is *independent* $A = A\%\{w\} = A^{\mathrm{X}}$ where $\{w\} = V$. *Uniform-cartesian histograms*, which are *scalar* multiples of the *cartesian*, $A = V_z^{\mathrm{C}}$ where $V_z^{\mathrm{C}} = \mathrm{scalar}(z/v) * V^{\mathrm{C}}$, $z = \mathrm{size}(A)$ and $v = |V^{\mathrm{C}}|$, are *independent*, $V_z^{\mathrm{C}} = V_z^{\mathrm{CX}}$.

A *completely effective pluri-variate independent histogram*, $A^{\mathrm{XF}} = V^{\mathrm{C}}$ where $|V| > 1$, for which all of the *variables* are *pluri-valent*, $\forall w \in V\ (|U_w| > 1)$, must be *non-causal*,

$$\begin{aligned}&\forall K \subset V\ (K \notin \{\emptyset, V\} \implies \\ &\qquad \{(S\ \%\ K,\ S\ \%\ (V \setminus K)) : S \in A^{\mathrm{XFS}}\}\ \notin\ K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}})\end{aligned}$$

The set of *substrate histories* $\mathcal{H}_{U,V,z}$ is the set of *histories* having *event identifiers* $\{1 \ldots z\}$, fixed *size* z and fixed *variables* V,

$$\begin{aligned}\mathcal{H}_{U,V,z} &:= \{1 \ldots z\} :\to V^{\mathrm{CS}} \\ &= \{H : H \subseteq \{1 \ldots z\} \times V^{\mathrm{CS}},\ \mathrm{dom}(H) = \{1 \ldots z\},\ |H| = z\}\end{aligned}$$

The cardinality of the *substrate histories* is $|\mathcal{H}_{U,V,z}| = v^z$ where $v = |V^{\mathrm{C}}|$. If the *volume*, v, is finite, the set of *substrate histories* is finite, $|\mathcal{H}_{U,V,z}| < \infty$.

The corresponding set of *integral substrate histograms* $\mathcal{A}_{U,\mathrm{i},V,z}$ is the set of *complete integral histograms* in *variables* V with *size* z,

$$\begin{aligned}\mathcal{A}_{U,\mathrm{i},V,z} &:= \{\mathrm{histogram}(H) : H \in \mathcal{H}_{U,V,z}\} \\ &= \{A : A \in V^{\mathrm{CS}} :\to \{0 \ldots z\},\ \mathrm{size}(A) = z\}\end{aligned}$$

Note that the *histogram* function is redefined here to return *complete histograms*, $\mathrm{histogram}(H) := \{(S, |X|) : (S, X) \in H^{-1}\}\ +\ V^{\mathrm{CS}} \times \{0\}$.

The cardinality of *integral substrate histograms* is the cardinality of weak compositions,

$$|\mathcal{A}_{U,\mathrm{i},V,z}| = \frac{(z + v - 1)!}{z!\ (v - 1)!}$$

If the *volume*, v, is finite, the set of *integral substrate histograms* is finite, $|\mathcal{A}_{U,\mathrm{i},V,z}| < \infty$.

2.3.2 Entropy and alignment

The *entropy* of a *non-zero histogram* $A \in \mathcal{A}$ is defined as the expected negative logarithm of the normalised *counts*,

$$\mathrm{entropy}(A) := -\sum_{S \in A^{\mathrm{FS}}} \hat{A}_S \ln \hat{A}_S$$

The *sized entropy* is $z \times \mathrm{entropy}(A)$ where $z = \mathrm{size}(A)$. The *entropy* of a *singleton* is zero, $z \times \mathrm{entropy}(\{(\cdot, z)\}) = 0$. *Entropy* is highest in *cartesian histograms*, which are *uniform* and have maximum *effective volume*. The maximum *sized entropy* is $z \times \mathrm{entropy}(V_z^{\mathrm{C}}) = z \ln v$ where $v = |V^{\mathrm{C}}|$.

Given a *histogram* A and a set of query *variables* $K \subset V$, the label *entropy* is the degree to which the *histogram* is ambiguous or *non-causal* in the query *variables*, K. It is the sum of the *sized entropies* of the *contingent slices reduced* to the label *variables*, $V \setminus K$,

$$\sum_{R \in (A\%K)^{\mathrm{FS}}} (A\%K)_R \times \mathrm{entropy}(A * \{R\}^{\mathrm{U}} \ \% \ (V \setminus K))$$

When the *histogram*, A, is *causal* in the query *variables*, $\mathrm{split}(K, A^{\mathrm{FS}}) \in K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$, the label *entropy* is zero because each *slice* is an *effective singleton*, $\forall R \in (A\%K)^{\mathrm{FS}}\ (|A^{\mathrm{F}} * \{R\}^{\mathrm{U}}| = 1)$. In this case the label *state* is unique for every *effective* query *state*. By contrast, when the label *variables* are *independent* of the query *variables*, $A = Z * \hat{A}\%K * \hat{A}\%(V \setminus K)$, the label *entropy* is maximised.

The *multinomial coefficient* of a *non-zero integral histogram* $A \in \mathcal{A}_{\mathrm{i}}$ is

$$\frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \quad \in \quad \mathbf{N}_{>0}$$

where $z = \mathrm{size}(A) > 0$. In the case where the *histogram* is *non-integral* the *multinomial coefficient* is defined by the unit-translated gamma function,

$$\frac{\Gamma_! z}{\prod_{S \in A^{\mathrm{S}}} \Gamma_! A_S}$$

Given an arbitrary *substrate history* $H \in \mathcal{H}_{U,V,z}$ and its *histogram* $A = \mathrm{histogram}(H)$, the cardinality of *histories* having the same *histogram*, A, is the *multinomial coefficient*,

$$|\{G : G \in \mathcal{H}_{U,V,z},\ \mathrm{histogram}(G) = A\}| \quad = \quad \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!}$$

In the case where the *counts* are not small, $z \gg \ln z$, the logarithm of the *multinomial coefficient* approximates to the *sized entropy*,

$$\ln \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \quad \approx \quad z \times \mathrm{entropy}(A)$$

so the *entropy*, $\mathrm{entropy}(A)$, is a measure of the probability of the *histogram* of a randomly chosen *history*. *Singleton histograms* are least probable and *uniform histograms* are most probable.

The *sized relative entropy* between a *histogram* and its *independent* is the *sized mutual entropy*,

$$\sum_{S \in A^{\mathrm{FS}}} A_S \ln \frac{A_S}{A^{\mathrm{X}}_S}$$

It can be shown that the *size* scaled expected logarithm of the *independent* with respect to the *histogram* equals the *size* scaled expected logarithm of the *independent* with respect to the *independent*,

$$\sum_{S \in A^{\mathrm{FS}}} A_S \ln A^{\mathrm{X}}_S = \sum_{S \in A^{\mathrm{XFS}}} A^{\mathrm{X}}_S \ln A^{\mathrm{X}}_S$$

so the *sized mutual entropy* is the difference between the *sized independent entropy* and the *sized histogram entropy*,

$$\sum_{S \in A^{\mathrm{FS}}} A_S \ln \frac{A_S}{A^{\mathrm{X}}_S} \quad = \quad z \times \mathrm{entropy}(A^{\mathrm{X}}) - z \times \mathrm{entropy}(A)$$

The *sized mutual entropy* can be viewed as a measure of the probability of the *independent*, A^{X}, relative to the *histogram*, A, given arbitrary *substrate history*. Equivalently, *sized mutual entropy* can be viewed as a measure of the surprisal of the *histogram*, A, relative to the *independent*, A^{X}. That is, *sized mutual entropy* is a measure of the dependency between the *variables* in the *histogram*, A.

The *sized mutual entropy* is the *sized relative entropy* so it is always positive,

$$z \times \mathrm{entropy}(A^{\mathrm{X}}) - z \times \mathrm{entropy}(A) \quad \geq \quad 0$$

and so the *independent entropy* is always greater than or equal to the *histogram entropy*

$$\mathrm{entropy}(A^{\mathrm{X}}) \quad \geq \quad \mathrm{entropy}(A)$$

That is, *histograms* of *substrate histories* arbitrarily chosen from a uniform distribution are probably *independent* or nearly *independent*. The expected *sized mutual entropy* is low.

An example of a dependency between *variables* is where a *histogram* A is *causal* in a subset of its *variables* $K \subset V$. In this case the *histogram* cannot be *independent*, $A \neq A^{\mathrm{X}}$, and so the *sized mutual entropy* must be greater than zero,

$$\{(S \,\%\, K,\ S \,\%\, (V \setminus K)) : S \in A^{\mathrm{FS}}\} \ \in\ K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}} \implies$$
$$z \times \mathrm{entropy}(A^{\mathrm{X}}) - z \times \mathrm{entropy}(A) \ >\ 0$$

The *alignment* of a *histogram* $A \in \mathcal{A}$ is defined

$$\mathrm{algn}(A) \;:=\; \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S$$

where $\Gamma_!$ is the unit-translated gamma function.

In the case where both the *histogram* and its *independent* are *integral*, $A, A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, then the *alignment* is the difference between the sum log-factorial *counts* of the *histogram* and its *independent*,

$$\mathrm{algn}(A) \;=\; \sum_{S \in A^{\mathrm{S}}} \ln A_S! - \sum_{S \in A^{\mathrm{XS}}} \ln A^{\mathrm{X}}_S!$$

Alignment is the logarithm of the ratio of the *independent multinomial coefficient* to the *multinomial coefficient*,

$$\mathrm{algn}(A) \;=\; \ln \left(\frac{z!}{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!} \Big/ \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \right)$$

so *alignment* is the logarithm of the probability of the *independent*, A^{X}, relative to the *histogram*, A. Equivalently, *alignment* is the logarithm of the surprisal of the *histogram*, A, relative to the *independent*, A^{X}. *Alignment* is a measure of the dependency between the *variables* in the *histogram*, A.

Alignment is approximately equal to the *sized mutual entropy*,

$$\begin{aligned}\mathrm{algn}(A) \;&\approx\; z \times \mathrm{entropy}(A^{\mathrm{X}}) - z \times \mathrm{entropy}(A) \\ &=\; \sum_{S \in A^{\mathrm{FS}}} A_S \ln \frac{A_S}{A^{\mathrm{X}}_S}\end{aligned}$$

so the *histogram* of an arbitrary *history* usually has low *alignment*. Note that, because *sized entropy* is only an approximation to the logarithm of the *multinomial coefficient*, especially at low *sizes*, *alignment* is the better measure of the surprisal of the *histogram*, A, relative to the *independent*, A^{X}, than *sized mutual entropy*.

The *alignment* of an *independent histogram*, $A = A^{\mathrm{X}}$, is zero. In particular, *scalar histograms*, $V = \emptyset$, *mono-variate histograms*, $|V| = 1$, *uniform cartesian histograms*, $A = V^{\mathrm{C}}_z$, and *effective singleton histograms*, $|A^{\mathrm{F}}| = 1$, all have zero *alignment*.

The maximum *alignment* of a *histogram* A occurs when the *histogram* is both *uniform* and *fully diagonalised*. No pair of *effective states* shares any *value*, $\forall S,T \in A^{\mathrm{FS}}\ (S \neq T \implies S \cap T = \emptyset)$, and all *counts* are equal along the *diagonal*, $\forall S,T \in A^{\mathrm{FS}}\ (A_S = A_T)$. The maximum *alignment* of a *regular histogram* with *dimension* $n = |V|$ and *valency* d is

$$d \ln \Gamma_! \frac{z}{d} \ - \ d^n \ln \Gamma_! \frac{z}{d^n}$$

The maximum *alignment* is approximately $z \ln d^{n-1} = z \ln v/d$, where $v = d^n$. It can be compared to the maximum *sized entropy* of the *'co-histogram' reduced* by one *variable* along the *diagonal*.

Although *alignment* varies against *sized entropy*, $\mathrm{algn}(A) \sim -z \times \mathrm{entropy}(A)$, the maximum *alignment* does not occur when the *entropy* is minimised. At minimum *entropy* the *histogram* is a *singleton*, but the *alignment* is zero because *singletons* are *independent*.

An example of an *aligned histogram* A is where the *histogram* is *causal* in a subset of its *variables* $K \subset V$. In this case the *histogram* cannot be *independent*, $A \neq A^{\mathrm{X}}$, and so the *alignment* must be greater than zero,

$$\{(S\%K,\ S\%(V \setminus K)) : S \in A^{\mathrm{FS}}\} \in K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}} \implies \mathrm{algn}(A) > 0$$

At maximum *alignment* the *histogram* is *fully diagonalised*, so all pairs of *variables* are necessarily bijectively *causal* or equational,

$$\forall u,w \in V\ (\{(S\%\{u\}, S\%\{w\}) : S \in A^{\mathrm{FS}}\} \ \in \ \{u\}^{\mathrm{CS}} \to \{w\}^{\mathrm{CS}})$$

2.3.3 Encoding and compression

A *substrate history probability function* $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is a normalised distribution over *substrate histories*, $\sum(P_H : H \in \mathcal{H}_{U,V,z}) = 1$. The entropy of the *probability function* is $\mathrm{entropy}(P)$. Note that *history probability function* entropy is not to be confused with *histogram entropy*. A *history probability function* is a distribution over *histories*, $\mathcal{H}_{U,V,z} \to \mathbf{Q}_{\geq 0}$, whereas a *histogram* is a distribution of *events* over *states*, $V^{\mathrm{CS}} \to \mathbf{Q}_{\geq 0}$.

History coders define the conversion of lists of *histories*, $\mathcal{L}(\mathcal{H})$, to and from the natural numbers, $\mathbf{N}$. A *substrate history coder* $C \in \mathrm{coders}(\mathcal{H}_{U,V,z})$ defines an *encode* function of any list of *substrate histories* into a positive integer, $\mathrm{encode}(C) \in \mathcal{L}(\mathcal{H}_{U,V,z}) :\to \mathbf{N}$, and the corresponding *decode* function of the integer back into the list of *histories*, $\mathrm{decode}(C) \in \mathbf{N} \times \mathbf{N} \to: \mathcal{L}(\mathcal{H}_{U,V,z})$, given the length of the list.

A third function is the *space* function, $\mathrm{space}(C) \in \mathcal{H}_{U,V,z} :\to \ln \mathbf{N}_{>0}$, which defines the logarithm of the cardinality of the encoding states of a *substrate history*. The encoding integer of a single *history* is always less than this cardinality, $\forall H \in \mathcal{H}_{U,V,z}\ (\mathrm{encode}(C)(\{(1,H)\}) < \exp(\mathrm{space}(C)(H)))$. The *space* of an encoded list of *histories* is the sum of the *spaces* of the *histories*. The *space* function is also denoted $C^{\mathrm{s}} = \mathrm{space}(C)$.

Given a *substrate history probability function* $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$, the *expected substrate history space* is $\sum(P_H C^{\mathrm{s}}(H) : H \in \mathcal{H}_{U,V,z})$. The *expected space* is always greater than or equal to the *probability function* entropy (or Shannon entropy in nats), $\sum(P_H C^{\mathrm{s}}(H) : H \in \mathcal{H}_{U,V,z}) \ \geq\ \mathrm{entropy}(P)$.

A *minimal history coder* $C_{\mathrm{m},U,V,z} \in \mathrm{coders}(\mathcal{H}_{U,V,z})$ encodes the *history* by encoding the index of an enumeration of the entire set of *substrate histories*, $\mathrm{encode}(C_{\mathrm{m},U,V,z})(\{(1,H)\}) \in \{0 \ldots v^z - 1\}$. The *space* is fixed, $C^{\mathrm{s}}_{\mathrm{m},U,V,z}(H) = \ln |\mathcal{H}_{U,V,z}| = z \ln v$. In the case where the *probability function* is uniform, $P = \mathcal{H}_{U,V,z} \times \{1/v^z\}$, the *expected space* equals the *probability function* entropy, $\sum(P_H C^{\mathrm{s}}_{\mathrm{m},U,V,z}(H) : H \in \mathcal{H}_{U,V,z}) = \mathrm{entropy}(P) = z \ln v$. In other words, when the *history* is arbitrary then the *minimal history coder* has the least *expected space*.

There are two *canonical history coders*, the *index history coder* C_{H} and the *classification coder* C_{G}. The *index substrate history coder* $C_{\mathrm{H},U,V,z} \in \mathrm{coders}(\mathcal{H}_{U,V,z})$ is the simpler of the two. It encodes each *history* by indexing the *volume* for each *event*. The *space* of an index into a *volume* $v = |V^{\mathrm{CS}}|$ is $\ln v$. So the total *space* of any *substrate history* $H \in \mathcal{H}_{U,V,z}$ is

$$C^{\mathrm{s}}_{\mathrm{H},U,V,z}(H) \quad = \quad z \ln v$$

The *space* is fixed because it does not depend on the *histogram*, A. The *index history space* equals the *minimal history space*, $C^{\mathrm{s}}_{\mathrm{H},U,V,z}(H) = C^{\mathrm{s}}_{\mathrm{m},U,V,z}(H) = z \ln v$, but the *encode* functions are different. In the case of an arbitrary *history*, or uniform *history probability function*, the *index history coder* also has least *expected space*.

The *classification substrate history coder* $C_{\mathrm{G},U,V,z} \in \mathrm{coders}(\mathcal{H}_{U,V,z})$ encodes each *history* in two steps. First the *histogram* is encoded by choosing one of the *integral substrate histograms*, $\mathcal{A}_{U,\mathrm{i},V,z}$. The choice has fixed *space*

$$\ln |\mathcal{A}_{U,\mathrm{i},V,z}| \quad = \quad \ln \frac{(z+v-1)!}{z!\ (v-1)!}$$

Given the *histogram*, A, the cardinality of *classifications* equals the *multinomial coefficient*. Now the choice of the *classification*, H^{-1}, is encoded in a *space* equal to the logarithm of the *multinomial coefficient*,

$$\ln \frac{z!}{\prod_{S \in A^{\mathrm{s}}} A_S!}$$

The total *space* to encode the *history* in the *classification substrate history coder* is

$$C^{\mathrm{s}}_{\mathrm{G},U,V,z}(H) \quad = \quad \ln \frac{(z+v-1)!}{z!\ (v-1)!} \ + \ \ln \frac{z!}{\prod_{S \in A^{\mathrm{s}}} A_S!}$$

The *space* is not fixed because it depends on the *histogram*, A.

The *classification space* may be approximated in terms of *sized entropy*,

$$C^{\mathrm{s}}_{\mathrm{G},U,V,z}(H) \quad \approx \quad (z+v)\ln(z+v) \ - \ z \ln z \ - \ v \ln v \ + \ z \times \mathrm{entropy}(A)$$

The maximum *sized entropy*, $z \times \mathrm{entropy}(A)$, is $z \ln v$, so when the *entropy* is high the *classification space* is greater than the *index space*, $C^{\mathrm{s}}_{\mathrm{G},U,V,z}(H) > C^{\mathrm{s}}_{\mathrm{H},U,V,z}(H)$, but when the *entropy* is low the *classification space* is less than the *index space*, $C^{\mathrm{s}}_{\mathrm{G},U,V,z}(H) < C^{\mathrm{s}}_{\mathrm{H},U,V,z}(H)$. The break-even *sized entropy* is approximately

$$z \times \mathrm{entropy}(A) \quad \approx \quad z \ln v \ - \ ((z+v)\ln(z+v) \ - \ z \ln z \ - \ v \ln v)$$

In the case where the *size* is much less than the *volume*, $z \ll v$, the break-even *sized entropy* is approximately $z \times \mathrm{entropy}(A) \approx z \ln z$.

2.4 Induction without model

Induction may be defined as the determination of the *likely* properties of *unknown history probability functions*.

Let P be a *substrate history probability function*, $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$. Let the domain of the *probability function*, $\mathrm{dom}(P) = \mathcal{H}_{U,V,z}$, be *known*. The simplest case of *induction* is that nothing else is *known* about the *probability function*, P. If the *probability function* is assumed to be the normalisation of the distribution of a finite *history* valued function of undefined particle, $\mathcal{X} \to \mathcal{H}$, and this particle function is assumed to be chosen arbitrarily, then the *maximum likelihood estimate* $\tilde{P}$ for the *probability function*, P, maximises the entropy, $\mathrm{entropy}(\tilde{P})$, at the mode. So the *likely history probability function*, $\tilde{P}$, is the uniform distribution,

$$\tilde{P} \quad = \quad \mathcal{H}_{U,V,z} \times \{1/v^z\}$$

That is, the *likely substrate histories* are arbitrary or random.

The next case is where a *history* $H \in \mathcal{H}_{U,V,z}$ is *known* to be *necessary*, $P(H) = 1$. In this case the *probability function*, P, is,

$$\begin{aligned} P &= \{(H,1)\} \ \cup \\ &\quad\ \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \neq H\} \end{aligned}$$

If the *history*, H, is *known*, then the *probability function*, P, is *known*. The *maximum likelihood estimate* equals the *probability function*, $\tilde{P} = P$. The entropy is zero, $\mathrm{entropy}(\tilde{P}) = 0$.

2.4.1 Classical induction

In *classical induction* the *history probability functions* are constrained by *histogram*.

Let his = histogram. Now consider the case where the *histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ is *known* to be *necessary*, $\sum(P(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) = 1$. The *maximum likelihood estimate* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned} \tilde{P} &= \{(H,1) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A\}^{\wedge}\ \cup \\ &\quad\ \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A\} \\ &= \{(H, 1/\frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!}) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A\}\ \cup \\ &\quad\ \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A\} \end{aligned}$$

where $()^{\wedge}$ = normalise. That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *histories* with the *histogram*, $\mathrm{his}(H) = A$, are uniformly probable and all other *histories*, $\mathrm{his}(G) \neq A$, are impossible, $\tilde{P}(G) = 0$. If the *histogram*, A, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*. Note that the *likely history probability function* entropy varies with the *histogram entropy*, $\mathrm{entropy}(\tilde{P}) \sim \mathrm{entropy}(A)$.

Next consider the case where either *histogram* A or *histogram* B are *known* to be *necessary*, $\sum(P(H) : H \in \mathcal{H}_{U,V,z},\ (\mathrm{his}(H) = A\ \vee\ \mathrm{his}(H) = B)) = 1$. The *maximum likelihood estimate* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} &= \{(H,1) : H \in \mathcal{H}_{U,V,z},\ (\mathrm{his}(H) = A\ \vee\ \mathrm{his}(H) = B)\}^{\wedge}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A,\ \mathrm{his}(G) \neq B\} \\
&= \{(H, 1/\left(\frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} + \frac{z!}{\prod_{S \in B^{\mathrm{S}}} B_S!}\right)) : \\
&\qquad\qquad H \in \mathcal{H}_{U,V,z},\ (\mathrm{his}(H) = A\ \vee\ \mathrm{his}(H) = B)\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A,\ \mathrm{his}(G) \neq B\}
\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *histories* with either *histogram*, A or B, are uniformly probable and all other *histories*, $\mathrm{his}(G) \neq A$ and $\mathrm{his}(G) \neq B$, are impossible, $\tilde{P}(G) = 0$. If the *histograms*, A and B, are *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

Given a *history* $H_E \in \mathcal{H}_{U,V,z_E}$, of *size* $z_E = |H_E|$, consider the case where its subsets of *size* z are *known* to be necessary, $\sum(P(H) : H \subseteq H_E,\ |H| = z) = 1$. The given *history*, H_E, is called the *distribution history*. A subset $H \subseteq H_E$ is a *sample history drawn* from the *distribution history*, H_E. The *maximum likelihood estimate* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} &= \{(H,1) : H \subseteq H_E,\ |H| = z\}^{\wedge}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\} \\
&= \{(H, 1/\binom{z_E}{z}) : H \subseteq H_E,\ |H| = z\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\}
\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn histories* $H \subseteq H_E$ of *size* $|H| = z$ are uniformly probable and all other *histories*, $G \nsubseteq H_E$, are impossible, $\tilde{P}(G) = 0$. If the *distribution histogram*, H_E, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

Now consider the case where the *drawn histogram* A is *known* to be *necessary*, $\sum(P(H) : H \subseteq H_E,\ \mathrm{his}(H) = A) = 1$. The *maximum likelihood* *estimate* which maximises the entropy, entropy$(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} &= \{(H,1) : H \subseteq H_E,\ \mathrm{his}(H) = A\}^{\wedge}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A\} \\
&= \{(H, 1/\prod_{S \in A^{\mathrm{S}}} \binom{E_S}{A_S}) : H \subseteq H_E,\ \mathrm{his}(H) = A\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A\}
\end{aligned}$$

where the *distribution histogram* $E = \mathrm{his}(H_E)$.

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn histories* $H \subseteq H_E$ with the *histogram*, $\mathrm{his}(H) = A$, are uniformly probable and all other *histories*, $G \not\subseteq H_E$ or $\mathrm{his}(G) \neq A$, are impossible, $\tilde{P}(G) = 0$. If the *histogram*, A, is *known* and the *distribution histogram*, H_E, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

The *historical distribution* $Q_{\mathrm{h},U}$ is defined

$$Q_{\mathrm{h},U}(E,z)(A) \ := \ \prod_{S \in A^{\mathrm{S}}} \binom{E_S}{A_S} = \prod_{S \in A^{\mathrm{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!}$$

where $A \leq E$. The *frequency* of *histogram* A in the *historical distribution*, $Q_{\mathrm{h},U}$, parameterised by *draw* (E,z), is the cardinality of *histories drawn without replacement* having *histogram* A,

$$Q_{\mathrm{h},U}(E,z)(A) \ = \ |\{H : H \subseteq H_E,\ \mathrm{his}(H) = A\}|$$

The *historical probability distribution* is normalised,

$$\hat{Q}_{\mathrm{h},U}(E,z)(A) \ := \ 1/\binom{z_E}{z} \times Q_{\mathrm{h},U}(E,z)(A)$$

The *likely history probability function*, $\tilde{P}$, can be re-written in terms of the *historical distribution*,

$$\begin{aligned}
\tilde{P} &= \{(H, 1/Q_{\mathrm{h},U}(E,z)(A)) : H \subseteq H_E,\ \mathrm{his}(H) = A\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A\}
\end{aligned}$$

So the *likely history probability function* entropy, $\text{entropy}(\tilde{P})$, is maximised when the *historical distribution frequency*, $Q_{\text{h},U}(E,z)(A)$, is maximised.

Consider the case where the *histogram*, A, is *known*, but the *distribution histogram*, E, is *unknown* and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *historical probability distribution* is a *probability function*, $\hat{Q}_{\text{h},U}(E,z) \in \mathcal{P}$, parameterised by the *distribution histogram*, E, so there is a corresponding *likelihood function* $L_{\text{h},U}(A) \in \mathcal{A}_{U,\text{i},V,z_E} \to \mathbf{Q}_{\geq 0}$ such that $L_{\text{h},U}(A)(E) = \hat{Q}_{\text{h},U}(E,z)(A)$. The *maximum likelihood estimate* $\tilde{E}$ for the *distribution histogram*, E, is a modal value of this *likelihood function*,

$$\begin{aligned}\tilde{E} &\in \text{maxd}(L_{\text{h},U}(A)) \\ &= \text{maxd}(\{(D, Q_{\text{h},U}(D,z)(A)) : D \in \mathcal{A}_{U,\text{i},V,z_E}\})\end{aligned}$$

The *likely distribution histogram*, $\tilde{E}$, is *known* if the *distribution histogram size*, z_E, is *known* and the *histogram*, A, is *known*. If it is assumed that the *distribution histogram* equals the *likely distribution histogram*, $E = \tilde{E}$, then the *likely history probability* is *known*, $\tilde{P}(H) = 1/Q_{\text{h},U}(\tilde{E},z)(A)$ where $\text{his}(H) = A$.

The *multinomial distribution* $Q_{\text{m},U}$ is defined

$$Q_{\text{m},U}(E,z)(A) \;:=\; \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!} \prod_{S \in A^{\text{S}}} E_S^{A_S}$$

where $A^{\text{F}} \leq E^{\text{F}}$. The *frequency* of *histogram* A in the *multinomial distribution*, $Q_{\text{m},U}$, parameterised by *draw* (E,z), is the cardinality of *histories drawn with replacement* having *histogram* A,

$$Q_{\text{m},U}(E,z)(A) \;=\; |\{L : L \in H_E^z,\ \text{his}(\{((i,x),S) : (i,(x,S)) \in L\}) = A\}|$$

where $H_E^z \in \mathcal{L}(H_E)$ is the set of lists of the *distribution history events* of length z.

The *multinomial probability distribution* is normalised,

$$\begin{aligned}\hat{Q}_{\text{m},U}(E,z)(A) &:= \frac{1}{z_E^z} \times Q_{\text{m},U}(E,z)(A) \\ &= \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!} \prod_{S \in A^{\text{S}}} \hat{E}_S^{A_S}\end{aligned}$$

so the *multinomial probability*, $\hat{Q}_{\text{m},U}(E,z)(A) = \hat{Q}_{\text{m},U}(\hat{E},z)(A)$, does not depend on the *distribution histogram size*, z_E.

As the *distribution histogram size*, z_E, tends to infinity, the *historical probability* tends to the *multinomial probability*. That is, for large *distribution histogram size*, $z_E \gg z$, the *historical probability* may be approximated by the *multinomial probability*, $\hat{Q}_{\mathrm{h},U}(E,z)(A) \approx \hat{Q}_{\mathrm{m},U}(E,z)(A)$.

In the case where the *distribution histogram* is *known* to be *cartesian*, $E = V^{\mathrm{C}}_{z_E}$, but the *distribution histogram size*, z_E, is *unknown*, except that it is *known* to be large, $z_E \gg z$, then the case where the *drawn histogram*, A, is *known* to be *necessary*, $\sum(P(H) : H \subseteq H_E,\ \mathrm{his}(H) = A) = 1$, approximates to the case where the *substrate histogram*, A, is *known* to be *necessary*, $\sum(P(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) = 1$. That is,

$$\begin{aligned}
\tilde{P} \ &= \ \{(H, 1/\prod_{S \in A^{\mathrm{S}}} \binom{V^{\mathrm{C}}_{z_E}(S)}{A(S)}) : H \subseteq H_E,\ \mathrm{his}(H) = A\} \ \cup \\
&\qquad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\} \ \cup \\
&\qquad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A\} \\
&\approx \ \{(H, 1/\frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!}) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A\} \ \cup \\
&\qquad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A\}
\end{aligned}$$

In this case, the *likely history probability function* entropy varies with the *histogram entropy*, $\mathrm{entropy}(\tilde{P}) \sim \mathrm{entropy}(A)$.

In the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $\tilde{E}$ for the *distribution probability histogram*, $\hat{E}$, may be approximated by a modal value of a *likelihood function* which depends on the *multinomial distribution* instead,

$$\tilde{E} \ \in \ \mathrm{maxd}(\{(D, Q_{\mathrm{m},U}(D,z)(A)) : D \in \mathcal{A}_{U,V,1}\})$$

The *mean* of the *multinomial probability distribution* is the *sized distribution histogram*,

$$\mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E,z)) \ = \ \mathrm{scalar}(z) * \hat{E}$$

so the *maximum likelihood estimate*, $\tilde{E}$, for the *distribution probability histogram*, $\hat{E}$, is the *sample probability histogram*, $\hat{A}$,

$$\tilde{E} \ = \ \hat{A}$$

If it is assumed that the *distribution probability histogram* equals the *likely distribution probability histogram*, $\hat{E} = \tilde{E} = \hat{A}$, then the *likely history probability* varies against the *sample-distributed multinomial probability*, $\tilde{P}(H) \sim 1/\hat{Q}_{\mathrm{m},U}(\hat{A}, z)(A)$.

The *sample-distributed multinomial log-likelihood* is

$$\ln \hat{Q}_{\mathrm{m},U}(A, z)(A) = \ln z! - z \ln z - \sum_{S \in A^{\mathrm{S}}} \ln A_S! + \sum_{S \in A^{\mathrm{FS}}} A_S \ln A_S$$

which varies against the sum of the logarithms of the *counts*

$$\ln \hat{Q}_{\mathrm{m},U}(A, z)(A) \quad \sim \quad - \sum_{S \in A^{\mathrm{FS}}} \ln A_S$$

So the *log-likelihood* varies weakly against the *histogram entropy*,

$$\ln \hat{Q}_{\mathrm{m},U}(A, z)(A) \quad \sim \quad - \operatorname{entropy}(A)$$

If it is assumed that the *distribution probability histogram* equals the *likely distribution probability histogram*, $\hat{E} = \tilde{E} = \hat{A}$, then the *likely history probability function* entropy varies against the *histogram entropy*, $\operatorname{entropy}(\tilde{P}) \sim - \operatorname{entropy}(A)$, in contrast to the case where the *distribution histogram* is *cartesian*.

The *Fisher information* of a *probability function* varies with the negative curvature of the *likelihood function* near the *maximum likelihood estimate* of the parameter. So the *Fisher information* is a measure of the sensitivity of the *likelihood function* with respect to the *maximum likelihood estimate*. The *Fisher information* of the *multinomial probability distribution*, $\hat{Q}_{\mathrm{m},U}(E, z)$, is the *sum sensitivity*

$$\operatorname{sum}(\operatorname{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E, z))) \quad = \quad \sum_{S \in V^{\mathrm{CS}}} \frac{z}{\hat{E}_S(1 - \hat{E}_S)}$$

The *sum sensitivity* varies against the *sized entropy*,

$$\operatorname{sum}(\operatorname{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E, z))) \quad \sim \quad -z \times \operatorname{entropy}(E)$$

So, in the case of *sample-distributed multinomial probability distribution*, $\hat{Q}_{\mathrm{m},U}(A, z)$, the *sum sensitivity* varies weakly with the *log-likelihood*,

$$\begin{aligned} \operatorname{sum}(\operatorname{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A, z))) \quad &\sim \quad -z \times \operatorname{entropy}(A) \\ &\sim \quad \ln \hat{Q}_{\mathrm{m},U}(A, z)(A) \end{aligned}$$

If it is assumed that the *distribution probability histogram* equals the *likely distribution probability histogram*, $\hat{E} = \tilde{E} = \hat{A}$, then, as the *likely history probability function* entropy, $\text{entropy}(\tilde{P})$, increases, the *sensitivity* to the *distribution histogram*, $\tilde{E}$, increases.

The lower the *entropy* of the *sample* the more *likely* the *normalised sample histogram*, $\hat{A}$, equals the *normalised distribution histogram*, $\hat{E}$, but the larger the *likely* difference between them if they are not equal.

Now consider the case where either the *drawn histogram* A or the *drawn histogram* B are *known* to be *necessary*, $\sum(P(H) : H \subseteq H_E,\ (\text{his}(H) = A \lor \text{his}(H) = B)) = 1$. The *maximum likelihood estimate* which maximises the entropy, $\text{entropy}(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} &= \{(H,1) : H \subseteq H_E,\ (\text{his}(H) = A \ \lor\ \text{his}(H) = B)\}^{\wedge} \ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\} \ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \text{his}(G) \neq A,\ \text{his}(G) \neq B\} \\
&= \{(H, 1/(Q_{\text{h},U}(E,z)(A) + Q_{\text{h},U}(E,z)(B))) : \\
&\qquad\qquad H \subseteq H_E,\ (\text{his}(H) = A \ \lor\ \text{his}(H) = B)\} \ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\} \ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \text{his}(G) \neq A,\ \text{his}(G) \neq B\}
\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn histories* $H \subseteq H_E$ with either *histogram*, A or B, are uniformly probable and all other *histories*, $G \not\subseteq H_E$ or $\text{his}(G) \neq A$ and $\text{his}(G) \neq B$, are impossible, $\tilde{P}(G) = 0$. If the *histograms*, A and B, are *known* and the *distribution histogram*, H_E, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

The *likely probability* of *drawing histogram* A from *necessary drawn histograms* A or B is

$$\sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \text{his}(H) = A) \ = \ \frac{Q_{\text{h},U}(E,z)(A)}{Q_{\text{h},U}(E,z)(A) + Q_{\text{h},U}(E,z)(B)}$$

The *likely history probability function* entropy, $\text{entropy}(\tilde{P})$, is maximised when the sum of the *historical frequencies*, $Q_{\text{h},U}(E,z)(A) + Q_{\text{h},U}(E,z)(B)$, is maximised.

Consider the case where the *drawn histograms*, A and B, are *known*, but the *distribution histogram*, E, is *unknown* and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *maximum likelihood estimate* $\tilde{E}$ for the *distribution histogram*, E, is a modal value of the *likelihood function*,

$$\tilde{E} \in \mathrm{maxd}(\{(D, Q_{\mathrm{h},U}(D,z)(A) + Q_{\mathrm{h},U}(D,z)(B)) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\})$$

The *likely distribution histogram*, $\tilde{E}$, is *known* if the *distribution histogram size*, z_E, is *known* and the *drawn histograms*, A and B, are *known*. If it is assumed that the *distribution histogram* equals the *likely distribution histogram*, $E = \tilde{E}$, then the *likely history probability* is *known*, $\tilde{P}(H) = 1/(Q_{\mathrm{h},U}(\tilde{E},z)(A) + Q_{\mathrm{h},U}(\tilde{E},z)(B))$ where $\mathrm{his}(H) = A$ or $\mathrm{his}(H) = B$.

In the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $\tilde{E}$ for the *distribution probability histogram*, $\hat{E}$, may be approximated by a modal value of a *likelihood function* which depends on the *multinomial distribution* instead,

$$\tilde{E} \in \mathrm{maxd}(\{(D, Q_{\mathrm{m},U}(D,z)(A) + Q_{\mathrm{m},U}(D,z)(B)) : D \in \mathcal{A}_{U,V,1}\})$$

Now the *likely distribution histogram*, $\tilde{E}$, is *known* if there is a computable solution and the *drawn histograms*, A and B, are *known*.

Consider the case where the *histogram* is *uniformly possible*. Instead of assuming the *substrate history probability function* $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ to be the distribution of an arbitrary *history* valued function of undefined particle, $\mathcal{X} \to \mathcal{H}$, assume that it is the distribution of an arbitrary *history* valued function, $\mathcal{X} \to \mathcal{H}$, given an arbitrary *histogram* valued function, $\mathcal{X} \to \mathcal{A}$. In this case, the *history* valued function is chosen arbitrarily from the constrained subset

$$\big\{\{((x,A,y),H) : (x,(A,G)) \in F,\ (y,H) \in G,\ \mathrm{his}(H) = A\} : F \in \mathcal{X} \to (\mathcal{A} \times (\mathcal{X} \to \mathcal{H}))\big\} \subset \mathcal{X} \to \mathcal{H}$$

In the case where there is no *distribution history*, the *maximum likelihood estimate* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}\tilde{P} &= \Big(\bigcup\{\{(H,1) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A\}^\wedge : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}\Big)^\wedge \\ &= \{(H, 1/|\mathcal{A}_{U,\mathrm{i},V,z}| \times 1/\frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!}) : H \in \mathcal{H}_{U,V,z},\ A = \mathrm{his}(H)\}\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *histograms* are uniformly probable, $\forall A \in \mathcal{A}_{U,\mathrm{i},V,z}\ (\sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) =$

$1/|\mathcal{A}_{U,\mathrm{i},V,z}|$), and then all *histories* with the same *histogram*, $\mathrm{his}(H) = A$, are uniformly probable. The *likely probability function*, $\tilde{P}$, is *known*.

In the case where there is a *distribution history* H_E, the *maximum likelihood estimate* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} &= \Big(\bigcup\{\{(H,1) : H \subseteq H_E,\ \mathrm{his}(H) = A\}^{\wedge} : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}\Big)^{\wedge} \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\} \\
&= \Big(\bigcup\{\{(H, 1/Q_{\mathrm{h},U}(E,z)(A)) : H \subseteq H_E,\ \mathrm{his}(H) = A\} : \\
&\qquad\qquad A \in \mathcal{A}_{U,\mathrm{i},V,z}\}\Big)^{\wedge} \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\}
\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn histograms*, $A \leq E$, are uniformly probable, and then all *drawn histories* $H \subseteq H_E$ with the same *histogram*, $\mathrm{his}(H) = A$, are uniformly probable. If the *distribution histogram*, H_E, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

Consider the case where a *drawn sample* A is *known*, but the *distribution histogram*, E, is *unknown* and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *maximum likelihood estimate* $\tilde{E}$ for the *distribution histogram*, E, is the same as for *necessary histogram*,

$$\tilde{E} \in \mathrm{maxd}(\{(D, Q_{\mathrm{h},U}(D,z)(A)) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\})$$

The *likely distribution histogram*, $\tilde{E}$, is *known* if the *distribution histogram size*, z_E, is *known* and the *histogram*, A, is *known*. If it is assumed that the *distribution histogram* equals the *likely distribution histogram*, $E = \tilde{E}$, then the *likely history probability* is *known*, $\tilde{P}(H) = 1/|\{A : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq \tilde{E}\}| \times 1/Q_{\mathrm{h},U}(\tilde{E},z)(A)$ where $\mathrm{his}(H) = A$.

In the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $\tilde{E}$ for the *distribution probability histogram*, $\hat{E}$, may be approximated by a modal value of a *likelihood function* which depends on the *multinomial distribution* instead,

$$\tilde{E} \in \mathrm{maxd}(\{(D, Q_{\mathrm{m},U}(D,z)(A)) : D \in \mathcal{A}_{U,V,1}\})$$

Again, the *maximum likelihood estimate*, $\tilde{E}$, for the *distribution probability histogram*, $\hat{E}$, is the *sample probability histogram*, $\hat{A}$,

$$\tilde{E} \;=\; \hat{A}$$

If it is assumed that the *distribution probability histogram* equals the *likely distribution probability histogram*, $\hat{E} = \tilde{E} = \hat{A}$, then the *likely history probability* varies against the *sample-distributed multinomial probability*, $\tilde{P}(H) \sim 1/|\mathcal{A}_{U,\mathrm{i},V,z}| \times 1/\hat{Q}_{\mathrm{m},U}(\hat{A},z)(A)$.

So the properties of *uniform possible histogram* are similar to *necessary histogram* except that more *histories* are possible but less probable.

2.4.2 Aligned induction

In *aligned induction* the *history probability functions* are constrained by *independent histogram*.

The *independent histogram* valued function of *integral substrate histograms* $Y_{U,\mathrm{i},V,z}$ is defined

$$Y_{U,\mathrm{i},V,z} \;:=\; \{(A, A^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The finite set of *iso-independents* of *independent histogram* A^{X} is

$$Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \;=\; \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} = A^{\mathrm{X}}\}$$

Given any subset of the *integral substrate histograms* $I \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$ that contains the *histogram*, $A \in I$, the degree to which the subset is said to be *aligned-like* is called the *iso-independence*. The *iso-independence* is defined as the ratio of (i) the cardinality of the intersection between the *integral substrate histograms* subset and the set of *integral iso-independents*, and (ii) the cardinality of the union,

$$\frac{1}{|\mathcal{A}_{U,\mathrm{i},V,z}|} \;\leq\; \frac{|I \;\cap\; Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|I \;\cup\; Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} \;\leq\; 1$$

Consider the case where the *independent* A^{X} of *drawn histories* is *known* to be *necessary*, $\sum(P(H) : H \subseteq H_E,\ \mathrm{his}(H)^{\mathrm{X}} = A^{\mathrm{X}}) = 1$. The *maximum* *likelihood estimate* which maximises the entropy, $\text{entropy}(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} &= \{(H,1) : H \subseteq H_E,\ \text{his}(H)^{\mathrm{X}} = A^{\mathrm{X}}\}^{\wedge}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \text{his}(G)^{\mathrm{X}} \neq A^{\mathrm{X}}\} \\
&= \{(H, 1/\textstyle\sum(Q_{\mathrm{h},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))) : \\
&\qquad\qquad\qquad\qquad H \subseteq H_E,\ \text{his}(H)^{\mathrm{X}} = A^{\mathrm{X}}\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\}\ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \text{his}(G)^{\mathrm{X}} \neq A^{\mathrm{X}}\}
\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn histories* $H \subseteq H_E$ with the *independent*, $\text{his}(H)^{\mathrm{X}} = A^{\mathrm{X}}$, are uniformly probable and all other *histories*, $G \not\subseteq H_E$ or $\text{his}(G)^{\mathrm{X}} \neq A^{\mathrm{X}}$, are impossible, $\tilde{P}(G) = 0$. If the *independent*, A^{X}, is *known* and the *distribution histogram*, H_E, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

The *likely probability* of *drawing histogram* A from *necessary drawn independent* A^{X} is

$$\sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \text{his}(H) = A) = \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum Q_{\mathrm{h},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}$$

The *likely history probability function* entropy, $\text{entropy}(\tilde{P})$, is maximised when the sum of the *iso-independent historical frequencies*, $\sum Q_{\mathrm{h},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, is maximised.

Consider the case where the *independent*, A^{X}, is *known*, but the *distribution histogram*, E, is *unknown* and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *maximum likelihood estimate* $\tilde{E}$ for the *distribution histogram*, E, is a modal value of the *likelihood function*,

$$\tilde{E} \in \text{maxd}(\{(D, \textstyle\sum(Q_{\mathrm{h},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\})$$

The *likely distribution histogram*, $\tilde{E}$, is *known* if the *distribution histogram size*, z_E, is *known* and the *independent*, A^{X}, is *known*. If it is assumed that the *distribution histogram* equals the *likely distribution histogram*, $E = \tilde{E}$, then the *likely history probability* is *known*, $\tilde{P}(H) = 1/\sum(Q_{\mathrm{h},U}(\tilde{E},z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))$ where $\text{his}(H)^{\mathrm{X}} = A^{\mathrm{X}}$.

In the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $\tilde{E}$ for the *distribution probability histogram*, $\hat{E}$, may be approximated by a modal value of a *likelihood function* which depends on the *multinomial distribution* instead,

$$\tilde{E} \in \text{maxd}(\{(D, \sum(Q_{\text{m},U}(D,z)(B) : B \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}}))) : D \in \mathcal{A}_{U,V,1}\})$$

which has a solution $\tilde{E} = \hat{A}^{\text{X}}$. So the *maximum likelihood estimate*, $\tilde{E}$, for the *distribution probability histogram*, $\hat{E}$, is the *independent probability histogram*, $\hat{A}^{\text{X}}$,

$$\tilde{E} = \hat{A}^{\text{X}}$$

In the case where the *independent* is *integral*, $A^{\text{X}} \in \mathcal{A}_{\text{i}}$, the sum of the *iso-independent independent-distributed multinomial probabilities* varies with the *independent independent-distributed multinomial probability*,

$$\sum(Q_{\text{m},U}(A^{\text{X}},z)(B) : B \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})) \sim Q_{\text{m},U}(A^{\text{X}},z)(A^{\text{X}})$$

So, if it is assumed that the *distribution probability histogram* equals the *likely distribution probability histogram*, $\hat{E} = \tilde{E} = \hat{A}^{\text{X}}$, then the *likely history probability* varies against the *independent-distributed multinomial probability* of the *independent*, $\tilde{P}(H) \sim 1/\hat{Q}_{\text{m},U}(A^{\text{X}},z)(A^{\text{X}})$.

In this case, the *likely probability* of *drawing histogram* A from *necessary drawn independent* A^{X} is approximately

$$\begin{aligned}
&\sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \text{his}(H) = A) \\
&\qquad\approx \frac{Q_{\text{m},U}(A^{\text{X}},z)(A)}{\sum Q_{\text{m},U}(A^{\text{X}},z)(B) : B \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})} \\
&\qquad\sim \frac{Q_{\text{m},U}(A^{\text{X}},z)(A)}{Q_{\text{m},U}(A^{\text{X}},z)(A^{\text{X}})}
\end{aligned}$$

The negative logarithm of the ratio of the *histogram independent-distributed multinomial probability* to the *independent independent-distributed multinomial probability* equals the *alignment*,

$$-\ln \frac{Q_{\text{m},U}(A^{\text{X}},z)(A)}{Q_{\text{m},U}(A^{\text{X}},z)(A^{\text{X}})} = \text{algn}(A)$$

So the logarithm of the *likely probability* of *drawing histogram* A from *necessary drawn independent* A^{X} varies against the *alignment*,

$$\ln \sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) \quad \sim \quad -\ \mathrm{algn}(A)$$

The *independent*, A^{X}, which has zero *alignment*, $\mathrm{algn}(A^{\mathrm{X}}) = 0$, is the most *probable histogram*, $\forall B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})\ (Q_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A^{\mathrm{X}}) \geq Q_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B))$. As the *alignment* increases, $\mathrm{algn}(A) > 0$, the *likely histogram probability*, $Q_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A)/\sum(Q_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))$, decreases.

The *likely history probability function* entropy varies with the *independent entropy*, $\mathrm{entropy}(\tilde{P}) \sim \mathrm{entropy}(A^{\mathrm{X}})$.

Define the *dependent histogram* $A^{\mathrm{Y}} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram* A conditional that it is an *iso-independent*,

$$\{A^{\mathrm{Y}}\} \quad = \quad \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D, z)(A)}{\sum Q_{\mathrm{m},U}(D, z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}) : D \in \mathcal{A}_{U,V,z}\})$$

Note that the *dependent*, A^{Y}, is not always computable, but an approximation to any accuracy can be made to it. In the case where the *histogram* is *independent*, the *dependent* equals the *independent*, $A = A^{\mathrm{X}} \implies A^{\mathrm{Y}} = A = A^{\mathrm{X}}$. The *dependent alignment* is greater than or equal to the *histogram alignment*, $\mathrm{algn}(A^{\mathrm{Y}}) \geq \mathrm{algn}(A) \geq \mathrm{algn}(A^{\mathrm{X}}) = 0$. In the case where the *histogram* is *uniformly diagonalised*, the *histogram alignment*, $\mathrm{algn}(A)$, is at the maximum, and the *dependent* equals the *histogram*, $A^{\mathrm{Y}} = A$.

Now consider the case where, given *necessary drawn independent* A^{X}, it is *known*, in addition, that the *sample histogram* A is the most *probable histogram*, regardless of its *alignment*. That is, the *likely probability* of *drawing histogram* A from *necessary drawn independent* A^{X},

$$\sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) \ = \ \frac{Q_{\mathrm{h},U}(E, z)(A)}{\sum Q_{\mathrm{h},U}(E, z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}$$

is maximised.

In the case where the *sample*, A, is *known*, but the *distribution histogram*,

E, is *unknown*, the *maximum likelihood estimate* $\tilde{E}$ for the *distribution histogram*, E, is a modal value of the *likelihood function*,

$$\tilde{E} \in \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{h},U}(D,z)(A)}{\sum Q_{\mathrm{h},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\})$$

The *likely distribution histogram*, $\tilde{E}$, is *known* if the *distribution histogram size*, z_E, is *known* and the *sample*, A, is *known*. If it is assumed that the *distribution histogram* equals the *likely distribution histogram*, $E = \tilde{E}$, then the *likely history probability* is *known*, $\tilde{P}(H) = 1/\sum(Q_{\mathrm{h},U}(\tilde{E},z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))$ where $\mathrm{his}(H)^{\mathrm{X}} = A^{\mathrm{X}}$.

If the *histogram* is *independent*, $A = A^{\mathrm{X}}$, then the additional constraint of *probable sample* makes no change to the *maximum likelihood estimate*, $\tilde{E}$,

$$\begin{aligned} A = A^{\mathrm{X}} \implies & \\ & \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{h},U}(D,z)(A)}{\sum Q_{\mathrm{h},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\}) \\ = & \ \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{h},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\}) \end{aligned}$$

If the *histogram* is not *independent*, $\mathrm{algn}(A) > 0$, however, then the *likely history probability function* entropy, $\mathrm{entropy}(\tilde{P})$, is lower than it is in the case of *necessary independent* unconstrained by *probable sample*.

In the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $\tilde{E}$ for the *distribution probability histogram*, $\hat{E}$, is now approximated by a modal value of the conditional *likelihood function*,

$$\tilde{E} \in \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}) : D \in \mathcal{A}_{U,V,1}\})$$

The solution to this is the *normalised dependent*, $\tilde{E} = \hat{A}^{\mathrm{Y}}$. The *maximum likelihood estimate* is near the *sample*, $\tilde{E} \sim \hat{A}$, only in as much as it is far from the *independent*, $\tilde{E} \nsim \hat{A}^{\mathrm{X}}$. This may be compared to the case unconstrained by *probable sample* where the *maximum likelihood estimate* equals the *independent*, $\tilde{E} = \hat{A}^{\mathrm{X}}$. In the *probable sample* case the *sized maximum likelihood estimate* is *aligned*, $\mathrm{algn}(A^{\mathrm{Y}}) > 0$, so there are fewer ways to *draw* the *iso-independents* and the *likely history probability function* entropy, $\mathrm{entropy}(\tilde{P})$, is lower. At maximum *alignment*, where the *histogram* is *uniformly diagonalised*, the *dependent* equals the *histogram*, $A^{\mathrm{Y}} = A$, and the *likely history* *probability function* entropy, $\mathrm{entropy}(\tilde{P})$, is least.

The *iso-independent conditional multinomial probability distribution* is defined,

$$\hat{Q}_{\mathrm{m,y},U}(E,z)(A) \quad := \quad \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|} \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum Q_{\mathrm{m},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}$$

So the optimisation can be rewritten,

$$\tilde{E} \;\; \in \;\; \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{m,y},U}(D,z)(A)) : D \in \mathcal{A}_{U,V,1}\})$$

The logarithm of the *independent-distributed iso-independent conditional multinomial probability* varies against the *alignment*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{X}},z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \quad \sim \quad -\ \mathrm{algn}(A)$$

Conversely, the logarithm of the *dependent-distributed iso-independent conditional multinomial probability* varies with the *alignment*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}},z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \quad \sim \quad \mathrm{algn}(A)$$

That is, the *log-likelihood* varies with the *sample alignment*,

$$\ln \hat{Q}_{\mathrm{m,y},U}(A^{\mathrm{Y}},z)(A) \quad \sim \quad \mathrm{algn}(A)$$

In the case where the *alignment* is low the *sum sensitivity* varies with the *alignment*

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,y},U}(A^{\mathrm{Y}},z))) \quad \sim \quad \mathrm{algn}(A)$$

and in the case where the *alignment* is high the *sum sensitivity* varies against the *alignment*

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,y},U}(A^{\mathrm{Y}},z))) \quad \sim \quad -\ \mathrm{algn}(A)$$

At intermediate *alignments* the *sum sensitivity* is independent of the *alignment.*

So, in the *probable sample* case, if it is assumed that the *distribution probability histogram* equals the *likely distribution probability histogram*, $\hat{E} = \tilde{E} = \hat{A}^{\mathrm{Y}}$, then the *likely history probability function* entropy varies against the *alignment*, $\mathrm{entropy}(\tilde{P}) \sim -\ \mathrm{algn}(A)$.

As the *alignment*, $\mathrm{algn}(A)$, increases towards its maximum, the *likely distribution probability histogram* tends to the *histogram*, $\tilde{E} = \hat{A}^{\mathrm{Y}} \sim \hat{A}$, and the *log-likelihood*, $\ln \hat{Q}_{\mathrm{m,y},U}(A^{\mathrm{Y}}, z)(A)$, increases, but the *sensitivity* to *distribution histogram*, E, decreases. In other words, the more *aligned* the *sample* the more *likely* the *normalised sample histogram*, $\hat{A}$, equals the *normalised distribution histogram*, $\hat{E}$, and the smaller the *likely* difference between them if they are not equal.

Consider the case where the *independent* is *uniformly possible*. Assume that the *substrate history probability function* $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is the distribution of an arbitrary *history* valued function, $\mathcal{X} \to \mathcal{H}$, given an arbitrary *independent* valued function, $\mathcal{X} \to \mathcal{A}$. In this case, the *history* valued function is chosen arbitrarily from the constrained subset

$$\big\{\{((x,A,y),H) : (x,(A,G)) \in F,\ (y,H) \in G,\ \mathrm{his}(H)^{\mathrm{X}} = A\} \ : \\ F \in \mathcal{X} \to (\mathcal{A} \times (\mathcal{X} \to \mathcal{H}))\big\} \ \subset\ \mathcal{X} \to \mathcal{H}$$

Uniformly possible independent is a weaker constraint than *uniformly possible histogram*, so the subset of *history* valued functions is larger.

In the case where there is a *distribution history* H_E, the *maximum likelihood estimate* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} \ &=\ \Big(\bigcup\big\{\{(H,1) : H \subseteq H_E,\ \mathrm{his}(H)^{\mathrm{X}} = A\}^{\wedge} : A \in \mathrm{ran}(Y_{U,\mathrm{i},V,z})\big\}\Big)^{\wedge} \ \cup \\
&\quad\ \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\} \\
&=\ \Big(\bigcup\big\{\{(H, 1/\textstyle\sum(Q_{\mathrm{h},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))) : \\
&\qquad\qquad H \subseteq H_E,\ \mathrm{his}(H)^{\mathrm{X}} = A\} \ :\ A \in \mathrm{ran}(Y_{U,\mathrm{i},V,z})\big\}\Big)^{\wedge} \ \cup \\
&\quad\ \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\}
\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn independents* are uniformly probable, and then all *drawn histories* $H \subseteq H_E$ with the same *independent*, $\mathrm{his}(H)^{\mathrm{X}} = A$, are uniformly probable. If the *distribution histogram*, H_E, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

The properties of *uniformly possible independent* are the same as for *necessary independent*, except that the probabilities are scaled. So, in the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *likely history probability* varies against the *independent-distributed multinomial probability* of the *independent*,

$$\tilde{P}(H) \quad \sim \quad 1/|\mathrm{ran}(Y_{U,\mathrm{i},V,z})| \times 1/\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A^{\mathrm{X}})$$

That is, more *histories* are possible but less probable.

2.5 Models

2.5.1 Transforms

Transforms are the simplest *models*. All *models* can be converted to *transforms*.

Given a *histogram* $X \in \mathcal{A}$ and a subset of its *variables* $W \subseteq \mathrm{vars}(X)$, the pair $T = (X, W)$ forms a *transform*. The *variables*, W, are the *derived variables*. The complement $V = \mathrm{vars}(X) \setminus W$ are the *underlying variables*. The set of all *transforms* is

$$\mathcal{T} \quad := \quad \{(X, W) : X \in \mathcal{A},\ W \subseteq \mathrm{vars}(X)\}$$

The *transform histogram* is $X = \mathrm{his}(T)$. The *transform derived* is $W = \mathrm{der}(T)$. The *transform underlying* is $V = \mathrm{und}(T)$. The set of *underlying variables* of a *transform* is also called the *substrate*.

The *null transform* is $(X, \emptyset)$. The *full transform* is $(X, \mathrm{vars}(X))$.

Given a *histogram* $A \in \mathcal{A}$, the *multiplication* of the *histogram*, A, by the *transform* $T \in \mathcal{T}$ equals the *multiplication* of the *histogram*, A, by the *transform histogram* $X = \mathrm{his}(T)$ followed by the *reduction* to the *derived variables* $W = \mathrm{der}(T)$,

$$A * T \ = \ A * (X, W) \quad := \quad A * X \ \% \ W$$

If the *histogram variables* are a superset of the *underlying variables*, $\mathrm{vars}(A) \supseteq \mathrm{und}(T)$, then the *histogram*, A, is called the *underlying histogram* and the *multiplication*, $A * T$, is called the *derived histogram*. The *derived histogram variables* equals the *derived variables*, $\mathrm{vars}(A * T) = \mathrm{der}(T)$.

The application of the *null transform* of the *cartesian* is the *scalar*, $A * (V^{\mathrm{C}}, \emptyset) = A\%\emptyset = \mathrm{scalar}(\mathrm{size}(A))$, where $V = \mathrm{vars}(A)$. The application of the *full transform* of the *cartesian* is the *histogram*, $A * (V^{\mathrm{C}}, V) = A\%V = A$.

Given a *histogram* $A \in \mathcal{A}$ and a *transform* $T \in \mathcal{T}$, the *formal histogram* is defined as the *independent derived*, $A^{\mathrm{X}} * T$. The *abstract histogram* is defined as the *derived independent*, $(A * T)^{\mathrm{X}}$.

In the case where the *formal* and *abstract* are equal, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, the *abstract* equals the *independent abstract*, $(A*T)^{\mathrm{X}} = A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, and so only depends on the *independent*, A^{X}, not on the *histogram*, A. The *formal* equals the *formal independent*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, and so is itself *independent*.

A *transform* $T \in \mathcal{T}$ is *functional* if there is a *causal* relation between the *underlying variables* $V = \mathrm{und}(T)$ and the *derived variables* $W = \mathrm{der}(T)$,

$$\mathrm{split}(V, X^{\mathrm{FS}}) \;\; \in \;\; V^{\mathrm{CS}} \to W^{\mathrm{CS}}$$

where $X = \mathrm{his}(T)$. The set of *functional transforms* $\mathcal{T}_{\mathrm{f}} \subset \mathcal{T}$ is the subset of all *transforms* that are *causal*.

A *functional transform* $T \in \mathcal{T}_{\mathrm{f}}$ has an *inverse*,

$$T^{-1} \;\; := \;\; \{((S\%V, c), S\%W) : (S, c) \in X\}^{-1}$$

A *transform* T is *one functional* in *system* U if the *reduction* of the *transform histogram* to the *underlying variables* equals the *cartesian histogram*, $X\%V = V^{\mathrm{C}}$. So the *causal* relation is a *derived state* valued left total function of *underlying state*, $\mathrm{split}(V, X^{\mathrm{S}}) \in V^{\mathrm{CS}} :\to W^{\mathrm{CS}}$. The set of *one functional transforms* $\mathcal{T}_{U,\mathrm{f},1} \subset \mathcal{T}_{\mathrm{f}}$ is

$$\begin{aligned} \mathcal{T}_{U,\mathrm{f},1} \;\; = \;\; & \{(\{(S \cup R, 1) : (S, R) \in Q\}, W) : \\ & \qquad V, W \subseteq \mathrm{vars}(U),\ V \cap W = \emptyset,\ Q \in V^{\mathrm{CS}} :\to W^{\mathrm{CS}}\} \end{aligned}$$

The application of a *one functional transform* to an *underlying histogram* preserves the *size*, $\mathrm{size}(A * T) = \mathrm{size}(A)$.

The *one functional transform inverse* is a *unit component* valued function of *derived state*, $T^{-1} \in W^{\mathrm{CS}} \to \mathrm{P}(V^{\mathrm{C}})$. That is, the range of the *inverse* corresponds to a *partition* of the *cartesian states* into *components*, $\mathrm{ran}(T^{-1}) \in \mathrm{B}(V^{\mathrm{C}})$.

The application of a *one functional transform* T to its *underlying cartesian* V^{C} is the *component cardinality histogram*, $V^{\mathrm{C}} * T = \{(R, |C|) : (R, C) \in T^{-1}\}$. The *effective cartesian derived volume* is less than or equal to the *derived volume*, $|(V^{\mathrm{C}} * T)^{\mathrm{F}}| = |T^{-1}| \leq |W^{\mathrm{C}}|$.

A *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ may be applied to a *history* $H \in \mathcal{H}$ in the *underlying variables* of the *transform*, $\mathrm{vars}(H) = \mathrm{und}(T)$, to construct a *derived history*,

$$H * T \ := \ \{(x,R) : (x,S) \in H,\ \{R\} = (\{S\}^{\mathrm{U}} * T)^{\mathrm{FS}}\}$$

The *size* is unchanged, $|H*T| = |H|$, and the *event identifiers* are conserved, $\mathrm{dom}(H * T) = \mathrm{dom}(H)$.

Given a *partition* $P \in \mathrm{B}(V^{\mathrm{CS}})$ of the *cartesian states* of *variables* V, a *one functional transform* can be constructed. The *partition transform* is

$$P^{\mathrm{T}} \ := \ (\{(S \cup \{(P,C)\},1) : C \in P,\ S \in C\}, \{P\})$$

The set of *derived variables* of the *partition transform* is a singleton of the *partition variable*, $\mathrm{der}(P^{\mathrm{T}}) = \{P\}$. The *derived volume* is the *component* cardinality, $|\{P\}^{\mathrm{C}}| = |P|$. The *underlying variables* are the given *variables*, $\mathrm{und}(P^{\mathrm{T}}) = V$.

The *unary partition transform* is $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$. The *self partition transform* is $T_{\mathrm{s}} = V^{\mathrm{CS}\{\}\mathrm{T}}$.

Given a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$, the *natural converse* is

$$T^{\dagger} \ := \ (X/(X\%W), V)$$

where $(X,W) = T$ and $V = \mathrm{und}(T)$. The *natural converse* may be expressed in terms of *components*,

$$T^{\dagger} \ := \ (\sum_{(R,C) \in T^{-1}} \{R\}^{\mathrm{U}} * \hat{C},\ V)$$

Given a *histogram* $A \in \mathcal{A}$ in the *underlying variables*, $\mathrm{vars}(A) = V$, the *naturalisation* is the application of the *natural converse transform* to the *derived histogram*, $A * T * T^{\dagger}$. The *naturalisation* can be rewritten $A * X \ \% \ W * X \ / \ (X\%W) \ \% \ V$. The *naturalisation* is in the *underlying variables*, $\mathrm{vars}(A * T * T^{\dagger}) = V$. The *size* is conserved, $\mathrm{size}(A * T * T^{\dagger}) = \mathrm{size}(A)$. The *naturalisation derived* equals the *derived*, $A * T * T^{\dagger} * T = A * T$.

The *naturalisation* equals the *sum* of the *scaled components*, $A * T * T^{\dagger} = \sum \mathrm{scalar}((A * T)_R) * \hat{C} \ : \ (R,C) \in T^{-1}$. So each *component* is *uniform*, $\forall (R,C) \in T^{-1} \ (|\mathrm{ran}(A * T * T^{\dagger} * C)| = 1)$.

The *naturalisation* of the *unary partition transform*, $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, is the *sized cartesian*, $A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger} = V_z^{\mathrm{C}}$, where $z = \mathrm{size}(A)$. The *naturalisation* of the *self partition transform*, $T_{\mathrm{s}} = V^{\mathrm{CS}\{\}\mathrm{T}}$, is the *histogram*, $A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger} = A$.

A *histogram* is *natural* when it equals its *naturalisation*, $A = A * T * T^{\dagger}$. The *cartesian* is *natural*, $V^{\mathrm{C}} = V^{\mathrm{C}} * T * T^{\dagger}$.

Given a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ with *underlying variables* $V = \mathrm{und}(T)$, and a *histogram* $A \in \mathcal{A}$ in the same *variables*, $\mathrm{vars}(A) = V$, the *sample converse* is

$$(\hat{A} * X, V)$$

where $X = \mathrm{his}(T)$.

Related to the *sample converse*, the *actual converse* is defined as the *summed normalised* application of the *components* to the *sample histogram*,

$$T^{\odot A} \ := \ (\sum_{(R,C) \in T^{-1}} \{R\}^{\mathrm{U}} * (A * C)^{\wedge},\ V)$$

The application of the *actual converse transform* to the *derived histogram* equals the *histogram*, $A * T * T^{\odot A} = A$.

Given a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ with *underlying variables* $V = \mathrm{und}(T)$, and a *histogram* $A \in \mathcal{A}$ in the same *variables*, $\mathrm{vars}(A) = V$, the *independent converse* is defined as the *summed normalised independent* application of the *components* to the *sample histogram*,

$$T^{\dagger A} \ := \ (\sum_{(R,C) \in T^{-1}} \{R\}^{\mathrm{U}} * (A * C)^{\wedge \mathrm{X}},\ V)$$

The *idealisation* is the application of the *independent converse transform* to the *derived histogram*, $A * T * T^{\dagger A}$. The *idealisation* is in the *underlying variables*, $\mathrm{vars}(A * T * T^{\dagger A}) = V$. The *size* is conserved, $\mathrm{size}(A * T * T^{\dagger A}) = \mathrm{size}(A)$. The *idealisation derived* equals the *derived*, $A * T * T^{\dagger A} * T = A * T$.

The *idealisation* equals the *sum* of the *independent components*, $A * T * T^{\dagger A} = \sum (A * C)^{\mathrm{X}} : (R, C) \in T^{-1}$. So each *component* is *independent*, $\forall (R, C) \in T^{-1}\ (A * T * T^{\dagger A} * C = (A * T * T^{\dagger A} * C)^{\mathrm{X}} = (A * C)^{\mathrm{X}})$.

The *idealisation* of the *unary partition transform*, $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, is the *sized cartesian*, $A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A} = V_z^{\mathrm{C}}$. The *idealisation* of the *self partition transform*, $T_{\mathrm{s}} = V^{\mathrm{CS}\{\}\mathrm{T}}$, is the *histogram*, $A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger A} = A$.

The *idealisation independent* equals the *independent*, $(A * T * T^{\dagger A})^{\mathrm{X}} = A^{\mathrm{X}}$. The *idealisation formal* equals the *formal*, $(A * T * T^{\dagger A})^{\mathrm{X}} * T = A^{\mathrm{X}} * T$. The *idealisation abstract* equals the *abstract*, $(A * T * T^{\dagger A} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$.

A *histogram* is *ideal* when it equals its *idealisation*, $A = A * T * T^{\dagger A}$. The *cartesian* is *ideal*, $V^{\mathrm{C}} = V^{\mathrm{C}} * T * T^{\dagger V^{\mathrm{C}}}$.

The sense in which a *transform* is a simple *model* can be seen by considering queries on a *sample histogram*. Let *histogram* A have a set of *variables* $V = \mathrm{vars}(A)$ which is partitioned into query *variables* $K \subset V$ and label *variables* $V \setminus K$. Let $T = (X, W)$ be a *one functional transform* having *underlying variables* equal to the query *variables*, $\mathrm{und}(T) = K$. Given a query *state* $Q \in K^{\mathrm{CS}}$ that is *ineffective* in the *sample*, $Q \notin (A\%K)^{\mathrm{FS}}$, but is *effective* in the *sample derived*, $R \in (A * T)^{\mathrm{FS}}$ where $\{R\} = (\{Q\}^{\mathrm{U}} * T)^{\mathrm{FS}}$, the *probability histogram* for the label is

$$(\{Q\}^{\mathrm{U}} * T * (\hat{A} * X, V))^{\wedge} \ \% \ (V \setminus K) \quad \in \quad \mathcal{A} \cap \mathcal{P}$$

where the *sample converse transform* is $(\hat{A} * X, V)$. This can be expressed more simply in terms of the *actual converse*,

$$\{Q\}^{\mathrm{U}} * T * T^{\odot A} \ \% \ (V \setminus K) \quad \in \quad \mathcal{A} \cap \mathcal{P}$$

The query of the *sample* via *model* can also be written without the *transforms*, $(\{Q\}^{\mathrm{U}} * X \ \% \ W * X * \hat{A})^{\wedge} \ \% \ (V \setminus K)$. The query *state*, Q, in the query *variables*, K, is raised to the query *derived state*, R, in the *derived variables*, W, then lowered to *effective sample states*, in the *sample variables*, V, and finally *reduced* to label *states*, in the label *variables*, $V \setminus K$. Even though the *sample* itself does not contain the query, $\{Q\}^{\mathrm{U}} * \hat{A} = \emptyset$, the *sample derived* does contain the query *derived*, $\{R\}^{\mathrm{U}} * (\hat{A} * T) \neq \emptyset$, and so the resultant labels are those of the corresponding *effective component*, $(A * C)^{\wedge} \ \% \ (V \setminus K)$, where $(R, C) \in T^{-1}$.

The set of *substrate histories* $\mathcal{H}_{U,V,z}$ is defined above as the set of *histories* having *event identifiers* $\{1 \ldots z\}$, fixed *size* z and fixed *variables* V,

$$\mathcal{H}_{U,V,z} \quad := \quad \{1 \ldots z\} :\to V^{\mathrm{CS}}$$

The corresponding set of *integral substrate histograms* $\mathcal{A}_{U,\mathrm{i},V,z}$ is the set of *complete integral histograms* in *variables* V with *size* z,

$$\mathcal{A}_{U,\mathrm{i},V,z} \ := \ \{A : A \in V^{\mathrm{CS}} :\to \{0 \ldots z\},\ \mathrm{size}(A) = z\}$$

The set of *substrate transforms* $\mathcal{T}_{U,V}$ is the subset of *one functional transforms*, $\mathcal{T}_{U,V} \subset \mathcal{T}_{U,\mathrm{f},1}$, that have *underlying variables* V and *derived variables* which are *partitions*,

$$\mathcal{T}_{U,V} \ = \ \{(\prod_{(X,\cdot) \in F} X,\ \bigcup_{(\cdot,W) \in F} W) : F \subseteq \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\}\}$$

Let v be the *volume* of the *substrate*, $v = |V^{\mathrm{C}}|$. The cardinality of the *substrate transforms set* is $|\mathcal{T}_{U,V}| = 2^{\mathrm{bell}(v)}$, where $\mathrm{bell}(n)$ is Bell's number, which has factorial computation complexity. If the *volume*, v, is finite, the set of *substrate transforms* is finite, $|\mathcal{T}_{U,V}| < \infty$.

2.5.2 Transform entropy

Let T be a *one functional transform*, $T \in \mathcal{T}_{U,\mathrm{f},1}$, having *underlying variables* $V = \mathrm{und}(T)$. Let A be a *histogram*, $A \in \mathcal{A}$, in the *underlying variables*, $\mathrm{vars}(A) = V$, having *size* $z = \mathrm{size}(A) > 0$. The *underlying volume* is $v = |V^{\mathrm{C}}|$.

The *derived entropy* or *component size entropy* is

$$\mathrm{entropy}(A * T) \ := \ -\sum_{(R,\cdot) \in T^{-1}} (\hat{A} * T)_R \times \ln\ (\hat{A} * T)_R$$

The *derived entropy* is positive and less than or equal to the logarithm of the *size*, $0 \leq \mathrm{entropy}(A * T) \leq \ln z$.

Complementary to the *derived entropy* is the *expected component entropy*,

$$\begin{aligned}\mathrm{entropyComponent}(A, T) \ &:= \ \sum_{(R,C) \in T^{-1}} (\hat{A} * T)_R \times \mathrm{entropy}(A * C) \\ &= \ \sum_{(R,\cdot) \in T^{-1}} (\hat{A} * T)_R \times \mathrm{entropy}(\{R\}^{\mathrm{U}} * T^{\odot A})\end{aligned}$$

The *cartesian derived entropy* or *component cardinality entropy* is

$$\mathrm{entropy}(V^{\mathrm{C}} * T) \ := \ -\sum_{(R,\cdot) \in T^{-1}} (\hat{V}^{\mathrm{C}} * T)_R \times \ln\ (\hat{V}^{\mathrm{C}} * T)_R$$

The *cartesian derived entropy* is positive and less than or equal to the logarithm of the *volume*, $0 \leq \text{entropy}(V^{\mathrm{C}} * T) \leq \ln v$.

The *cartesian derived derived sum entropy* or *component size cardinality sum entropy* is

$$\text{entropy}(A * T) + \text{entropy}(V^{\mathrm{C}} * T)$$

The *component size cardinality cross entropy* is the negative *derived histogram expected normalised cartesian derived count logarithm*,

$$\text{entropyCross}(A * T, V^{\mathrm{C}} * T) \quad := \quad - \sum_{(R,\cdot) \in T^{-1}} (\hat{A} * T)_R \times \ln \; (\hat{V}^{\mathrm{C}} * T)_R$$

The *component size cardinality cross entropy* is greater than or equal to the *derived entropy*, $\text{entropyCross}(A * T, V^{\mathrm{C}} * T) \geq \text{entropy}(A * T)$.

The *component cardinality size cross entropy* is the negative *cartesian derived expected normalised derived histogram count logarithm*,

$$\text{entropyCross}(V^{\mathrm{C}} * T, A * T) \quad := \quad - \sum_{(R,\cdot) \in T^{-1}} (\hat{V}^{\mathrm{C}} * T)_R \times \ln \; (\hat{A} * T)_R$$

The *component cardinality size cross entropy* is greater than or equal to the *cartesian derived entropy*, $\text{entropyCross}(V^{\mathrm{C}} * T, A * T) \geq \text{entropy}(V^{\mathrm{C}} * T)$.

The *component size cardinality sum cross entropy* is,

$$\text{entropy}(A * T + V^{\mathrm{C}} * T)$$

The *component size cardinality sum cross entropy* is positive and less than or equal to the logarithm of the sum of the *size* and *volume*, $0 \leq \text{entropy}(A * T + V^{\mathrm{C}} * T) \leq \ln(z + v)$. The *component size cardinality sum cross entropy* is greater than or equal to the *derived entropy*, $\text{entropy}(A * T + V^{\mathrm{C}} * T) \geq \text{entropy}(A * T)$, and greater than or equal to the *cartesian derived entropy*, $\text{entropy}(A * T + V^{\mathrm{C}} * T) \geq \text{entropy}(V^{\mathrm{C}} * T)$.

In all cases the *cross entropy* is maximised when high *size components* are low *cardinality components*, $(\hat{A} * T)_R \gg (\hat{V}^{\mathrm{C}} * T)_R$ or $\text{size}(A * C)/z \gg |C|/v$, and low *size components* are high *cardinality components*, $(\hat{A} * T)_R \ll (\hat{V}^{\mathrm{C}} * T)_R$ or $\text{size}(A * C)/z \ll |C|/v$, where $(R, C) \in T^{-1}$.

The *cross entropy* is minimised when the *normalised derived histogram* equals the *normalised cartesian derived*, $\hat{A} * T = \hat{V}^{\mathrm{C}} * T$ or $\forall (R, C) \in T^{-1}$ $(\mathrm{size}(A * C)/z = |C|/v)$. In this case the *cross entropy* equals the corresponding *component entropy*.

The *component size cardinality relative entropy* is the *component size cardinality cross entropy* minus the *component size entropy*,

$$\begin{aligned}
&\mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T) \\
&\quad := \sum_{(R,\cdot) \in T^{-1}} (\hat{A} * T)_R \times \ln \frac{(\hat{A} * T)_R}{(\hat{V}^{\mathrm{C}} * T)_R} \\
&\quad = \mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) \ - \ \mathrm{entropy}(A * T)
\end{aligned}$$

The *component size cardinality relative entropy* is positive, $\mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T) \geq 0$.

The *component cardinality size relative entropy* is the *component cardinality size cross entropy* minus the *component cardinality entropy*,

$$\begin{aligned}
&\mathrm{entropyRelative}(V^{\mathrm{C}} * T, A * T) \\
&\quad := \sum_{(R,\cdot) \in T^{-1}} (\hat{V}^{\mathrm{C}} * T)_R \times \ln \frac{(\hat{V}^{\mathrm{C}} * T)_R}{(\hat{A} * T)_R} \\
&\quad = \mathrm{entropyCross}(V^{\mathrm{C}} * T, A * T) \ - \ \mathrm{entropy}(V^{\mathrm{C}} * T)
\end{aligned}$$

The *component cardinality size relative entropy* is positive, $\mathrm{entropyRelative}(V^{\mathrm{C}} * T, A * T) \geq 0$.

The *size-volume* scaled *component size cardinality sum relative entropy* is the *size-volume* scaled *component size cardinality sum cross entropy* minus the *size-volume* scaled *component size cardinality sum entropy*,

$$\begin{aligned}
&(z + v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\
&\qquad\qquad - z \times \mathrm{entropy}(A * T) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)
\end{aligned}$$

The *size-volume* scaled *component size cardinality sum relative entropy* is positive, $(z + v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \ - z \times \mathrm{entropy}(A * T) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T) \geq 0$.

In all cases the *relative entropy* is maximised when (a) the *cross entropy* is maximised and (b) the *component entropy* is minimised. That is, the *relative entropy* is maximised when both (i) the *component size entropy*,

$\text{entropy}(A*T)$, and (ii) the *component cardinality entropy*, $\text{entropy}(V^{\text{C}}*T)$, are low, but low in different ways so that the *component size cardinality sum cross entropy*, $\text{entropy}(A*T+V^{\text{C}}*T)$, is high.

Let *histogram* A have a set of *variables* $V = \text{vars}(A)$ which is partitioned into query *variables* $K \subset V$ and label *variables* $V \setminus K$. Let $T \in \mathcal{T}_{U,\text{f},1}$ be a *one functional transform* having *underlying variables* equal to the query *variables*, $\text{und}(T) = K$. As shown above, given a query *state* $Q \in K^{\text{CS}}$ that is *effective* in the *sample derived*, $R \in (A*T)^{\text{FS}}$ where $\{R\} = (\{Q\}^{\text{U}}*T)^{\text{FS}}$, the *probability histogram* for the label is

$$\{Q\}^{\text{U}}*T*T^{\odot A} \ \% \ (V \setminus K) \ \ \in \ \ \mathcal{A} \cap \mathcal{P}$$

If the normalised *histogram*, $\hat{A} \in \mathcal{A} \cap \mathcal{P}$, is treated as a *probability function* of a single-*state* query, the *expected entropy* of the *modelled transformed conditional product*, or label *entropy*, is

$$\begin{aligned} &\sum_{(R,C) \in T^{-1}} (\hat{A}*T)_R \times \text{entropy}(A*C \ \% \ (V \setminus K)) \\ &= \sum_{(R,\cdot) \in T^{-1}} (\hat{A}*T)_R \times \text{entropy}(\{R\}^{\text{U}}*T^{\odot A} \ \% \ (V \setminus K)) \end{aligned}$$

This is similar to the definition of the *expected component entropy*, above,

$$\begin{aligned} \text{entropyComponent}(A,T) &:= \sum_{(R,C) \in T^{-1}} (\hat{A}*T)_R \times \text{entropy}(A*C) \\ &= \sum_{(R,\cdot) \in T^{-1}} (\hat{A}*T)_R \times \text{entropy}(\{R\}^{\text{U}}*T^{\odot A}) \end{aligned}$$

but now the *component* is *reduced* to the label *variables*, $V \setminus K$.

In the case where the relation between the *derived variables* and the label *variables* is functional or *causal*,

$$\text{split}(W, (A*\text{his}(T) \ \% \ (W \cup V \setminus K))^{\text{FS}}) \ \ \in \ \ W^{\text{CS}} \to (V \setminus K)^{\text{CS}}$$

the label *entropy* is zero,

$$\sum_{(R,C) \in T^{-1}} (\hat{A}*T)_R \times \text{entropy}(A*C \ \% \ (V \setminus K)) \ \ = \ \ 0$$

So label *entropy* is a measure of the ambiguity in the relation between the *derived variables* and the label *variables*. Negative label *entropy* may be viewed as the degree to which the *derived variables* of the *model* predict the label *variables*. In the cases of low label *entropy*, or high *causality*, the *derived variables* and the label *variables* are correlated and therefore *aligned*, $\text{algn}(A * \text{his}(T)\ \%\ (W \cup V \setminus K)) > 0$. In these cases the *derived histogram* tends to the *diagonal*, $\text{algn}(A * T) > 0$.

2.5.3 Functional definition sets

This section may be skipped until section 'Artificial neural networks'.

A *functional definition set* $F \in \mathcal{F}$ is a set of *unit functional transforms*, $\forall T \in F\ (T \in \mathcal{T}_{\mathrm{f}})$. *Functional definition sets* are also called *fuds*. *Fuds* are constrained such that *derived variables* can appear in only one *transform*. That is, the sets of *derived variables* are disjoint,

$$\forall F \in \mathcal{F}\ \forall T_1, T_2 \in F\ (T_1 \neq T_2 \implies \text{der}(T_1) \cap \text{der}(T_2) = \emptyset)$$

The set of *fud histograms* is $\text{his}(F) := \{\text{his}(T) : T \in F\}$. The set of *fud variables* is $\text{vars}(F) := \bigcup\{\text{vars}(X) : X \in \text{his}(F)\}$. The *fud derived* is $\text{der}(F) := \bigcup_{T \in F} \text{der}(T) \setminus \bigcup_{T \in F} \text{und}(T)$. The *fud underlying* is $\text{und}(F) := \bigcup_{T \in F} \text{und}(T) \setminus \bigcup_{T \in F} \text{der}(T)$. The set of *underlying variables* of a *fud* is also called the *substrate*.

A *functional definition set* is a *model*, so it can be converted to a *functional transform*,

$$F^{\mathrm{T}} \ := \ (\prod \text{his}(F)\ \%\ (\text{der}(F) \cup \text{und}(F)),\ \text{der}(F))$$

The resultant *transform* has the same *derived* and *underlying variables* as the *fud*, $\text{der}(F^{\mathrm{T}}) = \text{der}(F)$ and $\text{und}(F^{\mathrm{T}}) = \text{und}(F)$.

The set of *one functional definition sets* $\mathcal{F}_{U,1}$ in *system* U is the subset of the *functional definition sets*, $\mathcal{F}_{U,1} \subset \mathcal{F}$, such that all *transforms* are *one functional* and the *fuds* are not *circular*. The *transform* of a *one functional definition set* is a *one functional transform*, $\forall F \in \mathcal{F}_{U,1}\ (F^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1})$.

A *dependent variable* of a *one functional definition set* $F \in \mathcal{F}_{U,1}$ is any *variable* that is not a *fud underlying variable*, $\text{vars}(F) \setminus \text{und}(F)$. Each *dependent variable depends* on an *underlying* subset of the *fud*, $\text{depends} \in \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathcal{F}$ where $\forall w \in \text{vars}(F) \setminus \text{und}(F)\ (\text{depends}(F, \{w\}) \subseteq F)$.

Each *dependent variable* is in a *layer*. The *layer* is the length of the longest path of *underlying transforms* to the *dependent variable*. Given *fud* $F \in \mathcal{F}_{U,1}$, let l be the highest *layer*, $l = \mathrm{layer}(F, \mathrm{der}(F))$, where $\mathrm{layer} \in \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathbf{N}$ is defined in terms of $\mathrm{depends} \in \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathcal{F}$. Let F_i be the subset of the *fud* in a particular *layer*, $F_i = \{T : T \in F,\ \mathrm{layer}(F, \mathrm{der}(T)) = i\}$. Then $F = \bigcup_{i \in \{1 \ldots l\}} F_i$.

A *one functional definition set* $F \in \mathcal{F}_{U,1}$ is *non-overlapping* if the sets of *underlying transforms* of each of the *fud derived variables* are disjoint, $\forall v, w \in \mathrm{der}(F)\ (v \neq w\ \wedge\ \mathrm{vars}(\mathrm{depends}(F, \{v\})) \cap \mathrm{vars}(\mathrm{depends}(F, \{w\})) = \emptyset)$. A *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ is *non-overlapping* if it is equal to the *transform* of a *non-overlapping fud*, $T = F^{\mathrm{T}}$.

Given a set of *substrate variables* V, the set of *substrate functional definition sets* $\mathcal{F}_{U,V}$ is the subset of *one functional definition sets*, $\mathcal{F}_{U,V} \subset \mathcal{F}_{U,1}$, that (i) have *underlying variables* which are subsets of the *substrate*, $\forall F \in \mathcal{F}_{U,V}\ (\mathrm{und}(F) \subseteq V)$, and (ii) consist of *partition transforms*, $\forall F \in \mathcal{F}_{U,V}\ \forall T \in F\ \exists P \in \mathrm{B}(\mathrm{und}(T)^{\mathrm{CS}})\ (T = P^{\mathrm{T}})$. In addition, *partition circularities* are excluded by ensuring that the *partitions* are unique in the *fud* when *flattened* to *substrate*.

Let v be the *volume* of the *substrate*, $v = |V^{\mathrm{C}}|$. If the *volume*, v, is finite, the set of *substrate fuds* is finite, $|\mathcal{F}_{U,V}| < \infty$.

Avoiding *partition circularities* is computationally expensive. The *infinite-layer substrate functional definition sets* $\mathcal{F}_{\infty,U,V}$ is the superset of the *substrate functional definition sets*, $\mathcal{F}_{\infty,U,V} \supset \mathcal{F}_{U,V}$, that drop the exclusion of *partition circularities*. The *infinite-layer substrate fud set* is defined recursively,

$$\mathcal{F}_{\infty,U,V} \quad = \quad \{F : F \subseteq \mathrm{powinf}(U,V)(\emptyset),\ \mathrm{und}(F) \subseteq V\}$$

where

$$\mathrm{powinf}(U,V)(F) := F \cup G \cup \mathrm{powinf}(U,V)(F \cup G) : \\ G = \{P^{\mathrm{T}} : K \subseteq \mathrm{vars}(F) \cup V,\ P \in \mathrm{B}(K^{\mathrm{CS}})\}$$

The cardinality of the *infinite-layer substrate fud set* is infinite, $|\mathcal{F}_{\infty,U,V}| = \infty$.

2.5.4 Decompositions

This section may be skipped until section 'Tractable and practicable aligned induction'.

A *functional definition set decomposition* is a *model* that consists of a tree of *fuds* that are *contingent* on *components*.

The set of *functional definition set decompositions* $\mathcal{D}_{\mathrm{F}}$ is a subset of the trees of pairs of (i) *states*, $\mathcal{S}$, and (ii) *functional definition sets*, $\mathcal{F}$

$$\mathcal{D}_{\mathrm{F}} \subset \mathrm{trees}(\mathcal{S} \times \mathcal{F})$$

Let D be a *fud decomposition*, $D \in \mathcal{D}_{\mathrm{F}}$. The set of *fuds* is $\mathrm{fuds}(D) := \{F : ((\cdot, F), \cdot) \in \mathrm{nodes}(D)\}$. The *underlying* is $\mathrm{und}(D) := \bigcup\{\mathrm{und}(F) : F \in \mathrm{fuds}(D)\}$. The set of *underlying variables* of a *decomposition* is also called the *substrate*.

Fud decompositions are constrained such that each of the *states* in child pairs are *states* in the *derived variables* of the parent *fud*,

$$\forall D \in \mathcal{D}_{\mathrm{F}}\ \forall((\cdot, F), E) \in \mathrm{nodes}(D)\ \forall((S, \cdot), \cdot) \in E\ (S \in \mathrm{dom}((F^{\mathrm{T}})^{-1}))$$

The root nodes have no parent and so their *states* are constrained to be null, $\forall D \in \mathcal{D}_{\mathrm{F}}\ \ \forall((S, \cdot), \cdot) \in D\ (S = \emptyset)$. Given a *fud decomposition* $D \in \mathcal{D}_{\mathrm{F}}$ having *underlying variables* $V = \mathrm{und}(D)$, each *fud* $F \in \mathrm{fuds}(D)$ is *contingent* on the *component* $C \in \mathrm{B}(V^{\mathrm{C}})$ implied by the union of the ancestor *derived states* in the *derived variables* of the union of the ancestor *fuds*. Let L be a path in the *fud decomposition*, $L \in \mathrm{paths}(D)$. Then for each child *fud* $(\cdot, F) = L_i$, where $i \in \{2 \ldots |L|\}$, the union of the ancestor *derived states* is $R = \bigcup\{S : j \in \{1 \ldots i-1\},\ (S, \cdot) = L_j\}$, the union of the ancestor *fuds* is $G = \bigcup\{H : j \in \{1 \ldots i-1\},\ (\cdot, H) = L_j\}$, and so the *contingent component* is $(G^{\mathrm{T}})^{-1}(R)$. In the case where the *underlying* of the ancestor *fud*, G, is the whole *substrate*, $\mathrm{und}(G) = V$, then the *component* is $C = (G^{\mathrm{T}})^{-1}(R) \subseteq V^{\mathrm{C}}$.

The function $\mathrm{cont} \in \mathcal{D}_{\mathrm{F}} \to \mathrm{P}(\mathcal{A} \times \mathcal{F})$ returns the set of *component-fud* pairs of the *fud decomposition*. When the *fud decomposition*, D, is applied to a *histogram* $A \in \mathcal{A}$ in *variables* $\mathrm{vars}(A) = V$, each *fud transform* is applied to the *contingent slice*, $A * C * F^{\mathrm{T}}$ where $(C, F) \in \mathrm{cont}(D)$. Two *fuds* on the same path $(\cdot, F_1) \in L_j$ and $(\cdot, F_2) \in L_i$ where $L \in \mathrm{paths}(D)$ and $j < i$, are such that the *fud* $(C_1, F_1) \in \mathrm{cont}(D)$ nearer the root has a *component* which is a superset of the *component* of the *fud* $(C_2, F_2) \in \mathrm{cont}(D)$ nearer the leaves, $C_1 \supset C_2$. So the *slice* nearer the root is greater than or equal to the *slice* nearer the leaves, $A * C_1 \geq A * C_2$. That is, the *fuds* are more and more selectively *contingent* along the *fud decomposition's* paths, and so are applied to smaller and smaller *slices*.

In the case where each of the *slice derived* are *diagonalised*, $\forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{diagonal}(A * C * F^{\mathrm{T}}))$, the *fud decomposition*, D, is a *contingent, layered, redundant model* of the *sample histogram*, A.

A *fud decomposition* is a *model*, so it can be converted to a *functional transform*, $D^{\mathrm{T}} \in \mathcal{T}_{\mathrm{f}}$. The *partition* of the *fud decomposition transform* is equal to the set of *components* corresponding to those *fud derived states* that are not parent *derived states* in the *decomposition* tree, $\bigcup\{\mathrm{dom}((F^{\mathrm{T}})^{-1}) \setminus \{S : ((S,\cdot),\cdot) \in E\} : ((\cdot,F),E) \in \mathrm{nodes}(D)\}$. The resultant *transform* has the same *underlying variables* as the *fud decomposition*, $\mathrm{und}(D^{\mathrm{T}}) = \mathrm{und}(D)$.

Given a set of *substrate variables* V, the set of *substrate fud decompositions* $\mathcal{D}_{\mathrm{F},U,V}$ is a subset of *fud decompositions*, $\mathcal{D}_{\mathrm{F},U,V} \subset \mathcal{D}_{\mathrm{F}}$, that contain only *substrate fuds*, $\forall D \in \mathcal{D}_{\mathrm{F},U,V}\ \forall F \in \mathrm{fuds}(D)\ (F \in \mathcal{F}_{U,V})$. In addition, each *fud* is unique in a path, $\forall D \in \mathcal{D}_{\mathrm{F},U,V}\ \forall L \in \mathrm{paths}(D)\ (|\{F : (\cdot,(\cdot,F)) \in L\}| = |L|)$.

Let v be the *volume* of the *substrate*, $v = |V^{\mathrm{C}}|$. If the *volume*, v, is finite, the set of *substrate fud decompositions* is finite, $|\mathcal{D}_{\mathrm{F},U,V}| < \infty$.

Similarly, the *infinite-layer substrate fud decompositions* $\mathcal{D}_{\mathrm{F},\infty,U,V}$ is the superset of the *substrate fud decompositions*, $\mathcal{D}_{\mathrm{F},\infty,U,V} \supset \mathcal{D}_{\mathrm{F},U,V}$, that contain only *infinite-layer substrate fuds*, $\forall D \in \mathcal{D}_{\mathrm{F},\infty,U,V}\ \forall F \in \mathrm{fuds}(D)\ (F \in \mathcal{F}_{\infty,U,V})$. The cardinality of the *infinite-layer substrate fud decomposition set* is infinite, $|\mathcal{D}_{\mathrm{F},\infty,U,V}| = \infty$.

2.6 Induction with model

2.6.1 Classical induction

Given *substrate transform* $T \in \mathcal{T}_{U,V}$, the *derived histogram* valued *integral substrate histograms* function $D_{U,\mathrm{i},T,z}$ is defined

$$D_{U,\mathrm{i},T,z} \ := \ \{(A, A * T) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The finite set of *iso-deriveds* of *derived histogram* $A * T$ is

$$D^{-1}_{U,\mathrm{i},T,z}(A * T) \ = \ \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B * T = A * T\}$$

The degree to which an *integral iso-set* $I \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$ that contains the *histogram*, $A \in I$, is said to be *law-like* is called the *iso-derivedence*. The *iso-derivedence* is defined as the ratio of (i) the cardinality of the intersection between the *integral iso-set* and the set of *integral iso-deriveds*, and (ii) the cardinality of the union,

$$\frac{1}{|\mathcal{A}_{U,\mathrm{i},V,z}|} \leq \frac{|I \ \cap \ D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|I \ \cup \ D^{-1}_{U,\mathrm{i},T,z}(A * T)|} \leq 1$$

In *classical modelled induction* the *history probability functions* are constrained by *derived histogram*.

Let P be a *substrate history probability function*, $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$. Given a *history* $H_E \in \mathcal{H}_{U,V,z_E}$, of *size* $z_E = |H_E|$, consider the case where the *derived histogram* $A * T$ of *drawn histories* is *known* to be *necessary*, $\sum(P(H) : H \subseteq H_E,\ \mathrm{his}(H) * T = A * T) = 1$. The *maximum likelihood estimate* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} &= \{(H,1) : H \subseteq H_E,\ \mathrm{his}(H) * T = A * T\}^{\wedge} \ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\} \ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) * T \neq A * T\} \\
&= \{(H, 1/\textstyle\sum(Q_{\mathrm{h},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T))) : \\
&\qquad\qquad H \subseteq H_E,\ \mathrm{his}(H) * T = A * T\} \ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\} \ \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) * T \neq A * T\}
\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn histories* $H \subseteq H_E$ with the *derived*, $\mathrm{his}(H) * T = A * T$, are uniformly probable and all other *histories*, $G \nsubseteq H_E$ or $\mathrm{his}(G) * T \neq A * T$, are impossible, $\tilde{P}(G) = 0$. If (i) the *transform*, T, is *known*, (ii) the *derived*, $A * T$, is *known* and (iii) the *distribution histogram*, H_E, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

The *likely probability* of *drawing histogram* A from *necessary drawn derived* $A * T$ is

$$\sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) = \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum Q_{\mathrm{h},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}$$

The *likely history probability function* entropy, $\mathrm{entropy}(\tilde{P})$, is maximised when the sum of the *iso-derived historical frequencies*, $\sum Q_{\mathrm{h},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)$, is maximised.

Consider the case where the *transform*, T, is *known* and the *derived*, $A * T$, is *known*, but the *distribution histogram*, E, is *unknown* and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *maximum likelihood estimate* $\tilde{E}$ for the *distribution histogram*, E, is a modal value of the *likelihood function*,

$$\tilde{E} \in \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{h},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T))) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\})$$

The *likely distribution histogram*, $\tilde{E}$, is *known* if the *distribution histogram size*, z_E, is *known*, the *transform*, T, is *known* and the *derived*, $A * T$, is *known*. If it is assumed that the *distribution histogram* equals the *likely distribution histogram*, $E = \tilde{E}$, then the *likely history probability* is *known*, $\tilde{P}(H) = 1/\sum(Q_{\mathrm{h},U}(\tilde{E},z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A*T))$ where $\mathrm{his}(H)*T = A*T$.

In the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $\tilde{E}$ for the *distribution probability histogram*, $\hat{E}$, may be approximated by a modal value of a *likelihood function* which depends on the *multinomial distribution* instead,

$$\tilde{E} \in \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T))) : D \in \mathcal{A}_{U,V,1}\})$$

The *normalised naturalisation*, $\hat{A} * T * T^{\dagger}$, is a solution. The *naturalisation*, $A * T * T^{\dagger}$, is the *independent analogue* of the *iso-deriveds*. So the *maximum likelihood estimate*, $\tilde{E}$, for the *distribution probability histogram*, $\hat{E}$, is the *naturalisation probability histogram*, $\hat{A} * T * T^{\dagger}$,

$$\tilde{E} = \hat{A} * T * T^{\dagger}$$

In the case where the *naturalisation* is *integral*, $A * T * T^{\dagger} \in \mathcal{A}_{\mathrm{i}}$, the sum of the *iso-derived naturalisation-distributed multinomial probabilities* varies with the *naturalisation naturalisation-distributed multinomial probability*,

$$\sum Q_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T) \sim Q_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})$$

So, if it is assumed that the *distribution probability histogram* equals the *likely distribution probability histogram*, $\hat{E} = \tilde{E} = \hat{A} * T * T^{\dagger}$, then the *likely history probability* varies against the *naturalisation-distributed multinomial probability* of the *naturalisation*, $\tilde{P}(H) \sim 1/\hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})$.

The cardinality of the set of *integral iso-deriveds* may be stated explicitly as the product of the weak compositions of the *components*,

$$|D^{-1}_{U,\mathrm{i},T,z}(A*T)| \;=\; \prod_{(R,C) \in T^{-1}} \frac{((A*T)_R + |C|-1)!}{(A*T)_R!\ (|C|-1)!}$$

So the *integral iso-deriveds log-cardinality* varies against the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}&\ln |D^{-1}_{U,\mathrm{i},T,z}(A*T)| \;\sim\\ &\quad -\ \big((z+v) \times \mathrm{entropy}(A*T + V^{\mathrm{C}}*T)\\ &\qquad\qquad -z \times \mathrm{entropy}(A*T)\ - v \times \mathrm{entropy}(V^{\mathrm{C}}*T)\big)\end{aligned}$$

where *size* $z = \mathrm{size}(A) = \mathrm{size}(A*T)$ and *volume* $v = |V^{\mathrm{C}}|$. In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *integral iso-deriveds log-cardinality* varies against the *size* scaled *component size cardinality relative entropy*,

$$\ln |D^{-1}_{U,\mathrm{i},T,z}(A*T)| \;\sim\; -z \times \mathrm{entropyRelative}(A*T, V^{\mathrm{C}}*T)$$

So the logarithm of the *likely probability* of *drawing histogram* A from *necessary drawn derived* $A*T$ varies with the *relative entropy*,

$$\begin{aligned}&\ln \sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A)\ \sim\\ &\qquad\qquad z \times \mathrm{entropyRelative}(A*T, V^{\mathrm{C}}*T)\end{aligned}$$

The *naturalisation*, $A*T*T^{\dagger}$, is the most *probable histogram*, $\forall B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)\ (Q_{\mathrm{m},U}(A*T*T^{\dagger},z)(A*T*T^{\dagger}) \geq Q_{\mathrm{m},U}(A*T*T^{\dagger},z)(B))$. In the case where the *histogram* is *natural*, $A = A*T*T^{\dagger}$, then, as the *relative entropy*, $\mathrm{entropyRelative}(A*T, V^{\mathrm{C}}*T)$, increases, the *likely histogram probability*, $Q_{\mathrm{m},U}(A,z)(A)/\sum(Q_{\mathrm{m},U}(A,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A*T))$, increases.

The *likely history probability function* entropy varies with the *naturalisation entropy*, $\mathrm{entropy}(\tilde{P}) \sim \mathrm{entropy}(A*T*T^{\dagger})$, and against the *relative entropy*, $\mathrm{entropy}(\tilde{P}) \sim -\ \mathrm{entropyRelative}(A*T, V^{\mathrm{C}}*T)$.

Consider the case where a *drawn histogram* A is *known*, but neither the *distribution histogram*, E, is *known* nor the *transform*, T, is *known*, and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *maximum* *likelihood estimate* $(\tilde{E},\tilde{T})$ for the pair of the *distribution histogram*, E, and the *transform*, T, is a modal value of the *likelihood function*,

$$\begin{aligned}(\tilde{E},\tilde{T}) \in\ & \\ &\mathrm{maxd}(\{((D,M),\sum(Q_{\mathrm{h},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},M,z}(A * M))) : \\ &\qquad\qquad D \in \mathcal{A}_{U,\mathrm{i},V,z_E},\ M \in \mathcal{T}_{U,V}\})\end{aligned}$$

All solutions are such that the *transform maximum likelihood estimate* is *unary*, $\tilde{T} = T_{\mathrm{u}}$ where $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$. This is the trivial case where the set of *iso-derived histograms* is the entire set of *substrate histograms*, $D^{-1}_{U,\mathrm{i},T_{\mathrm{u}},z}(A * T_{\mathrm{u}}) = \mathcal{A}_{U,\mathrm{i},V,z}$. In this case *necessary derived*, $H \subseteq H_E$ and $\mathrm{his}(H) * T_{\mathrm{u}} = A * T_{\mathrm{u}}$, reduces to *drawn history*, $H \subseteq H_E$. If it is assumed that the *transform* equals the *likely transform*, $T = \tilde{T} = T_{\mathrm{u}}$, then the *likely history probability function* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}\tilde{P} \ &=\ \{(H, 1/\binom{z_E}{z}) : H \subseteq H_E,\ |H| = z\}\ \cup \\ &\quad\ \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\}\end{aligned}$$

That is, in the case of *unknown transform*, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn histories* $H \subseteq H_E$ of *size* $|H| = z$ are uniformly probable and all other *histories*, $G \nsubseteq H_E$, are impossible, $\tilde{P}(G) = 0$.

Define the *derived-dependent* $A^{\mathrm{D}(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-derived*,

$$\begin{aligned}\{A^{\mathrm{D}(T)}\} \ =\ & \\ &\mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *derived-dependent*, $A^{\mathrm{D}(T)}$, is the *dependent analogue* of the *iso-deriveds*. Note that the *derived-dependent*, $A^{\mathrm{D}(T)}$, is not always computable, but an approximation to any accuracy can be made to it. In the case where the *histogram* is *natural*, the *derived-dependent* equals the *naturalisation*, $A = A * T * T^{\dagger} \implies A^{\mathrm{D}(T)} = A = A * T * T^{\dagger}$.

Now consider the case where, given *necessary drawn derived* $A * T$, it is *known*, in addition, that the *sample histogram* A is the most *probable histogram* of the *iso-derived*. That is, the *likely probability* of *drawing histogram*

A from *necessary drawn derived* $A * T$,

$$\sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) \ = \ \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum Q_{\mathrm{h},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}$$

is maximised.

In the case where the *transform*, T, is *known* and the *sample*, A, is *known*, but the *distribution histogram*, E, is *unknown*, the *maximum likelihood estimate* $\tilde{E}$ for the *distribution histogram*, E, is a modal value of the *likelihood function*,

$$\tilde{E} \ \in \ \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{h},U}(D,z)(A)}{\sum Q_{\mathrm{h},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\})$$

The *likely distribution histogram*, $\tilde{E}$, is *known* if the *distribution histogram size*, z_E, is *known*, the *transform*, T, is *known* and the *sample*, A, is *known*. If it is assumed that the *distribution histogram* equals the *likely distribution histogram*, $E = \tilde{E}$, then the *likely history probability* is *known*, $\tilde{P}(H) = 1/\sum(Q_{\mathrm{h},U}(\tilde{E},z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T))$ where $\mathrm{his}(H) * T = A * T$.

If the *histogram* is *natural*, $A = A * T * T^{\dagger}$, then the additional constraint of *probable sample* makes no change to the *maximum likelihood estimate*, $\tilde{E}$,

$$\begin{aligned}
&A = A * T * T^{\dagger} \implies \\
&\qquad \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{h},U}(D,z)(A)}{\sum Q_{\mathrm{h},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\}) \\
&\quad = \ \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{h},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T))) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\})
\end{aligned}$$

If the *histogram* is not *natural*, $A \neq A * T * T^{\dagger}$, however, then the *likely history probability function* entropy, $\mathrm{entropy}(\tilde{P})$, is lower than it is in the case of *necessary derived* unconstrained by *probable sample*.

In the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $\tilde{E}$ for the *distribution probability histogram*, $\hat{E}$, is now approximated by a modal value of the conditional *likelihood function*,

$$\tilde{E} \ \in \ \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}) : D \in \mathcal{A}_{U,V,1}\})$$

The solution to this is the *normalised derived-dependent*, $\tilde{E} = \hat{A}^{\mathrm{D}(T)}$. The *maximum likelihood estimate* is near the *sample*, $\tilde{E} \sim \hat{A}$, only in as much as it is far from the *naturalisation*, $\tilde{E} \not\sim \hat{A} * T * T^{\dagger}$.

The *iso-derived conditional multinomial probability distribution* is defined

$$\hat{Q}_{\mathrm{m,d},T,U}(E,z)(A) \ := \ \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},T,z})|} \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}$$

So the optimisation can be rewritten,

$$\tilde{E} \ \in \ \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{m,d},T,U}(D,z)(A)) : D \in \mathcal{A}_{U,V,1}\})$$

In the case where the *histogram* is *natural*, $A = A * T * T^{\dagger}$, the *log likelihood* varies against the *iso-derived log-cardinality*,

$$\begin{aligned} \ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \ &\propto \ \ln \frac{Q_{\mathrm{m},U}(A,z)(A)}{\sum Q_{\mathrm{m},U}(A,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)} \\ &\sim \ -\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \end{aligned}$$

So the *log likelihood* varies with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned} &\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \ \sim \\ &\quad (z+v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\ &\qquad\qquad -z \times \mathrm{entropy}(A * T) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T) \end{aligned}$$

In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *log likelihood* varies with the *size* scaled *component size cardinality relative entropy*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \ \sim \ z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

In other words, the *log likelihood* is maximised where (i) the *derived entropy*, $\mathrm{entropy}(A * T)$, is minimised, and (ii) the *cross entropy*, $\mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T)$, is maximised, so that high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*.

If the *histogram* is *natural*, $A = A * T * T^{\dagger}$, and the *component size cardinality relative entropy* is high, $\mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) > \ln |T^{-1}|$, it can also be shown that the *log likelihood* varies against the *derived multinomial probability*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \ \sim \ - \ \ln \hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)$$

In this case the *sum sensitivity* of the *iso-derived conditional multinomial probability distribution* varies with the *underlying-derived multinomial probability distribution sum sensitivity difference*,

$$\begin{aligned}&\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m,d},T,U}(A,z))) \sim \\ &\quad\quad \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(A,z))) - \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(A*T,z)))\end{aligned}$$

and so is less than or equal to the *sum sensitivity* of the *multinomial probability distribution*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m,d},T,U}(A,z))) \quad\leq\quad \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(A,z)))$$

Furthermore, the *sum sensitivity* varies against the *log-likelihood*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m,d},T,U}(A,z))) \quad\sim\quad -\ln \hat{Q}_{\text{m,d},T,U}(A,z)(A)$$

That is, in the high *relative entropy natural* case, the maximisation of the *log-likelihood* also tends to minimise the *sum sensitivity* to the *maximum likelihood estimate*. This is opposite to the relationship between the *sum sensitivity* and the *log-likelihood* in *classical non-modelled induction*, which was found to be weakly positively correlated.

As the *relative entropy*, entropyRelative$(A*T, V^{\text{C}}*T)$, increases, the *log-likelihood*, $\ln \hat{Q}_{\text{m,d},T,U}(A,z)(A)$, increases, but the *sensitivity* to *distribution histogram*, E, decreases. In other words, the higher the *sample relative entropy* the more *likely* the *normalised sample histogram*, $\hat{A}$, equals the *normalised distribution histogram*, $\hat{E}$, and the smaller the *likely* difference between them if they are not equal.

Given *necessary derived* and *probable sample*, consider the case where a *drawn histogram* A is *known*, but neither the *distribution histogram*, E, is *known* nor the *transform*, T, is *known*, and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *maximum likelihood estimate* $(\tilde{E},\tilde{T})$ for the pair of the *distribution histogram*, E, and the *transform*, T, is a modal value of the *likelihood function*,

$$\begin{aligned}&(\tilde{E},\tilde{T}) \in \\ &\quad \text{maxd}(\{((D,M), \frac{Q_{\text{h},U}(D,z)(A)}{\sum Q_{\text{h},U}(D,z)(B) : B \in D^{-1}_{U,\text{i},M,z}(A*M)}) : \\ &\quad\quad\quad\quad\quad\quad\quad\quad D \in \mathcal{A}_{U,\text{i},V,z_E},\ M \in \mathcal{T}_{U,V}\})\end{aligned}$$

All solutions are such that the *transform maximum likelihood estimate* is *self*, $\tilde{T} = T_{\text{s}}$ where $T_{\text{s}} = V^{\text{CS}\{\}\text{T}}$. This is the trivial case where the set of *iso-derived histograms* is just the *sample*, $D^{-1}_{U,\mathrm{i},T_\mathrm{s},z}(A * T_\mathrm{s}) = \{A\}$. In this case *necessary derived*, $\mathrm{his}(H) * T_\mathrm{s} = A * T_\mathrm{s}$, reduces to *necessary histogram*, $\mathrm{his}(H) = A$. If it is assumed that the *transform* equals the *likely transform*, $T = \tilde{T} = T_\mathrm{s}$, then the *likely history probability function* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}\tilde{P} \;\; = \;\; & \{(H, 1/Q_{\mathrm{h},U}(E,z)(A)) : H \subseteq H_E,\ \mathrm{his}(H) = A\} \ \cup \\ & \{(G, 0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\} \ \cup \\ & \{(G, 0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G) \neq A\}\end{aligned}$$

That is, in the case of *unknown transform*, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn histories* $H \subseteq H_E$ with the *histogram*, $\mathrm{his}(H) = A$, are uniformly probable and all other *histories*, $G \not\subseteq H_E$ or $\mathrm{his}(G) \neq A$, are impossible, $\tilde{P}(G) = 0$.

In this case the *maximum likelihood estimate*, $\tilde{E}$, for the *distribution probability histogram*, $\hat{E}$, is the *sample probability histogram*, $\hat{A}$,

$$\tilde{E} \;\; = \;\; \hat{A} \;\; = \;\; \hat{A} * T_\mathrm{s} * T_\mathrm{s}^\dagger$$

Consider the case where the *derived* is *uniformly possible*. Given *substrate transform* $T \in \mathcal{T}_{U,V}$, assume that the *substrate history probability function* $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is the distribution of an arbitrary *history* valued function, $\mathcal{X} \to \mathcal{H}$, given an arbitrary *derived* valued function, $\mathcal{X} \to \mathcal{A}$. In this case, the *history* valued function is chosen arbitrarily from the constrained subset

$$\begin{aligned}&\big\{\{((x,A',y),H) : (x,(A',G)) \in F,\ (y,H) \in G,\ \mathrm{his}(H) * T = A'\} \ : \\ &\qquad\qquad\qquad F \in \mathcal{X} \to (\mathcal{A} \times (\mathcal{X} \to \mathcal{H}))\big\} \ \subset \ \mathcal{X} \to \mathcal{H}\end{aligned}$$

Uniformly possible derived is a weaker constraint than *uniformly possible histogram*, so the subset of *history* valued functions is larger.

This subset of the *substrate history probability functions* can be generalised for all *substrate transforms* as the subset derived from

$$\bigcup_{T \in \mathcal{T}_\mathrm{f}} (\mathcal{X} \to (\mathcal{A} \times_T (\mathcal{X} \to \mathcal{H})))$$

where $\mathcal{T}_\mathrm{f}$ is the set of all *functional transforms*, and the fibre product $\times_T$ is defined

$$\begin{aligned}&\mathcal{A} \times_T (\mathcal{X} \to \mathcal{H}) \ := \\ &\qquad \{(A',G) : (A',G) \in \mathcal{A} \times (\mathcal{X} \to \mathcal{H}),\ \forall (\cdot,H) \in G\ (\mathrm{his}(H) * T = A')\}\end{aligned}$$

In the case where there is a *distribution history* H_E and a *substrate transform* $T \in \mathcal{T}_{U,V}$, the *maximum likelihood estimate* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} &= \Big(\bigcup\{\{(H,1) : H \subseteq H_E,\ \mathrm{his}(H) * T = A'\}^{\wedge} : A' \in \mathrm{ran}(D_{U,\mathrm{i},T,z})\}\Big)^{\wedge} \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\} \\
&= \Big(\bigcup\{\{(H, 1/\textstyle\sum(Q_{\mathrm{h},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A'))) : \\
&\qquad\qquad H \subseteq H_E,\ \mathrm{his}(H) * T = A'\} \ :\ A' \in \mathrm{ran}(D_{U,\mathrm{i},T,z})\}\Big)^{\wedge} \cup \\
&\quad \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\}
\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn deriveds* are uniformly probable, and then all *drawn histories* $H \subseteq H_E$ with the same *derived*, $\mathrm{his}(H) * T = A'$, are uniformly probable. If the *distribution histogram*, H_E, is *known* and the *substrate transform*, T, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

The properties of *uniformly possible derived* are the same as for *necessary derived*, except that the probabilities are scaled. So, in the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *likely history probability* varies against the *naturalisation-distributed multinomial probability* of the *naturalisation*,

$$\tilde{P}(H) \ \sim\ 1/|\mathrm{ran}(D_{U,\mathrm{i},T,z})| \times 1/\hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})$$

That is, more *histories* are possible but less probable.

Now consider the case where, given *uniform possible derived*, it is *known*, in addition, that the *sample histogram* A is the most *probable histogram* of its *iso-derived*.

The *iso-derived conditional multinomial probability distribution*, is defined above as

$$\hat{Q}_{\mathrm{m,d},T,U}(E,z)(A) \ :=\ \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},T,z})|}\frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}$$

The *iso-derived conditional multinomial probability* already includes the *uniform possible* scaling factor of $1/|\mathrm{ran}(D_{U,\mathrm{i},T,z})|$.

The cardinality of the *derived*, $|\text{ran}(D_{U,\text{i},T,z})|$, is equal to the cardinality of the *derived substrate histograms*,

$$|\text{ran}(D_{U,\text{i},T,z})| \quad = \quad \frac{(z+w'-1)!}{z!\ (w'-1)!}$$

where $w' = |T^{-1}|$. So the additional term, $-\ln|\text{ran}(D_{U,\text{i},T,z})|$, in the *uniform possible log likelihood*, $\ln \hat{Q}_{\text{m,d},T,U}(E,z)(A)$, varies against the *derived volume*, w', where the *derived volume* is less than the *size*, $w' < z$, otherwise against the *size* scaled *log derived volume*, $z \ln w'$,

$$-\ln|\text{ran}(D_{U,\text{i},T,z})| \quad \sim \quad -\ ((w'\ :\ w' < z)\ +\ (z \ln w'\ :\ w' \geq z))$$

In the case where the *sample* is *natural*, $A = A * T * T^{\dagger}$, the *uniform possible log likelihood* varies (i) against the *derived volume*, w', where the *derived volume* is less than the *size*, $w' < z$, otherwise against the *size* scaled *log derived volume*, $z \ln w'$, and (ii) with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}
&\ln \hat{Q}_{\text{m,d},T,U}(A,z)(A)\ \sim \\
&\qquad -\ ((w'\ :\ w' < z)\ +\ (z \ln w'\ :\ w' \geq z)) \\
&\qquad +\ (z+v) \times \text{entropy}(A * T + V^{\text{C}} * T) \\
&\qquad\qquad\qquad\qquad -z \times \text{entropy}(A * T)\ \ -v \times \text{entropy}(V^{\text{C}} * T)
\end{aligned}$$

In other words, the *log likelihood* is maximised where (i) the *derived volume*, w', is minimised, (ii) the *derived entropy*, $\text{entropy}(A * T)$, is minimised, and (iii) the *cross entropy*, $\text{entropyCross}(A * T, V^{\text{C}} * T)$, is maximised, so that high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*.

As in the case of *necesary derived* and *probable sample*, above, if the *histogram* is *natural*, $A = A * T * T^{\dagger}$, and the *component size cardinality relative entropy* is high, $\text{entropyCross}(A * T, V^{\text{C}} * T) > \ln w'$, the *sum sensitivity* of the *iso-derived conditional multinomial probability distribution* is less than or equal to the *sum sensitivity* of the *multinomial probability distribution*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m,d},T,U}(A,z))) \quad \leq \quad \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(A,z)))$$

and varies against the *log-likelihood*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m,d},T,U}(A,z))) \quad \sim \quad -\ln \hat{Q}_{\text{m,d},T,U}(A,z)(A)$$

Given *uniform possible derived* and *probable sample*, consider the case where a *drawn histogram* A is *known*, but neither the *distribution histogram*, E, is *known* nor the *transform*, T, is *known*, and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. In the case where the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $(\tilde{E}, \tilde{T})$ for the pair of the *distribution histogram*, E, and the *transform*, T, is approximated by a modal value of the conditional *likelihood function*,

$$(\tilde{E}, \tilde{T}) \in \text{maxd}(\{((D, M), \hat{Q}_{\text{m,d},M,U}(D, z)(A)) : D \in \mathcal{A}_{U,V,1},\ M \in \mathcal{T}_{U,V}\})$$

If there is a unique maximum for the *distribution probability histogram*, $\tilde{E}$, this can be rewritten in terms of the *derived-dependent*,

$$\tilde{T} \in \text{maxd}(\{(M, \hat{Q}_{\text{m,d},M,U}(A^{\text{D}(M)}, z)(A)) : M \in \mathcal{T}_{U,V}\})$$

The *derived-dependent*, $A^{\text{D}(T)}$, is not always computable, but an approximation to any accuracy can be made to it, so a computable approximation of the *maximum likelihood estimate*, $\tilde{T}$, can be made for the *unknown transform*, T. In some cases the *likely transform*, $\tilde{T}$, is not trivial, $\tilde{T} \neq T_{\text{u}}$ and $\tilde{T} \neq T_{\text{s}}$.

If it is also *known* that the *sample* is *natural*, the optimisation can be restricted to *natural transforms*, $A = A * T * T^{\dagger} \implies A^{\text{D}(T)} = A$. In this case the optimisation is

$$\tilde{T} \in \text{maxd}(\{(M, \hat{Q}_{\text{m,d},M,U}(A, z)(A)) : M \in \mathcal{T}_{U,V},\ A = A * M * M^{\dagger}\})$$

or

$$\begin{aligned}\tilde{T} \in \text{maxd}(\{(M, \frac{1}{|\text{ran}(D_{U,\text{i},M,z})|}\frac{Q_{\text{m},U}(A, z)(A)}{\sum Q_{\text{m},U}(A, z)(B) : B \in D^{-1}_{U,\text{i},M,z}(A * M)}) :\\ M \in \mathcal{T}_{U,V},\ A = A * M * M^{\dagger}\})\end{aligned}$$

The numerator is constant, so the optimisation can be simplified,

$$\begin{aligned}\tilde{T} \in \text{mind}(\{(M, |\text{ran}(D_{U,\text{i},M,z})| \sum Q_{\text{m},U}(A, z)(B) : B \in D^{-1}_{U,\text{i},M,z}(A * M)) :\\ M \in \mathcal{T}_{U,V},\ A = A * M * M^{\dagger}\})\end{aligned}$$

In this case the *maximum likelihood estimate*, $\tilde{E}$, for the *distribution probability histogram*, $\hat{E}$, is the *sample probability histogram*, $\hat{A}$,

$$\tilde{E} = \hat{A} = \hat{A} * \tilde{T} * \tilde{T}^{\dagger}$$

Note that, although computable, this optimisation is intractable because the cardinality of the *substrate transforms*, $|\mathcal{T}_{U,V}|$, is factorial in the *volume*, v. Tractable optimisations require the computation to be at most polynomial.

Note, also, that, although the *sensitivity* to *distribution*, E, is defined above for *uniform possible derived*, the *sensitivity* to *model*, T, is not yet defined.

2.6.2 Specialising coder induction

It is shown above that there are two *canonical history coders*, the *index history coder* C_{H} and the *classification coder* C_{G}. Given *variables* V and *size* z, the *index substrate history coder*, $C_{\mathrm{H},U,V,z}$, encodes each *substrate history* $H \in \mathcal{H}_{U,V,z}$ in a fixed *space* of $C^{\mathrm{s}}_{\mathrm{H},U,V,z}(H) = z \ln v$, where *volume* $v = |V^{\mathrm{C}}|$. By contrast, the *classification substrate history coder*, $C_{\mathrm{G},U,V,z}$, encodes each *history* in a *space* which depends on the *histogram* $A = \mathrm{his}(H)$,

$$C^{\mathrm{s}}_{\mathrm{G},U,V,z}(H) \;=\; \ln \frac{(z+v-1)!}{z!\ (v-1)!} \;+\; \ln \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!}$$

When the *histogram entropy*, $\mathrm{entropy}(A)$, is high the *classification space* is greater than the *index space*, $C^{\mathrm{s}}_{\mathrm{G},U,V,z}(H) > C^{\mathrm{s}}_{\mathrm{H},U,V,z}(H)$, but when the *entropy* is low the *classification space* is less than the *index space*, $C^{\mathrm{s}}_{\mathrm{G},U,V,z}(H) < C^{\mathrm{s}}_{\mathrm{H},U,V,z}(H)$. In the case where the *size* is much less than the *volume*, $z \ll v$, the break-even *sized entropy* is approximately $z \times \mathrm{entropy}(A) \approx z \ln z$.

Given *substrate transform* $T \in \mathcal{T}_{U,V}$, the *specialising derived substrate history coder*, $C_{\mathrm{G},\mathrm{H},U,T,z}$, is intermediate between the *classification coder*, $C_{\mathrm{G},U,V,z}$, and the *index coder*, $C_{\mathrm{H},U,V,z}$. Given a *substrate history* $H \in \mathcal{H}_{U,V,z}$, the *derived history*, $H * T$, is encoded in a *classification coder*, $C_{\mathrm{G},U,W,z}$, where *derived variables* $W = \mathrm{der}(T)$. Then each *sub-history* H_C, corresponding to a *component* of the *partition*, $H_C \subseteq H$, where $(R,C) \in T^{-1}$, is encoded in a *index coder*, C_{H,U,C,z_C}, where $z_C = (A * T)_R$. The *specialising space* is

$$\begin{aligned} C^{\mathrm{s}}_{\mathrm{G},\mathrm{H},U,T,z}(H) \;=\;& \ln \frac{(z+w'-1)!}{z!\ (w'-1)!} \;+\; \ln \frac{z!}{\prod_{(R,\cdot) \in T^{-1}} (A * T)_R!} \;+ \\ & \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C| \end{aligned}$$

where $w' = |T^{-1}|$.

In the case where the *transform* is *self*, $T = T_{\mathrm{s}}$ where $T_{\mathrm{s}} = V^{\mathrm{CS}\{\}\mathrm{T}}$, then the *specialising space* equals the *classification space*, $C^{\mathrm{s}}_{\mathrm{G,H},U,T_{\mathrm{s}},z}(H) = C^{\mathrm{s}}_{\mathrm{G},U,V,z}(H)$. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$ where $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, then the *specialising space* equals the *index space*, $C^{\mathrm{s}}_{\mathrm{G,H},U,T_{\mathrm{u}},z}(H) = C^{\mathrm{s}}_{\mathrm{H},U,V,z}(H)$.

The *specialising space* depends only on the *transform*, T, and the *derived*, $A * T$. Define the *specialising space* function $\mathrm{sp}(T)(A * T) := C^{\mathrm{s}}_{\mathrm{G,H},U,T,z}(H)$.

The *specialising space* varies (i) with the *derived volume*, w', where the *derived volume* is less than the *size*, $w' < z$, otherwise with the *size* scaled *log derived volume*, $z \ln w'$, and (ii) against the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned} C^{\mathrm{s}}_{\mathrm{G,H},U,T,z}(H) \quad \sim \quad & (w' \ : \ w' < z) \ + \ (z \ln w' \ : \ w' \geq z) \\ & - \ z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T) \end{aligned}$$

In general, the *specialising space* is less than either of the two *canonical spaces* where the *derived entropy*, $\mathrm{entropy}(A * T)$, is low, but the *expected component entropy*, $\mathrm{entropyComponent}(A, T)$, is high. So the *specialising space* is minimised when (a) the *derived volume*, w', is minimised, (b) the *derived entropy*, $\mathrm{entropy}(A * T)$, is minimised, (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*, and (d) the *expected component entropy* is maximised.

In *specialising induction* the *history probability functions* are constrained by *specialising space* which in turn depends on *derived histogram*.

In the discussion of Occam's Razor, above, it was shown that, of a subset of the micro-state valued functions of distinguishable particle, the *maximum likelihood estimate* of the implied *probability function* is the *probability function* with the greatest entropy.

Consider a system of r undefined particles where the micro-state is a *substrate history*, $H \in \mathcal{H}_{U,V,z}$. The set of *substrate history* valued functions having exactly r particles with integer identifier is $\{1 \ldots r\} :\to \mathcal{H}_{U,V,z} \subset \mathcal{X} \to \mathcal{H}$. Given *substrate transform* $T \in \mathcal{T}_{U,V}$, let the subset $S \subset \{1 \ldots r\} :\to \mathcal{H}_{U,V,z}$ be such that the expected *specialising space* is a constant, $\forall R \in S\ (\sum (C^{\mathrm{s}}_{\mathrm{G,H},U,T,z}(H)/r \ : \ (\cdot, H) \in R) = \epsilon)$. Of this subset, S, the implied *probability function* with the greatest entropy, $\tilde{P} \in \mathrm{maxd}(\{(N, \mathrm{entropy}(N)) : R \in S,\ N = \{(H, |C|/r) : (H, C) \in R^{-1}\}\})$, approximates to a Boltzmann distribution.

Given *substrate transform* $T \in \mathcal{T}_{U,V}$, the *maximum likelihood estimate* $\tilde{P}$ of the *substrate history probability function* $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$, which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned}
\tilde{P} &= \{(H, \exp(-C^{\mathrm{s}}_{\mathrm{G,H},U,T,z}(H))) : H \in \mathcal{H}_{U,V,z}\}^{\wedge} \\
&= \{(H, \exp(-\mathrm{sp}(T)(\mathrm{his}(H) * T))) : H \in \mathcal{H}_{U,V,z}\}^{\wedge} \\
&= \{(H, \frac{\exp(-\mathrm{sp}(T)(\mathrm{his}(H) * T))}{\sum \exp(-\mathrm{sp}(T)(\mathrm{his}(G) * T)) : G \in \mathcal{H}_{U,V,z}}) : H \in \mathcal{H}_{U,V,z}\}
\end{aligned}$$

where exp is the exponential function. The *likely* probability of a *history*, $\tilde{P}(H)$, is inversely proportional to the bounding integer, for which the *space* is the logarithm, of the integer encoding of the *history* in the *specialising coder*. That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *substrate histories* $H \in \mathcal{H}_{U,V,z}$ with the same *specialising space*, $C^{\mathrm{s}}_{\mathrm{G,H},U,T,z}(H)$, are equally probable and all *histories* are possible, $\tilde{P}(H) > 0$. If the *transform*, T, is *known*, then the *likely probability function*, $\tilde{P}$, is *known* and an approxmation to the *expected specialising space*, ϵ, is *known*.

The *specialising space*, $\mathrm{sp}(T)(\mathrm{his}(H) * T) = C^{\mathrm{s}}_{\mathrm{G,H},U,T,z}(H)$, depends only on the *transform*, T, and the *derived*, $\mathrm{his}(H) * T$, so all *substrate histories* with the same *derived*, $\mathrm{his}(H) * T = A * T$, are equally probable. All *histories* are possible, $\tilde{P}(H) > 0$, so *specialising coder induction* is similar to *uniformly possible derived induction*, above, except that the *deriveds* are not necesssarily equally probable.

The *likely history probability function* entropy, $\mathrm{entropy}(\tilde{P})$, is maximised when the expected numerator, $\exp(-\mathrm{sp}(T)(\mathrm{his}(H) * T))$, is minimised. The *expected specialising space* is $\sum(\tilde{P}(H) \times \mathrm{sp}(T)(\mathrm{his}(H) * T) : H \in \mathcal{H}_{U,V,z}) \approx \epsilon$, so the *likely history probability function* entropy varies with the *expected specialising space*, $\mathrm{entropy}(\tilde{P}) \sim \epsilon$.

Now consider the case where, given *specialising*, it is *known*, in addition, that the *sample histogram* A is the most *probable histogram*. That is, the *likely probability* of *histogram* A,

$$\sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) = \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \times \frac{\exp(-\mathrm{sp}(T)(A * T))}{\sum \exp(-\mathrm{sp}(T)(\mathrm{his}(G) * T)) : G \in \mathcal{H}_{U,V,z}}$$

is maximised.

The *specialising probability distribution* is defined

$$\hat{Q}_{\mathrm{G,H},T,U}(z) \;\;:=\;\; \{(A, \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \times \exp(-\mathrm{sp}(T)(A * T))) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}^{\wedge}$$

The *specialising log likelihood* varies (i) with the *size* scaled *underlying entropy* (ii) against the *derived volume*, w', where the *derived volume* is less than the *size*, $w' < z$, otherwise against the *size* scaled *log derived volume*, $z \ln w'$, and (iii) with the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{G,H},T,U}(z)(A) \;\;\sim\;\; & \;\; z \times \mathrm{entropy}(A) \\ & - ((w' \;:\; w' < z) \;+\; (z \ln w' \;:\; w' \geq z)) \\ & + z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)\end{aligned}$$

In other words, the *log likelihood* is maximised where (i) the *derived volume*, w', is minimised, (ii) the *derived entropy*, $\mathrm{entropy}(A * T)$, is minimised, (iii) the *cross entropy*, $\mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T)$, is maximised, so that high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*, and (iv) the *expected component entropy*, $\mathrm{entropyComponent}(A, T)$, is maximised.

In the case of *probable sample*, the *likely history probability function* entropy varies against the *relative entropy*, $\mathrm{entropy}(\tilde{P}) \sim -\,\mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$. Similarly, the *expected specialising space* varies against the *relative entropy*, $\epsilon \sim -\,\mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$.

Given *specialising* and *probable sample*, consider the case where the *histogram* A is *known*, but the *transform*, T, is *unknown*, and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *maximum likelihood estimate* $\tilde{T}$ for the *transform*, T, is approximated by a modal value of the *specialising likelihood*,

$$\tilde{T} \;\;\in\;\; \mathrm{maxd}(\{(M, \hat{Q}_{\mathrm{G,H},M,U}(z)(A)) : M \in \mathcal{T}_{U,V}\})$$

Note that, as in the case of *uniform possible derived induction*, although computable, this optimisation is intractable because the cardinality of the *substrate transforms*, $|\mathcal{T}_{U,V}|$, is factorial in the *volume*, v.

Unlike *uniform possible derived induction*, in *specialising induction* there is no *distribution history*, H_E, and so no *sensitivity* to *distribution*, E. A *sensitivity* to *model*, T, can be defined, however, as the negative logarithm of the cardinality of the *maximum likelihood estimate models*,

$$- \ln |\max(\{(M, \hat{Q}_{\mathrm{G,H},M,U}(z)(A)) : M \in \mathcal{T}_{U,V}\})|$$

That is, as the cardinality of the modal *models* of the *log likelihood* function increases, the *sensitivity* to *model* decreases. It can be shown that the *sensitivity* to *model* varies against the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned} &- \ln |\max(\{(M, \hat{Q}_{\mathrm{G,H},M,U}(z)(A)) : M \in \mathcal{T}_{U,V}\})| \sim \\ &\quad -\big((z+v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\ &\quad\quad -z \times \mathrm{entropy}(A * T) \; - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)\big) \end{aligned}$$

So the *sensitivity* to *model* varies against the *log likelihood*,

$$- \ln |\max(\{(M, \hat{Q}_{\mathrm{G,H},M,U}(z)(A)) : M \in \mathcal{T}_{U,V}\})| \quad \sim \quad - \ln \hat{Q}_{\mathrm{G,H},T,U}(z)(A)$$

As the *relative entropy*, $\mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$, increases, the *log-likelihood*, $\ln \hat{Q}_{\mathrm{G,H},T,U}(z)(A)$, increases, but the *sensitivity* to *model*, T, decreases. In other words, the higher the *sample relative entropy* the more *likely* the *maximum likelihood estimate*, $\tilde{T}$, equals the *model*, T, and the smaller the *likely* difference between them if they are not equal.

It is shown above, in the case of *uniform possible derived* and *natural sample*, $A = A * T * T^{\dagger}$, that the *log likelihood* varies against the *derived volume* and with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned} &\ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A) \sim \\ &\quad - \; ((w' \; : \; w' < z) \; + \; (z \ln w' \; : \; w' \geq z)) \\ &\quad + \; (z+v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\ &\quad\quad -z \times \mathrm{entropy}(A * T) \; - v \times \mathrm{entropy}(V^{\mathrm{C}} * T) \end{aligned}$$

so the *iso-derived conditional log likelihood* varies with the *specialising log likelihood*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A) \quad \sim \quad \ln \hat{Q}_{\mathrm{G,H},T,U}(z)(A)$$

and the *iso-derived conditional model sensitivity* varies against the *iso-derived conditional log likelihood*,

$$\begin{aligned} &- \ln |\max(\{(M, \hat{Q}_{\mathrm{m,d},M,U}(A, z)(A)) : M \in \mathcal{T}_{U,V}, \; A = A * M * M^{\dagger}\})| \sim \\ &\quad\quad - \ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A) \end{aligned}$$

The *iso-derived conditional model sensitivity* may be compared to the *iso-derived conditional distribution sensitivity* which also varies against the *iso-derived conditional log likelihood*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m,d},T,U}(A,z))) \quad \sim \quad -\ln \hat{Q}_{\text{m,d},T,U}(A,z)(A)$$

That is, in *classical modelled induction*, the *log likelihood* is maximised and the *sensitivities* to both *distribution* and *model* are minimised where (i) the *derived volume* is minimised, (ii) the *derived entropy* is minimised, (iii) the *cross entropy* is maximised, so that high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*, and (iv) the *expected component entropy* is maximised.

2.6.3 Artificial neural networks

In the discussion of *classical modelled induction*, above, it is shown that, given *uniform possible derived* and *probable sample* $A \in \mathcal{A}_{U,V,z}$, where the *sample* is *natural*, $A = A * T * T^{\dagger}$, the *maximum likelihood estimate* $\tilde{T}$ for *unknown transform* $T \in \mathcal{T}_{U,V}$, is

$$\tilde{T} \quad \in \quad \text{maxd}(\{(M, \hat{Q}_{\text{m,d},M,U}(A,z)(A)) : M \in \mathcal{T}_{U,V},\ A = A * M * M^{\dagger}\})$$

Similarly, given *specialising* and *probable sample*, the *maximum likelihood estimate*, $\tilde{T}$, for the *transform*, T, is approximated by a modal value of the *specialising likelihood*,

$$\tilde{T} \quad \in \quad \text{maxd}(\{(M, \hat{Q}_{\text{G,H},M,U}(z)(A)) : M \in \mathcal{T}_{U,V}\})$$

In both cases, although computable, the optimisations are intractable because the cardinality of the *substrate transforms*, $|\mathcal{T}_{U,V}|$, is factorial in the *volume*, v. In order to make the optimisation tractable and then practicable, the search must be restricted to a subset of the *models*.

Artificial neural network induction is an example of practicable *classical modelled induction*. Here the *models* are artificial neural networks which correspond to *functional definition sets* of *transforms* representing the neurons. The optimisation consists of a sequence of these networks. The graph of the network remains constant, but the weights between neurons of successive networks are altered to decrease a loss function step by step. The weights of the initial network are chosen at random. The optimisation proceeds until the loss falls below a threshold. The *fud* of the terminating network is then the practicable *model*. The network graph is chosen depending on the given *sample*. In some cases of configuration the *entropy* properties of the resultant *model* are those of *classical induction*.

The *one functional transforms*, $\mathcal{T}_{U,\mathrm{f},1}$, are *derived state* valued left total functions of *underlying state*,

$$\forall T \in \mathcal{T}_{U,\mathrm{f},1}\ (\mathrm{split}(V, X^{\mathrm{S}}) \in V^{\mathrm{CS}} :\to W^{\mathrm{CS}})$$

where $(X, W) = T$ and $V = \mathrm{und}(T)$. In order to construct a coordinate from a *state* define $()^{\square} \in \mathcal{S} \to \mathcal{L}(\mathcal{W})$ as

$$S^{\square} := \{(i, u) : ((v, u), i) \in \mathrm{order}(D_{\mathcal{V}\times\mathcal{W}}, S)\}$$

where $D_{\mathcal{V}\times\mathcal{W}}$ is an *order* on the *variables* and *values*. The converse function to construct a *state* from a coordinate $()^{V} \in \mathcal{L}(\mathcal{W}) \to \mathcal{S}$ is

$$S^{V} := \{(v, S_i) : (v, i) \in \mathrm{order}(D_{\mathcal{V}}, V)\}$$

Now *one functional transforms* may be represented as *derived value* coordinate valued left total functions of *underlying value* coordinate,

$$\begin{aligned}\{(S^{\square}, R^{\square}) : (S, R) \in \mathrm{split}(V, X^{\mathrm{S}})\} &\in \{S^{\square} : S \in V^{\mathrm{CS}}\} :\to \{R^{\square} : R \in W^{\mathrm{CS}}\} \\ &\subset \mathcal{W}^{n} \to \mathcal{W}^{m}\end{aligned}$$

where $n = |V|$ and $m = |W|$.

So an alternative definition for a *one functional transform* is a tuple of (i) the *underlying variables*, V, (ii) the *derived variables*, W, and (iii) a *derived value* coordinate valued left total function of *underlying value* coordinate, f,

$$\begin{aligned}&\mathcal{T}_{U,\mathrm{f},1} = \\ &\quad \{(V, W, f) : V, W \in \mathrm{P}(\mathrm{vars}(U)),\ V \cap W = \emptyset, \\ &\qquad\qquad f \in \{S^{\square} : S \in V^{\mathrm{CS}}\} :\to \{R^{\square} : R \in W^{\mathrm{CS}}\}\}\end{aligned}$$

The *histogram* of a *function-defined one functional transform* $T = (V, W, f) \in \mathcal{T}_{U,\mathrm{f},1}$ is

$$\mathrm{histogram}(T) \ := \ \{S \cup f(S^{\square})^{W} : S \in V^{\mathrm{CS}}\} \times \{1\}$$

In the special case where the *transform* is *mono-derived-variate*, $T = (V, \{w\}, f)$, the function may be simplified to $f \in \{S^{\square} : S \in V^{\mathrm{CS}}\} :\to U_w$, and the *histogram* is

$$\mathrm{histogram}(T) := \{S \cup \{(w, f(S^{\square}))\} : S \in V^{\mathrm{CS}}\} \times \{1\}$$

In the further special case of *mono-derived-variate transform* where its *variables* are real, $\forall v \in V\ (U_v = \mathbf{R})$ and $U_w = \mathbf{R}$, then the function is a real valued left total function of a real coordinate, $f \in \mathbf{R}^n :\to \mathbf{R}$. Here the *cartesian states* are $V^{\mathrm{CS}} = \prod_{v \in V}(\{v\} \times \mathbf{R})$, so the *histogram* is

$$\begin{aligned}
\mathrm{histogram}(T) &= \{S \cup \{(w, f(S^{\square}))\} : S \in \prod_{v \in V}(\{v\} \times \mathbf{R})\} \times \{1\} \\
&= \{S^V \cup \{(w, f(S))\} : S \in \mathbf{R}^n\} \times \{1\}
\end{aligned}$$

The *cartesian volume* is infinite, $|V^{\mathrm{C}}| = |\mathbf{R}^n|$, so the cardinality of the *histogram* is infinite, $|\mathrm{histogram}(T)| = |\mathbf{R}^n|$.

The reals form a metric space so a real valued function of real coordinates may be discretised given a finite subset of the reals $D \subset \mathbf{R} : |D| < \infty$. The discretised function is

$$\mathrm{discrete}(D, n)(f) := \{(X, \mathrm{nearest}(D, f(X))) : X \in D^n\} \in D^n :\to D$$

where $\mathrm{nearest} \in \mathrm{P}(\mathbf{R}) \times \mathbf{R} \to \mathbf{R}$ is defined

$$\mathrm{nearest}(D, r) := t : \{t\} \in \mathrm{mind}(\{(s, (|r - s|, s)) : s \in D\})$$

The cardinality of the discretised *transform's histogram* is finite,

$$|\mathrm{histogram}((V, \{w\}, \mathrm{discrete}(D, n)(f)))| = |D^n| = |D|^n$$

An example of a *transform* defined by a real valued function occurs in the function composition of artificial neural networks. Here a *transform* represents a model of a neuron called a perceptron, $T = (V, w, f_\sigma(Q))$, where the *dimension* is $n = |V|$ and the function $f_\sigma(Q) \in \mathbf{R}^n :\to \mathbf{R}$ is parameterised by (i) some differentiable function $\sigma \in \mathbf{R} :\to \mathbf{R}$, called the activation function, and (ii) a vector of weights, $Q \in \mathbf{R}^{n+1}$, and is defined

$$f_\sigma(Q)(S) := \sigma(\sum_{i \in \{1 \dots n\}} Q_i S_i \ + \ Q_{n+1})$$

The function composition of artificial neural networks may be represented by *fuds* of these *transforms*. Define nets as a subset of the set of lists of tuples of the graph and real weights,

$$\mathrm{nets} \ := \ \{G : G \in \mathcal{L}(\mathrm{P}(\mathcal{V}) \times \mathcal{V} \times \mathcal{L}(\mathbf{R})),\ \forall (\cdot, (V, \cdot, Q)) \in G\ (|Q| = |V| + 1)\}$$

Define the set of *transforms*, $\mathrm{fud}(\sigma) \in \mathrm{nets} \to \mathrm{P}(\mathcal{T}_{\mathrm{f}})$ as

$$\begin{aligned}
&\mathrm{fud}(\sigma)(G) \ := \\
&\quad \{(\{S^V \cup \{(w, f_\sigma(Q)(S))\} : S \in \mathbf{R}^n\} \times \{1\}, \{w\}) : \\
&\qquad\qquad (\cdot, (V, w, Q)) \in G,\ n = |V|\}
\end{aligned}$$

The *fud* search is restricted to the *neural net substrate fud set*, $\mathcal{F}_{\infty,U,V,\sigma} = \mathcal{F}_{\infty,U,V} \cap (\mathrm{fud}(\sigma) \circ \mathrm{nets})$.

An example of a *neural net substrate fud* $F \in \mathcal{F}_{\infty,U,V,\sigma}$ has $l = \mathrm{layer}(F, \mathrm{der}(F))$ *layers* of fixed *breadth* equal to the *underlying dimension*, $\forall i \in \{1 \ldots l\}\ (|F_i| = n)$ where $n = |V|$ and $F_i = \{T : T \in F,\ \mathrm{layer}(F, \mathrm{der}(T)) = i\}$, such that the *underlying* of each *transform* is the *derived* of the *layer* below, $\forall T \in F_1\ (\mathrm{und}(T) = V)$ and $\forall i \in \{2 \ldots l\}\ \forall T \in F_i\ (\mathrm{und}(T) = \mathrm{der}(F_{i-1}))$.

The optimisation of artificial neural networks can be divided into unsupervised and supervised types. In the supervised case there is additional *knowledge*. First, there exists an *unknown distribution histogram* E from which the *known sample histogram*, A, is *drawn*, $A < E$. Secondly, the *substrate* can be partitioned into query *variables* $K \subset V$ and label *variables*, $V \setminus K$, such that the *distribution histogram*, E, is *causal* between the query *variables* and the label *variables*,

$$\mathrm{split}(K, E^{\mathrm{FS}}) \in K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$$

and so the *sample histogram*, A, is also *causal*,

$$\mathrm{split}(K, A^{\mathrm{FS}}) \in K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$$

That is, in the supervised case, there is a functional relation such that there is exactly one label *state* for every *effective* query *state*. In an optimisation, a *fud* $F \in \mathcal{F}_{\infty,U,K,\sigma}$ has its *underlying variables* restricted to the query *variables*, $\mathrm{und}(F) \subseteq K$. The optimisation maximises the *causality* between the *derived variables* and the label *variables* by minimising the loss function. At the optimum there is no error and the relation is functional,

$$\mathrm{split}(W_F, (A * X_F \ \% \ (W_F \cup V \setminus K))^{\mathrm{FS}}) \in W_F^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$$

where $X_F = \mathrm{histogram}(F^{\mathrm{T}})$ and $W_F = \mathrm{der}(F)$. At zero loss the label *state* is implied for all query *states* that are *effective* in the *sample derived*,

$$\mathrm{split}(K, (K^{\mathrm{C}} * F^{\mathrm{T}} * (A * X_F) \ \% \ (V \setminus K))^{\mathrm{FS}}) \in K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$$

That is, a query *state* $Q \in K^{\mathrm{CS}}$ that is *effective* in the *sample derived* $R \in (A * F^{\mathrm{T}})^{\mathrm{FS}}$, where $\{R\} = (\{Q\}^{\mathrm{U}} * F^{\mathrm{T}})^{\mathrm{FS}}$, but that is not necessarily *effective* in the *sample* itself, $Q \notin (A\%K)^{\mathrm{FS}}$, still has an implied label *state*, $\{L\} = (A * X_F * \{R\}^{\mathrm{U}} \ \% \ (V \setminus K))^{\mathrm{FS}}$ where $L \in (V \setminus K)^{\mathrm{CS}}$.

In the case where the *derived variables* of the *fud* is a *literal frame* of the label *variables*, $W_F :\leftrightarrow: (V \setminus K)$ and $\forall v \in (V \setminus K)\ (U_v \subseteq \mathbf{R})$, the least squares loss function $\mathrm{lsq} \in \mathcal{A} \times \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathbf{R}$ is

$$\mathrm{lsq}(A,F,K) \;:=\; \sum_{(S,c) \in A * X_F} \big(c \times \sum_{i \in \{1 \ldots m\}} ((S\%W_F)^{[]}_i - (S\%(V \setminus K))^{[]}_i)^2\big)$$

where $m = |W_F| = |(V \setminus K)|$. The loss function is a continuous real valued function and so its derivative with respect to each weight can be defined. In this case the optimisation is least squares gradient descent.

If the optimisation of artificial neural networks is of the unsupervised type, there is no *knowledge* of a *causal* label. Here the method of least squares gradient descent is still used but the label is simply a copy of the *substrate*, V, itself. Usually the network graph is constrained so that a middle *layer* $a \in \{2 \ldots l-1\}$ has narrower *breadth* than the *substrate*, $|F_a| < n$.

In the computations of *alignment* and *entropy* that follow, the *derived variables* are *discretised* to the *values* of the label *variables*, $D = \bigcup\{U_v : v \in (V \setminus K)\}$.

In some cases of *sample* and network optimisation configuration, the negative least squares loss (a) varies against the *effective derived volume*

$$-\,\mathrm{lsq}(A,F_D,K) \;\sim\; -\,|(A * F_D^{\mathrm{T}})^{\mathrm{F}}|$$

(b) varies against the *derived entropy* of the *fud transform*,

$$-\,\mathrm{lsq}(A,F_D,K) \;\sim\; -\,\mathrm{entropy}(A * F_D^{\mathrm{T}})$$

(c) varies with the *component size cardinality relative entropy*,

$$-\,\mathrm{lsq}(A,F_D,K) \;\sim\; \mathrm{entropyRelative}(A * F_D^{\mathrm{T}}, V^{\mathrm{C}} * F_D^{\mathrm{T}})$$

and (d) varies with the *expected component entropy*,

$$-\,\mathrm{lsq}(A,F_D,K) \;\sim\; \mathrm{entropyComponent}(A, F_D^{\mathrm{T}})$$

The initial *fud* F_R has arbitrary weights, so is likely to have a high least squares loss. That is, far from the *derived variables* and the label *variables* being *causally* related, $W_D^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$, they are likely to be *independent*,

$$\mathrm{algn}(A * X_{F_R} * \{W_D^{\mathrm{CS}\{\}\mathrm{T}}, (V \setminus K)^{\mathrm{CS}\{\}\mathrm{T}}\}^{\mathrm{T}}) \;\approx\; 0$$

where $\{W_D^{\mathrm{CS}\{\}\mathrm{T}}, (V \setminus K)^{\mathrm{CS}\{\}\mathrm{T}}\}$ is the *fud* of the *self transforms* of the (i) *discretised derived variables* and (ii) label *variables*.

As the optimisation proceeds from the initial *fud*, F_R, to the optimal *fud* F, the loss decreases and the relation between the top *layer* and the label becomes more *causal*,

$$\mathrm{algn}(A * X_F * \{W_D^{\mathrm{CS}\{\}\mathrm{T}}, (V \setminus K)^{\mathrm{CS}\{\}\mathrm{T}}\}^{\mathrm{T}}) \ > \ 0$$

The negative least squares loss varies with the *alignment* of the *self partition transforms*, so varies against the *derived entropy* of the *fud transform*,

$$\begin{aligned} -\ \mathrm{lsq}(A, F_D, K) &\sim \mathrm{algn}(A * X_F * \{W_D^{\mathrm{CS}\{\}\mathrm{T}}, (V \setminus K)^{\mathrm{CS}\{\}\mathrm{T}}\}^{\mathrm{T}}) \\ &\sim -z \times \mathrm{entropy}(A * F_D^{\mathrm{T}}) \end{aligned}$$

That is, as the loss, $\mathrm{lsq}(A, F_D, K)$, is minimised, the *derived entropy*, $\mathrm{entropy}(A * F_D^{\mathrm{T}})$, tends to be minimised. The minimisation of *derived entropy* is a property of *classical induction*.

The negative least squares loss only varies with the *component size cardinality relative entropy*, $\mathrm{entropyRelative}(A * F_D^{\mathrm{T}}, V^{\mathrm{C}} * F_D^{\mathrm{T}})$, in the case where the *histogram*, A, is clustered by the label *variables*. This requires *alignment* within the query *variables*, $\mathrm{algn}(A\%K) > 0$. Clustering may be described as follows.

Consider the case of a *multi-variate* set of real *valued* query *variables* K, where $k = |K| \geq 2$ and $\forall x \in K\ (U_x \subseteq \mathbf{R})$, and a *neural net fud* $F \in \mathcal{F}_{\infty,U,K,\sigma}$ consisting of two *transforms*, $F = \{T_1, T_2\}$, each having the query *variables* as the *underlying*, $\mathrm{und}(T_1) = \mathrm{und}(T_2) = K$. For a coordinate $S \in \mathbf{R}^k$ the weights of the *transforms* form a pair of hyperplanes,

$$\sum_{i \in \{1 \ldots k\}} Q_{1,i} S_i \ + \ Q_{1,k+1} \ = \ 0$$

and

$$\sum_{i \in \{1 \ldots k\}} Q_{2,i} S_i \ + \ Q_{2,k+1} \ = \ 0$$

where $Q_1, Q_2 \in \mathbf{R}^{k+1}$ are the weights correspondng to T_1, T_2. If the hyperplanes of the arbitrarily weighted initial *fud*, F_R, intersect, the acute angle between them is expected to be $45°$. That is, given an activation function,

σ, which is a step function, or a binary set of *discrete values*, $D = \{0, 1\}$, the probability distribution of the *component cardinalities* of the initial *fud* is bi-modal. If $(\cdot, C_1), (\cdot, C_2) \in (F^{\mathrm{T}}_{R,\{0,1\}})^{-1}$ are such that $|C_1| < |C_2|$, then it is expected that $3|C_1| = |C_2|$. So the *component cardinality entropy* of the initial *fud* is expected to be less than maximal,

$$\mathrm{entropy}(K^{\mathrm{C}} * F^{\mathrm{T}}_{R,D}) \ < \ \mathrm{entropy}(W^{\mathrm{C}}_D)$$

The *derived entropy* of the initial *fud* is expected to be approximately equal to the *component cardinality entropy*,

$$\mathrm{entropy}(A * F^{\mathrm{T}}_{R,D}) \ \approx \ \mathrm{entropy}(K^{\mathrm{C}} * F^{\mathrm{T}}_{R,D})$$

and so the *component size cardinality relative entropy* of the initial *fud* is expected to be small,

$$\mathrm{entropyRelative}(A * F^{\mathrm{T}}_{R,D}, K^{\mathrm{C}} * F^{\mathrm{T}}_{R,D}) \ \approx \ 0$$

If the *histogram*, A, is approximately uniformly distributed over the *volume*, then the *component size cardinality relative entropy* remains small during the optimisation,

$$\mathrm{entropyRelative}(A * F^{\mathrm{T}}_{D}, K^{\mathrm{C}} * F^{\mathrm{T}}_{D}) \ \approx \ 0$$

In contrast, consider the case where the *histogram*, A, is not uniformly distributed, but clustered by label *state*. Let $Y_L \subset K^{\mathrm{CS}}$ be the set of the centres of the clusters for *effective* label *state* $L \in (A\%(V \setminus K))^{\mathrm{FS}}$. The maximum radius $r_L \in \mathbf{R}_{>0}$ is such that

$$\forall S \in A^{\mathrm{FS}} \ \Diamond L = S\%(V \setminus K) \ \exists Q \in Y_L \ (\sum_{i \in \{1 \dots k\}} (Q^{[]}_i - S^{[]}_i)^2 \ \leq \ r_L^2)$$

Let r_C be the radius of *component* C. In the case where the *histogram* is clustered such that the cluster radius of a label *state* is much smaller than the least initial *component* radius, $\forall (\cdot, C) \in (F^{\mathrm{T}}_{R,\{0,1\}})^{-1} \ (r_L \ll r_C)$, then optimised rotations of the hyperplanes, that sweep up nearby clusters in the same label *state*, tend to be such that the magnitude of the change in the fractional *component size*, $|(A * F^{\mathrm{T}}_{2,D})(R) - (A * F^{\mathrm{T}}_{1,D})(R)|/z$, is greater than magnitude of the change in the fractional *component cardinality*, $|(K^{\mathrm{C}} * F^{\mathrm{T}}_{2,D})(R) - (K^{\mathrm{C}} * F^{\mathrm{T}}_{1,D})(R)|/|K^{\mathrm{C}}|$. So, in the clustered case, as the optimisation decreases the *derived entropy*, $\mathrm{entropy}(A * F^{\mathrm{T}}_D)$, the *component sizes* and *component cardinalities* become less synchronised and the *component size* *cardinality relative entropy* increases,

$$\begin{aligned} -\ \mathrm{lsq}(A, F_D, K) &\sim -z \times \mathrm{entropy}(A * F_D^{\mathrm{T}}) \\ &\sim z \times \mathrm{entropyRelative}(A * F_D^{\mathrm{T}}, K^{\mathrm{C}} * F_D^{\mathrm{T}}) \\ &= z \times \mathrm{entropyRelative}(A * F_D^{\mathrm{T}}, V^{\mathrm{C}} * F_D^{\mathrm{T}}) \end{aligned}$$

The same reasoning applies to *fuds* consisting of more than two *transforms*, $|F| > 2$, but note that at higher *fud* cardinalities the initial *component cardinality entropy*, $\mathrm{entropy}(K^{\mathrm{C}} * F_{R,D}^{\mathrm{T}})$, tends to be multi-modal and so approximates more closely to the *uniform cartesian derived entropy*, $\mathrm{entropy}(W_D^{\mathrm{C}})$. So there is less freedom for the *relative entropy* of the *fud* to increase during optimisation. In the case of *multi-layer fuds*, however, the *breadth* can be constrained and so the *relative entropy* of deeper, narrrower *fuds* may be higher than in shallower, wider *fuds* of the same cardinality.

In general, in the clustered case, the optimised *fud* is such that high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*. The maximisation of *relative entropy* is a property of *classical induction*.

The *accuracy* of the approximation of *artificial neural network induction* to *classical induction* can be defined as the ratio of the practicable *model sample-distributed iso-derived conditional log likelihood* to the maximum *model sample-distributed iso-derived conditional log likelihood*,

$$0 \ < \ \frac{\hat{Q}_{\mathrm{m,d},F^{\mathrm{T}},U}(A,z)(A)}{\hat{Q}_{\mathrm{m,d},\tilde{T},U}(A,z)(A)} \ \leq \ 1$$

The *accuracy* varies against the *sensitivity* to *model*,

$$\frac{\hat{Q}_{\mathrm{m,d},F^{\mathrm{T}},U}(A,z)(A)}{\hat{Q}_{\mathrm{m,d},\tilde{T},U}(A,z)(A)} \ \sim \ -(-\ln|\max(\{(M, \hat{Q}_{\mathrm{m,d},M,U}(A,z)(A)) : M \in \mathcal{T}_{U,V}\})|)$$

and so varies with the *log-likelihood*,

$$\frac{\hat{Q}_{\mathrm{m,d},F^{\mathrm{T}},U}(A,z)(A)}{\hat{Q}_{\mathrm{m,d},\tilde{T},U}(A,z)(A)} \ \sim \ \ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A)$$

That is, although the *model* obtained from *least squares gradient descent* is merely an approximation, in the cases where the *log-likelihood* is high, and so the *sensitivity* to *model* is low, the approximation may be reasonably close nonetheless.

2.6.4 Aligned induction

Given *substrate transform* $T \in \mathcal{T}_{U,V}$, the *abstract histogram* valued *integral substrate histograms* function $Y_{U,\mathrm{i},T,\mathrm{W},z}$ is defined

$$Y_{U,\mathrm{i},T,\mathrm{W},z} \;:=\; \{(A, (A * T)^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The finite set of *iso-abstracts* of *abstract histogram* $(A * T)^{\mathrm{X}}$ is

$$Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \;=\; \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}$$

The degree to which an *integral iso-set* $I \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$ that contains the *histogram*, $A \in I$, is said to be *entity-like* is called the *iso-abstractence*. The *iso-abstractence* is defined as the ratio of (i) the cardinality of the intersection between the *integral iso-set* and the set of *integral iso-abstracts*, and (ii) the cardinality of the union,

$$\frac{1}{|\mathcal{A}_{U,\mathrm{i},V,z}|} \;\leq\; \frac{|I \;\cap\; Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}{|I \;\cup\; Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \;\leq\; 1$$

Law-like iso-sets are subsets of the set of *iso-abstracts*,

$$D^{-1}_{U,\mathrm{i},T,z}(A * T) \;\subseteq\; Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

and so are also *entity-like*.

The *formal histogram* valued *integral substrate histograms* function $Y_{U,\mathrm{i},T,\mathrm{V},z}$ is defined

$$Y_{U,\mathrm{i},T,\mathrm{V},z} \;:=\; \{(A, A^{\mathrm{X}} * T) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The finite set of *iso-formals* of *formal histogram* $A^{\mathrm{X}} * T$ is

$$Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \;=\; \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T\}$$

Aligned-like iso-sets are subsets of the set of *iso-formals*,

$$Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \;\subseteq\; Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)$$

The *formal-abstract* pair valued *integral substrate histograms* function $Y_{U,\mathrm{i},T,z}$ is defined

$$Y_{U,\mathrm{i},T,z} \;:=\; \{(A, (A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The finite set of *iso-transform-independents* of $(A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})$ is

$$\begin{aligned} Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})) &= \\ &\{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\} \end{aligned}$$

The *iso-transform-independents* is the intersection of the *iso-formals* and the *iso-abstracts*,

$$Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})) = Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

In *aligned modelled induction* the *history probability functions* are constrained by *formal* and *abstract histograms*.

Let P be a *substrate history probability function*, $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$. Given a *history* $H_E \in \mathcal{H}_{U,V,z_E}$, of *size* $z_E = |H_E|$, consider the case where both the *formal histogram* $A^{\mathrm{X}} * T$ of *drawn histories* is *known* to be *necessary* and the *abstract histogram* $(A * T)^{\mathrm{X}}$ of *drawn histories* is *known* to be *necessary*, $\sum(P(H) : H \subseteq H_E,\ \mathrm{his}(H)^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (\mathrm{his}(H) * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}) = 1$. The *maximum likelihood estimate* which maximises the entropy, $\mathrm{entropy}(\tilde{P})$, is

$$\begin{aligned} \tilde{P} &= \{(H, 1) : \\ &\qquad H \subseteq H_E,\ \mathrm{his}(H)^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (\mathrm{his}(H) * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}^{\wedge} \cup \\ &\quad \{(G, 0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\} \cup \\ &\quad \{(G, 0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G)^{\mathrm{X}} * T \neq A^{\mathrm{X}} * T\} \cup \\ &\quad \{(G, 0) : G \in \mathcal{H}_{U,V,z},\ (\mathrm{his}(G) * T)^{\mathrm{X}} \neq (A * T)^{\mathrm{X}}\} \\ &= \{(H, 1/\textstyle\sum(Q_{\mathrm{h},U}(E, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})))) : \\ &\qquad H \subseteq H_E,\ \mathrm{his}(H)^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (\mathrm{his}(H) * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\} \cup \\ &\quad \{(G, 0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\} \cup \\ &\quad \{(G, 0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G)^{\mathrm{X}} * T \neq A^{\mathrm{X}} * T\} \cup \\ &\quad \{(G, 0) : G \in \mathcal{H}_{U,V,z},\ (\mathrm{his}(G) * T)^{\mathrm{X}} \neq (A * T)^{\mathrm{X}}\} \end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn histories* $H \subseteq H_E$ with both the *formal*, $\mathrm{his}(H)^{\mathrm{X}} * T = A^{\mathrm{X}} * T$ and the *abstract*, $(\mathrm{his}(H) * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$, are uniformly probable and all other *histories*, $G \nsubseteq H_E$ or $\mathrm{his}(G)^{\mathrm{X}} * T \neq A^{\mathrm{X}} * T$ or $(\mathrm{his}(G) * T)^{\mathrm{X}} \neq (A * T)^{\mathrm{X}}$, are impossible, $\tilde{P}(G) = 0$. If (i) the *transform*, T, is *known*, (ii) the *formal*, $A^{\mathrm{X}} * T$, is *known*, (iii) the *abstract*, $(A * T)^{\mathrm{X}}$, is *known* and (iv) the *distribution histogram*, H_E, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

The *likely probability* of *drawing histogram* A from *necessary drawn formal* $A^{\mathrm{X}} * T$ and *necessary drawn abstract* $(A * T)^{\mathrm{X}}$ is

$$\sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) \ = \ \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum Q_{\mathrm{h},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))}$$

The *likely history probability function* entropy, $\mathrm{entropy}(\tilde{P})$, is maximised when the sum of the *iso-transform-independent historical frequencies*,

$$\sum Q_{\mathrm{h},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))$$

is maximised.

Consider the case where the *transform*, T, is *known*, the *formal*, $A^{\mathrm{X}} * T$, is *known*, and the *abstract*, $(A * T)^{\mathrm{X}}$, is *known*, but the *distribution histogram*, E, is *unknown* and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *maximum likelihood estimate* $\tilde{E}$ for the *distribution histogram*, E, is a modal value of the *likelihood function*,

$$\tilde{E} \ \in \ \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{h},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})))) : D \in \mathcal{A}_{U,\mathrm{i},V,z_E}\})$$

The *likely distribution histogram*, $\tilde{E}$, is *known* if the *distribution histogram size*, z_E, is *known*, the *transform*, T, is *known*, the *formal*, $A^{\mathrm{X}} * T$, is *known*, and the *abstract*, $(A * T)^{\mathrm{X}}$, is *known*. If it is assumed that the *distribution histogram* equals the *likely distribution histogram*, $E = \tilde{E}$, then the *likely history probability* is *known*, $\tilde{P}(H) = 1/\sum(Q_{\mathrm{h},U}(\tilde{E},z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})))$ where $\mathrm{his}(H)^{\mathrm{X}} * T = A^{\mathrm{X}} * T$ and $(\mathrm{his}(H) * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$.

In the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $\tilde{E}$ for the *distribution probability histogram*, $\hat{E}$, may be approximated by a modal value of a *likelihood function* which depends on the *multinomial distribution* instead,

$$\tilde{E} \ \in \ \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})))) : D \in \mathcal{A}_{U,V,1}\})$$

If it is *known*, in addition, that the *formal* equals the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, then the *normalised naturalised abstract*, $(\hat{A} * T)^{\mathrm{X}} * T^{\dagger}$, is a solution.

In this case the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, or *naturalised formal*, $A^{\mathrm{X}} * T * T^{\dagger} = (A * T)^{\mathrm{X}} * T^{\dagger}$, is the *independent analogue* of the *iso-transform-independents*. So the *maximum likelihood estimate*, $\tilde{E}$, for the *distribution probability histogram*, $\hat{E}$, is the *naturalised abstract probability histogram*, $(\hat{A} * T)^{\mathrm{X}} * T^{\dagger}$,

$$\tilde{E} \ = \ (\hat{A} * T)^{\mathrm{X}} * T^{\dagger}$$

Formal-abstract equivalence, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, is also called *mid transform*. In this case the *abstract* equals the *independent abstract*, $(A * T)^{\mathrm{X}} = A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, and so does not depend on the *histogram alignment*, $\mathrm{algn}(A)$. The *formal* equals the *formal independent*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, and so does not depend on its own *alignment*, $\mathrm{algn}(A^{\mathrm{X}} * T) = 0$.

The *naturalised abstract* is the *independent analogue* of the *iso-transform-independents*, so, in the case where the *naturalised abstract* is *integral*, $(A * T)^{\mathrm{X}} * T^{\dagger} \in \mathcal{A}_{\mathrm{i}}$, the sum of the *iso-transform-independent naturalised-abstract-distributed multinomial probabilities* varies with the *naturalised-abstract naturalised abstract-distributed multinomial probability*,

$$\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})) \ \sim \\ Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)((A * T)^{\mathrm{X}} * T^{\dagger})$$

So, if it is assumed that the *distribution probability histogram* equals the *likely distribution probability histogram*, $\hat{E} = \tilde{E} = (\hat{A} * T)^{\mathrm{X}} * T^{\dagger}$, then the *likely history probability* varies against the *naturalised-abstract-distributed multinomial probability* of the *naturalised abstract*, $\tilde{P}(H) \sim 1/\hat{Q}_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)((A * T)^{\mathrm{X}} * T^{\dagger})$. The *likely history probability function* entropy varies with the *naturalised abstract entropy*, $\mathrm{entropy}(\tilde{P}) \sim \mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\dagger})$.

Given *necessary formal*, *necessary abstract* and *mid transform*, consider the case where a *drawn histogram* A is *known*, but neither the *distribution histogram*, E, is *known* nor the *transform*, T, is *known*, and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. The *maximum likelihood estimate* $(\tilde{E}, \tilde{T})$ for the pair of the *distribution histogram*, E, and the *transform*, T, is a modal value of the *likelihood function*,

$$(\tilde{E}, \tilde{T}) \ \in \\ \mathrm{maxd}(\{((D, M), \sum (Q_{\mathrm{h},U}(D, z)(B) : B \in Y^{-1}_{U,\mathrm{i},M,z}((A^{\mathrm{X}} * M, (A * M)^{\mathrm{X}})))) : \\ D \in \mathcal{A}_{U,\mathrm{i},V,z_E}, \ M \in \mathcal{T}_{U,V}, \ A^{\mathrm{X}} * M = (A * M)^{\mathrm{X}}\})$$

In some cases of *drawn sample*, A, the *transform maximum likelihood estimate*, $\tilde{T}$, is not trivial. That is, the *transform maximum likelihood estimate* is not necessarily *unary*, $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, nor *self*, $T_{\mathrm{s}} = V^{\mathrm{CS}\{\}\mathrm{T}}$. In the cases where the *transform maximum likelihood estimate* is trivial, $\tilde{T} \in \{T_{\mathrm{u}}, T_{\mathrm{s}}\}$, *aligned modelled induction* reduces to *aligned non-modelled induction*,

$$\begin{aligned}
\tilde{P} \;=\; & \{(H,1) : H \subseteq H_E,\ \mathrm{his}(H)^{\mathrm{X}} = A^{\mathrm{X}}\}^{\wedge} \ \cup \\
& \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \nsubseteq H_E\} \ \cup \\
& \{(G,0) : G \in \mathcal{H}_{U,V,z},\ \mathrm{his}(G)^{\mathrm{X}} \neq A^{\mathrm{X}}\}
\end{aligned}$$

Define the *transform-dependent* $A^{\mathrm{Y}(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-transform-independent*,

$$\begin{aligned}
\{A^{\mathrm{Y}(T)}\} \;=\; & \\
& \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))}) : \\
& \qquad D \in \mathcal{A}_{U,V,z}\})
\end{aligned}$$

The *transform-dependent*, $A^{\mathrm{Y}(T)}$, is the *dependent analogue* of the *iso transform independents*. Note that the *transform-dependent*, $A^{\mathrm{Y}(T)}$, is not always computable, but an approximation to any accuracy can be made to it. In the case where the *formal* equals the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, and the *histogram* equals the *naturalised abstract*, the *transform-dependent* equals the *naturalised abstract*, $A = (A * T)^{\mathrm{X}} * T^{\dagger} \implies A^{\mathrm{Y}(T)} = A = (A * T)^{\mathrm{X}} * T^{\dagger}$.

Now consider the case where, given *necessary formal*, *necessary abstract* and *mid transform*, it is *known*, in addition, that the *sample histogram* A is the most *probable histogram* of the *iso-transform-independents*. That is, the *likely probability* of *drawing histogram* A from *necessary formal-abstract* $(A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})$,

$$\begin{aligned}
& \sum(\tilde{P}(H) : H \in \mathcal{H}_{U,V,z},\ \mathrm{his}(H) = A) \;=\; \\
& \qquad \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum Q_{\mathrm{h},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))}
\end{aligned}$$

is maximised.

In the case where the *transform*, T, is *known* and the *sample*, A, is *known*, but the *distribution histogram*, E, is *unknown*, the *maximum likelihood estimate* $\tilde{E}$ for the *distribution histogram*, E, is a modal value of the *likelihood function*,

$$\begin{aligned}\tilde{E} &\in \text{maxd}(\{(D, \frac{Q_{\text{h},U}(D,z)(A)}{\sum Q_{\text{h},U}(D,z)(B) : B \in Y^{-1}_{U,\text{i},T,z}((A^{\text{X}} * T, (A * T)^{\text{X}}))}) : \\ &\qquad D \in \mathcal{A}_{U,\text{i},V,z_E}\})\end{aligned}$$

The *likely distribution histogram*, $\tilde{E}$, is *known* if the *distribution histogram size*, z_E, is *known*, the *transform*, T, is *known* and the *sample*, A, is *known*. If it is assumed that the *distribution histogram* equals the *likely distribution histogram*, $E = \tilde{E}$, then the *likely history probability* is *known*, $\tilde{P}(H) = 1/\sum(Q_{\text{h},U}(\tilde{E},z)(B) : B \in Y^{-1}_{U,\text{i},T,z}((A^{\text{X}} * T, (A * T)^{\text{X}})))$ where $\text{his}(H)^{\text{X}} * T = A^{\text{X}} * T$ and $(\text{his}(H) * T)^{\text{X}} = (A * T)^{\text{X}}$.

If the *histogram* is *naturalised abstract*, $A = (A * T)^{\text{X}} * T^{\dagger}$, then the additional constraint of *probable sample* makes no change to the *maximum likelihood estimate*, $\tilde{E}$,

$$\begin{aligned}&A = (A * T)^{\text{X}} * T^{\dagger} \implies \\ &\quad \text{maxd}(\{(D, \frac{Q_{\text{h},U}(D,z)(A)}{\sum Q_{\text{h},U}(D,z)(B) : B \in Y^{-1}_{U,\text{i},T,z}((A^{\text{X}} * T, (A * T)^{\text{X}}))}) : \\ &\qquad D \in \mathcal{A}_{U,\text{i},V,z_E}\}) \\ &= \text{maxd}(\{(D, \sum(Q_{\text{h},U}(D,z)(B) : B \in Y^{-1}_{U,\text{i},T,z}((A^{\text{X}} * T, (A * T)^{\text{X}})))) : \\ &\qquad D \in \mathcal{A}_{U,\text{i},V,z_E}\})\end{aligned}$$

If the *histogram* is not *naturalised abstract*, $A \neq (A * T)^{\text{X}} * T^{\dagger}$, however, then the *likely history probability function* entropy, $\text{entropy}(\tilde{P})$, is lower than it is in the case of *necessary formal-abstract* unconstrained by *probable sample*.

In the case where the *distribution histogram*, E, is *unknown*, and the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $\tilde{E}$ for the *distribution probability histogram*, $\hat{E}$, is now approximated by a modal value of the conditional *likelihood function*,

$$\begin{aligned}\tilde{E} &\in \text{maxd}(\{(D, \frac{Q_{\text{m},U}(D,z)(A)}{\sum Q_{\text{m},U}(D,z)(B) : B \in Y^{-1}_{U,\text{i},T,z}((A^{\text{X}} * T, (A * T)^{\text{X}}))}) : \\ &\qquad D \in \mathcal{A}_{U,V,1}\})\end{aligned}$$

The solution to this is the *normalised transform-dependent*, $\tilde{E} = \hat{A}^{\text{Y}(T)}$. The *maximum likelihood estimate* is near the *sample*, $\tilde{E} \sim \hat{A}$, only in as much as it is far from the *naturalised abstract*, $\tilde{E} \nsim (\hat{A} * T)^{\mathrm{X}} * T^{\dagger}$.

The *iso-transform-independent conditional multinomial probability distribution* is defined

$$\begin{aligned}\hat{Q}_{\mathrm{m,y},T,U}(E,z)(A) &:= \\ &\frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},T,z})|}\frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum Q_{\mathrm{m},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))}\end{aligned}$$

So the optimisation can be rewritten,

$$\tilde{E} \in \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{m,y},T,U}(D,z)(A)) : D \in \mathcal{A}_{U,V,1}\})$$

Consider the case where the *distribution* equals the *transform-dependent*, $\hat{E} = \hat{A}^{\mathrm{Y}(T)}$. First, the logarithm of the *iso-transform-independent conditional multinomial probability* of the *histogram*, A, with respect to the *dependent analogue* or *transform-dependent*, $A^{\mathrm{Y}(T)}$, varies against the logarithm of the *iso-transform-independent conditional multinomial probability* with respect to the *independent analogue* or *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$,

$$\begin{aligned}&\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))} \\ \sim\quad &-\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger},z)(A)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger},z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))}\end{aligned}$$

Second, the negative logarithm of the *iso-transform-independent conditional multinomial probability* of the *histogram*, A, with respect to the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, varies with the negative logarithm of the *lifted iso-transform-independent conditional multinomial probability* of the *derived*, $A * T$, with respect to the *abstract*, $(A * T)^{\mathrm{X}}$,

$$\begin{aligned}&-\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger},z)(A)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger},z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))} \\ \sim\quad &-\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}},z)(A * T)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}},z)(B') : B' \in Y'^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))}\end{aligned}$$

where $Y'^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})) = \{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))\}$.

Third, the negative logarithm of the *lifted iso-transform-independent conditional multinomial probability* with respect to the *abstract*, $(A * T)^{\mathrm{X}}$, varies with the negative logarithm of the *relative multinomial probability* with respect to the *abstract*, $(A * T)^{\mathrm{X}}$, which is the *derived alignment*,

$$\begin{aligned}
& -\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(B') : B' \in Y'^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))} \\
\sim\ & -\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)((A * T)^{\mathrm{X}})} \\
=\ & \mathrm{algn}(A * T)
\end{aligned}$$

So the *log-likelihood* varies with the *derived alignment*,

$$\ln \hat{Q}_{\mathrm{m},\mathrm{y},T,U}(A^{\mathrm{Y}(T)}, z)(A) \quad \sim \quad \mathrm{algn}(A * T)$$

The *mid transform* constraint allows the *log-likelihood*, which is a function of the *histogram*, A, to be *lifted* to the *derived alignment*, which is a function of the *derived*, $A * T$. So a *model* optimisation need only search in the *derived volume*, $|T^{-1}|$, which is typically much smaller than the *underlying volume*, $|T^{-1}| \ll |V^{\mathrm{C}}|$. It is this relation between the *log-likelihood* and the *derived alignment* that makes *aligned induction* practicable.

The case of *classical modelled induction*, where the *derived* is *necessary*, may be termed *law-like* because the set of *iso-derived*, $D^{-1}_{U,\mathrm{i},T,z}(A * T)$, is *law-like*. All *drawn histories* $H \subseteq H_E$, are such that their *derived histograms* are fixed, $\mathrm{his}(H) * T = A * T$.

By contrast, the case of *aligned modelled induction*, where the *abstract* is *necessary*, may be termed *entity-like* because the set of *iso-abstracts*, $Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$, is *entity-like*. All *drawn histories* are such that their *abstract histograms* are fixed, $(\mathrm{his}(H) * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$. That is, the *derived variables* are separately *necessary*, $\forall u \in W\ (\mathrm{his}(H) * T \ \% \ \{u\} = A * T \ \% \ \{u\})$. *Necessary abstract* is a weaker constraint than *necessary derived* because the *iso-abstracts* are a superset of the *iso-derived*, $D^{-1}_{U,\mathrm{i},T,z}(A * T) \subseteq Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$. In fact, *aligned induction* is stricter than pure *entity-like* because the *formal* is *necessary* too, $\mathrm{his}(H)^{\mathrm{X}} * T = A^{\mathrm{X}} * T$, and so *aligned induction* is also *aligned-like*, $Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \subseteq Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)$. *Aligned induction*, however, is not necessarily *law-like*, $\mathrm{his}(H) * T \neq A * T$, and so does not always approximate to *classical induction*. *Mid transform* is stricter still, but this constraint does not necessarily increase *law-likeness*, but merely allows *lifting*.

Consider the case where, given *necessary formal*, *necessary abstract*, *mid* *transform* and *probable sample*, it is *known*, in addition, that the *sample histogram* is *ideal*, $A = A * T * T^{\dagger A}$. The *idealisation independent* equals the *independent*, $(A*T*T^{\dagger A})^{\mathrm{X}} = A^{\mathrm{X}}$, so the *idealisation* is *aligned-like*. The *ideal sample* approximates to the *independent analogue* of the *iso-derived*, which is the *naturalisation*, $A \approx A*T*T^{\dagger}$, and so, if it is also the case that *derived alignment* is high, $\mathrm{algn}(A * T) \gg 0$, the *iso-transform-independent conditional multinomial log-likelihood* with respect to the *dependent analogue* or *transform-dependent*, $A^{\mathrm{Y}(T)}$, varies with the *iso-derived conditional multinomial log-likelihood* with respect to the *independent analogue* or *naturalisation*, $A * T * T^{\dagger}$,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{m,y},T,U}(A^{\mathrm{Y}(T)}, z)(A) &\sim \ln \hat{Q}_{\mathrm{m,d},T,U}(A * T * T^{\dagger}, z)(A) \\ &\sim \ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A)\end{aligned}$$

So the *log likelihood* varies with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{m,y},T,U}(A^{\mathrm{Y}(T)}, z)(A) \sim \\ &\quad (z+v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\ &\qquad\qquad -z \times \mathrm{entropy}(A * T) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)\end{aligned}$$

and the *maximum likelihood estimate derived* approximates to the *normalised sample derived*,

$$\begin{aligned}\tilde{E} * T &= \hat{A}^{\mathrm{Y}(T)} * T \\ &\approx \hat{A} * T\end{aligned}$$

In the case where the *underlying alignment* is intermediate, $\mathrm{algn}(A) \gg 0$, and the *component size cardinality relative entropy* is high, $\mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) > \ln |T^{-1}|$, the *sum sensitivity* varies against the *log likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,y},T,U}(A^{\mathrm{Y}(T)}, z))) \ \sim\ -\ \ln \hat{Q}_{\mathrm{m,y},T,U}(A^{\mathrm{Y}(T)}, z)(A)$$

and the *model sensitivity* varies against the *log likelihood*,

$$\begin{aligned}&-\ \ln |\max(\{(M, \hat{Q}_{\mathrm{m,y},M,U}(A^{\mathrm{Y}(M)}, z)(A)) : M \in \mathcal{T}_{U,V}, \\ &\qquad\qquad A^{\mathrm{X}} * M = (A * M)^{\mathrm{X}},\ A = A * M * M^{\dagger A}\})| \\ &\qquad\qquad\qquad\qquad \sim\ -\ \ln \hat{Q}_{\mathrm{m,y},T,U}(A^{\mathrm{Y}(T)}, z)(A)\end{aligned}$$

That is, given *mid-ideal transform*, the maximisation of the *derived alignment* tends to make the properties of *aligned modelled induction* similar to those of *classical modelled induction*.

Given *necessary formal-abstract*, *mid-ideal transform* and *probable sample*, consider the case where a *drawn histogram* A is *known*, but neither the *distribution histogram*, E, is *known* nor the *transform*, T, is *known*, and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. In the case where the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $(\tilde{E}, \tilde{T})$ for the pair of the *distribution histogram*, E, and the *transform*, T, is approximated by a modal value of the conditional *likelihood function*,

$$
\begin{aligned}
(\tilde{E},\tilde{T}) \in \\
&\mathrm{maxd}(\{((D,M), \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},M,z}((A^{\mathrm{X}} * M, (A * M)^{\mathrm{X}}))}) : \\
&\quad D \in \mathcal{A}_{U,V,1},\ M \in \mathcal{T}_{U,V},\ A^{\mathrm{X}} * M = (A * M)^{\mathrm{X}},\ A = A * M * M^{\dagger A}\})
\end{aligned}
$$

So the *likely distribution* equals the *likely transform-dependent*, $\tilde{E} = \hat{A}^{\mathrm{Y}(\tilde{T})}$, and the *likely model* is such that

$$
\begin{aligned}
\tilde{T} \in \\
&\mathrm{maxd}(\{(M, \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(M)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(M)},z)(B) : B \in Y^{-1}_{U,\mathrm{i},M,z}((A^{\mathrm{X}} * M, (A * M)^{\mathrm{X}}))}) : \\
&\quad M \in \mathcal{T}_{U,V},\ A^{\mathrm{X}} * M = (A * M)^{\mathrm{X}},\ A = A * M * M^{\dagger A}\})
\end{aligned}
$$

The *log-likelihood* varies with the *derived alignment*, so an approximation to the *likely model* is

$$
\begin{aligned}
\tilde{T} \in\ &\mathrm{maxd}(\{(M, \mathrm{algn}(A * M)) : \\
&\quad M \in \mathcal{T}_{U,V},\ A^{\mathrm{X}} * M = (A * M)^{\mathrm{X}},\ A = A * M * M^{\dagger A}\})
\end{aligned}
$$

This optimisation is still intractable, because the cardinality of the *substrate transforms*, $|\mathcal{T}_{U,V}|$, is factorial in the *volume*, v. The computation of the *derived alignment*, $\mathrm{algn}(A*M)$, is tractable, however, and so limited searches can be made tractable and then practicable.

In *classical modelled induction* the constraint must be weakened from *necessary derived* to *uniform possible derived* if the *likely model* is to be non-trivial, $\tilde{T} \notin \{T_{\mathrm{u}}, T_{\mathrm{s}}\}$. *Uniform possible* is not required for *aligned modelled induction* because the *likely model* is sometimes non-trivial when constrained by *necessary formal-abstract*, which is already weaker than *necessary derived*.

Consider, however, the case where the *formal-abstract* pair is *uniformly possible*. Given *substrate transform* $T \in \mathcal{T}_{U,V}$, assume that the *substrate history* *probability function* $P \in (\mathcal{H}_{U,V,z} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is the distribution of an arbitrary *history* valued function, $\mathcal{X} \to \mathcal{H}$, given an arbitrary *formal-abstract* valued function, $\mathcal{X} \to \mathcal{A}^2$. In this case, the *history* valued function is chosen arbitrarily from the constrained subset

$$\begin{aligned}\big\{\{((x,A',B',y),H) : (x,((A',B'),G)) \in F,\ (y,H) \in G,\\ \text{his}(H)^{\mathrm{X}} * T = A',\ (\text{his}(H) * T)^{\mathrm{X}} = B'\} \ :\\ F \in \mathcal{X} \to (\mathcal{A}^2 \times (\mathcal{X} \to \mathcal{H}))\big\} \ \subset\ \mathcal{X} \to \mathcal{H}\end{aligned}$$

In the case of *mid transform*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, the constrained subset is simpler,

$$\begin{aligned}\big\{\{((x,A',y),H) : (x,(A',G)) \in F,\ (y,H) \in G,\\ \text{his}(H)^{\mathrm{X}} * T = (\text{his}(H) * T)^{\mathrm{X}} = A'\} \ :\\ F \in \mathcal{X} \to (\mathcal{A} \times (\mathcal{X} \to \mathcal{H}))\big\} \ \subset\ \mathcal{X} \to \mathcal{H}\end{aligned}$$

This subset of the *substrate history probability functions* can be generalised for all *substrate transforms* as the subset derived from

$$\bigcup_{T \in \mathcal{T}_{\mathrm{f}}} (\mathcal{X} \to (\mathcal{A} \times_T (\mathcal{X} \to \mathcal{H})))$$

where $\mathcal{T}_{\mathrm{f}}$ is the set of all *functional transforms*, and the fibre product $\times_T$ is defined

$$\begin{aligned}\mathcal{A} \times_T (\mathcal{X} \to \mathcal{H}) \ :=&\\ &\{(A',G) : (A',G) \in \mathcal{A} \times (\mathcal{X} \to \mathcal{H}),\\ &\quad \forall(\cdot,H) \in G\ (\text{his}(H)^{\mathrm{X}} * T = (\text{his}(H) * T)^{\mathrm{X}} = A')\}\end{aligned}$$

In the case of *uniform possible formal-abstract*, where there is a *distribution history* H_E and a *substrate transform* $T \in \mathcal{T}_{U,V}$, the *maximum likelihood estimate* which maximises the entropy, $\text{entropy}(\tilde{P})$, is

$$\begin{aligned}\tilde{P} \ =&\ \Big(\bigcup\{\{(H,1) : H \subseteq H_E,\ \text{his}(H)^{\mathrm{X}} * T = A',\ (\text{his}(H) * T)^{\mathrm{X}} = B'\}^{\wedge} \ :\\ &\qquad (A',B') \in \text{ran}(Y_{U,\mathrm{i},T,z})\}\Big)^{\wedge} \ \cup\\ &\ \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\}\\ =&\ \Big(\bigcup\{\{(H,1/\textstyle\sum(Q_{\mathrm{h},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,z}((A',B'))) :\\ &\qquad H \subseteq H_E,\ \text{his}(H)^{\mathrm{X}} * T = A',\ (\text{his}(H) * T)^{\mathrm{X}} = B'\} \ :\\ &\qquad (A',B') \in \text{ran}(Y_{U,\mathrm{i},T,z})\}\Big)^{\wedge} \ \cup\\ &\ \{(G,0) : G \in \mathcal{H}_{U,V,z},\ G \not\subseteq H_E\}\end{aligned}$$

That is, the *maximum likelihood estimate*, $\tilde{P}$, is such that all *drawn formal-abstracts* are uniformly probable, and then all *drawn histories* $H \subseteq H_E$ with the same *formal-abstract*, $\mathrm{his}(H)^{\mathrm{X}} * T = A'$ and $(\mathrm{his}(H) * T)^{\mathrm{X}} = B'$, are uniformly probable. If the *distribution histogram*, H_E, is *known* and the *substrate transform*, T, is *known*, then the *likely probability function*, $\tilde{P}$, is *known*.

The properties of *uniformly possible formal-abstract* are the same as for *necessary formal-abstract*, except that the probabilities are scaled by the fraction $1/|\mathrm{ran}(Y_{U,\mathrm{i},T,z})|$.

Given *uniform possible formal-abstract*, *mid-ideal transform* and *probable sample*, consider the case where a *drawn histogram* A is *known*, but neither the *distribution histogram*, E, is *known* nor the *transform*, T, is *known*, and hence the *likely history probability function*, $\tilde{P}$, is *unknown*. In the case where the *distribution histogram size*, z_E, is also *unknown*, except that it is *known* to be large, $z_E \gg z$, then the *maximum likelihood estimate* $(\tilde{E}, \tilde{T})$ for the pair of the *distribution histogram*, E, and the *transform*, T, is approximated by a modal value of the conditional *likelihood function*,

$$\begin{aligned}(\tilde{E}, \tilde{T}) \in\ & \mathrm{maxd}(\{((D, M), \hat{Q}_{\mathrm{m,y},M,U}(D, z)(A)) : \\ & \quad D \in \mathcal{A}_{U,V,1},\ M \in \mathcal{T}_{U,V},\ A^{\mathrm{X}} * M = (A * M)^{\mathrm{X}},\ A = A * M * M^{\dagger A}\})\end{aligned}$$

So the *likely distribution* equals the *likely transform-dependent*, $\tilde{E} = \hat{A}^{\mathrm{Y}(\tilde{T})}$, and the *likely model* is such that

$$\begin{aligned}\tilde{T} \in\ & \mathrm{maxd}(\{(M, \hat{Q}_{\mathrm{m,y},M,U}(A^{\mathrm{Y}(M)}, z)(A)) : \\ & \quad M \in \mathcal{T}_{U,V},\ A^{\mathrm{X}} * M = (A * M)^{\mathrm{X}},\ A = A * M * M^{\dagger A}\})\end{aligned}$$

The *log-likelihood* varies with the *derived alignment*, so an approximation to the *likely model* is

$$\begin{aligned}\tilde{T} \in\ & \mathrm{maxd}(\{(M, \mathrm{algn}(A * M)) : \\ & \quad M \in \mathcal{T}_{U,V},\ A^{\mathrm{X}} * M = (A * M)^{\mathrm{X}},\ A = A * M * M^{\dagger A}\})\end{aligned}$$

Note, however, that this approximation is looser than in the *necessary formal-abstract* case because the scaling fraction, $1/|\mathrm{ran}(Y_{U,\mathrm{i},\tilde{T},z})|$, is ignored.

2.6.5 Tractable and practicable aligned induction

In the discussion of *aligned induction* above it is shown that, given *necessary formal-abstract*, *mid-ideal transform* and *probable sample*, the *maximum* *likelihood estimate* $\tilde{T}$ for the *transform*, T, is approximated by a maximisation of the *derived alignment*,

$$\begin{aligned}\tilde{T} &\in \text{maxd}(\{(M, \text{algn}(A * M)) : \\ &\qquad M \in \mathcal{T}_{U,V},\ A^{\text{X}} * M = (A * M)^{\text{X}},\ A = A * M * M^{\dagger A}\})\end{aligned}$$

This optimisation is intractable because the cardinality of the *substrate transforms*, $|\mathcal{T}_{U,V}|$, is factorial in the *volume*, v. Consider how limited searches can be made tractable and then practicable.

Given *sample histogram* $A \in \mathcal{A}_{U,\text{i},V,z}$, the tractable *limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer* is defined

$$\begin{aligned}&I'^{*}_{z,\text{Sd,D,F},\infty,\text{n,q}}(A) = \\ &\quad \{(M, I^{*}_{\approx\mathbf{R}}(\textstyle\sum \text{algn}(A * C * F^{\text{T}})/w_F^{1/m_F} : (C,F) \in \text{cont}(M))) : \\ &\qquad M \in \mathcal{D}_{\text{F},\infty,U,V} \cap \text{trees}(\mathcal{S} \times (\mathcal{F}_{\text{n}} \cap \mathcal{F}_{\text{q}})), \\ &\qquad\quad \forall (C,F) \in \text{cont}(M)\ (\text{algn}(A * C * F^{\text{T}}) > 0)\}\end{aligned}$$

where *derived variables* $W_F = \text{der}(F)$, *derived volume* $w_F = |W_F^{\text{C}}|$, *derived dimension* $m_F = |W_F|$ and $I^{*}_{\approx\mathbf{R}}$ computes an approximation to a real number. The geometric average of the *fud derived valencies* is w_F^{1/m_F}.

Here the *model* has been extended from *transforms*, $M \in \mathcal{T}_{U,V}$, to *functional definition set decompositions*, $M \in \mathcal{D}_{\text{F},\infty,U,V}$. At the same time the set of *fud decompositions* has been restricted to those having (a) *fuds* that are *non-overlapping*, $\mathcal{F}_{\text{n}}$, (b) *fuds* with a *limited-underlying*, *limited-derived*, *limited-layer* and *limited-breadth* structure, $\mathcal{F}_{\text{q}} = \mathcal{F}_{\text{u}} \cap \mathcal{F}_{\text{d}} \cap \mathcal{F}_{\text{h}} \cap \mathcal{F}_{\text{b}}$, and (c) *fuds* with *derived alignment*, $\text{algn}(A * C * F^{\text{T}}) > 0$. The tractable optimal *model* is

$$D_{\text{Sd}} \in \text{maxd}(I'^{*}_{z,\text{Sd,D,F},\infty,\text{n,q}}(A))$$

The maximisation of the *contingent fud derived alignment valency-density*, $\text{algn}(A{*}C{*}F^{\text{T}})/w_F^{1/m_F}$, of the *non-overlapping fud* $(C,F) \in \text{cont}(D_{\text{Sd}})$ for the *sample slice* $A{*}C$, tends to *mid fud transform*, $(A{*}C)^{\text{X}}{*}F^{\text{T}} \approx (A{*}C{*}F^{\text{T}})^{\text{X}}$. Then the maximisation of the *summed alignment valency-density*, $\sum \text{algn}(A * C * F^{\text{T}})/w_F^{1/m_F} : (C,F) \in \text{cont}(D_{\text{Sd}})$, for all of the *contingent slices*, tends to *mid-ideal fud decomposition transform*, $A \approx A * D_{\text{Sd}}^{\text{T}} * D_{\text{Sd}}^{\text{T}\dagger A}$. The *summed alignment valency-density* varies with the *derived alignment*, $\text{algn}(A * D_{\text{Sd}}^{\text{T}})$, so the tractable *model* approximates to the *likely model*, $D_{\text{Sd}}^{\text{T}} \approx \tilde{T}$, depending on the limits chosen.

The *derived alignment accuracy* of the approximation can be defined as the exponential of the difference in *derived alignments*,

$$0 \ < \ \frac{\exp(\mathrm{algn}(A * D_{\mathrm{Sd}}^{\mathrm{T}}))}{\exp(\mathrm{algn}(A * \tilde{T}))} \ \leq \ 1$$

This definition of *accuracy* is consistent with the gradient of the likelihood function at the mode, so the *derived alignment accuracy* varies against the *sensitivity* to *model*,

$$\begin{aligned}\frac{\exp(\mathrm{algn}(A * D_{\mathrm{Sd}}^{\mathrm{T}}))}{\exp(\mathrm{algn}(A * \tilde{T}))} \ &\sim \\ &-(-\ln|\max(\{(M, \mathrm{algn}(A * M)) : M \in \mathcal{T}_{U,V}, \\ &\qquad A^{\mathrm{X}} * M = (A * M)^{\mathrm{X}},\ A = A * M * M^{\dagger A}\})|)\end{aligned}$$

The *log-likelihood* varies against the *sensitivity* to *model*, so the *derived alignment accuracy* varies with the *derived alignment*,

$$\frac{\exp(\mathrm{algn}(A * D_{\mathrm{Sd}}^{\mathrm{T}}))}{\exp(\mathrm{algn}(A * \tilde{T}))} \ \sim \ \mathrm{algn}(A * T)$$

That is, although the *model* obtained from the tractable *summed alignment valency-density inducer* is merely an approximation, in the cases where the *log-likelihood* or *derived alignment* is high, and so the *sensitivity* to *model/distribution* is low, the approximation may be reasonably close nonetheless.

The maximisation of *derived alignment* tends to make the properties of *mid-ideal aligned induction* similar to those of *natural classical induction*. This is also the case for the tractable optimisation, so the tractable *model* approximates to the *likely classical model*, $D_{\mathrm{Sd}}^{\mathrm{T}} \approx \tilde{T}$, where

$$\tilde{T} \ \in \ \mathrm{maxd}(\{(M, \hat{Q}_{\mathrm{m,d},M,U}(A, z)(A)) : M \in \mathcal{T}_{U,V},\ A = A * M * M^{\dagger}\})$$

That this is true may be seen by considering the *entropy* properties. The correlations for *summed alignment valency-density* are similar to those for *iso-derived log-likelihood*. The *summed alignment valency-density* (a) varies against the *derived volume* $w' = |(D_{\mathrm{Sd}}^{\mathrm{T}})^{-1}|$,

$$\mathrm{algnValDensSum}(U)(A, D_{\mathrm{Sd}}) \ \sim \ 1/w'$$

(b) varies against the *derived entropy*,

$$\mathrm{algnValDensSum}(U)(A, D_{\mathrm{Sd}}) \quad \sim \quad -z \times \mathrm{entropy}(A * D_{\mathrm{Sd}}^{\mathrm{T}})$$

(c) varies with the *component size cardinality relative entropy*,

$$\mathrm{algnValDensSum}(U)(A, D_{\mathrm{Sd}}) \quad \sim \quad z \times \mathrm{entropyRelative}(A * D_{\mathrm{Sd}}^{\mathrm{T}}, V^{\mathrm{C}} * D_{\mathrm{Sd}}^{\mathrm{T}})$$

and (d) varies with the *expected component entropy*,

$$\mathrm{algnValDensSum}(U)(A, D_{\mathrm{Sd}}) \quad \sim \quad z \times \mathrm{entropyComponent}(A, D_{\mathrm{Sd}}^{\mathrm{T}})$$

where

$$\begin{aligned} \mathrm{algnValDensSum}(U)(A, D) \ &:= \\ &\sum \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} : (C, F) \in \mathrm{cont}(D) \end{aligned}$$

The maximisation of the *derived alignment valency-density*, $\mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F}$, of the *contingent fud* $(C, F) \in \mathrm{cont}(D_{\mathrm{Sd}})$, tends to *diagonalise* the *mid fud transform*, $\mathrm{diagonal}(A * C * F^{\mathrm{T}})$, so minimising the *fud derived entropy*, $\mathrm{entropy}(A * C * F^{\mathrm{T}})$, and hence minimising the overall *decomposition transform derived entropy*, $\mathrm{entropy}(A * D_{\mathrm{Sd}}^{\mathrm{T}})$. The *component cardinality entropy*, $\mathrm{entropy}(C * F^{\mathrm{T}})$, also decreases but is synchronised with the *derived entropy*, $\mathrm{entropy}(A * C * F^{\mathrm{T}})$, so the *mid component size cardinality relative entropy* tends to remain small, $\mathrm{entropyRelative}(A * C * F^{\mathrm{T}}, C * F^{\mathrm{T}}) \approx 0$. The maximisation of the *valency-density*, however, shortens the *diagonal* and so the *off-diagonal derived states* tend to be *ineffective*. The recursive *slicing* during the *decomposition* then removes the *ineffective components*, concentrating the *effective derived states* in smaller *components*, and so maximising the overall *decomposition transform component size cardinality relative entropy*, $\mathrm{entropyRelative}(A * D_{\mathrm{Sd}}^{\mathrm{T}}, V^{\mathrm{C}} * D_{\mathrm{Sd}}^{\mathrm{T}})$, when fully *idealised*.

The *limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, limits the optimisation to make *aligned induction* tractable. By additionally imposing a sequence on the search and other constraints, tractable *induction* is made practicable in the *highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}$. (The details of the implementation are not defined here.) Now, given a set of search parameters P, the *fud decomposition* is

$$D_{\mathrm{Scsd},P} \in \mathrm{maxd}(I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}(A))$$

The set of practicable searched *models* is approximately a subset of the tractable searched *models*, so the practicable *derived alignment* is less than or equal to the tractable *derived alignment*,

$$\mathrm{algn}(A * D_{\mathrm{Scsd},P}^{\mathrm{T}}) \quad \leq \quad \mathrm{algn}(A * D_{\mathrm{Sd}}^{\mathrm{T}})$$

Even so, in the cases where the *log-likelihood* or *derived alignment* is high, and so both the *sensitivity* to *model* and the *sensitivity* to *distribution* are low, the approximation to the *maximum likelihood estimate*, $D_{\mathrm{Scsd},P}^{\mathrm{T}} \approx \tilde{T}$, may be reasonably close nonetheless.

The *highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}$, is an example of practicable *aligned induction*. *Artificial neural network induction* is an example of practicable *classical induction*. Let the *ANN classical model* $F_{\mathrm{gr,lsq},P}^{\mathrm{T}} \approx \tilde{T}$ be obtained by *least squares gradient descent* given a *sample* A subject to the constraints of (i) real *valued variables*, (ii) *causal histogram*, (iii) a *literal frame*, and (iv) clustered *histogram*. The *ANN classical induction* is supervised, requiring that there is a *causal* relation between query *variables* $K \subset V$ and label *variables*, $V \setminus K$,

$$\mathrm{split}(K, A^{\mathrm{FS}}) \quad \in \quad K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$$

At the optimum there is no error and the relation between the *classical derived variables* and the label *variables* is functional,

$$\mathrm{split}(W, (A * X \ \% \ (W \cup V \setminus K))^{\mathrm{FS}}) \quad \in \quad W^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$$

where $(X, W) = F_{\mathrm{gr,lsq},P}^{\mathrm{T}}$.

By contrast, *aligned induction* is unsupervised, so no label is required. *Aligned induction*, however, must have *alignments* between the *underlying variables*,

$$\mathrm{algn}(A) \quad > \quad 0$$

If there is a label, the *aligned induction model* does not necessarily have a *causal* relation between the *derived variables* and the label *variables*, so the label *entropy* may be non-zero,

$$\sum_{(R,C) \in T^{-1}} (A * T)_R \times \mathrm{entropy}(A * C \ \% \ (V \setminus K)) \quad > \quad 0$$

or

$$\sum_{(R,\cdot) \in T^{-1}} (A * T)_R \times \mathrm{entropy}(\{R\}^{\mathrm{U}} * T^{\odot A} \ \% \ (V \setminus K)) \quad > \quad 0$$

where $T = D^{\mathrm{T}}_{\mathrm{Scsd},P}$.

The *ANN classical induction* also requires that the *sample*, A, is clustered. This implies that the query *variables*, K, are real-*valued*, so that there is a metric. The practicable *aligned inducer* requires that the *underlying variables* be discrete, so they must be bucketed if they are in fact continuous.

The *ANN fud*, $F_{\mathrm{gr,lsq},P}$, has a fixed graph so that the *derived variables* have a *literal frame* mapping to the label *variables* in the loss function. This graph is defined a priori in the parameter set, P, and depends on the query *variables*, K, and the label *variables*, $V \setminus K$. The *aligned inducer model*, $D_{\mathrm{Scsd},P}$, is a *fud decomposition* in which the *fuds* are built upwards from the *substrate*, and the only parameters are limits to gross *fud* structure. In addition, a *decomposition* allows *fuds* to be built on *contingent slices*, $A * C$ where $(C, F) \in \mathrm{cont}(D_{\mathrm{Scsd},P})$, which depend on the *components* corresponding to *effective derived states* of ancestor *fuds*. In this way, the *derived variables* near the root of the *decomposition* are most general, applying to the largest *slices*, while the *derived variables* near the leaves of the *decomposition* are most specific, applying to the smallest *slices* as the *alignments* are removed in the *idealisation*. So in the *decomposition*, $D_{\mathrm{Scsd},P}$, each *contingent fud derived*, $A * C * F^{\mathrm{T}}$, may be meaningful in the problem domain. By contrast, the *ANN fud derived variables* apply to the entire query *volume*, K^{C}, and so the *derived*, $A * F^{\mathrm{T}}_{\mathrm{gr,lsq},P}$, is less context specific.

3 Terminology

3.1 Systems

The set of all *systems* $\mathcal{U}$ is the set of all functional relations between the set of all *variables* $\mathcal{V}$ and non-empty subsets of the set of all *values* $\mathcal{W}$,

$$\mathcal{U} = \mathcal{V} \to (\mathrm{P}(\mathcal{W}) \setminus \{\emptyset\})$$

where P is the powerset function.

The function vars $\in \mathcal{U} \to \mathrm{P}(\mathcal{V})$ is the set of *variables* in a *system* $U \in \mathcal{U}$. It is a synonym for the domain of U,

$$\mathrm{vars}(U) := \mathrm{dom}(U) = \{v : (v, W) \in U\}$$

The set of *values* is U_v for some *variable* $v \in \mathrm{vars}(U)$. The *valency* of a *variable* is the cardinality of its *values*, $|U_v|$. The *values* of a *variable* are unordered.

The set $\mathcal{U}$ of all *systems* is defined such that each *variable* must have at least one *value*,

$$\forall U \in \mathcal{U}\ \forall v \in \mathrm{vars}(U)\ (|U_v| \geq 1)$$

In a *system* of finite *variables*, $\forall v \in \mathrm{vars}(U)\ (|U_v| < \infty)$, each *variable* has a set of discrete *values*.

For any subset of *variables* in a *system*, $V \subseteq \mathrm{vars}(U)$, define the parameterised function *volume*, $\mathrm{volume}(U) \in \mathrm{P}(\mathrm{vars}(U)) \to \mathbf{N}_{>0}$

$$\mathrm{volume}(U)(V) := \prod_{v \in V} |U_v|$$

In the case that $V = \emptyset$, define $\mathrm{volume}(U)(\emptyset) := 1$.

3.2 States

The set of all *states* $\mathcal{S}$ is the set of all functional relations between the set of all *variables* $\mathcal{V}$ and the set of all *values* $\mathcal{W}$,

$$\mathcal{S} = \mathcal{V} \to \mathcal{W}$$

State $S \in \mathcal{S}$ is a functional relation between *variables* and *values*,

$$\forall S \in \mathcal{S}\ (|\mathrm{vars}(S)| = |S|)$$

where the function $\mathrm{vars} \in \mathcal{S} \to \mathrm{P}(\mathcal{V})$ is the set of *variables* in the *state*

$$\mathrm{vars}(S) := \mathrm{dom}(S) = \{v : (v, w) \in S\}$$

So a *variable* is an index of the *state*. The cardinality of the set of *variables* $|\mathrm{vars}(S)|$ is the *dimension*. The *variables* of a set of *states* is $\mathrm{vars} \in \mathrm{P}(\mathcal{S}) \to \mathrm{P}(\mathcal{V})$ defined $\mathrm{vars}(Q) := \bigcup\{\mathrm{vars}(S) : S \in Q\}$.

The parameterised set $\mathcal{S}_U$ where $\mathcal{S}_U \subseteq \mathcal{S}$ of all *states* in a particular *system* U is further constrained such that the *variables* of the *state* is a subset of the *variables* of its *system*

$$\forall U \in \mathcal{U}\ \forall S \in \mathcal{S}_U\ (\mathrm{vars}(S) \subseteq \mathrm{vars}(U))$$

Also, each *value* of any *variable-value* pair in a *state* must be an element of the set of *values* for that *variable* in the *system*

$$\forall U \in \mathcal{U}\ \forall S \in \mathcal{S}_U\ \forall (v, w) \in S\ (w \in U_v)$$

For any subset of *variables* in a *system*, $V \subseteq \mathrm{vars}(U)$, define a parameterised *cartesian* set of *states*, $\mathrm{cartesian}(U) \in \mathrm{P}(\mathrm{vars}(U)) \to \mathrm{P}(\mathcal{S}_U)$

$$\mathrm{cartesian}(U)(V) := \prod_{v \in V} \{(v, w) : w \in U_v\}$$

or

$$\mathrm{cartesian}(U)(V) := \prod_{v \in V} \{v\} \times U_v$$

and $\mathrm{cartesian}(U)(\emptyset) := \{\emptyset\}$. So $\mathrm{volume}(U)(V) = |\mathrm{cartesian}(U)(V)|$.

The function $\mathrm{filter} \in \mathrm{P}(\mathcal{V}) \times \mathcal{S} \to \mathcal{S}$ is defined

$$\mathrm{filter}(V, S) := \{(v, u) : (v, u) \in S,\ v \in V\}$$

Define the shorthand $(\%) \in \mathcal{S} \times \mathrm{P}(\mathcal{V}) \to \mathcal{S}$ as $S\%V := \mathrm{filter}(V, S)$. The application of a filter is also known as a *reduction*.

The function $\mathrm{split} \in \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{S}) \to \mathrm{P}(\mathcal{S} \times \mathcal{S})$ is defined

$$\mathrm{split}(V, Q) := \{(\mathrm{filter}(V, S), \mathrm{filter}(\mathrm{vars}(S) \setminus V, S)) : S \in Q\}$$

Two *states* $S, T \in \mathcal{S}$ are said to *join* if their union is also a *state*, $S \cup T \in \mathcal{S}$. That is, a *join* is functional,

$$\begin{aligned} S \cup T \in \mathcal{S} \quad &\iff \quad |\mathrm{vars}(S) \cup \mathrm{vars}(T)| = |S \cup T| \\ &\iff \quad \forall v \in \mathrm{vars}(S) \cap \mathrm{vars}(T)\ (S_v = T_v) \end{aligned}$$

States in disjoint *variables* always *join*, $\forall S, T \in \mathcal{S}\ (\mathrm{vars}(S) \cap \mathrm{vars}(T) = \emptyset \implies S \cup T \in \mathcal{S})$. *States* in the same *variables* only *join* if they are equal, $\forall S, T \in \mathcal{S}\ (\mathrm{vars}(S) = \mathrm{vars}(T) \implies (S \cup T \in \mathcal{S} \iff S = T))$.

The *literal reframing* $\mathrm{reframe} \in (\mathcal{V} \leftrightarrow \mathcal{V}) \times \mathcal{S} \to \mathcal{S}$ is defined

$$\begin{aligned} &\mathrm{reframe}(X, S) := \\ &\quad \{(X_v, u) : (v, u) \in S,\ v \in \mathrm{dom}(X)\} \cup \{(v, u) : (v, u) \in S,\ v \notin \mathrm{dom}(X)\} \end{aligned}$$

which is defined if $\mathrm{ran}(X) \cap (\mathrm{vars}(S) \setminus \mathrm{dom}(X)) = \emptyset$. If the *state* S is in some *system* U, $S \in \mathcal{S}_U$, then the *reframed state* is in U, $\mathrm{reframe}(X, S) \in \mathcal{S}_U$, if $\forall (v, w) \in X\ (U_v \subseteq U_w)$.

The *non-literal reframing* $\text{reframe} \in (\mathcal{V} \leftrightarrow (\mathcal{V} \times (\mathcal{W} \leftrightarrow \mathcal{W})) \times \mathcal{S} \to \mathcal{S}$ is defined

$$\begin{aligned}
&\text{reframe}(X, S) := \\
&\quad \{(w, W_u) : (v, u) \in S,\ v \in \text{dom}(X),\ (w, W) = X_v,\ u \in \text{dom}(W)\} \cup \\
&\quad \{(w, u) : (v, u) \in S,\ v \in \text{dom}(X),\ (w, W) = X_v,\ u \notin \text{dom}(W)\} \cup \\
&\quad \{(v, u) : (v, u) \in S,\ v \notin \text{dom}(X)\}
\end{aligned}$$

which is defined if $\text{dom}(\text{ran}(X)) \cap (\text{vars}(S) \setminus \text{dom}(X)) = \emptyset$. If the *state* S is in some *system* U, $S \in \mathcal{S}_U$, then the *reframed state* is in U, $\text{reframe}(X, S) \in \mathcal{S}_U$, if $\forall (w, W) \in \text{ran}(X)\ (\text{ran}(W) \subseteq U_w)$.

When there exists a *literal* or *non-literal frame* between sets of *variables* $V, W \subset \text{vars}(U)$ in a *system*, $|V| = |W| \wedge (\exists Q \in V \cdot W\ \forall (v, w) \in Q\ (|U_v| = |U_w|))$, they are said to have the same *geometry*.

3.3 Histories

An *event identifier* is any member of the universal set $\mathcal{X}$. An *event* is a pair of an *event identifier* and a *state*, $\mathcal{X} \times \mathcal{S}$. A *history* H is a *state* valued function of *event identifiers*. The set of all *histories* $\mathcal{H}$ is a subset of all functional relations of *events*

$$\mathcal{H} \subset \mathcal{X} \to \mathcal{S}$$

Note that the *event identifier* in a *history* need not form a contiguous sequence nor a chronological series. There is no order required or implied with respect to the *event identifier*.

The *size* of a *history* is defined, $\text{size} \in \mathcal{H} \to \mathbf{N}$,

$$\text{size}(H) := |H|$$

The set of *states* of a *history* is defined, $\text{states} \in \mathcal{H} \to \text{P}(\mathcal{S})$. It is the range of the *history*

$$\text{states}(H) := \text{ran}(H) = \{S : (x, S) \in H\}$$

Define the shorthand $()^{\text{S}} \in \mathcal{H} \to \text{P}(\mathcal{S})$ as $H^{\text{S}} := \text{states}(H)$.

The set of *event identifiers* of a *history* is the domain of the *history*, $\text{ids} \in \mathcal{H} \to \text{P}(\mathcal{X})$

$$\text{ids}(H) := \text{dom}(H) = \{x : (x, S) \in H\}$$

A *history's states* must all have exactly the same set of *variables*

$$\forall H \in \mathcal{H}\ \forall S \in \mathrm{states}(H)\ (\mathrm{vars}(S) = \mathrm{vars}(H))$$

where $\mathrm{vars} \in \mathcal{H} \to \mathrm{P}(\mathcal{V})$ is defined

$$\mathrm{vars}(H) := \{v : S \in \mathrm{states}(H),\ v \in \mathrm{vars}(S)\}$$

The parameterised subset $\mathcal{H}_U$ where $\mathcal{H}_U \subseteq \mathcal{H}$ is the set of all *histories* in *system* U such that

$$\forall U \in \mathcal{U}\ \forall H \in \mathcal{H}_U\ (\mathrm{states}(H) \subseteq \mathcal{S}_U)$$

The parameterised *volume* of a *history* is the *volume* of its *variables* $\mathrm{vars}(H) \subseteq \mathrm{vars}(U)$, defined $\mathrm{volume}(U) \in \mathcal{H}_U \to \mathbf{N}_{>0}$

$$\mathrm{volume}(U)(H) := \mathrm{volume}(U)(\mathrm{vars}(H))$$

Define the *reduction* of a *history* as the *reduction* of its *events*, $\mathrm{reduce} \in \mathrm{P}(\mathcal{V}) \to (\mathcal{H} \to \mathcal{H})$ as

$$\mathrm{reduce}(V)(H) := \{(x, S\%V) : (x, S) \in H\}$$

Define $H\%V := \mathrm{reduce}(V)(H)$.

The *addition* operation of *histories* is defined as the disjoint union if both *histories* have the same *variables*. Define $(+) \in \mathcal{H} \times \mathcal{H} \to \mathcal{H}$ as

$$H_1 + H_2 \ := \ \{((x,\cdot), S) : (x, S) \in H_1\} \ \cup \ \{((\cdot, y), T) : (y, T) \in H_2\}$$

where $\mathrm{vars}(H_1) = \mathrm{vars}(H_2)$. The *size* of the *sum* equals the sum of the *sizes*, $|H_1 + H_2| = |H_1| + |H_2|$.

The *multiplication* operation of *histories* is defined as the product where the *states join*. Define $(*) \in \mathcal{H} \times \mathcal{H} \to \mathcal{H}$ as

$$\begin{aligned} H_1 * H_2 \ := \ \{((x,y), S \cup T) : (x, S) \in H_1,\ (y, T) \in H_2, \\ \forall v \in \mathrm{vars}(S) \cap \mathrm{vars}(T)\ (S_v = T_v)\} \end{aligned}$$

The *size* of the *product* equals the product of the *sizes* if the *variables* are disjoint, $\mathrm{vars}(H_1) \cap \mathrm{vars}(H_2) = \emptyset \implies |H_1 * H_2| = |H_1| \times |H_2|$. The *variables* of the *product* is the union of the *variables* if the *size* is non-zero, $H_1 * H_2 \neq \emptyset \implies \mathrm{vars}(H_1 * H_2) = \mathrm{vars}(H_1) \cup \mathrm{vars}(H_2)$.

3.4 Histograms

3.4.1 Definition

A *histogram* A is a functional relation between *states* and positive rational *counts*. The set of all *histograms* $\mathcal{A}$ is a subset

$$\mathcal{A} \subset \mathcal{S} \to \mathbf{Q}_{\geq 0}$$

The *histogram* is a functional relation so any *state* may appear no more than once in the *histogram*. The *histogram* is indexable by *state*

$$\forall A \in \mathcal{A}\ (|\text{states}(A)| = |A|)$$

where states $\in \mathcal{A} \to \mathrm{P}(\mathcal{S})$ is defined

$$\text{states}(A) := \text{dom}(A) = \{S : (S, c) \in A\}$$

Define the shorthand $()^{\mathrm{S}} \in \mathcal{A} \to \mathrm{P}(\mathcal{S})$ as $A^{\mathrm{S}} := \text{states}(A)$.

Similar to *history*, a *histogram's states* must all have exactly the same set of *variables*, so defining vars $\in \mathcal{A} \to \mathrm{P}(\mathcal{V})$,

$$\text{vars}(A) := \{v : S \in \text{states}(A),\ v \in \text{vars}(S)\}$$

require that

$$\forall A \in \mathcal{A}\ \forall S \in \text{states}(A)\ (\text{vars}(S) = \text{vars}(A))$$

Also, the *counts* of a *histogram* must always be positive,

$$\forall A \in \mathcal{A}\ \forall S \in \text{states}(A)\ (A_S \geq 0)$$

The *size* of a *histogram* is size $\in \mathcal{A} \to \mathbf{Q}_{\geq 0}$,

$$\text{size}(A) := \sum A_S : S \in \text{states}(A)$$

and $\text{size}(\emptyset) := 0$. The *size* must always be greater than or equal to zero,

$$\forall A \in \mathcal{A}\ (\text{size}(A) \geq 0)$$

The *empty histogram* A equals the empty set, $A = \emptyset$. Its *size* is zero, $\text{size}(\emptyset) = 0$. The *empty histogram* has no *variables*, $\text{vars}(\emptyset) = \emptyset$.

The *scalar histogram* of some positive *count* $c \in \mathbf{Q}_{\geq 0}$ is defined $A = \{(\emptyset, c)\}$. Define the constructor scalar $\in \mathbf{Q}_{\geq 0} \to \mathcal{A}$ such that $\text{scalar}(c) := \{(\emptyset, c)\}$. A *scalar histogram* has no *variables*, $\forall c \in \mathbf{Q}_{\geq 0}\ (\text{vars}(\text{scalar}(c)) = \emptyset)$.

A *trimmed histogram* has only non-zero *counts*. Define trim $\in \mathcal{A} \to \mathcal{A}$

$$\text{trim}(A) := \{(S,c) : (S,c) \in A,\ c > 0\}$$

Histogram A is *congruent* to *histogram* B, if both have the same *variables* and *size*. Define congruent $\in \mathcal{A} \times \mathcal{A} \to \mathbf{B}$

$$\text{congruent}(A,B) := (\text{vars}(A) = \text{vars}(B)) \wedge (\text{size}(A) = \text{size}(B))$$

Histogram A is *equivalent* to *histogram* B, if the non-zero *count states* are equal,

$$A \equiv B := \text{trim}(A) = \text{trim}(B)$$

Note that this definition implies that *zero histograms*, which are those that are such that all *counts* are zero, even if in different *variables*, are all *equivalent*.

A *sub-histogram* A of a *histogram* B is such that the *trimmed states* of A are a subset of the *states* of B and the *counts* of A are less than or equal to those of B, $(\leq) \in \mathcal{A} \times \mathcal{A} \to \mathbf{B}$

$$A \leq B := (\text{st}(\text{tm}(A)) \subseteq \text{st}(B)) \wedge (\forall S \in \text{st}(\text{tm}(A))\ (A_S \leq B_S))$$

where st = states and tm = trim. The relation is a pre-order. The *empty histogram* $A = \emptyset$ is a *sub-histogram* of all *histograms*, $\forall B \in \mathcal{A}\ (\emptyset \leq B)$. *Equivalent histograms* are *sub-histograms* of each other, $A \equiv B \implies (A \leq B) \wedge (B \leq A)$. The *super-histogram* operator is typed $(\geq) \in \mathcal{A} \times \mathcal{A} \to$ B and defined $A \geq B := B \leq A$. The *proper sub-histogram* is defined $A < B := (A \leq B) \wedge \neg(A \equiv B)$, and the *proper super-histogram* is defined $A > B := (A \geq B) \wedge \neg(A \equiv B)$. It is not necessary that the relation between A and B is *sub-histogram* or *super-histogram*. It may be the case that neither holds $\exists A, B \in \mathcal{A}\ (\text{vars}(A) = \text{vars}(B) \wedge \neg(A \leq B \vee A \geq B))$.

A *system* can be implied by a *histogram*. Define implied $\in \mathcal{A} \to \mathcal{U}$ as

$$\text{implied}(A) := \{(v, \{S_v : S \in A^{\text{S}}\}) : v \in \text{vars}(A)\}$$

Given a *system* U, the parameterised subset $\mathcal{A}_U$ where $\mathcal{A}_U \subseteq \mathcal{A}$ is the set of all *histograms* in U such that

$$\forall A \in \mathcal{A}_U\ (\text{states}(A) \subseteq \mathcal{S}_U)$$

Similar to *history*, the parameterised *volume* of a *histogram* is the *volume* of its *variables* $\text{vars}(A) \subseteq \text{vars}(U)$, defined $\text{volume}(U) \in \mathcal{A}_U \to \mathbf{N}_{>0}$

$$\text{volume}(U)(A) := \text{volume}(U)(\text{vars}(A))$$

If a pair of *histograms* A and B, in the same *system* U, have a *variables* mapping X such that $\exists X \in \text{vars}(A) \cdot \text{vars}(B)\ \forall (v,u) \in X\ (U_v = U_u)$ then the *variables* of A and the *variables* of B are said to be *literal frames* of each other mapped by X.

The function $\text{reframe} \in (\mathcal{V} \leftrightarrow \mathcal{V}) \times \mathcal{A} \to \mathcal{A}$ is defined

$$\text{reframe}(X, A) := \{(\text{reframe}(X, S), c) : (S, c) \in A\}$$

which is defined if $\text{ran}(X) \cap (\text{vars}(A) \setminus \text{dom}(X)) = \emptyset$. If the *histogram* A is in some *system* U, $A \in \mathcal{A}_U$, then $\text{reframe}(X, A) \in \mathcal{A}_U$ if $\forall (v, w) \in X\ (U_v \subseteq U_w)$.

If a pair of *histograms* A and B, in the same *system* U, have a *variables* mapping X such that $\exists X \subset \{(v, (w, W)) : Q \in \text{vars}(A) \cdot \text{vars}(B),\ (v, w) \in Q,\ W \in U_v \cdot U_w\}\ (X \in \mathcal{V}_U \leftrightarrow (\mathcal{V}_U \times (\mathcal{W}_U \leftrightarrow \mathcal{W}_U)))$ then the *variables* of A and the *variables* of B are said to be *non-literal frames* of each other mapped by X.

The function $\text{reframe} \in (\mathcal{V} \leftrightarrow (\mathcal{V} \times (\mathcal{W} \leftrightarrow \mathcal{W})) \times \mathcal{A} \to \mathcal{A}$ is defined

$$\text{reframe}(X, A) := \{(\text{reframe}(X, S), c) : (S, c) \in A\}$$

which is defined if $\text{reframe}(X, S)$ is defined for each *state*. In other words, $\text{reframe}(X, A)$ is defined if $\text{dom}(\text{ran}(X)) \cap (\text{vars}(A) \setminus \text{dom}(X)) = \emptyset$. If the *histogram* A is in some *system* U, $A \in \mathcal{A}_U$, then the *reframed histogram* is in U, $\text{reframe}(X, A) \in \mathcal{A}_U$, if $\forall (w, W) \in \text{ran}(X)\ (\text{ran}(W) \subseteq U_w)$.

The function $\text{resize} \in \mathbf{Q}_{\geq 0} \times \mathcal{A} \to \mathcal{A}$ is defined

$$\text{resize}(z, A) := \{(S, cz/z_A) : (S, c) \in A,\ z_A = \text{size}(A)\}$$

which is defined if $\text{size}(A) > 0$. The *resize* is such that $\text{size}(\text{resize}(z, A)) = z$.

Define the *ceiling* and *floor* functions that return *integral histograms*. Define $\text{ceiling} \in \mathcal{A} \to \mathcal{A}_{\text{i}}$

$$\text{ceiling}(A) := \{(S, d) : (S, c) \in A,\ d \in \mathbf{N},\ d \geq c,\ d - c < 1\}$$

where the *integral histograms* is the set $\mathcal{A}_{\text{i}} = \mathcal{A} \cap (\mathcal{S} \to \mathbf{N})$. Define $\text{floor} \in \mathcal{A} \to \mathcal{A}_{\text{i}}$

$$\text{floor}(A) := \{(S, d) : (S, c) \in A,\ d \in \mathbf{N},\ d \leq c,\ c - d < 1\}$$

Thus $\text{size}(\text{floor}(A)) \leq \text{size}(A) \leq \text{size}(\text{ceiling}(A))$.

3.4.2 Unit histograms

A *unit histogram* $A^{\mathrm{U}} \in \mathcal{A}$ is a special case in which all its *counts* equal 1. Define $\mathrm{unit} \in \mathcal{A} \to \mathcal{A}_{\mathrm{i}}$

$$\mathrm{unit}(A) := \mathrm{states}(A) \times \{1\} = \{(S, 1) : (S, c) \in A\}$$

where the *integral histograms* is the set $\mathcal{A}_{\mathrm{i}} = \mathcal{A} \cap (\mathcal{S} \to \mathbf{N})$. Define the shorthand $A^{\mathrm{U}} := \mathrm{unit}(A)$. Thus $\forall S \in \mathrm{states}(A^{\mathrm{U}})\ (A^{\mathrm{U}}_S = 1)$. In such a *histogram*, $\mathrm{size}(A^{\mathrm{U}}) = |\mathrm{states}(A^{\mathrm{U}})| = |A^{\mathrm{U}}|$. *Unit histograms* provide a useful shorthand for the *states*. Define a convenience function to promote a set of *states* to a *unit histogram*, define $\mathrm{unit} \in \mathrm{P}(\mathcal{S}) \to (\mathcal{S} \to \{1\})$ as $\mathrm{unit}(Q) := Q \times \{1\}$, with a shorthand $Q^{\mathrm{U}} := \mathrm{unit}(Q)$. Depending on the argument set of *states* Q the function Q^{U} may be a *histogram*, $\exists Q \in \mathrm{P}(\mathcal{S})\ (\mathrm{unit}(Q) \in \mathcal{A}_{\mathrm{i}})$.

A *zero histogram* $A^{\mathrm{Z}} \in \mathcal{A}$ is a special case in which all its *counts* equal 0. Define $\mathrm{zero} \in \mathcal{A} \to \mathcal{A}_{\mathrm{i}}$

$$\mathrm{zero}(A) := \mathrm{states}(A) \times \{0\}$$

and the shorthand $A^{\mathrm{Z}} := \mathrm{zero}(A)$.

There are a couple of useful variations on the theme of *unit histograms*, *unit effective histogram* $A^{\mathrm{F}} \in \mathcal{A}_{\mathrm{i}}$ and *unit cartesian histogram* $A^{\mathrm{C}} \in \mathcal{A}_{U,\mathrm{i}}$.

The *unit effective histogram* A^{F} of a *histogram* A only includes *states* where the *count* is non-zero. Defining $\mathrm{effective} \in \mathcal{A} \to \mathcal{A}_{\mathrm{i}}$

$$\mathrm{effective}(A) := \{(S, 1) : (S, c) \in A,\ c > 0\} = \mathrm{unit}(\mathrm{trim}(A))$$

Define the shorthand $A^{\mathrm{F}} := \mathrm{effective}(A)$.

Histogram equivalence can be defined in terms of *effective histograms*, $A \equiv B := \{(S, A_S) : S \in A^{\mathrm{FS}}\} = \{(T, B_T) : T \in B^{\mathrm{FS}}\}$.

The function $\mathrm{cartesian}(U) \in \mathrm{P}(\mathrm{vars}(U)) \to \mathrm{P}(\mathcal{S}_U)$ is the *cartesian* set of *states* for some set of *variables* in a *system*. Define a shorthand V^{C} which is the *histogram* for this set, $()^{\mathrm{C}} \in \mathrm{P}(\mathrm{vars}(U)) \to \mathcal{A}_{U,\mathrm{i}}$

$$V^{\mathrm{C}} := \mathrm{cartesian}(U)(V) \times \{1\}$$

where the context of *system* U is implicit. Define a similar shorthand for the *unit cartesian histogram* $A^{\mathrm{C}} \in \mathcal{A}_{U,\mathrm{i}}$ which includes all *states* of $\mathrm{vars}(A)$ in *system* U, $()^{\mathrm{C}} \in \mathcal{A}_U \to \mathcal{A}_{U,\mathrm{i}}$

$$A^{\mathrm{C}} := (\mathrm{vars}(A))^{\mathrm{C}}$$

assuming the context of *system* U. Clearly the *unit cartesian histogram* A^{C} does not depend on the *counts* in A, only on the *variables* of A in the *system* U. Also $\forall A \in \mathcal{A}_U\ (A^{\mathrm{F}} \leq A^{\mathrm{C}})$ and $\forall A \in \mathcal{A}_U\ (|A^{\mathrm{F}}| \leq |A^{\mathrm{C}}|)$ and $\forall A \in \mathcal{A}_U\ (|A^{\mathrm{C}}| = \mathrm{volume}(U)(A))$.

A *histogram* is *complete* in some *system* when the *unit histogram* equals the *unit cartesian histogram*, $A^{\mathrm{U}} = A^{\mathrm{C}}$.

The *unit effective complement histogram* of a *histogram* A is $A^{\mathrm{C}} \setminus A^{\mathrm{F}}$.

3.4.3 Arithmetic operators

The *addition* operation of *histograms* is defined if both *histograms* have the same *variables*. For *histograms* A and B, if $\mathrm{vars}(A) = \mathrm{vars}(B)$, define $(+) \in \mathcal{A} \times \mathcal{A} \to \mathcal{A}$

$$
\begin{aligned}
&A + B := \\
&\quad \{(S,c) : (S,c) \in A,\ S \notin \mathrm{states}(B)\} \cup \\
&\quad \{(S,c+d) : (S,c) \in A,\ (T,d) \in B,\ S = T\} \cup \\
&\quad \{(T,d) : (T,d) \in B,\ T \notin \mathrm{states}(A)\}
\end{aligned}
$$

Clearly $\mathrm{size}(A+B) = \mathrm{size}(A) + \mathrm{size}(B)$. *Completeness* is cumulative for *addition*, $(A+B)^{\mathrm{U}} = A^{\mathrm{U}} \cup B^{\mathrm{U}}$. *Addition* is associative and commutative.

Negative *counts* are not allowed, so *subtraction* is not constructed in terms of *addition* of a *negative* but instead as a binary operation $(-) \in \mathcal{A} \times \mathcal{A} \to \mathcal{A}$

$$
\begin{aligned}
&A - B := \\
&\quad \{(S,c) : (S,c) \in A,\ S \notin \mathrm{states}(B)\} \cup \\
&\quad \{(S,c-d) : (S,c) \in A,\ (T,d) \in B,\ S = T,\ c \geq d\} \cup \\
&\quad \{(S,0) : (S,c) \in A,\ (T,d) \in B,\ S = T,\ c < d\} \cup \\
&\quad \{(T,0) : (T,d) \in B,\ T \notin \mathrm{states}(A)\}
\end{aligned}
$$

Subtraction explicitly prevents negative *counts*. *Completeness* is cumulative for *subtraction*, $(A-B)^{\mathrm{U}} = A^{\mathrm{U}} \cup B^{\mathrm{U}}$. Construct a zero, the identity *histogram* for *addition*, $A^{\mathrm{C}} - A^{\mathrm{C}}$. To make a *histogram complete* add the zero, $A + A^{\mathrm{C}} - A^{\mathrm{C}}$.

The *sub-histogram* relation can be defined in terms of *addition* and *subtraction* operators, $A \leq B \iff (B - A + A) \equiv B$.

To define the *multiplication* operation of *histograms* cross the *states* of each *histogram*, keeping only those that *join*, where a pair of *states* S and T *join* if their union is a *state*, $|\text{vars}(S \cup T)| = |S \cup T|$. In other words, the intersection of the *variables* of the *states* must also be the intersection of the *variable-value* pairs. Defining $(*) \in \mathcal{A} \times \mathcal{A} \to \mathcal{A}$,

$$A * B := \{(S \cup T, cd) : (S, c) \in A,\ (T, d) \in B,\ |\text{vars}(S \cup T)| = |S \cup T|\}$$

Define $A * \emptyset := \emptyset$ and $\emptyset * A := \emptyset$.

Multiplication unions the *variables* of *non-empty histograms*, $\text{vars}(A * B) = \text{vars}(A) \cup \text{vars}(B)$. If both A and B are *unit*, their product $A * B$ is also *unit*. If the *variables* of A and B are disjoint, then the cross is cartesian and $|A*B| = |A| \times |B|$ and $\text{size}(A*B) = \text{size}(A) \times \text{size}(B)$. *Multiplication* is *complete* if the *variables* of A and B are disjoint and A and B are both *complete*. *Multiplication* is associative and commutative. *Multiplication* and *addition* together are distributive. If negative *counts* were allowed, *histograms* would obey the algebra of fields. The identity *histogram* for *multiplication* is the *unit cartesian histogram*, A^{C}. *Multiplication* of a *histogram* A by a scalar w can be accomplished by promoting the scalar to a *histogram* with an empty *state*, $A * \{(\emptyset, c)\}$ or $A * \text{scalar}(c)$, so that $\text{size}(A * \text{scalar}(c)) = \text{size}(A) \times c$.

Division is calculated by defining the *reciprocal* of a *histogram* as the reciprocal of its *counts*, $(1/) \in \mathcal{A} \to \mathcal{A}$,

$$1/A := \{(S, 1/c) : (S, c) \in A,\ c > 0\}$$

Define $1/\emptyset := \emptyset$. Define division $(/) \in \mathcal{A} \times \mathcal{A} \to \mathcal{A}$, as $B/A := B * (1/A)$. The *reciprocal* $1/A$ ignores any *count* which is zero. The *reciprocal* $1/A$ is defined even if $\text{size}(A) = 0$. The *reciprocal* $1/A$ is as *complete* as the *effective states* of A, $(1/A)^{\mathrm{U}} = A^{\mathrm{F}} \subseteq A^{\mathrm{U}}$. If A is a *zero histogram* then the *reciprocal* is *empty* $(1/A) = \emptyset$. A *unit histogram* is its own *reciprocal* $1/A^{\mathrm{U}} = A^{\mathrm{U}}$. Division is often used when normalising by a scalar.

3.4.4 Reduction

Define *reduction* of a *histogram* as the reduction of the set of its *variables* to some subset, $\text{reduce} \in \mathrm{P}(\mathcal{V}) \to (\mathcal{A} \to \mathcal{A})$

$$\text{reduce}(V)(A) := \{(S, \sum(c : (T, c) \in A,\ T \supseteq S)) : S \in \{R\%V : R \in A^{\mathrm{S}}\}\}$$

Define the shorthand operator $(\%) \in \mathcal{A} \times \mathrm{P}(\mathcal{V}) \to \mathcal{A}$ as $A\%V := \text{reduce}(V)(A)$. By definition, $\forall V \in \mathrm{P}(\mathcal{V})\ (\text{vars}(A\%V) = V \cap \text{vars}(A))$. *Reduction* leaves the *size* of a *histogram* unchanged, $\text{size}(A\%V) = \text{size}(A)$, but the number of *states* may be fewer, $|A\%V| \leq |A|$. *Reduction* of a *histogram* by its own *variables* leaves the *histogram* unchanged, $A\%\text{vars}(A) = A$. *Reduction* of a *histogram* by the empty set, leaves a *scalar*, $A\%\emptyset = \{(\emptyset, c)\}$ where $c = \text{size}(A)$. Sometimes it shall be assumed that the set Z is a relation between all *histograms* and the *scalars* of their *sizes*, $Z = \{(A, \text{scalar}(\text{size}(A))) : A \in \mathcal{A}\}$. For example, $Z_X = X\%\emptyset = \{(\emptyset, \text{size}(X))\}$.

A *histogram* A is *completely effective* in some set of *variables* V if $(A\%V)^{\text{F}} = (A\%V)^{\text{C}}$. Similarly, A is *cartesian* in some set of *variables* V if $A\%V = (A\%V)^{\text{C}}$. If A is *cartesian* in V then it is also *completely effective* in V.

3.4.5 Histogram expressions

A *histogram expression* $N \in \mathcal{E}_U$ in *system* U is an expression consisting of the *histogram arithmetic operators*, the *histogram reduction operator*, constant *histograms* and free variable identifiers. A *histogram expression* can be *evaluated* by substitution of the free variable identifiers by *histograms* and then evaluation of the dependent *operators* to yield a resultant *histogram*. The *histogram expression application* is denoted $N(A) \in \mathcal{A}_U$, $N(A, B) \in \mathcal{A}_U$, $N(A, B, C) \in \mathcal{A}_U$, etc where $A, B, C \in \mathcal{A}_U$.

A *model* $M \in \mathcal{M}_U \subset \mathcal{E}_U$ is a special case of a *histogram expression* which has a single free variable identifier. A *model's* resultant *histogram*, $M(A)$, is in the same *variables* as the argument *histogram*, $\text{vars}(M(A)) = \text{vars}(A)$. A further constraint is that at some point during the *evaluation* none of the *variables* of the argument *histogram* remain, having been removed by *reduction*. Thus a *model* can be thought of as a path from an argument *histogram* in some set of underlying *variables* via an intermediate set of derived *variables* and back to the given set of *variables*. The underlying *variables* are called the *substrate*.

See appendix 'Histogram expressions' for more formal definitions.

3.4.6 Types of histogram

An *empty histogram* A equals the empty set, $A = \emptyset$. Its *size* is zero, $\text{size}(\emptyset) = 0$. An *empty histogram* has no *variables*, $\text{vars}(\emptyset) = \emptyset$. Its *volume* is one in all *systems*, $|\emptyset^{\text{C}}| = 1$.

A *zero histogram* A is a *non-empty histogram* that has unique zero *count*, $A \neq \emptyset \wedge \text{trim}(A) = \emptyset$. Its *size* is zero. The only *zero histogram* that has no *variables* is the *scalar histogram* $\{(\emptyset, 0)\}$. All *zero histograms* are *equivalent* to the *empty histogram*, $\forall A \in \mathcal{A}\ (\text{trim}(A) = \emptyset \implies A \equiv \emptyset)$.

A *uniform histogram* A has unique non-zero *count*, $|\text{ran}(\text{trim}(A))| = 1$. A *bi-form histogram* A has two non-zero *counts*, $|\text{ran}(\text{trim}(A))| = 2$.

The *counts* of an *integral histogram* are positive integers, $\text{ran}(A) \subset \mathbf{N}$. The set of all *integral histograms* $\mathcal{A}_\text{i}$ is

$$\mathcal{A}_\text{i} = \mathcal{A} \cap (\mathcal{S} \to \mathbf{N}) = \{A : A \in \mathcal{A},\ \text{ran}(A) \subset \mathbf{N}\}$$

A *singleton histogram* A has unique non-zero *state*, $|\text{trim}(A)| = 1$.

A *regular histogram* A of *variables* $V = \text{vars}(A) \neq \emptyset$ in *system* U has unique *valency* of its *variables*, $|\{|U_v| : v \in V\}| = 1$. The *volume* of a *regular histogram* is d^n where *valency* d is such that $\{d\} = \{|U_v| : v \in V\}$ and *dimension* $n = |V|$. (The use of d here suggests diagonal rather than dimension.)

If a *histogram* A has no *variables*, $\text{vars}(A) = \emptyset$, then either the *histogram* is *empty* $A = \emptyset$ or it is a *scalar* $A = \{(\emptyset, c)\}$ of some positive *count* $c \in \mathbf{Q}_{\geq 0}$. Define constructor $\text{scalar} \in \mathbf{Q}_{\geq 0} \to \mathcal{A}$ such that $\text{scalar}(c) := \{(\emptyset, c)\}$.

When a *histogram* has exactly one *variable*, $|\text{vars}(A)| = 1$, it is *mono-variate*.

Pluri-variate histograms can be classified by *incidence*, $\text{incidence} \in \mathcal{A} \times \mathcal{S} \times \mathbf{N} \to \mathcal{A}$

$$\text{incidence}(A, S, i) := \{(T, d) : (T, d) \in \text{trim}(A),\ |S \cap T| = i\}$$

where i is the *degree of incidence* of *state* S. If *state* $S \in \text{states}(A)$, the maximum cardinality of *incidence* $|\text{incidence}(A, S, i)|$ for a particular *degree of incidence* i of a *regular histogram* of *variables* V of *valency* d and *volume* d^n where *dimension* $n = |V|$ is

$$\frac{n!}{i!(n-i)!}(d-1)^{n-i}$$

The maximum cardinality of *incidence* having some *degree of incidence* is $d^n - (d-1)^n$.

A *pluri-variate histogram*, $|\text{vars}(A)| > 1$, is *causal* if the *effective states* can be functionally split. Define causal $\in \mathcal{A} \to \mathbf{B}$

$$\text{causal}(A) := |V| > 1 \wedge (\exists K \subset V \ (K \neq V \implies \text{split}(K, A^{\text{FS}}) \in \mathcal{S} \to \mathcal{S}))$$

where $V = \text{vars}(A)$. If A is *causal* the *variables* K are said to *cause* the remaining *variables* $V \setminus K$.

A *histogram* is *planar* if all *effective states* are *incident* on some *value* u of *variable* w, $A^{\text{F}} \subseteq \{S : S \in V^{\text{CS}},\ (w,u) \in S\}^{\text{U}}$. Define planar $\in \mathcal{A} \to \mathbf{B}$

$$\text{planar}(A) := \exists (w,u) \in \bigcup A^{\text{FS}} \ \forall S \in A^{\text{FS}} \ ((w,u) \in S)$$

A *pluri-variate histogram*, $|\text{vars}(A)| > 1$, is *anti-planar* if each *reduction* has at least two *effective states*. Define antiplanar $\in \mathcal{A} \to \mathbf{B}$

$$\text{antiplanar}(A) := |V| > 1 \wedge (\forall w \in V \ (|(A\%\{w\})^{\text{F}}| > 1))$$

A *pluri-variate histogram*, $|\text{vars}(A)| > 1$, is *diagonalised* if no pair of *effective states* shares any *value*, $\forall S,T \in A^{\text{FS}} \ (S \neq T \implies S \cap T = \emptyset)$ or $\forall (S,c) \in \text{trim}(A) \ (\text{trim}(A) \setminus \{(S,c)\} = \text{incidence}(A,S,0))$. Define diagonal $\in \mathcal{A} \to \mathbf{B}$

$$\text{diagonal}(A) := |V| > 1 \wedge (\forall S,T \in A^{\text{FS}} \ (S \neq T \implies |S \cap T| = 0))$$

A *diagonalised histogram* A is *fully diagonalised* if its non-zero cardinality equals its minimum *valency* of its *variables*, $|\text{trim}(A)| = \text{minr}(\{(v,|U_v|) : v \in \text{vars}(A)\})$. Define diagonalFull$(U) \in \mathcal{A} \to \mathbf{B}$ in *system* U

$$\text{diagonalFull}(U)(A) \ := \ \text{diagonal}(A) \ \wedge |A^{\text{F}}| = \text{minr}(\{(v,|U_v|) : v \in V\})$$

where $V = \text{vars}(A)$. A *full diagonal* has the maximum cardinality of *effective states* of *diagonal histogram* for the given *variables*, $\text{diagonalFull}(U)(A) \implies |A^{\text{F}}| = \text{maxr}(\{(B,|B|) : B \subseteq V^{\text{C}},\ \text{diagonal}(B)\})$. In a *fully diagonalised regular histogram* of *valency* d, where $\{d\} = \{|U_v| : v \in V\}$, the cardinality of non-zero *states* is $|A^{\text{F}}| = d$. The cardinality of the subsets of a *regular cartesian* which are *fully diagonalised* is $|\{A : A \subseteq A^{\text{C}},\ \text{diagonalFull}(U)(A)\}| = (d!)^{n-1}$ where $\{d\} = \{|U_v| : v \in V\}$, $n = |V|$ and $V \neq \emptyset$.

An *anti-diagonal histogram* A is such that all pairs of *effective states* share at least one *value*. Define antidiagonal $\in \mathcal{A} \to \mathbf{B}$

$$\text{antidiagonal}(A) := \forall S,T \in A^{\text{FS}} \ (|S \cap T| > 0)$$

A *full anti-diagonal histogram* A is such that all *states* that are *incident* on some *value* u of *variable* w are *effective*, $A^{\mathrm{F}} = \{S : S \in V^{\mathrm{CS}},\ (w,u) \in S\}^{\mathrm{U}}$. Thus a *full anti-diagonal histogram* is *planar*. The cardinality of a *full anti-diagonal regular histogram* of *valency* d and *dimension* n is d^{n-1}. The cardinality of *fully anti-diagonal planar* subsets of a *cartesian*, V^{C}, is $\sum_{w \in V} |U_w|$. If V is *regular* this is nd.

A *histogram* is a *line* if each pair of *non-zero states* differs in no more than one of the *variables*, $\forall S,T \in A^{\mathrm{FS}}\ (|S \cap T| >= n-1)$ where $V = \mathrm{vars}(A)$ and $n = |V|$. Define line $\in \mathcal{A} \to \mathbf{B}$

$$\mathrm{line}(A) \ :=\ \forall S,T \in A^{\mathrm{FS}}\ (|S \cap T| >= n-1)$$

A *pluri-variate histogram* is a *crown* if (i) it is *anti-planar*, and (ii) each pair of *non-zero states* differs in exactly two of the *variables*, $\forall S,T \in A^{\mathrm{FS}}\ (S \neq T \implies |S \cap T| = n-2)$ where $V = \mathrm{vars}(A)$ and $n = |V|$. Define crown $\in \mathcal{A} \to \mathbf{B}$

$$\mathrm{crown}(A) \ :=\ \mathrm{antiplanar}(A) \wedge (\forall S,T \in A^{\mathrm{FS}}\ (S \neq T \implies |S \cap T| = n-2))$$

A *crown histogram* A is a *full crown* if its non-zero cardinality equals the *dimension*. Define crownFull $\in \mathcal{A} \to \mathbf{B}$

$$\mathrm{crownFull}(A) \ :=\ \mathrm{crown}(A)\ \wedge |A^{\mathrm{F}}| = n$$

where $V = \mathrm{vars}(A)$. *Crown histograms* are also called *orthogonal*. The *pivot* of a *full crown* is the *zero state* which is the union of the intersections $\bigcup\{S \cap T : S,T \in A^{\mathrm{FS}}\} \in A^{\mathrm{CS}} \setminus A^{\mathrm{FS}}$.

A *histogram* is an *axial histogram* A if there exists exactly one *pivot state* $P \in V^{\mathrm{CS}}$ such that *non-zero states* differ in no more than one of the *variables*, $\forall S \in A^{\mathrm{FS}}\ (|S \cap P| >= n-1)$ where $V = \mathrm{vars}(A)$ and $n = |V|$

$$\mathrm{axial}(A) := |\{P : P \in (\prod_{w \in V} A\%\{w\})^{\mathrm{FS}},\ (\forall S \in A^{\mathrm{FS}}\ (|S \cap P| >= n-1))\}| = 1$$

So an *axial histogram* is intermediate between a *line* and a *crown*. The *pivot state* the union of the intersections $P = \bigcup\{S \cap T : S,T \in A^{\mathrm{FS}}\}$. If A is *tri-variate*, $n = 3$, then some *non-full anti-diagonal histograms*, $\forall S,T \in A^{\mathrm{FS}}\ (|S \cap T| \geq 1)$, are *axial*. An *anti-diagonal histogram* is less *orthogonal* than *axial* which in turn is less *orthogonal* than *crown*. An *axial histogram* A is *full* if its *effective* cardinality equals one plus the sum of the *valencies* less one, $|A^{\mathrm{F}}| = 1 + \sum_{w \in V}(|U_w| - 1)$. A *full axial regular histogram* has *effective* cardinality of $1 + n(d-1)$. The cardinality of *fully axial* subsets of a *cartesian*, V^{C}, is the *volume*, $\prod_{w \in V} |U_w|$. If V is *regular* this is d^n.

A *skeletal histogram* is such that all *reductions* to pairs of variables are *linear* or *axial*. Define skeletal $\in \mathcal{A} \to \mathbf{B}$

$$\text{skeletal}(A) := \forall K \subseteq V \ (|K| = 2 \implies \text{line}(A\%K) \vee \text{axial}(A\%K))$$

where $V = \text{vars}(A)$. All *axial histograms* are *skeletal*, $\text{axial}(A) \implies \text{skeletal}(A)$. If a *skeletal histogram* is not itself *axial*, then it has no *pivot state*

$$\text{skeletal}(A) \wedge \neg\text{axial}(A) \implies \bigcup\{S \cap T : S, T \in A^{\text{FS}}\} \notin A^{\text{CS}}$$

A *singleton histogram* can be defined in terms of *effective states* too, $\forall S, T \in A^{\text{FS}} \ (|S \cap T| = n)$.

A *histogram* is *pivoted* if there exists exactly one *effective state* $P \in A^{\text{FS}}$ which shares no *value* with any other *effective state*, $\forall S \in A^{\text{FS}} \ (S \neq P \implies S \cap P = \emptyset)$. The *state* P is called the *pivot*. Define pivot $\in \mathcal{A} \to \mathbf{B}$

$$\text{pivot}(A) := |\{P : P \in A^{\text{FS}}, \ (\forall S \in A^{\text{FS}} \ (S \neq P \implies S \cap P = \emptyset))\}| = 1$$

A *histogram* is a *full pivot* if $|A^{\text{F}}| = 1 + \prod_{w \in V}(|U_w| - 1)$. If the *valency* of a *regular full pivoted histogram* is two, $d = 2$, then the *histogram* is a *full diagonal*. Thus a *pivoted histogram* can be viewed as weakly *diagonal*. A *full pivoted histogram* is less *diagonal* than a *full diagonal* subset because it is more *cartesian*. A *full pivoted regular histogram* has *effective* cardinality of $1 + (d-1)^n$. The cardinality of *full pivoted* subsets of a *cartesian*, V^{C}, is the *volume*, $\prod_{w \in V} |U_w|$. If V is *regular* this is d^n.

A *histogram* is *anti-pivoted* if there exists exactly one *zero state* $P \in (A^{\text{C}} \setminus A^{\text{F}})$ which shares at least one *value* with all *effective states*, $\forall S \in A^{\text{FS}} \ (S \cap P \neq \emptyset)$. Define antipivot $\in \mathcal{A} \to \mathbf{B}$

$$\text{antipivot}(A) := |\{P : P \in (\prod_{w \in V} A\%\{w\})^{\text{FS}} \setminus A^{\text{FS}}, \ (\forall S \in A^{\text{FS}} \ (S \cap P \neq \emptyset))\}| = 1$$

An *anti-pivoted histogram*, A, is the *unit effective complement* of a *pivoted histogram*, $\text{antipivot}(A) \implies \text{pivot}(A^{\text{C}} \setminus A^{\text{F}})$. Thus the cardinality of the *full anti-pivot* equals the *volume* minus the cardinality of the *full pivot*, $|A^{\text{F}}| = |A^{\text{C}}| - (1 + \prod_{w \in V}(|U_w| - 1))$. If the *dimension* equals two, $n = 2$, the *effective full anti-pivot*, A^{F}, unioned with the *effective singleton* at the *pivot state*, $P \notin A^{\text{FS}}$, equals the *effective full axial*, $A^{\text{F}} \cup \{P\}^{\text{U}}$. Thus the cardinality of the *bi-variate full anti-pivot* equals the cardinality of the superset *full axial* less one, $|A^{\text{F}}| = \sum_{w \in V}(|U_w| - 1)$. A *bi-variate full anti-pivoted histogram* is more *orthogonal* than its *full axial* superset because it is less *cartesian*. A *full anti-pivoted regular histogram* has *effective* cardinality of $d^n - (1 + (d-1)^n)$. The cardinality of *full anit-pivoted* subsets of a *cartesian*, V^{C}, is the *volume*, $\prod_{w \in V} |U_w|$. If V is *regular* this is d^n.

Histograms may be classified in terms of their *counts*. A *histogram* is *unit* if $A = A^{\mathrm{U}}$. A *histogram* is *complete* in some *system* when $A^{\mathrm{U}} = A^{\mathrm{C}}$. A *histogram* is *unit or zero* if $A \supseteq A^{\mathrm{F}}$. A *non-empty histogram* is *zero* or *none* if $A^{\mathrm{F}} = \emptyset$. A *histogram* is *one* or *cartesian* if $A = A^{\mathrm{C}}$. A *histogram* is *one or zero* in some *system* if it is *complete* and *unit or zero* $(A^{\mathrm{U}} = A^{\mathrm{C}}) \wedge (A \supseteq A^{\mathrm{F}})$. A *histogram* is *not zero* or *any* if $A^{\mathrm{F}} \neq \emptyset$. A *histogram* is *none zero* or *all* or *completely effective* in some *system* when $A^{\mathrm{F}} = A^{\mathrm{C}}$.

A *histogram* is a *cartesian sub-volume* if the cartesian product of the *values* for each *variable* in the *trimmed states* is equal to the *trimmed states*

$$\prod \{\{\mathrm{filter}(\{v\}, S) : S \in \mathrm{states}(\mathrm{trim}(A))\} : v \in \mathrm{vars}(A)\} = \mathrm{states}(\mathrm{trim}(A))$$

Full planar histograms, *linear histograms* and *singleton histograms* are all *cartesian sub-volumes*.

A *cardinal substrate histogram* is such that the *variables* and *values* are integral, $V = \{1 \ldots |V|\}$ where $V = \mathrm{vars}(A)$, and $\forall w \in V\ \lozenge W = \{S_w : S \in A^{\mathrm{S}}\}\ (W = \{1 \ldots |W|\})$. Define the set of *cardinal substrate histograms* $\mathcal{A}_{\mathrm{c}}$

$$\begin{aligned}\mathcal{A}_{\mathrm{c}} = \{A : A \in \mathcal{A},\ V = \mathrm{vars}(A),\ V = \{1 \ldots |V|\},\\ \forall w \in V\ \lozenge W = \{S_w : S \in A^{\mathrm{S}}\}\ (W = \{1 \ldots |W|\})\}\end{aligned}$$

A *cardinal substrate histogram* in *system* U, $A \in \mathcal{A}_U \cap \mathcal{A}_{\mathrm{c}}$, is such that $\forall w \in V\ (U_w = \{1 \ldots |U_w|\})$.

A *histogram* $A \in \mathcal{A}_U$ in *system* U that is not necessarily a *cardinal substrate histogram*, $A \in \mathcal{A}_{\mathrm{c}} \vee A \notin \mathcal{A}_{\mathrm{c}}$, can be *reframed* to a *cardinal substrate histogram* of the same *geometry* given enumerations in the *variables* and *values*. Let $X \in \mathcal{V} \leftrightarrow (\mathbf{N}_{>0} \times (\mathcal{W} \leftrightarrow \mathbf{N}_{>0}))$ such that $\{(w, i) : (w, (i, \cdot)) \in X\} \in \mathrm{enums}(V)$ where $V = \mathrm{vars}(A)$, and $\forall (\cdot, (\cdot, Y)) \in X\ (Y \in \mathrm{enums}(\mathrm{dom}(Y)))$, so that $\mathrm{reframe}(X, A)$ is defined and $\mathrm{reframe}(X, A) \in \mathcal{A}_{\mathrm{c}}$. The *frame* mapping, X, is called a *cardinal substrate permutation*. There are $|V|! \prod_{w \in V} |U_w|!$ *cardinal substrate permutations* of a *histogram* $A \in \mathcal{A}_U$ in *system* U. If *histogram* A is *regular* having *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$ then the cardinality of *cardinal substrate permutations* is $n!(d!)^n$.

Given some *slice state* $R \in K^{\text{CS}}$, where $K \subset V$ and $V = \text{vars}(A)$, the *slice histogram*, $A * \{R\}^{\text{U}} \subset A$, is said to be *contingent* on the *incident slice state*, $A * \{R\}^{\text{U}} = \text{incidence}(A, R, |R|)$. For example, if the *slice histogram* is *diagonalised*, $\text{diagonal}(A * \{R\}^{\text{U}} \ \% \ (V \setminus K))$, then the *histogram*, A, is said to be *contingently diagonalised*. Let $\text{slices} \in \text{P}(\mathcal{V}) \times \mathcal{A} \to (\mathcal{S} \to \mathcal{A})$ be defined

$$\text{slices}(K, A) := \{(R, A * \{R\}^{\text{U}}) : R \in (A\%K)^{\text{S}}\}$$

A *histogram* $P \in \mathcal{A}$ is a *probability function* if its size is 1, $\text{size}(P) = 1 \implies P \in \mathcal{P}$, where the set of *probability functions*, $\mathcal{P}$, is defined in appendix 'Probability functions'. In this case the *histogram* is called a *probability histogram*.

In the case where a *histogram* $P' \in \mathcal{A}$ has a *size* less than or equal to 1, it is a *weak probability function*, $\text{size}(P') \leq 1 \implies P' \in \mathcal{P}'$, and is called a *weak probability histogram*.

The normalisations of two *histograms* may be compared as *probability histograms* by calculating the *relative entropy*. Let $\hat{A} = \text{normalise}(A)$. Let $A_1, A_2 \in \mathcal{A}$ be such that $\text{vars}(A_1) = \text{vars}(A_2)$, $\text{size}(A_1) > 0$ and $\text{size}(A_2) > 0$, then $\hat{A}_1, \hat{A}_2 \in \mathcal{A} \cap \mathcal{P}$ and the *relative entropy* of A_2 with respect to A_1 is

$$\text{entropyRelative}(A_1, A'_2) = \sum_{S \in A_1^{\text{FS}}} \hat{A}_1(S) \ln \frac{\hat{A}_1(S)}{\hat{A}'_2(S)}$$

where $A'_2 = A_2 + (A_1^{\text{F}} - A_2^{\text{F}})^\wedge$. The *relative entropy* is with respect to the *effective states* of *histogram* A_1. The *histogram* A_2 is stuffed with one *event* uniformly distributed over the *ineffective states* with respect to the *histogram* A_1, $(A_1^{\text{F}} - A_2^{\text{F}})^\wedge$. In the case where *histogram* A_2 is as *effective* as the *histogram* A_1, then no stuffing is needed, $A_2^{\text{F}} \geq A_1^{\text{F}} \implies A'_2 = A_2$. The *relative entropy* is zero when the *histograms* are equal, $\text{entropyRelative}(A_1, A_1) = 0$. Maximum *relative entropy* occurs when there is no *effective* intersection $A_1^{\text{F}} \cap A_2^{\text{F}} = \emptyset$ and the *histogram* A_1 is *uniform* over all but one *state* $S \in V^{\text{CS}}$, $\text{entropyRelative}(\text{resize}(z, V^{\text{C}} - \{S\}^{\text{U}}), \text{resize}(z, \{S\}^{\text{U}})) = \ln(z+1)/(v-1)$ where $v = |V^{\text{C}}|$ and $z \geq v \geq 2$.

3.4.7 Classifications

A *histogram* may be obtained from a *history* by counting the *event identifiers* of each *state*, $\text{histogram} \in \mathcal{H} \to \mathcal{A}$,

$$\text{histogram}(H) := \text{count}(\text{flip}(H)) = \text{count}(\{(S, x) : (x, S) \in H\})$$

That is, the *histogram*, histogram(H), is the distribution of *events* over *states* of the *history* H.

Let $A = \text{histogram}(H)$ where $H \in \mathcal{H}_U$, then $A \in \mathcal{A}_U$, $|A| \leq |H|$, $\text{states}(A) = \text{states}(H)$, $\text{vars}(A) = \text{vars}(H)$, $\text{size}(A) = \text{size}(H)$, and $\text{volume}(U)(A) = \text{volume}(U)(H)$. If each *state* appears only once in a *history* the resultant *histogram* will be *unit* and $|A| = |H|$.

The converse operation to construct a *history* from a *histogram* can be implemented only if the *counts* in the *histogram* are all integral, $\forall (S,c) \in A\ (c \in \mathbf{N})$, or $A \in \mathcal{S} \to \mathbf{N}$. If that is the case, then a *history* can be created, for example $\{((S,i),S) : (S,c) \in A,\ c > 0,\ i \in \{1 \dots c\}\}$.

All *histories* have the equivalent *histogram* in the set of *integral histograms*

$$\forall H \in \mathcal{H}\ (\text{histogram}(H) \in \mathcal{A}_{\text{i}})$$

A converse operation can be defined history $\in \mathcal{A}_{\text{i}} \to \mathcal{H}$

$$\text{history}(A) := \{((S,i),S) : (S,c) \in \text{trim}(A),\ i \in \{1 \dots c\}\}$$

But note that in general $\text{history}(\text{histogram}(H)) \neq H$.

Now consider the intermediate step between *history* and *classification*. Classify a *history* H by its *states*, to make a functional relation between the *states* and subsets of the *history*, $J \in \mathcal{S} \to \mathcal{H}$

$$J = \{(S, \{(x,T) : (x,T) \in H,\ T = S\}) : S \in \text{states}(H)\}$$

This structure is constrained $\forall (S,F) \in J\ (\text{states}(F) = \{S\})$ and $\sum(|F| : F \in \text{ran}(J)) = |\text{ids}(H)|$. The second constraint means that the *event identifiers* are unique in the whole structure J. The next step from *histories* to *classifications* is to throw away the duplicated *state*.

Let $\mathcal{G}$ be the set of *classifications*, $\mathcal{G} \subset \mathcal{S} \to (\text{P}(\mathcal{X}) \setminus \{\emptyset\})$. The set of *states* of a *classification* is defined, states $\in \mathcal{G} \to \text{P}(\mathcal{S})$ as $\text{states}(G) := \text{dom}(G)$. A *classification's states* must all have exactly the same set of *variables*

$$\forall G \in \mathcal{G}\ \forall S \in \text{states}(G)\ (\text{vars}(S) = \text{vars}(G))$$

where vars $\in \mathcal{G} \to \text{P}(\mathcal{V})$ is defined

$$\text{vars}(G) := \{v : S \in \text{states}(G),\ v \in \text{vars}(S)\}$$

In addition, the *classification* partitions its *event identifiers*,

$$\forall G \in \mathcal{G}\ \forall (S,C),(T,D) \in G\ (S \neq T \implies C \cap D = \emptyset)$$

or

$$\forall G \in \mathcal{G}\ (\mathrm{ran}(G) \in \mathrm{B}(\mathrm{ids}(G)))$$

where $\mathrm{ids} \in \mathcal{G} \to \mathrm{P}(\mathcal{X})$ is defined $\mathrm{ids}(G) := \bigcup \mathrm{ran}(G)$. Define $\mathrm{size} \in \mathcal{G} \to \mathbf{N}$ as $\mathrm{size}(G) := |\mathrm{ids}(G)|$.

A *classification* is the inverse of a *history*. That is, a *classification* is an *event identifier* component valued function of *state*. Define $\mathrm{classification} \in \mathcal{H} \to \mathcal{G}$

$$\begin{aligned}\mathrm{classification}(H) &:= \mathrm{inverse}(H)\\ &= \{(S, \{x : (x,T) \in H,\ T = S\}) : S \in \mathrm{states}(H)\}\end{aligned}$$

Another inverse restores the *history*, define $\mathrm{history} \in \mathcal{G} \to \mathcal{H}$

$$\mathrm{history}(G) := \{(x,S) : (S,C) \in G,\ x \in C\}$$

Thus *classifications* and *histories* are isomorphic

$$\forall H \in \mathcal{H}\ (\mathrm{history}(\mathrm{classification}(H)) = H)$$

$$\forall G \in \mathcal{G}\ (\mathrm{classification}(\mathrm{history}(G)) = G)$$

or $G \cong H$.

To construct a *histogram* from a *classification*, define $\mathrm{histogram} \in \mathcal{G} \to \mathcal{A}_{\mathrm{i}}$

$$\mathrm{histogram}(G) := \{(S, |C|) : (S,C) \in G\}$$

So the construction of a *histogram* from a *history* can also be defined

$$\mathrm{histogram}(H) = \{(S, |C|) : (S,C) \in H^{-1}\}$$

where $H^{-1} := \mathrm{classification}(H)$.

The *histogram* of a *reduction* of a *history* equals the *reduction* of the *histogram* of the *history*,

$$\mathrm{histogram}(H\ \%\ V) = \mathrm{histogram}(H)\ \%\ V$$

The *histogram* of an *addition* of *histories* equals the *addition* of the *histograms* of the *histories*,

$$\mathrm{histogram}(H_1 + H_2) = \mathrm{histogram}(H_1) + \mathrm{histogram}(H_2)$$

Multiplication is also homomorphic,

$$\mathrm{histogram}(H_1 * H_2) = \mathrm{histogram}(H_1) * \mathrm{histogram}(H_2)$$

3.4.8 Histogram entropy

The definitions of *probability function* and *entropy* in the context of *histograms* that follow are discussed more generally in appendices 'Probability functions' and 'Entropy and Gibbs' inequality', below.

Let $A \in \mathcal{A}$ be a *non-zero histogram* of *size* $z = \mathrm{size}(A) > 0$. The *normalised histogram*,

$$\hat{A} \;=\; \mathrm{normalise}(A) \;=\; \{(S, c/z) : (S, c) \in A\}$$

is a *probability function*, $\hat{A} \in \mathcal{P}$. That is, the *normalised counts* are between zero and one, $\mathrm{ran}(\hat{A}) \subset \mathbf{Q}_{[0,1]}$, and sum to one, $\sum_{S \in A^{\mathrm{S}}} \hat{A}_S = 1$.

Entropy is defined for any *probability function*, so may be defined for *non-zero histograms*, $\mathrm{entropy} \in \mathcal{A} \to \mathbf{Q}_{\geq 0} \ln \mathbf{Q}_{>0}$, as

$$\mathrm{entropy}(A) := -\sum_{S \in A^{\mathrm{FS}}} \hat{A}_S \ln \hat{A}_S$$

Entropy is undefined for *zero* or *empty histograms*, $A^{\mathrm{F}} = \emptyset$. The *scaled entropy* or *size* scaled *entropy* is $z \times \mathrm{entropy}(A)$.

Entropy is positive, $\mathrm{entropy}(A) \geq 0$. The minimum *entropy* occurs where the *histogram* is *singleton*,

$$\forall S \in V^{\mathrm{CS}}\ (\mathrm{entropy}(\{(S, z)\}) = 0)$$

That is, the least *uniform histograms*, which have singular *effective volume*, $|\{(S, z)\}^{\mathrm{F}}| = 1$, have the lowest *entropy*.

If the *histogram* is *integral*, $A \in \mathcal{A}_{\mathrm{i}}$, the *entropy* is less than or equal to the logarithm of the *size*, $\mathrm{entropy}(A) \leq \ln z$. Given a set of *variables* $V = \mathrm{vars}(A)$ the maximum *entropy* occurs where the *histogram* is the *scaled cartesian*, $\mathrm{entropy}(V^{\mathrm{C}}_z) = \ln z$, where $v = |V^{\mathrm{C}}|$, $z/v \in \mathbf{N}_{>0}$, and $V^{\mathrm{C}}_z = \mathrm{scalar}(z/v) * V^{\mathrm{C}}$. That is, the most *uniform histogram*, $\mathrm{ran}(V^{\mathrm{C}}_z) = \{z/v\}$, with the highest *effective volume*, $|(V^{\mathrm{C}}_z)^{\mathrm{F}}| = v$, has the highest *entropy*.

Any non-empty, finite $\mathcal{X}$-valued function of $\mathcal{Y}$ implies a distribution of $\mathcal{Y}$ over $\mathcal{X}$ and hence a *probability function* by normalisation,

$$\forall R \in \mathcal{Y} \to \mathcal{X}\ (0 < |R| < \infty \implies \{(x, |C|) : (x, C) \in R^{-1}\}^{\wedge} \in \mathcal{P})$$

Similarly, every non-empty, finite *history* implies an *integral histogram* which is the distribution of *event identifiers* over *states*. The *normalised histogram* is a *probability function*,

$$\forall H \in \mathcal{H}\ (0 < |H| < \infty \implies \{(S, |C|/|H|) : (S,C) \in H^{-1}\} \in \mathcal{A} \cap \mathcal{P})$$

Let $I \in \mathcal{H} \to \mathcal{A}$ be the *histogram* valued function of all possible *histories* of *size* z in *variables* V,

$$\begin{aligned} I &= \{(H, \{(S, |C|) : (S,C) \in H^{-1}\}) : H \in \{1 \ldots z\} :\to V^{\mathrm{CS}}\} \\ &= \{(H, \mathrm{histogram}(H)) : H \in \{1 \ldots z\} :\to V^{\mathrm{CS}}\} \end{aligned}$$

Let $W \in \mathcal{A} \to \mathbf{N}_{>0}$ be the cardinality of *histories* for each *histogram*,

$$\begin{aligned} W &= \{(A, |D|) : (A,D) \in I^{-1}\} \\ &= \{(A, \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!}) : (A, \cdot) \in I^{-1}\} \end{aligned}$$

The cardinality of *histories* for a *histogram* A is the *multinomial coefficient*,

$$W(A) = \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \in \mathbf{N}_{>0}$$

In the case where the *histogram counts* are large, $\forall (\cdot, c) \in A\ (c > 0 \implies c \gg \ln c)$, Stirling's approximation, $\ln n! = n \ln n - n + \mathrm{O}(\ln n)$, may be applied,

$$\begin{aligned} \ln W(A) &= \ln z! - \sum_{S \in A^{\mathrm{S}}} \ln A_S! \\ &\approx (z \ln z - z) - \sum_{S \in A^{\mathrm{FS}}} (A_S \ln A_S - A_S) \\ &= -z \sum_{S \in A^{\mathrm{FS}}} \hat{A}_S \ln \hat{A}_S \\ &= z \times \mathrm{entropy}(A) \end{aligned}$$

That is, the logarithm of the probability of a *histogram* of an arbitrary *history* of *size* z in *variables* V varies with the *entropy* of the *histogram*,

$$\ln \hat{W}(A) \sim \mathrm{entropy}(A)$$

The *history probability function*, $(\{1 \ldots z\} :\to V^{\mathrm{CS}}) \times \{1/v^z\} \in \mathcal{P}$, is uniform, but the corresponding *histogram probability function*, $\hat{W} \in \mathcal{P}$, is not uniform.

The least probable *histograms* are the *singletons*,

$$\mathrm{mind}(W) = \{\{(S,z)\} : S \in V^{\mathrm{CS}}\}$$

which have a cardinality of one, $\text{minr}(W) = 1$,

$$\{\{(S,z)\} : S \in V^{\text{CS}}\} \quad :\leftrightarrow: \quad \{\{1 \ldots z\} \times \{S\} : S \in V^{\text{CS}}\}$$

and zero *entropy*, $\forall A \in \text{mind}(W)\ (\text{entropy}(A) = 0)$.

The modal *histogram* is the *scaled cartesian*,

$$\text{maxd}(W) \quad = \quad \{V^{\text{C}}_z\}$$

which has a cardinality of $\text{maxr}(W) = z!/((z/v)!)^v$, and maximum *entropy*, $\ln \text{maxr}(\hat{W}) \sim \text{entropy}(V^{\text{C}}_z) = \ln z$.

3.5 Transforms

3.5.1 Definition

A *histogram* X which has its *variables* partitioned into two components, the *underlying variables* V and the *derived variables* W, such that $\text{vars}(X) = V \cup W$, forms a pair (X,W) called a *transform*. The *underlying* and *derived variables* are disjoint $V \cap W = \emptyset$. The set of all *transforms* $\mathcal{T}$ such that $\mathcal{T} \subset \mathcal{A} \times \text{P}(\mathcal{V})$ is

$$\mathcal{T} = \{(X,W) : X \in \mathcal{A},\ W \in \text{P}(\text{vars}(X))\}$$

Define various accessor functions, $\text{histogram} \in \mathcal{T} \to \mathcal{A}$ as $\text{histogram}((X,W)) := X$, and $\text{underlying} \in \mathcal{T} \to \text{P}(\mathcal{V})$ as $\text{underlying}((X,W)) := \text{vars}(X) \setminus W$, and $\text{derived} \in \mathcal{T} \to \text{P}(\mathcal{V})$ as $\text{derived}((X,W)) := W$.

The transform function is a special case of a *histogram expression* that applies the *transform* to some *histogram* A by *multiplying* by the *transform histogram* and then *reducing* by the *derived variables*, $\text{transform} \in \mathcal{T} \times \mathcal{A} \to \mathcal{A}$

$$\text{transform}((X,W),A) := A * X \ \%\ W$$

Extend the *histogram multiplication* operator to *transforms* to make a convenient shorthand, $(*) \in \mathcal{A} \times \mathcal{T} \to \mathcal{A}$

$$\begin{aligned} A * (X,W) \quad &:= \quad \text{transform}((X,W),A) \\ &= \quad A * X \ \%\ W \end{aligned}$$

Often the *variables* of A will be the same as the *underlying variables* of the *transform*, but this is not necessary.

There are a some special cases of *transforms* of a *histogram* A in *variables* $V = \mathrm{vars}(A)$. A *disjoint transform* $T = (X, \mathrm{vars}(X))$ has an empty set of *underlying variables* $\mathrm{underlying}(T) = \emptyset$. If $V \cap \mathrm{derived}(T) = \emptyset$ and because $V \cap \mathrm{underlying}(T) = \emptyset$

$$\mathrm{transform}((X, \mathrm{vars}(X)), A) = Z_A * X$$

where Z_A is the *scalar histogram* $Z_A = \{(\emptyset, \mathrm{size}(A))\}$.

On the other hand, the *null transform* $T = (X, \emptyset)$ has an empty set of *derived variables* $\mathrm{derived}(T) = \emptyset$

$$\mathrm{transform}((X, \emptyset), A) = Z_{A*X}$$

where $V = \mathrm{underlying}(T)$ and Z_{A*X} is the *scalar histogram* $Z_{A*X} = \{(\emptyset, \mathrm{size}(A * X))\}$.

The *empty transform*, $(\emptyset, \emptyset)$, is both *null* and *disjoint*, but when *applied* to *histogram* A produces the *empty histogram*, $A * (\emptyset, \emptyset) = \emptyset$.

The set of all *transforms* $\mathcal{T}_U \subset \mathcal{T}$ in a particular *system* U is defined explicitly

$$\mathcal{T}_U = \{(X, W) : Y \subseteq \mathrm{vars}(U),\ W \subseteq Y,\ X \in \mathrm{cartesian}(U)(Y) \to \mathbf{Q}_{\geq 0}\}$$

A *transform* $T \in \mathcal{T}_U$ in *system* U, having *transform histogram* $X = \mathrm{his}(T)$, *underlying variables* $V = \mathrm{und}(T)$ and *derived variables* $W = \mathrm{der}(T)$, is a *frame transform* if there exists either (i) a *literal frame* map $Y \in V \cdot W$, or (ii) a *non-literal frame* map $Y \in V \cdot (W \times (\mathcal{W}_U \leftrightarrow \mathcal{W}_U))$, such that $\mathrm{reframe}(Y, X\%V)$ is defined in *system* U. The *underlying dimension* of *frame transforms* equals the *derived dimension*, $|V| = |W|$. The *valencies* of the *variables* of the pairs of the *frame* map are equal. That is, for (i) the *literal* case, $\forall (v, w) \in Y\ (|U_v| = |U_w|)$, and for (ii) the *non-literal* case, $\forall (v, (w, \cdot)) \in Y\ (|U_v| = |U_w|)$. Hence the *underlying volume* equals the *derived volume*, $|V^{\mathrm{C}}| = |W^{\mathrm{C}}|$. A special case of a *frame transform* is a *reframe transform*, where the *derived histogram* equals the *reframe* of the *underlying histogram*, $X\%W = \mathrm{reframe}(Y, X\%V)$, or $(V^{\mathrm{C}} * T)^{\mathrm{F}} = \mathrm{reframe}(Y, V^{\mathrm{C}})$.

If a pair of *transforms* $R, T \in \mathcal{T}_U$ in the same *system* U have *variables* such that $\exists Y \in \mathrm{vars}(R) \cdot \mathrm{vars}(T)\ \forall (v, u) \in Y\ (U_v = U_u)$ then the *variables* of

R and the *variables* of T are said to be *literal frames* of each other mapped by Y. If underlying(R) = underlying(T) then the *variables* of R and the *variables* of T are *derived frames*.

The function reframe $\in (\mathcal{V} \leftrightarrow \mathcal{V}) \times \mathcal{T} \to \mathcal{T}$ is defined

$$\begin{aligned}\mathrm{reframe}(Y,T) &:= (\mathrm{reframe}(Y,\mathrm{his}(T)),\\ &\qquad \{Y_w : w \in \mathrm{der}(T) \cap \mathrm{dom}(Y)\} \cup (\mathrm{der}(T) \setminus \mathrm{dom}(Y)))\end{aligned}$$

his = histogram and der = derived. reframe(Y,T) is defined if the underlying reframe$(Y, \mathrm{his}(T))$ is defined.

Similarly for pairs of *transforms* having *variables* which are *non-literal frames* of each other. Define reframe $\in (\mathcal{V} \leftrightarrow (\mathcal{V} \times (\mathcal{W} \leftrightarrow \mathcal{W})) \times \mathcal{T} \to \mathcal{T}$

$$\begin{aligned}\mathrm{reframe}(Y,T) &:= (\mathrm{reframe}(Y,\mathrm{his}(T)),\\ &\qquad \{w' : w \in \mathrm{der}(T) \cap \mathrm{dom}(Y),\ (w',\cdot) = Y_w\} \cup\\ &\qquad (\mathrm{der}(T) \setminus \mathrm{dom}(Y)))\end{aligned}$$

reframe(Y,T) is defined if reframe$(Y, \mathrm{his}(T))$ is defined.

An important subset of the *transforms* is the set of *functional transforms*. The *functional transforms* $\mathcal{T}_\mathrm{f} \subset \mathcal{T}$ is the subset of all *transforms* which form a functional relation between the *underlying states* and the *derived states* having non-zero *count*

$$\mathcal{T}_\mathrm{f} = \{T : T \in \mathcal{T},\ X = \mathrm{his}(T),\ V = \mathrm{und}(T),\ \mathrm{split}(V, X^{\mathrm{FS}}) \in \mathcal{S} \to \mathcal{S}\}$$

where the his = histogram and und = underlying. The *histogram* of a *functional transform* is *causal*, causal(his(T)). The *underlying variables*, und(T), are said to *cause* the *derived variables*, der(T).

If a *transform* is *functional*, $T \in \mathcal{T}_\mathrm{f}$, then an inverse function can be defined. First define another function stateDeriveds $\in \mathcal{T} \to \mathrm{P}(\mathcal{S})$ as

$$\mathrm{stateDeriveds}((X,W)) := \mathrm{states}(X\%W)$$

Then define inverse $\in \mathcal{T}_\mathrm{f} \to (\mathcal{S} \to \mathcal{A})$

$$\begin{aligned}\mathrm{inverse}(T) &:= \{(R, \{(S \setminus R, c) : (S,c) \in X,\ S \supseteq R\}) : R \in \mathrm{std}(T)\}\\ &= \{(R, X * \{R\}^{\mathrm{U}} \ \%\ V) : R \in \mathrm{std}(T)\}\\ &= \{(R, B\%V) : (R,B) \in \mathrm{slices}(W,X)\}\end{aligned}$$

where std = stateDeriveds, $X = \text{his}(T)$, $W = \text{der}(T)$ and $V = \text{und}(T)$. So $\text{dom}(\text{inverse}(T)) = \text{std}(T)$ and $\sum \text{ran}(\text{inverse}(T)) = X \ \% \ V$.

The inverse function can be defined in terms of the relational inverse function, $\text{inverse} \in (\mathcal{X} \to \mathcal{Y}) \to (\mathcal{Y} \to \text{P}(\mathcal{X}))$

$$\text{inverse}(T) := \text{inverse}(\{((S\%V, c), S\%W) : (S, c) \in X\})$$

where $(X, W) = T$ and $V = \text{und}(T)$. The inverse function can be defined in terms of *incidence*

$$\text{inverse}(T)(R) = \text{incidence}(X, R, |R|) \ \% \ V$$

where $R \in \text{std}(T)$.

A *functional transform* $T \in \mathcal{T}_\text{f}$ is said to be *effective* with respect to a *histogram* $A \in \mathcal{A}$, where $\text{vars}(A) = \text{underlying}(T)$, if the *effective underlying states* of the *transform* are a superset of the *effective states* of the *histogram*, $(X\%V)^\text{F} \geq A^\text{F}$ where $X = \text{his}(T)$ and $V = \text{his}(T)$. Define the function $\text{effective} \in \mathcal{A} \times \mathcal{T}_\text{f} \to \mathcal{T}_\text{f}$ which returns the smallest cardinality *effective transform* of a *transform* T with respect to a *histogram* A as

$$\text{effective}(A, T) := (\sum\{(A * C)^\text{F} * \{R\}^\text{U} : (R, C) \in \text{inverse}(T)\}, \text{der}(T))$$

The function $\text{effective}(A, T)$ is undefined if there is no *effective* intersection $X = \emptyset$. The *transform* T of the *cartesian histogram* V^C is already minimally *effective*, $\text{effective}(V^\text{C}, T) = T$.

A *functional transform* $T \in \mathcal{T}_\text{f}$ has a set of *reduced transforms* with respect to a *histogram* $A \in \mathcal{A}$, where $\text{vars}(A) = \text{underlying}(T)$. The *transform* is *functional* and so has an *inverse*. If the *transform* is *effective* with respect to the given *histogram*, $(X\%V)^\text{F} \geq A^\text{F}$, the set of the *trimmed* applications of the elements of the range of the *inverse*, $\text{ran}(\text{inverse}(T)) \subset \mathcal{A}$, partitions the given *trimmed histogram*, $(X\%V)^\text{F} \geq A^\text{F} \implies \{\text{trim}(A * C) : C \in \text{ran}(\text{inverse}(T))\} \setminus \{\emptyset\} \in \text{B}(\text{trim}(A))$. Each of the *reduced transforms* has a subset of the *derived variables* such that the partition of the set of *trimmed* applications is unchanged. Define $\text{reductions} \in \mathcal{A} \times \mathcal{T}_\text{f} \to \text{P}(\mathcal{T}_\text{f})$

$$\begin{aligned}
&\text{reductions}(A, T) := \\
&\quad \{R : K \subseteq W,\ R = (X\%(V \cup K), K), \\
&\quad \ \{A^\text{F} * D : (\cdot, D) \in R^{-1}\} \setminus \{\emptyset\} = \{A^\text{F} * C : (\cdot, C) \in T^{-1}\} \setminus \{\emptyset\}\}
\end{aligned}$$

where $T^{-1} = \text{inverse}(T)$, $(X, W) = T$ and $V = \text{vars}(A) \supseteq \text{und}(T)$. The set of *reductions* contains the *transform* itself, $T \in \text{reductions}(A, T)$. The set of *reductions* of a *null transform* is the singleton of the *null transform*, $\text{reductions}(A, (X\%V, \emptyset)) = \{(X\%V, \emptyset)\}$. The application of a *reduction* has the same *size* as the *transformed histogram*, $\forall R \in \text{reductions}(A, T)\ (\text{size}(A * R) = \text{size}(A * T))$. If a *transformed histogram* is *diagonal*, $\text{diagonal}(A * T)$, then the set of *reductions* has cardinality equal to the cardinality of the powerset of the *derived variables*, $\text{diagonal}(A * T) \implies |\text{reductions}(A, T)| = |\text{P}(W)| - 1 = 2^{|W|} - 1$, which is at least the cardinality of the *derived variables*. That is, the *transform* can be *reduced* to any of the *derived variables*

$$\text{diagonal}(A * T) \implies \{(X\%(V \cup \{w\}), \{w\}) : w \in W\} \subseteq \text{reductions}(A, T)$$

The subset of *functional transforms* that contains only *transforms* in a particular *system* U is $\mathcal{T}_{U,\text{f}} = \mathcal{T}_U \cap \mathcal{T}_\text{f}$. A *functional transform* $T \in \mathcal{T}_{U,\text{f}}$ is *left total* if it is *completely effective* in its *underlying* $(X\%V)^\text{F} = V^\text{C}$, where $(X, W) = T$ and $V = \text{und}(T)$. Similarly T is *right total* if it is *completely effective* in its *derived* $(X\%W)^\text{F} = W^\text{C}$. A *full functional transform* T is (i) *left total*, (ii) *right total*, and (iii) such that the *underlying volume* equals the *derived volume*, $|V^\text{C}| = |W^\text{C}|$. A *full functional transform* is bijective between its *underlying states* and *derived states*, $\text{split}(V, X^\text{S}) \in V^\text{CS} \leftrightarrow W^\text{CS}$. A special case of a *full functional transform* is a *frame full functional transform*, where V and W are *frames* of each other, $\exists Y \in V \cdot W\ \forall (v, w) \in Y\ (|U_v| = |U_w|)$. In this case, not only are the *underlying volume* and the *derived volume* equal, $|V^\text{C}| = |W^\text{C}|$, but the *underlying dimension* equals the *derived dimension*, $|V| = |W|$, and the *underlying valencies* equal the *derived valencies*, $\forall (v, w) \in Y\ (|U_v| = |U_w|)$. A special case of a *frame full functional transform* is a *value full functional transform* which is a *reframe transform*, $X\%W = \text{reframe}(Y, X\%V)$. In this case the *derived states*, $(X\%W)^\text{S}$, are *reframed underlying states*, $(X\%V)^\text{S}$. That is, $\exists Y \in V \cdot W\ \forall (v, w) \in Y\ (\text{split}(\{v\}, (X\%\{v, w\})^\text{S}) \in \{v\}^\text{CS} \leftrightarrow \{w\}^\text{CS})$.

The subset of *transforms* that contains only *unit transforms* is defined $\mathcal{T}_\text{U} \subset \mathcal{T}$

$$\mathcal{T}_\text{U} = \{T : T \in \mathcal{T},\ X = \text{his}(T),\ X = X^\text{U}\}$$

The subset of *unit functional transforms* in a particular *system* U is $\mathcal{T}_{U,\text{f},\text{U}} = \mathcal{T}_U \cap \mathcal{T}_\text{f} \cap \mathcal{T}_\text{U}$.

Consider *unit functional transform* $T \in \mathcal{T}_{U,\text{f},\text{U}}$ that is also *left total* $(X\%V)^\text{F} = V^\text{C}$ where $X = \text{his}(T)$ and $V = \text{und}(T)$. *Left total unit functional transforms* are also known as *one functional transforms* $T \in \mathcal{T}_{U,\text{f},1}$. The *size* and cardinality of the *histogram* of the *transform* equals the *volume* of the *underlying variables* $\mathrm{size}(X) = |X| = |V^{\mathrm{C}}|$. A *one functional transform* is *size-conservative* when applied to a given argument *histogram* A, $\mathrm{size}(A * T) = \mathrm{size}(A)$, where $\mathrm{und}(T) = \mathrm{vars}(A)$. The *empty transform* is not *one functional*, $(\emptyset, \emptyset) \notin \mathcal{T}_{U,\mathrm{f},1}$.

A *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ is a functor (or monoid homomorphism) of the *histogram addition* operator, $(A * T) + (I * T) = (A + I) * T$ where $I \in \mathcal{A}$ and $\mathrm{vars}(I) = \mathrm{und}(T) = \mathrm{vars}(A)$. A *one functional transform* is also a functor of *histogram subtraction*, $(A * T) - (D * T) = (A - D) * T$ where $D \in \mathcal{A}$ and $\mathrm{vars}(D) = \mathrm{und}(T) = \mathrm{vars}(A)$ if the *subtraction* has an inverse *addition*, $A - D + D = A$, or $D \leq A$.

The *histogram* X of *one functional transform* T is *argument-conservative* where it is *multiplied* by an argument *histogram* A and then *reduced* by the *variables* of A, if the *derived variables* of T are disjoint, $\mathrm{vars}(A) \cap \mathrm{derived}(T) = \emptyset$. That is, $A * X \ \% \ \mathrm{vars}(A) = A$. Also, $\mathrm{size}(A * X) = \mathrm{size}(A)$ and $|A * X| = |A|$. These constraints hold even if the *variables* of A and the *underlying variables* of T do not overlap, $\mathrm{vars}(A) \cap \mathrm{vars}(T) = \emptyset$. One can think of the *histograms* of *left total unit functional transforms* as adding *derived variables* while conserving the given *histogram* as an invariant.

If *unit functional transform* $T \in \mathcal{T}_{U,\mathrm{f},\mathrm{U}}$ is not *left total*, then there are weaker constraints $A*X \ \% \ \mathrm{vars}(A) \subseteq A$, $\mathrm{size}(A*X) \leq \mathrm{size}(A)$ and $|A*X| \leq |A|$. By contrast if T is a *full functional transform*, but is not necessarily *unit*, it may not be *argument-conservative* but it does obey the constraint $|A * T| = |A|$.

The set of *one functional transforms* $\mathcal{T}_{U,\mathrm{f},1}$ can be constructed explicitly

$$\begin{aligned}
&\mathcal{T}_{U,\mathrm{f},1} = \\
&\quad \{(X, W) : V, W \in \mathrm{P}(\mathrm{vars}(U)),\ V \cap W = \emptyset, \\
&\quad Q \in \mathrm{cartesian}(U)(V) \to \mathrm{cartesian}(U)(W),\ |Q| = |V^{\mathrm{C}}|, \\
&\quad X = \{(S \cup R, 1) : (S, R) \in Q\}\}
\end{aligned}$$

The set of *one functional models* $\mathcal{M}_{U,\mathrm{f},1}$ is such that each *model* $M \in \mathcal{M}_{U,\mathrm{f},1}$ has a corresponding *one functional transform*, $\mathrm{transform}(M) \in \mathcal{T}_{U,\mathrm{f},1}$, where $\mathrm{transform} \in \bigcup_{U \in \mathcal{U}} (\mathcal{M}_{U,\mathrm{f},1} \to \mathcal{T}_{U,\mathrm{f},1})$. There is a shorthand defined $M^{\mathrm{T}} := \mathrm{transform}(M)$.

The *one functional transforms*, $\mathcal{T}_{U,\mathrm{f},1}$, are *derived state* valued left total functions of *underlying state*,

$$\forall T \in \mathcal{T}_{U,\mathrm{f},1}\ (\mathrm{split}(V, X^{\mathrm{S}}) \in V^{\mathrm{CS}} :\to W^{\mathrm{CS}})$$

where $(X, W) = T$ and $V = \mathrm{und}(T)$. In order to construct a coordinate from a *state* define $()^{\square} \in \mathcal{S} \to \mathcal{L}(\mathcal{W})$ as

$$S^{\square} := \{(i, u) : ((v, u), i) \in \mathrm{order}(D_{\mathcal{V}\times\mathcal{W}}, S)\}$$

where $D_{\mathcal{V}\times\mathcal{W}}$ is an *order* on the *variables* and *values*. The converse function to construct a *state* from a coordinate $()^{V} \in \mathcal{L}(\mathcal{W}) \to \mathcal{S}$ is

$$S^{V} := \{(v, S_i) : (v, i) \in \mathrm{order}(D_{\mathcal{V}}, V)\}$$

Now *one functional transforms* may be represented as *derived value* coordinate valued left total functions of *underlying value* coordinate,

$$\begin{aligned}\{(S^{\square}, R^{\square}) : (S, R) \in \mathrm{split}(V, X^{\mathrm{S}})\} &\in \{S^{\square} : S \in V^{\mathrm{CS}}\} :\to \{R^{\square} : R \in W^{\mathrm{CS}}\} \\ &\subset \mathcal{W}^{n} \to \mathcal{W}^{m}\end{aligned}$$

where $n = |V|$ and $m = |W|$.

So an alternative definition for a *one functional transform* is a tuple of (i) the *underlying variables*, V, (ii) the *derived variables*, W, and (iii) a *derived value* coordinate valued left total function of *underlying value* coordinate, f,

$$\begin{aligned}&\mathcal{T}_{U,\mathrm{f},1} = \\ &\quad \{(V, W, f) : V, W \in \mathrm{P}(\mathrm{vars}(U)),\ V \cap W = \emptyset, \\ &\qquad\qquad f \in \{S^{\square} : S \in V^{\mathrm{CS}}\} :\to \{R^{\square} : R \in W^{\mathrm{CS}}\}\}\end{aligned}$$

The *histogram* of a *function-defined one functional transform* $T = (V, W, f) \in \mathcal{T}_{U,\mathrm{f},1}$ is

$$\mathrm{histogram}(T) \ := \ \{S \cup f(S^{\square})^{W} : S \in V^{\mathrm{CS}}\} \times \{1\}$$

In the special case where the *transform* is *mono-derived-variate*, $T = (V, \{w\}, f)$, the function may be simplified to $f \in \{S^{\square} : S \in V^{\mathrm{CS}}\} :\to U_w$, and the *histogram* is

$$\mathrm{histogram}(T) := \{S \cup \{(w, f(S^{\square}))\} : S \in V^{\mathrm{CS}}\} \times \{1\}$$

In the further special case of *mono-derived-variate transform* where its *variables* are real, $\forall v \in V\ (U_v = \mathbf{R})$ and $U_w = \mathbf{R}$, then the function is a real valued left total function of a real coordinate, $f \in \mathbf{R}^n :\to \mathbf{R}$. Here the *cartesian states* are $V^{\mathrm{CS}} = \prod_{v \in V}(\{v\} \times \mathbf{R})$, so the *histogram* is

$$\begin{aligned} \mathrm{histogram}(T) &= \{S \cup \{(w, f(S^{\square}))\} : S \in \prod_{v \in V}(\{v\} \times \mathbf{R})\} \times \{1\} \\ &= \{S^V \cup \{(w, f(S))\} : S \in \mathbf{R}^n\} \times \{1\} \end{aligned}$$

The *cartesian volume* is infinite, $|V^{\mathrm{C}}| = |\mathbf{R}^n|$, so the cardinality of the *histogram* is infinite, $|\mathrm{histogram}(T)| = |\mathbf{R}^n|$.

The reals form a metric space so a real valued function of real coordinates may be discretised given a finite subset of the reals $D \subset \mathbf{R} : |D| < \infty$. The discretised function is

$$\mathrm{discrete}(D, n)(f) \ := \ \{(X, \mathrm{nearest}(D, f(X))) : X \in D^n\} \in D^n :\to D$$

where $\mathrm{nearest} \in \mathrm{P}(\mathbf{R}) \times \mathbf{R} \to \mathbf{R}$ is defined

$$\mathrm{nearest}(D, r) \ := \ t : \{t\} \in \mathrm{mind}(\{(s, (|r - s|, s)) : s \in D\})$$

The cardinality of the discretised *transform's histogram* is finite,

$$|\mathrm{histogram}((V, \{w\}, \mathrm{discrete}(D, n)(f)))| = |D^n| = |D|^n$$

An example of a *transform* defined by a real valued function occurs in the function composition of artificial neural networks. Here a *transform* represents a model of a neuron called a perceptron $T = (V, \{w\}, f(K, Q))$ where the function $f(K, Q) \in \mathbf{R}^n :\to \mathbf{R}$ is parameterised by (i) some differentiable function $K \in \mathbf{R} :\to \mathbf{R}$, called the activation function, and (ii) a vector of weights, $Q \in \mathbf{R}^{n+1}$, and is defined

$$f(K, Q)(X) \ := \ K(\sum_{i \in \{1 \ldots n\}} Q_i X_i \ + \ Q_{n+1})$$

A *functional transform* $T \in \mathcal{T}_{\mathrm{f}}$ may be applied to a *history* $H \in \mathcal{H}$ in the *underlying variables* of the *transform*, $\mathrm{vars}(H) = \mathrm{und}(T)$, to construct a *derived history.* Define $\mathrm{transform} \in \mathcal{T}_{\mathrm{f}} \times \mathcal{H} \to \mathcal{H}$ as

$$\mathrm{transform}(T, H) \ := \ \{(x, P_S) : (x, S) \in H\}$$

where $V = \mathrm{und}(T)$, and $P = \mathrm{split}(V, \mathrm{his}(T)^{\mathrm{FS}}) \in \mathcal{S} \to \mathcal{S}$. Let $H * T := \mathrm{transform}(T, H)$. So $\mathrm{vars}(H * T) = W$ where $W = \mathrm{der}(T)$. If the *transform* is *one functional*, $T \in \mathcal{T}_{U,\mathrm{f},1}$, the *size* is unchanged, $|H * T| = |H|$, and the *event identifiers* are conserved, $\mathrm{ids}(H * T) = \mathrm{ids}(H)$.

3.5.2 Converses

The *simple converse* of a *transform* $T \in \mathcal{T}$ is straightforwardly defined as the pair of the *reciprocal* of the *histogram* and the *underlying variables*. Define converseSimple $\in \mathcal{T} \to \mathcal{T}$ as

$$\text{converseSimple}(T) \ := \ (1/X, V)$$

where $X = \text{histogram}(T)$ and $V = \text{und}(T)$.

The *natural converse*, which is denoted by a dagger, is similar but also scales inversely by the *effective incident volume* of each of the *derived states*. Define converseNatural $\in \mathcal{T} \to \mathcal{T}$ as

$$\text{converseNatural}(T) \ := \ (\frac{X^{\text{F}}}{X\%W}, V)$$

where $(X, W) = T$ and $V = \text{und}(T)$. Denote the *natural converse* with a dagger, $T^{\dagger} = \text{converseNatural}(T)$.

In the case of the *natural converse* $T^{\dagger} = (X/(X\%W), V)$ of *unit functional transform* $T \in \mathcal{T}_{U,\text{f},\text{U}}$, the *incident volume* of any *state* $S \in \text{states}(X)$ is $(X\%W)_R = \text{incidence}(X, R, |W|)$ where $R = \text{filter}(W, S)$. The *counts* of the *natural converse histogram* of a *unit functional transform* are greater than zero and less than or equal to one, $\forall (S, c) \in X/(X\%W)\ (0 < c \leq 1)$. The *reduction* of the *natural converse histogram* of a *unit functional transform* to the *derived states* is a *unit histogram*, $(X/(X\%W))\%W \subseteq W^{\text{C}}$.

In the case of the *natural converse* $T^{\dagger} = (X/(X\%W), V)$ of a *one functional transform* $T \in \mathcal{T}_{U,\text{f},1}$, the *natural converse* may be expressed in terms of *components*,

$$T^{\dagger} \ := \ (\sum_{(R,C) \in T^{-1}} \{R\}^{\text{U}} * \hat{C},\ V)$$

where the normalisation is defined $\hat{A} = A/(A\%\emptyset)$.

There are other *converses* which are variations on the definition of *natural converse* that scale each *state* $S \in \text{states}(X)$ inversely by different *degrees of incidence* $\text{incidence}(X, R, i)$ where $R = \text{filter}(W, S)$ and $i \in \{0 \ldots |W| - 1\}$. For example, a complement, $|X| - \text{incidence}(X, R, |W|)$.

Converses may be parameterised by a normalised sample *histogram* $\hat{A} \in \mathcal{A}$, having *variables* $V = \text{vars}(A)$, which is such that $\text{size}(\hat{A}) = 1$. Given a *transform* $T = (X, W) \in \mathcal{T}$, having *underlying variables* equal to the sample *variables*, $\mathrm{und}(T) = V$, the *sample converse* $\mathrm{converseSample} \in \mathcal{A} \times \mathcal{T} \to \mathcal{T}$ is defined as

$$\mathrm{converseSample}(A, T) \ := \ (\hat{A} * X, V)$$

In the case of *unit functional transform* $T \in \mathcal{T}_{\mathrm{U,f,U}}$, and $A = (X/(X\%W))\%V$, the *sample converse* equals the *natural converse*,

$$\mathrm{converseSample}(X/(X\%W), T) \ = \ \mathrm{converseNatural}(T)$$

The *actual converse* is very similar to the *natural converse* except that the *normalised* application of the *component* to an argument *histogram* is used, rather than just the *normalised component*. The *actual converse* is defined $\mathrm{converseActual} \in \mathcal{A} \times \mathcal{T}_{\mathrm{f}} \to \mathcal{T}$

$$\begin{aligned}&\mathrm{converseActual}(B, T) := \\ &\quad (\sum \frac{B * C}{(B * C)\%\emptyset} * \{R\}^{\mathrm{U}} : (R, C) \in \mathrm{inverse}(T), \ V)\end{aligned}$$

where $\mathrm{size}(B) > 0$ and $\mathrm{vars}(B) = V = \mathrm{underlying}(T)$. Define notation

$$T^{\odot B} \ = \ \mathrm{converseActual}(B, T)$$

The argument *transform* must be *functional* $T \in \mathcal{T}_{\mathrm{f}}$. The *actual converse* may be expressed more concisely,

$$T^{\odot B} \ := \ (\sum_{(R,C) \in T^{-1}} \{R\}^{\mathrm{U}} * (B * C)^{\wedge}, \ V) \tag{1}$$

The *actual converse*, $T^{\odot A}$, equals the *sample converse*, $(\hat{A} * X, V)$, if each of the *components* are *normalised*,

$$\forall (R, C) \in T^{-1} \ (\{R\}^{\mathrm{U}} * X * \hat{A} = \{R\}^{\mathrm{U}} * (A * C)^{\wedge})$$

A *converse transform* T is *conversely functional* if the *transform* formed from its *underlying variables* is *functional*, $(\mathrm{his}(T), \mathrm{und}(T)) \in \mathcal{T}_{\mathrm{f}}$. The *converse* of a *full functional transform* is also a *full functional transform*. If a *full functional transform* is also *unit*, then it is its own *natural converse*. In fact, the *converse* of a *unit full functional transform* is an identity, $A * T * T^{\dagger} = A$, where $\mathrm{und}(T) = \mathrm{vars}(A)$.

An *action* $C = (L, R) = ((X, W), (Y, V))$ is special case of a *model histogram expression* which is a pair of *transforms* having the same *variables* $\mathrm{vars}(X) = \mathrm{vars}(Y)$ and such that the *underlying variables* for the first *transform* are the *derived variables* of the second *transform* and vice-versa,

$$\begin{aligned}\mathrm{underlying}(L) &= \mathrm{derived}(R)\\ \mathrm{derived}(L) &= \mathrm{underlying}(R)\end{aligned}$$

The set of all *actions* $\mathrm{actions} \subset \mathcal{T} \times \mathcal{T}$ is defined

$$\mathrm{actions} = \{((X, W), (Y, V)) : (X, W), (Y, V) \in \mathcal{T},\ V \cap W = \emptyset, \mathrm{vars}(X) = \mathrm{vars}(Y)\}$$

Define function $\mathrm{action} \in \mathrm{actions} \times \mathcal{A} \to \mathcal{A}$

$$\begin{aligned}\mathrm{action}(((X, W), (Y, V)), A) &:= \mathrm{transform}((Y, V), \mathrm{transform}((X, W), A))\\ &= A * (X, W) * (Y, V)\\ &= A * X \ \%\ W * Y \ \%\ V\end{aligned}$$

The *simple action* is defined as a *transform* and its *simple converse*

$$((X, W), (\frac{1}{X}, V))$$

where (X, W) is a *transform* with *derived variables* W and *underlying variables* V. Similarly, *sample actions* are the pair of the *transform* and its *sample converse*

$$(T, (\hat{A} * X, V))$$

Again, *natural actions* are the pair of the *transform* and its *natural converse*

$$(T, T^{\dagger}) = ((X, W), (\frac{X^{\mathrm{F}}}{X\%W}, V))$$

The *natural action expression* applied to a given argument A is the *naturalisation*,

$$\mathrm{action}((T, T^{\dagger}), A) = A * T * T^{\dagger} = A * X \ \%\ W * \frac{X^{\mathrm{F}}}{X\%W} \ \%\ V$$

The *natural action* conserves the *size* of the given argument A if the *transform* T is *one functional* $T \in \mathcal{T}_{U,\mathrm{f},1}$

$$\mathrm{size}(A * T * T^{\dagger}) = \mathrm{size}(A)$$

Note that there are some *functional transforms* that conserve *size* but are not *one functional*. These, however, must be *none zero* and *uniform* in each of the *derived states*, $\forall R \in \mathrm{states}(X\%W)\ (|\{c : (S,c) \in X,\ R \subseteq S\}| = 1)$. For each of the *actions* constructed from one of these *functional transforms* and the *natural converse*, there is an equivalent *action* constructed from a *one functional transform*, because the *uniform count* cancels out.

The set of *states* of a *one functional naturalisation* is a superset of the set of *states* of the given argument A

$$\mathrm{states}(A * T * T^{\dagger}) \supseteq \mathrm{states}(A)$$

Or to put it another way, the *one functional naturalisation* may be more *effective* $(A*T*T^{\dagger})^{\mathrm{F}} \geq A^{\mathrm{F}}$. In the extreme case of a *one full functional naturalisation* the argument *histogram* is unchanged, $A * T * T^{\dagger} = A$, because *one full functional natural converses* are inverses. At the other extreme the *null one functional naturalisation* scales the *cartesian* of the *underlying* $A * T * T^{\dagger} = Z_A * V^{\mathrm{C}}/(V^{\mathrm{C}}\%\emptyset)$.

The extremes between the *null* $|A*T*T^{\dagger}| = |V^{\mathrm{C}}|$ and the *full* $|A*T*T^{\dagger}| = |A|$, show that the cardinality of the *states* of the applied *action* lies between $|A|$ and $|V^{\mathrm{C}}|$, that is $|A| \leq |A * T * T^{\dagger}| \leq |A^{\mathrm{C}}|$.

In addition, *one functional naturalisations* are limited in the changes that can be made to the *counts* of the argument *histogram* A where the *variables* V of A are the *underlying variables* of the first *transform* of the *action*. Consider a single *state* $(S,c) \in A$, and a *one functional natural action* $(T, T^{\dagger})$ and the *derived state* Q such that $\{(Q,1)\} = \{(S,1)\} * T$. Let $(X,W) = T$. In the case where there is only one *incident state* on Q, $(X\%W)_Q = |X * \{(Q,1)\}| = 1$, then $(A * T)_Q = c$ and so $(A * T * T^{\dagger})_S = c$. In the case where the *incident volume* $(X\%W)_Q = 2$ then there is another *state* R other than S contributing to $(A * T)_Q$. That is, $X * \{(Q,1)\} = \{(Q \cup S, 1), (Q \cup R, 1)\}$ such that $0 \leq A_R \leq z - c$ where $z = \mathrm{size}(A)$. Thus $z/2 \geq (A * T * T^{\dagger})_S \geq c/2$. Finally, extend the *incident volume* to $(X\%W)_Q = v - 1$ where $v = |V^{\mathrm{C}}|$. Here $z/(v-1) \geq (A*T*T^{\dagger})_S \geq c/(v-1)$. Overall, $z/2 \geq (A*T*T^{\dagger})_S \geq c/(v-1)$ except in the case where $c \geq z/2$ and the *incident volume* $(X\%W)_Q = 1$. The *count* of a *state* is limited in its possible increase under the *action*. It cannot decrease to below its original *count* approximately inversely scaled by the *volume*.

The *action* of a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ and its *actual converse*, $(T, T^{\odot B})$, is *size* conserving if all of the *components* of T are *non-zero* when applied to B. Thus

$$\text{size}(A * T * T^{\odot B}) = \text{size}(A)$$

where $\forall C \in \text{ran}(\text{inverse}(T))\ (\text{size}(B * C) > 0)$. In fact, the *action* is *size-conserving* under the weaker condition that the *transform* applied to B is at least as *effective* as the *transform* applied to A, $(B * T)^{\text{F}} \geq (A * T)^{\text{F}}$.

When $B = A$, then the application of the *action* $(T, T^{\odot A})$ is *equivalent* to A

$$A * T * T^{\odot A} \equiv A$$

The application of the *action* $(T, T^{\odot A})$ to the *scaled cartesian* is called the *unnaturalisation*,

$$V_z^{\text{C}} * T * T^{\odot A}$$

where $V = \text{vars}(A)$, $z = \text{size}(A)$, $v = |V^{\text{C}}|$ and $V_z^{\text{C}} = \text{scalar}(z/v) * V^{\text{C}}$.

3.5.3 Transforms and probability

Let *probability histogram* $P \in \mathcal{A} \cap \mathcal{P}$ have *variables* $\text{vars}(P) = X \cup Y$ where X and Y are disjoint, $X \cap Y = \emptyset$. A *probability histogram* has unit *size*, $\text{size}(P) = 1$.

The *conditional probability histogram* given X is $P/(P\%X)$. The *conditional probability histogram* given Y is $P/(P\%Y)$. Bayes' theorem may be expressed in terms of *conditional probability histograms*. Let $P[Y|X] = P/(P\%X)$, $P[X|Y] = P/(P\%Y)$, $P[Y] = P\%Y$ and $P[X] = P\%X$, then

$$\begin{aligned} P[Y|X] &= \frac{P[X|Y]\ P[Y]}{P[X]} \\ &= \frac{P/(P\%Y) * (P\%Y)}{P\%X} \\ &= \frac{P}{P\%X} \\ &= P[Y|X] \end{aligned}$$

Let query *probability histogram* $Q \in \mathcal{A} \cap \mathcal{P}$ have *variables* $\text{vars}(Q) = X$. The *product* of the query *probability histogram*, Q, and the *probability histogram*, P, is a *weak probability histogram*,

$$Q * P \in \mathcal{A} \cap \mathcal{P}'$$

A *weak probability histogram* has *size* less than or equal to one, $\text{size}(Q * P) \leq 1$.

If the set of query *variables* is empty, $X = \emptyset$, then the query *histogram* is *scalar one*, $Q = \{(\emptyset, 1)\}$, and the *product* is the given *histogram*, $Q * P = P$.

If the *effective states* of the *histograms* do not intersect, then the *product* is the *empty histogram*, which has a *size* of 0,

$$Q^{\mathrm{F}} \cap (P\%X)^{\mathrm{F}} = \emptyset \implies Q * P = \emptyset$$

If the *probability histogram* is as *effective* as the query *probability histogram* and either *histogram* is an *effective singleton* then the *product* is a *probability histogram*,

$$(Q^{\mathrm{F}} \leq (P\%X)^{\mathrm{F}}) \wedge (|Q^{\mathrm{F}}| = 1) \implies Q * P \in \mathcal{A} \cap \mathcal{P}$$

The *transform* implied by P and Q is $T_P = (P, Y) \in \mathcal{T}$. The *transformed product* is

$$Q * T_P = Q * (P, Y) = Q * P \ \% \ Y \in \mathcal{A} \cap \mathcal{P}'$$

The *conditional transform* implied by P and Q is $T'_P = (P/(P\%X), Y) \in \mathcal{T}$. The *transformed conditional product* is

$$Q * T'_P = Q * \left(\frac{P}{P\%X}, Y\right) = \frac{Q * P}{P\%X} \ \% \ Y \in \mathcal{A} \cap \mathcal{P}'$$

In the case where the *reduction* of P is as *effective* as Q, the *transformed conditional product* is a *probability histogram*,

$$Q^{\mathrm{F}} \leq (P\%X)^{\mathrm{F}} \implies Q * T'_P = Q * \left(\frac{P}{P\%X}, Y\right) \in \mathcal{A} \cap \mathcal{P}$$

because $(P/(P\%X))\%X = (P\%X)^{\mathrm{F}}$. In this case the *conditional product* $R = Q * P/(P\%X)$ is such that $R\%X = Q$ and $R\%Y = Q * (P/(P\%X), Y)$.

If the set of *variables* Y is empty, $Y = \emptyset$, then $\mathrm{vars}(P) = X$, $P\%X = P$ and $P/(P\%X) = P^{\mathrm{F}}$. If in addition P is as *effective* as Q, $Q^{\mathrm{F}} \leq P^{\mathrm{F}}$, then the *transformed conditional product* is *scalar one*, $Q * (P/(P\%X), Y) = Q\%\emptyset = \{(\emptyset, 1)\}$.

If (i) the *reduced probability histogram* $P\%X$ is *uniform*, $P\%X = (P\%X)^{\mathrm{FS}} \times \{1/|(P\%X)^{\mathrm{FS}}|\}$, and (ii) the *probability histogram* is as *effective* as the query *probability histogram*, $Q^{\mathrm{F}} \leq (P\%X)^{\mathrm{F}}$, then the *transformed conditional product* is the normalised *transformed product probability histogram*,

$$\begin{aligned}&(|\mathrm{ran}(P\%X)| = 1) \wedge (Q^{\mathrm{F}} \leq (P\%X)^{\mathrm{F}}) \implies \\ &\quad Q * T'_P = Q * \left(\frac{P}{P\%X}, Y\right) = (Q * (P, Y))^{\wedge} = (Q * T_P)^{\wedge} \in \mathcal{A} \cap \mathcal{P}\end{aligned}$$

where $(Q * (P,Y))^{\wedge} = \text{normalise}(Q * (P,Y))$ and $(Q * T_P)^{\wedge} = \text{normalise}(Q * T_P)$.

If the *probability histogram* is the normalised *cartesian histogram* $P = (X \cup Y)^{\mathrm{C}\wedge} = \text{normalise}((X \cup Y)^{\mathrm{C}})$, then the *transformed product* is the normalised *cartesian* in *variables* Y,

$$P = (X \cup Y)^{\mathrm{C}\wedge} \implies Q * T_P = Q * ((X \cup Y)^{\mathrm{C}\wedge}, Y) = Y^{\mathrm{C}\wedge} \in \mathcal{A} \cap \mathcal{P}$$

where $Y^{\mathrm{C}\wedge} = \text{normalise}(Y^{\mathrm{C}}) = \text{scalar}(1/|Y^{\mathrm{C}}|) * Y^{\mathrm{C}}$. That is, the *transformed product*, $Q{*}T_P$, is a constant, $Y^{\mathrm{C}\wedge}$, and so independent of the query *probability histogram*, Q.

The opposite extreme to normalised *cartesian histogram*, $(X \cup Y)^{\mathrm{C}\wedge}$, is where the *probability histogram*, P, is *causal*, $\text{causal}(P)$. In particular, if the *variables* Y are a function of the *variables* X, $\text{split}(X, P^{\mathrm{FS}}) \in X^{\mathrm{CS}} \to Y^{\mathrm{CS}}$, then the *transform* implied by Q and P, $T_P = (P,Y)$, is *functional*, $T_P \in \mathcal{T}_{\mathrm{f}}$. In this case, the *conditional transform* is also *functional*, $T'_P \in \mathcal{T}_{\mathrm{f}}$.

If the *probability histogram*, P, is (i) *causal* between *variables* X and *variables* Y, so that $T_P \in \mathcal{T}_{\mathrm{f}}$, and (ii) *completely effective* in *variables* X, so that $(P\%X)^{\mathrm{F}} = X^{\mathrm{C}}$, then the *conditional transform* is *one functional*,

$$(T_P \in \mathcal{T}_{\mathrm{f}}) \wedge ((P\%X)^{\mathrm{F}} = X^{\mathrm{C}}) \implies T'_P = \left(\frac{P}{P\%X}, Y\right) \in \mathcal{T}_{U,\mathrm{f},1}$$

In this case the *reduction* of the *probability histogram* is necessarily as *effective* as the query *probability histogram*, $Q^{\mathrm{F}} \leq (P\%X)^{\mathrm{F}} = X^{\mathrm{C}}$, and hence the *transformed conditional product* is necessarily a *probability histogram*,

$$(P\%X)^{\mathrm{F}} = X^{\mathrm{C}} \implies Q * T'_P = Q * \left(\frac{P}{P\%X}, Y\right) \in \mathcal{A} \cap \mathcal{P}$$

So the *one functional probability histogram*, P, is a *one functional model*, $P \in \mathcal{M}_{U,\mathrm{f},1}$, such that $P^{\mathrm{T}} = T'_P$, with *underlying variables* $\text{und}(P^{\mathrm{T}}) = X$ and *derived variables* $\text{der}(P^{\mathrm{T}}) = Y$.

Conversely, all *one functional transforms* are *conditional* in the *underlying variables*,

$$\forall T \in \mathcal{T}_{U,\mathrm{f},1}\ (\text{his}(T) = \text{his}(T)\ /\ (\text{his}(T)\ \%\ \text{und}(T)))$$

because the *transform histogram reduction* to *underlying variables* is *cartesian*,

$$\forall T \in \mathcal{T}_{U,\mathrm{f},1}\ (\text{his}(T)\ \%\ \text{und}(T) = (\text{und}(T))^{\mathrm{C}})$$

If the *probability histogram*, P, is *causal* between the *variables* X and Y, $\mathrm{split}(X, P^{\mathrm{FS}}) \in X^{\mathrm{CS}} \to Y^{\mathrm{CS}}$, then the *transform* is *functional*, $T_P \in \mathcal{T}_{\mathrm{f}}$, and so may have *size-conserving converses*. The *natural converse* is $T_P^{\dagger} = (P/(P\%Y), X) \in \mathcal{T}$. The *natural converse* is the *conditional transform* implied by P and N, where N is a query *probability histogram* in *variables* Y, $N \in \mathcal{A} \cap \mathcal{P}$ and $\mathrm{vars}(N) = Y$. That is, the *transformed conditional product* of N and P is

$$N * T_P^{\dagger} = N * \left(\frac{P}{P\%Y}, X\right) \in \mathcal{A} \cap \mathcal{P}'$$

If the *probability histogram*, P, is as *effective* as N, then the *transformed conditional product* is a *probability histogram*,

$$N^{\mathrm{F}} \leq (P\%Y)^{\mathrm{F}} \implies N * T_P^{\dagger} = N * \left(\frac{P}{P\%Y}, X\right) \in \mathcal{A} \cap \mathcal{P}$$

The *natural action* applied to query *probability histogram* Q is

$$Q * T_P * T_P^{\dagger} = Q * (P, Y) * \left(\frac{P}{P\%Y}, X\right) \in \mathcal{A} \cap \mathcal{P}'$$

If the *probability histogram* is bijective, $\mathrm{split}(X, P^{\mathrm{FS}}) \in X^{\mathrm{CS}} \leftrightarrow Y^{\mathrm{CS}}$, then the *transform* is a *full functional transform*. The *simple converse* is also *functional*, $(1/P, X) \in \mathcal{T}_{\mathrm{f}}$, and the *simple action* leaves the query unchanged, $Q * (P, Y) * (1/P, X) = Q$, if $Q^{\mathrm{F}} \leq (P\%X)^{\mathrm{F}}$. The *natural converse* is also *functional*, $T_P^{\dagger} \in \mathcal{T}_{\mathrm{f}}$, and so it has a *natural converse*, which equals the *simple converse* implied by P and N,

$$T_P^{\dagger\dagger} = \left(\frac{P/(P\%Y)}{P\%X}, Y\right) = (1/P, Y)$$

So the *natural action* applied to N is

$$N * T_P^{\dagger} * T_P^{\dagger\dagger} = N * \left(\frac{P}{P\%Y}, X\right) * (1/P, Y) \in \mathcal{A} \cap \mathcal{P}'$$

The normalisation of a *non-zero* sample *histogram* $A \in \mathcal{A}_U$, having non-empty *variables* $V = \mathrm{vars}(A) \neq \emptyset$, is a *probability histogram*, $\hat{A} \in \mathcal{A} \cap \mathcal{P}$, because $\mathrm{size}(\hat{A}) = 1$, where $\hat{A} = A/(A\%\emptyset)$. The normalisation of a *non-zero* query *histogram* $Q \in \mathcal{A}_U$, having *variables* $K = \mathrm{vars}(Q)$ that are a subset of the sample *variables*, $K \subseteq V$, is a *probability histogram*, $\hat{Q} \in \mathcal{A} \cap \mathcal{P}$. The *transform* implied by A and Q is $T_A = (\hat{A}, (V \setminus K)) \in \mathcal{T}$. The *transformed product* is $\hat{Q} * T_A = \hat{Q} * (\hat{A}, (V \setminus K)) \in \mathcal{A} \cap \mathcal{P}'$. The *conditional transform* implied by A and Q is $T'_A = (A/(A\%K), (V \setminus K)) \in \mathcal{T}$. The *transformed conditional product* is $\hat{Q} * T'_A = \hat{Q} * (A/(A\%K), (V \setminus K)) \in \mathcal{A} \cap \mathcal{P}'$. In the case where the *reduction* of A is as *effective* as Q, the *transformed conditional product* is a *probability histogram*, $Q^{\mathrm{F}} \leq (A\%K)^{\mathrm{F}} \implies \hat{Q}*T'_A \in \mathcal{A}\cap\mathcal{P}$.

If the *effective states* of the query *histogram*, Q, and the sample *histogram*, A, do not intersect, then both the *transformed product* and *transformed conditional product* are empty, $Q^{\mathrm{F}} \cap (A\%K)^{\mathrm{F}} = \emptyset \implies \hat{Q} * T_A = \emptyset$ and $Q^{\mathrm{F}} \cap (A\%K)^{\mathrm{F}} = \emptyset \implies \hat{Q} * T'_A = \emptyset$. Less drastically, if the sample *histogram* is not as *effective* as the query *histogram*, $|Q^{\mathrm{F}} \cap (A\%K)^{\mathrm{F}}| < |Q^{\mathrm{F}}|$, then the *transformed conditional product* cannot be a *probability histogram*, $\hat{Q} * T'_A \notin \mathcal{P}$. However, if there exists a *one functional transform* $T = (M, W) \in \mathcal{T}_{U,\mathrm{f},1}$, having *underlying variables* $J = \mathrm{und}(T)$ which are a subset of the sample *variables*, $J \subseteq V$, then a *model* analog to the *product* may be computed via the intermediate *derived variables*, W. In the case where the *underlying variables* are a subset of the query *variables*, $J \subseteq K$, the *model* substitute for the *transformed product*, $\hat{Q} * T_A \in \mathcal{A} \cap \mathcal{P}'$, is

$$\hat{Q} * M \ \% \ W * M * \hat{A} \ \% \ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}'$$

This is equivalent to the application of the normalised *sample action*, $(T, (\hat{A} * M, V))$, to the query *probability histogram*, $\hat{Q}$, followed by *reduction* to $V \setminus K$,

$$\hat{Q} * T * (\hat{A} * M, V) \ \% \ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}'$$

Now the intersection of *effective states* is $(Q * T)^{\mathrm{F}} \cap (A * T)^{\mathrm{F}}$. The *states* for which there is no *effective derived* sample *state*, $(Q*T)^{\mathrm{F}} \setminus (A*T)^{\mathrm{F}}$, are said to be *over-fitted*. That is, *over-fitted states*, $(((Q*T)^{\mathrm{F}} \setminus (A*T)^{\mathrm{F}}) * T^{\dagger})^{\mathrm{S}} \subseteq K^{\mathrm{CS}}$, have zero probability.

The *modelled transformed product* procedure is: (i) the *transform*, $T = (M, W)$, is applied, raising the query into the *derived variables*, $\mathrm{vars}(\hat{Q} * (M, W)) = W$, then (ii) the *sample converse transform*, $(\hat{A} * M, V)$, is applied, lowering back to the sample *variables*, $\mathrm{vars}(\hat{Q} * T * (\hat{A} * M, V)) = V$, which is followed by (iii) the removal of the query *variables*, $\mathrm{vars}(\hat{Q} * T * (\hat{A} * M, V) \ \% \ (V \setminus K)) = V \setminus K$.

In the case where the *underlying variables* are a proper superset of the query *variables*, $J \supset K$, the query *histogram* can be expanded by assuming a uniform *probability function* in the additional *variables*,

$$(J \setminus K)^{\mathrm{C}\wedge} = (J \setminus K)^{\mathrm{CS}} \times \{1/|(J \setminus K)^{\mathrm{C}}|\} \in \mathcal{A} \cap \mathcal{P}$$

The expanded query *probability histogram* is $\hat{Q}_J = \hat{Q} * (J \setminus K)^{\mathrm{C}\wedge} \in \mathcal{A} \cap \mathcal{P}$. It has *variables* $\mathrm{vars}(\hat{Q}_J) = K \cup J$. Now the application of the *model transform* is functional, $\mathrm{split}(K \cup J, (\hat{Q}_J * M)^{\mathrm{S}}) \in (K \cup J)^{\mathrm{CS}} \to W^{\mathrm{S}}$. The *modelled transformed product* becomes

$$\hat{Q}_J * T * (\hat{A} * M, V) \ \% \ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}'$$

In the absence of a wholly *effective* intersection between the query and the sample, $|Q^{\mathrm{F}} \cap (A\%K)^{\mathrm{F}}| < |Q^{\mathrm{F}}|$, a *model* analog for the *transformed conditional product*, $\hat{Q} * T'_A = \hat{Q} * (A/(A\%K), (V \setminus K))$, can be defined by normalising the application of the normalised *sample action* to the query *probability histogram*,

$$(\hat{Q}_J * T * (\hat{A} * M, V))^{\wedge} \ \% \ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}$$

if the intersection of *derived effective states* is not empty, $(Q*T)^{\mathrm{F}} \cap (A*T)^{\mathrm{F}} \neq \emptyset$. The normalisation is equivalent to assuming that the *reduced* sample *probability histogram*, $\hat{A}\%K$, is *uniform*,

$$\hat{A}\%K = (\hat{A}\%K)^{\mathrm{FS}} \times \{1/|(\hat{A}\%K)^{\mathrm{F}}|\} \in \mathcal{A} \cap \mathcal{P}$$

The renormalisation means that neither (i) the expansion, $\hat{Q}_J = \hat{Q}*(J \setminus K)^{\mathrm{C}\wedge}$, nor (ii) the normalisations of the query or sample *histograms*, $\hat{Q}$ and $\hat{A}$, need be calculated, so the *modelled transformed conditional product* can be simplified to

$$(Q * T * M * A)^{\wedge} \ \% \ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}$$

if $(Q * T)^{\mathrm{F}} \cap (A * T)^{\mathrm{F}} \neq \emptyset$.

The *modelled transformed conditional product* may be expressed in terms of the *actual converse transform*,

$$Q * T * T^{\odot A} \ \% \ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}$$

where the *actual converse transform* is defined

$$T^{\odot A} \ := \ \Big(\sum_{(R,C) \in T^{-1}} \{R\}^{\mathrm{U}} * (A * C)^{\wedge}, \ V\Big) \qquad (2)$$

where the normalisation is defined $\hat{A} = A/(A\%\emptyset)$ so that normalised *zero histograms* are *empty*, $(V^{\mathrm{CZ}})^{\wedge} = \emptyset$.

If the *effective sample reduction* to *underlying variables* is *cartesian*, $(A\%K)^{\mathrm{F}} =$

K^{C}, then the *actual converse transform* is *conditional* in the *derived variables*,

$$\mathrm{his}(T^{\odot A}) \;\; = \;\; \mathrm{his}(T^{\odot A}) \;/\; (\mathrm{his}(T^{\odot A}) \;\%\; W) \qquad (3)$$

because *actual converse transform histogram reduction* to *derived variables* is *cartesian*, $\mathrm{his}(T^{\odot A}) \;\%\; W = W^{\mathrm{C}}$. Note that, strictly speaking, this is only the case where the *transform* is *non-overlapping* (see section 'Overlapping transforms' below).

In the case where the *sample histogram* is *natural*, $A = A * T * T^{\dagger}$, the *actual converse* equals the *natural converse*, $T^{\odot A} = T^{\dagger}$, and the query application simplifies to $Q * T * T^{\dagger} \;\%\; (V \setminus K)$, which does not depend on the *sample*, A, only on the *model*, T.

The *relative entropy* of the *modelled transformed conditional product*, $(Q * T * M * A)^{\wedge} \;\%\; (V \setminus K) \in \mathcal{A} \cap \mathcal{P}$, with respect to the *transformed conditional product*, $\hat{Q} * T'_A = \hat{Q} * (A/(A\%K), (V \setminus K)) \in \mathcal{A} \cap \mathcal{P}$, in case when the *reduction* of A is as *effective* as Q, $Q^{\mathrm{F}} \leq (A\%K)^{\mathrm{F}}$, is

$$\mathrm{entropyRelative}(Q * A \;/\; (A\%K) \;\%\; (V \setminus K),\;\; Q * T * M * A \;\%\; (V \setminus K))$$

or

$$\mathrm{entropyRelative}(Q * A \;/\; (A\%K) \;\%\; (V \setminus K),\;\; Q * T * T^{\odot A} \;\%\; (V \setminus K))$$

There is no need to stuff *ineffective states* because the *modelled transformed conditional product* is as *effective* as the *transformed conditional product*, $(Q * T * T^{\odot A})^{\mathrm{F}} \geq (Q * A)^{\mathrm{F}}$.

The *relative entropy* is zero when the *modelled transformed conditional product* equals the *transformed conditional product*. This is the case, for example, for the *self partition transform* (see below), $T = V^{\mathrm{CS\{\}T}}$, or the *full functional transform* (see below), $T = \{\{v\}^{\mathrm{CS\{\}T}} : v \in V\}^{\mathrm{T}}$.

In the case where the *transform* is *unary* (see below), $T = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, the *modelled transformed conditional product* is $\hat{A} \;\%\; (V \setminus K)$. This equals the *transformed conditional product* only if the *histogram*, A, is *partially independent* (see below), $\hat{A} = (\hat{A} \;\%\; K) * (\hat{A} \;\%\; (V \setminus K))$.

Let the difference between the sample *variables* and the query *variables*, $V \setminus K$, be called the label *variables*. If the *histogram*, A, is such that it is *causal* between the query *variables* and the label *variables*, $\mathrm{split}(K, A^{\mathrm{FS}}) \in K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$, then all queries consisting of one of the *effective states* have unique label *state*, $\forall Q \in (A\%K)^{\mathrm{FS}}\ (|\{Q\}^{\mathrm{U}} * A\ /\ (A\%K)\ \%\ (V \setminus K)| = 1)$, and so have zero *transformed conditional product entropy*, $\forall Q \in (A\%K)^{\mathrm{FS}}\ (\mathrm{entropy}(\{Q\}^{\mathrm{U}} * A\ /\ (A\%K)\ \%\ (V \setminus K)) = 0)$. In this case the *relative entropy* is the *cross entropy*, $-\ln((\{Q\}^{\mathrm{U}} * T * T^{\odot A}\ \%\ (V \setminus K))(R))$, where $\{R\} = (\{Q\}^{\mathrm{U}} * A\ /\ (A\%K)\ \%\ (V \setminus K))^{\mathrm{S}}$.

If the normalised *histogram*, $\hat{A} \in \mathcal{A} \cap \mathcal{P}$, is treated as a *probability function* of a single-*state* query, the *expected entropy* of the *transformed conditional product* is zero in the case of *causal histogram*,

$$\sum_{Q \in (A\%K)^{\mathrm{FS}}} \mathrm{size}(\{Q\}^{\mathrm{U}} * \hat{A}) \times \mathrm{entropy}(\{Q\}^{\mathrm{U}} * A\ /\ (A\%K)\ \%\ (V \setminus K)) = 0$$

This may be compared to the *expected entropy* of the *modelled transformed conditional product*, or label *entropy*,

$$\begin{aligned} & \sum_{(R,\cdot) \in T^{-1}} (\hat{A} * T)_R \times \mathrm{entropy}(\{R\}^{\mathrm{U}} * T^{\odot A}\ \%\ (V \setminus K)) \\ = & \sum_{(R,C) \in T^{-1}} (\hat{A} * T)_R \times \mathrm{entropy}(A * C\ \%\ (V \setminus K)) \end{aligned}$$

Let *non-zero* test *histogram* $B \in \mathcal{A}_U$ have *variables* equal to the sample *variables*, $\mathrm{vars}(B) = \mathrm{vars}(A) = V$, and be such that it is *causal* between the query *variables* and the label *variables*, $\mathrm{split}(K, B^{\mathrm{FS}}) \in K^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$. The test *histogram* implies a query *probability histogram*, $\hat{B}\%K \in \mathcal{A} \cap \mathcal{P}$.

Let $R \in V :\leftrightarrow: V_R$ be a mapping from the sample *variables*, V, to a disjoint *reframed* set, V_R, such that the *reframe* is *literal*, $\forall (v, w) \in R\ (U_w = U_v)$. The test *histogram* may be extended by dotting with the *reframe*,

$$B_R = \{(S \cup \mathrm{reframe}(R, S), c) : (S, c) \in B\} \in \mathcal{A}_U$$

The *reframe variables* are disjoint, $V_R \cap V = \emptyset$, and so the extended test *histogram* has double the *variables*, $|\mathrm{vars}(B_R)| = 2|V|$. The extended test *histogram* is still *causal*, $\mathrm{split}(K, B_R^{\mathrm{FS}}) \in K^{\mathrm{CS}} \to (V \setminus K \cup V_R)^{\mathrm{CS}}$.

Given the *one functional transform* $T = (M, W) \in \mathcal{T}_{U,\mathrm{f},1}$, such that the *reframe variables* are disjoint with the *derived variables*, $V_R \cap W = \emptyset$, the *modelled transformed conditional product* for the test *histogram*, B, is

$$(\hat{B}_R\%(K \cup V_R) * (J \setminus K)^{\mathrm{C}\wedge} * M\ \%\ (W \cup V_R) * M * \hat{A})^{\wedge}\ \%\ (V \setminus K \cup V_R) \in \mathcal{A} \cap \mathcal{P}$$

if $(B\%K * T)^{\mathrm{F}} \cap (A * T)^{\mathrm{F}} \neq \emptyset$.

This may be simplified to

$$(B_R\%(K \cup V_R) * M \ \% \ (W \cup V_R) * M * A)^{\wedge} \ \% \ (V \setminus K \cup V_R) \in \mathcal{A} \cap \mathcal{P}$$

The *modelled* label *variables*, $V \setminus K$, may be compared to the *reframed* test label *variables*, $V_R \setminus K_R$, for each $K \cong K_R$, to judge the accuracy of the *model* in terms of the test.

Note that in the case where an expansion is necessary, $J \setminus K \neq \emptyset$, the additional label *variables*, $J \setminus K$, necessarily contradict the *reframed* label *variables*, $J_R \setminus K_R$, unless they are all *mono-valent*. That is,

$$\mathrm{split}(J \setminus K, (\hat{B}_R\%(K \cup V_R) * (J \setminus K)^{\mathrm{C}\wedge} \ \% \ (J_R \setminus K_R \cup J \setminus K))^{\mathrm{FS}}) \notin (J \setminus K)^{\mathrm{CS}} \leftrightarrow (J_R \setminus K_R)^{\mathrm{CS}}$$

where $\exists v \in J \setminus K \ (|U_v| > 1)$.

In the case where no expansion is necessary, $J \setminus K = \emptyset$, and there is a single *effective* test *state*, $|B^{\mathrm{F}}| = 1$, the *modelled transformed conditional product* is

$$((B\%K) * T * M * A)^{\wedge} \ \% \ (V \setminus K) * (B_R^{\mathrm{F}}\%V_R) \in \mathcal{A} \cap \mathcal{P}$$

3.5.4 Transform entropy

Now consider *derived entropy*. Let T be a *one functional transform*, $T \in \mathcal{T}_{U,\mathrm{f},1}$, having *underlying variables* $V = \mathrm{und}(T)$. Let A be a *non-zero histogram*, $A \in \mathcal{A}_U$, in *variables* $V = \mathrm{vars}(A)$ having *size* $z = \mathrm{size}(A) > 0$. The *underlying volume* is $v = |V^{\mathrm{C}}|$.

The *normalised derived histogram* $\hat{A} * T \in \mathcal{P}$ is a *probability function*,

$$\begin{aligned} \hat{A} * T &= \{(R, q/z) : (R, q) \in A * T\} \\ &= \{(R, \mathrm{size}(A * C)/z) : (R, C) \in T^{-1}\} \end{aligned}$$

In the case where the *histogram* is *integral*, $A \in \mathcal{A}_{\mathrm{i}}$, then a *history* $H = \mathrm{history}(\mathrm{trim}(A))$ is implied such that $z = |H| = \mathrm{size}(A) > 0$. In this case the *normalised derived histogram* is $\hat{A} * T = \{(R, |D|/z) : (R, D) \in (H * T)^{-1}\}$.

The *normalised cartesian derived* $\hat{V}^{\mathrm{C}} * T \in \mathcal{P}$ is a *probability function*,

$$\begin{aligned} \hat{V}^{\mathrm{C}} * T &= \{(R, q/v) : (R, q) \in V^{\mathrm{C}} * T\} \\ &= \{(R, |C|/v) : (R, C) \in T^{-1}\} \end{aligned}$$

The *derived entropy* or *component size entropy* is the negative *derived histogram expected normalised derived histogram count logarithm*,

$$\text{entropy}(A * T) \;\; := \;\; -\,\text{expected}(\hat{A} * T)(\ln(\hat{A} * T))$$

where $\ln A = \{(S, \ln c) : (S,c) \in A,\ c > 0\}$. The *derived entropy* is positive and less than or equal to the logarithm of the *size*, $0 \leq \text{entropy}(A*T) \leq \ln z$.

Complementary to the *derived entropy* is the *size expected component entropy*,

$$\begin{aligned}\text{entropyComponent}(A,T) \;\; := \\ \text{expected}(\hat{A} * T)(\{(R, \text{entropy}(A * C)) : (R,C) \in T^{-1}\})\end{aligned}$$

or

$$\text{entropyComponent}(A,T) \;\; := \sum_{(R,C) \in T^{-1}} (\hat{A} * T)_R \times \text{entropy}(A * C)$$

The *size expected component entropy* can be expressed in terms of the *actual converse*,

$$\text{entropyComponent}(A,T) \;\; = \sum_{(R,\cdot) \in T^{-1}} (\hat{A} * T)_R \times \text{entropy}(\{R\}^{\mathrm{U}} * T^{\odot A})$$

The *component cardinality entropy* is the negative *cartesian derived expected normalised cartesian derived count logarithm*,

$$\text{entropy}(V^{\mathrm{C}} * T) \;\; := \;\; -\,\text{expected}(\hat{V}^{\mathrm{C}} * T)(\ln(\hat{V}^{\mathrm{C}} * T))$$

The *cartesian derived entropy* is positive and less than or equal to the logarithm of the *volume*, $0 \leq \text{entropy}(V^{\mathrm{C}} * T) \leq \ln v$.

The *cartesian derived derived sum entropy* or *component size cardinality sum entropy* is

$$\begin{aligned}\text{entropy}(A * T) + \text{entropy}(V^{\mathrm{C}} * T) \;\; := \\ (-\,\text{expected}(\hat{A} * T)(\ln(\hat{A} * T))) \; + \; (-\,\text{expected}(\hat{V}^{\mathrm{C}} * T)(\ln(\hat{V}^{\mathrm{C}} * T)))\end{aligned}$$

The *component size cardinality cross entropy* is the negative *derived histogram expected normalised cartesian derived count logarithm*,

$$\text{entropyCross}(A * T, V^{\mathrm{C}} * T) \;\; := \;\; -\,\text{expected}(\hat{A} * T)(\ln(\hat{V}^{\mathrm{C}} * T))$$

By Gibbs' inequality the *component size cardinality cross entropy* is greater than or equal to the *derived entropy*, $\text{entropyCross}(A*T, V^{\mathrm{C}}*T) \geq \text{entropy}(A*T)$.

The *component cardinality size cross entropy* is the negative *cartesian derived expected normalised derived histogram count logarithm*,

$$\text{entropyCross}(V^{\mathrm{C}} * T, A * T) \ := \ -\ \text{expected}(\hat{V}^{\mathrm{C}} * T)(\ln(\hat{A} * T))$$

The *component cardinality size cross entropy* is greater than or equal to the *cartesian derived entropy*, $\text{entropyCross}(V^{\mathrm{C}} * T, A * T) \geq \text{entropy}(V^{\mathrm{C}} * T)$.

The *component size cardinality sum cross entropy* is,

$$\begin{aligned}\text{entropy}(A * T + V^{\mathrm{C}} * T) \ :=& \\ &-\ \text{expected}((A * T + V^{\mathrm{C}} * T)^{\wedge})(\ln((A * T + V^{\mathrm{C}} * T)^{\wedge}))\end{aligned}$$

The *component size cardinality sum cross entropy* is positive and less than or equal to the logarithm of the sum of the *size* and *volume*, $0 \leq \text{entropy}(A * T + V^{\mathrm{C}} * T) \leq \ln(z + v)$. The *component size cardinality sum cross entropy* is greater than or equal to the *derived entropy*, $\text{entropy}(A * T + V^{\mathrm{C}} * T) \geq \text{entropy}(A * T)$, and greater than or equal to the *cartesian derived entropy*, $\text{entropy}(A * T + V^{\mathrm{C}} * T) \geq \text{entropy}(V^{\mathrm{C}} * T)$.

In all cases the *cross entropy* is maximised when high *size components* are low *cardinality components*, $(\hat{A} * T)_R \gg (\hat{V}^{\mathrm{C}} * T)_R$ or $\text{size}(A * C)/z \gg |C|/v$, and vice-versa, $(\hat{A} * T)_R \ll (\hat{V}^{\mathrm{C}} * T)_R$ or $\text{size}(A * C)/z \ll |C|/v$.

The *cross entropy* is minimised when the *normalised derived histogram* equals the *normalised cartesian derived*, $\hat{A} * T = \hat{V}^{\mathrm{C}} * T$ or $\forall (R, C) \in T^{-1}\ (\text{size}(A * C)/z = |C|/v)$. In this case the *cross entropy* equals the corresponding *component entropy*.

The *component size cardinality relative entropy* is the *component size cardinality cross entropy* minus the *component size entropy*,

$$\begin{aligned}&\text{entropyRelative}(A * T, V^{\mathrm{C}} * T) \\ &\qquad := \text{expected}(\hat{A} * T)\left(\ln \frac{\hat{A} * T}{\hat{V}^{\mathrm{C}} * T}\right) \\ &\qquad = \ \text{entropyCross}(A * T, V^{\mathrm{C}} * T)\ -\ \text{entropy}(A * T)\end{aligned}$$

The *component size cardinality relative entropy* is positive, $\text{entropyRelative}(A * T, V^{\mathrm{C}} * T) \geq 0$.

The *multinomial distribution*, $Q_{\mathrm{m},U}$, is described in section 'Multinomial distributions', below. The *size* scaled *component size cardinality relative entropy* approximates to the negative logarithm of the *derived multinomial probability* with respect to the *cartesian derived*,

$$z \times \text{entropyRelative}(A * T, V^{\mathrm{C}} * T) \quad \approx \quad -\ln \hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}} * T, z)(A * T)$$

The *component cardinality size relative entropy* is the *component cardinality size cross entropy* minus the *component cardinality entropy*,

$$\begin{aligned}
&\text{entropyRelative}(V^{\mathrm{C}} * T, A * T) \\
&\qquad := \text{expected}(\hat{V}^{\mathrm{C}} * T)\left(\ln \frac{\hat{V}^{\mathrm{C}} * T}{\hat{A} * T}\right) \\
&\qquad = \ \text{entropyCross}(V^{\mathrm{C}} * T, A * T) \ - \ \text{entropy}(V^{\mathrm{C}} * T)
\end{aligned}$$

The *component cardinality size relative entropy* is positive, $\text{entropyRelative}(V^{\mathrm{C}} * T, A * T) \geq 0$.

The *volume* scaled *component cardinality size relative entropy* approximates to the negative logarithm of the *cartesian derived multinomial probability* with respect to the *derived*,

$$v \times \text{entropyRelative}(V^{\mathrm{C}} * T, A * T) \quad \approx \quad -\ln \hat{Q}_{\mathrm{m},U}(A * T, v)(V^{\mathrm{C}} * T)$$

where the *derived* is as *effective* as the *cartesian derived*, $(A * T)^{\mathrm{F}} \geq (V^{\mathrm{C}} * T)^{\mathrm{F}} \implies |(A * T)^{\mathrm{F}}| = |T^{-1}|$.

The *size-volume* scaled *component size cardinality sum relative entropy* is the *size-volume* scaled *component size cardinality sum cross entropy* minus the *size-volume* scaled *component size cardinality sum entropy*,

$$\begin{aligned}
&(z + v) \times \text{entropy}(A * T + V^{\mathrm{C}} * T) \\
&\qquad\qquad -z \times \text{entropy}(A * T) \ - v \times \text{entropy}(V^{\mathrm{C}} * T)
\end{aligned}$$

The *size-volume* scaled *component size cardinality sum relative entropy* is positive, $(z + v) \times \text{entropy}(A * T + V^{\mathrm{C}} * T) \ - z \times \text{entropy}(A * T) \ - v \times \text{entropy}(V^{\mathrm{C}} * T) \geq 0$.

In all cases the *relative entropy* is maximised when (a) the *cross entropy* is maximised and (b) the *component entropy* is minimised. The *relative entropy* is always positive by Gibbs' inequality, see Appendix 'Entropy and Gibbs' inequality', below. So the *cross entropy* is greater than or equal to the *component entropy*.

3.6 Functional definition sets

A *functional definition set* $F \in \mathcal{F}$ is a set of *unit functional transforms* subject to the constraint that *derived variables* may appear in only one *transform*. That is, the sets of *derived variables* are disjoint. Then $\mathcal{F} \subset \mathrm{P}(\mathcal{T}_{\mathrm{f,U}})$ where $\mathcal{T}_{\mathrm{f,U}} = \mathcal{T}_{\mathrm{f}} \cap \mathcal{T}_{\mathrm{U}}$, and

$$\forall F \in \mathcal{F}\ \forall (A,W),(B,X) \in F\ ((A,W) \neq (B,X) \implies W \cap X = \emptyset)$$

Defining accessor functions, the domain of a *functional definition set* has a synonym histograms $\in \mathcal{F} \to \mathrm{P}(\mathcal{A})$

$$\mathrm{histograms}(F) := \mathrm{dom}(F) = \{A : (A,W) \in F\}$$

Define vars $\in \mathcal{F} \to \mathrm{P}(\mathcal{V})$

$$\mathrm{vars}(F) := \bigcup \{\mathrm{vars}(A) : A \in \mathrm{histograms}(F)\}$$

Define accessors of a *functional definition set* such that its *derived* and *underlying variables* are disjoint. That is, derived $\in \mathcal{F} \to \mathrm{P}(\mathcal{V})$,

$$\mathrm{derived}(F) := \bigcup_{T \in F} \mathrm{derived}(T) \setminus \bigcup_{T \in F} \mathrm{underlying}(T)$$

And underlying $\in \mathcal{F} \to \mathrm{P}(\mathcal{V})$,

$$\mathrm{underlying}(F) := \bigcup_{T \in F} \mathrm{underlying}(T) \setminus \bigcup_{T \in F} \mathrm{derived}(T)$$

The *underlying variables* of a *fud* are sometimes called the *substrate*.

A *functional definition set* is a *histogram expression* which can be simplified to an *equivalent transform*, transform $\in \mathcal{F} \to \mathcal{T}_{\mathrm{f,U}}$

$$\mathrm{transform}(F) := (\prod \mathrm{histograms}(F)\ \%\ (\mathrm{der}(F)\ \cup\ \mathrm{und}(F)), \mathrm{der}(F))$$

where der = derived and und = underlying. Also, define $F^{\mathrm{T}} = \mathrm{transform}(F)$. The resultant *equivalent transform* is also a *unit functional transform* with the same *derived* and *underlying variables*, $\mathrm{der}(F^{\mathrm{T}}) = \mathrm{der}(F)$ and $\mathrm{und}(F^{\mathrm{T}}) = \mathrm{und}(F)$.

In order to apply a *functional definition set* F to a *histogram* A apply the *equivalent transform*, $A * F^{\mathrm{T}}$. However, the *evaluation* of this *histogram expression* requires the computation of an intermediate *histogram* $\prod \mathrm{histograms}(F)$, with cardinality of nearly $|\mathrm{vars}(F)^{\mathrm{C}}|$, before the *multiplication* with A and subsequent *reduction* to the *derived variables*, $\mathrm{der}(F)$, takes place. (See 'Computation of functional definition sets', below.) This cardinality may be much greater than that of the given *histogram* $|A|$. An alternative method of application, for example, is to navigate through the *functional definition set reducing* any *non-derived variables* as soon as possible. Define $\mathrm{apply} \in \mathcal{F} \times \mathcal{A} \to \mathcal{A}$ as $\mathrm{apply}(F, A) := \mathrm{apply}(\mathrm{und}(F), \mathrm{der}(F), \mathrm{his}(F), A)$ where $\mathrm{apply} \in \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{A}) \times \mathcal{A} \to \mathcal{A}$ is defined recursively as

$$\begin{aligned}
&\mathrm{apply}(V, W, M, A) := \mathrm{if}(X \neq \emptyset, \mathrm{apply}(V, W, N, C), A\%W) : \\
&\quad X = \{(|B|, B, Q)) : D \in M,\ \mathrm{vars}(D) \cap (\mathrm{vars}(A) \cup V) \neq \emptyset, \\
&\qquad Q = M \setminus \{D\},\ B = (A * D)\ \%\ (W \cup \bigcup\{\mathrm{vars}(E) : E \in Q\})\}, \\
&\quad \{(\cdot, C, N)\} = \mathrm{mind}(\mathrm{order}(D_{\mathbf{N} \times \mathcal{A} \times \mathcal{X}}, X))
\end{aligned}$$

The enumeration $D_{\mathbf{N} \times \mathcal{A} \times \mathcal{X}}$ orders by *size*, *histogram* and then arbitrarily, $D_{\mathbf{N} \times \mathcal{A} \times \mathcal{X}} \in \mathrm{enums}(\mathbf{N} \times \mathcal{A} \times \mathcal{X})$. The apply function assumes that the *variables* V and W are connected transitively via the *variables* of the *histograms* of the *fud*. Only those *histograms* that are in the closure of the union of *variables* V and $\mathrm{vars}(A)$ are applied. In the case where there is no path from V to W, the function returns $A\%W$. All of the *histograms* of the *fud*, F, are in the closure of the *underlying*, $\mathrm{und}(F)$, so in the case where $V = \mathrm{und}(F)$ the application of the *fud* equals the application of the *transform*, $\mathrm{apply}(F, A) = A * F^{\mathrm{T}}$. Also there are other implementations depending on the computational constraints.

Following from the constraint on the *derived variables* define $\mathrm{definitions} \in \mathcal{F} \to (\mathcal{V} \to \mathcal{T}_{\mathrm{f,U}})$ as

$$\mathrm{definitions}(F) := \{(w, (A, W)) : (A, W) \in F,\ w \in W\}$$

The subset of the *functional definition set* which recursively contains all the *underlying transforms* for a given set of *variables* is defined $\mathrm{depends} \in \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathcal{F}$ as $\mathrm{depends}(F, W) := \mathrm{depends}(F, W, \emptyset)$ where $\mathrm{depends} \in \mathcal{F} \times \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{V}) \to \mathcal{F}$ is defined

$$\begin{aligned}
&\mathrm{depends}(F, W, X) := \\
&\quad \bigcup\{\{T\} \cup \mathrm{depends}(F, \mathrm{und}(T), X \cup \{w\}) : \\
&\qquad\qquad w \in W \cap \mathrm{dom}(\mathrm{def}(F)) \setminus X,\ T = \mathrm{def}(F)(w)\}
\end{aligned}$$

where def = definitions.

Fuds can contain both *null transforms*, $\mathrm{der}(T) = \emptyset$, and *disjoint transforms*, $\mathrm{und}(T) = \emptyset$, if the *transforms* are *functional*. *Null transforms* are always *functional* because there is only one *derived state*, $\emptyset$. *Disjoint transforms* are only *functional* if the *derived states* forms a singleton, $|\mathrm{std}(T)| = 1$. This is the case, for example, when the *derived variables* are all *mono-valent* in the *system* U, $\{|U_w| : w \in \mathrm{der}(T)\} = \{1\}$.

A *fud circularity* in a *functional definition set* can occur where a *defined variable* w appears in its own *depends fud* $w \in \mathrm{vars}(\mathrm{depends}(F, \{w\}) \setminus \{\mathrm{definitions}(F)(w)\})$. Define $\mathrm{circular} \in \mathcal{F} \to \mathbf{B}$ as

$$\begin{aligned}&\mathrm{circular}(F) := \\ &\quad \exists w \in \mathrm{dom}(\mathrm{def}(F))\ (w \in \mathrm{vars}(\mathrm{depends}(F, \{w\}) \setminus \{\mathrm{def}(F)(w)\}))\end{aligned}$$

A *fud circularity* cannot occur in the *definition transform* $\mathrm{definitions}(F)(w)$ because the *derived variables* and *underlying variables* are disjoint. The *fud circularity* must be in a *layer* below.

Contradictions created by *fud circularities* can prevent the *equivalent transform* F^{T} from being *one functional*, $F^{\mathrm{T}} \notin \mathcal{T}_{U,\mathrm{f},1}$, even if the *fud* contains only *one functional transforms*, $F \in \mathrm{P}(\mathcal{T}_{U,\mathrm{f},1})$.

The *one functional definition set* subset $\mathcal{F}_{U,1} \subset \mathcal{F}$ in *system* U is defined such that all *transforms* are *one functional* and the *fud* is not *circular*,

$$\forall F \in \mathcal{F}_{U,1}\ (F \in \mathrm{P}(\mathcal{T}_{U,\mathrm{f},1}) \wedge \neg\ \mathrm{circular}(F))$$

Thus the *equivalent transform* of a *one functional definition set* is a *one functional transform*, $\forall F \in \mathcal{F}_{U,1}\ (F^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1})$.

A *one functional definition set* $F \in \mathcal{F}_{U,1}$ implies a *one functional definition set* $G \in \mathcal{F}_{U,1}$ such that all the *transforms* in G have a single *derived variable*, $\forall T \in G\ (|\mathrm{der}(T)| = 1)$, and such that the *equivalent transforms* are equal, $G^{\mathrm{T}} = F^{\mathrm{T}}$. Define $\mathrm{mono} \in \mathcal{F}_{U,1} \to \mathcal{F}_{U,1}$

$$\mathrm{mono}(F) := \bigcup\{\{(X\%(\mathrm{vars}(X) \setminus W \cup \{w\}), \{w\}) : w \in W\} : (X, W) \in F\}$$

A *non-circular functional definition set* can be viewed as a tree of *variables*, $\mathrm{trees}(\mathcal{V})$. To construct the *variables* tree from the *functional definition set*, define treeVariable $\in \mathcal{F} \to \mathrm{trees}(\mathcal{V})$ as treeVariable$(F) := \{(v, \mathrm{treev}(F, v, \emptyset)) : v \in \mathrm{der}(F)\}$. Define treev $\in \mathcal{F} \times \mathcal{V} \times \mathrm{P}(\mathcal{V}) \to \mathrm{trees}(\mathcal{V})$ as

$$\begin{aligned}
&\mathrm{treev}(F, v, X) := \\
&\quad \mathrm{if}(v \in \mathrm{dom}(\mathrm{def}(F)), \\
&\qquad \{(w, \mathrm{treev}(F, w, X \cup \{w\})) : w \in \mathrm{und}(\mathrm{def}(F)(v)) \setminus X\}, \emptyset)
\end{aligned}$$

Thus $\mathrm{vars}(F) = \mathrm{elements}(\mathrm{treeVariable}(F))$.

Another tree is a tree of *transforms*. Define treeTransform $\in \mathcal{F} \to \mathrm{trees}(\mathcal{T}_{\mathrm{f,U}})$ as treeTransform$(F) := \{(T, \mathrm{treet}(F, T, \mathrm{und}(F))) : v \in \mathrm{der}(F),\ T = \mathrm{def}(F)(v)\}$. Define treet $\in \mathcal{F} \times \mathcal{T}_{\mathrm{f,U}} \times \mathrm{P}(\mathcal{V}) \to \mathrm{trees}(\mathcal{T}_{\mathrm{f,U}})$ as

$$\begin{aligned}
&\mathrm{treet}(F, T, X) := \\
&\quad \{(R, \mathrm{treet}(F, R, X \cup \{w\})) : w \in \mathrm{und}(T) \setminus X,\ R = \mathrm{def}(F)(w)\}
\end{aligned}$$

The depends-subset of a *non-circular functional definition set* could be defined

$$\begin{aligned}
&\mathrm{depends}(F, W) := \\
&\quad \bigcup \mathrm{elements}(\mathrm{nodes}(\mathrm{treeTransform}(F))(\mathrm{def}(F)(w))) : w \in W \cap \mathrm{dom}(\mathrm{def}(F))\}
\end{aligned}$$

The *layer* in a *non-circular functional definition set* is the length of the longest path to the leaves from any of a given set of *variables*. Define layer $\in \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathbf{N}$ as $\mathrm{layer}(F, W) := \mathrm{layer}(F, W, \mathrm{und}(F))$, and layer $\in \mathcal{F} \times \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{V}) \to \mathbf{N}$ as

$$\begin{aligned}
&\mathrm{layer}(F, W, X) := \\
&\quad \mathrm{maxr}(\{(w, \mathrm{layer}(F, \mathrm{und}(T), X \cup \{w\}) + 1) : \\
&\qquad\qquad w \in W \cap \mathrm{dom}(\mathrm{def}(F)) \setminus X,\ T = \mathrm{def}(F)(w)\} \cup \{(\emptyset, 0)\})
\end{aligned}$$

The *layer* can also be defined using generic tree semantics, $\mathrm{layer}(F, W) := \mathrm{maxr}(\{(w, |L|) : w \in W,\ L \in \mathrm{paths}(\mathrm{treeTransform}(\mathrm{depends}(F, \{w\})))\})$.

The *transforms* of a *non-circular fud* $F \in \mathcal{F}$ can be arranged in a list of *layer fuds* $L \in \mathcal{L}(\mathrm{P}(F))$ such that each *transform* is in the highest *layer* of its *derived variables*, $L = \mathrm{inverse}(\{(T, \mathrm{layer}(F, \mathrm{der}(T))) : T \in F\})$. The set of *layer fuds* partitions the *fud*, $\mathrm{set}(L) \in \mathrm{B}(F)$. The *transforms* in a *layer fud* depend only on *transforms* in lower *layers*, $\forall (i, G) \in L\ \forall T \in G\ (\mathrm{depends}(F, \mathrm{der}(T)) \subseteq \bigcup \mathrm{set}(\mathrm{take}(i, L)))$.

A *linear fud* is a *non-circular fud* such that the *underlying variables* of the *transforms* in each *layer fud* are the *derived variables* of the *layer fud* immediately below, $\forall i \in \{2 \ldots |L|\}\ (\mathrm{und}(L_i) \subseteq \mathrm{der}(L_{i-1}))$. Each of the *layer fuds*, $\mathrm{set}(L) \subset \mathrm{P}(F)$, can be combined into a single *transform*. Thus a *linear fud* may be represented as a list of *transforms*, $\{(i, G^{\mathrm{T}}) : (i, G) \in L\} \in \mathcal{L}(\mathcal{T}_{\mathrm{f,U}})$.

The *top transform*, if it exists, is the *transform* in a *fud* that has the same *derived variables* as the *fud*, $\mathrm{top} \in \mathcal{F} \to \mathcal{T}_{\mathrm{f,U}}$

$$\mathrm{top} := T$$

where $\exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$.

A *functional definition set* is defined as *non-overlapping* if the *underlying variables* of the *depends fuds* of the *derived variables* are disjoint. Define $\mathrm{overlap} \in \mathcal{F} \to \mathbf{B}$

$$\begin{aligned}&\mathrm{overlap}(F) := \\ &\quad \exists v, w \in \mathrm{der}(F)\ (v \neq w \wedge (\mathrm{vars}(\mathrm{dep}(F, \{v\})) \cap \mathrm{vars}(\mathrm{dep}(F, \{w\})) \neq \emptyset))\end{aligned}$$

where dep = depends. The *empty fud* is *non-overlapping*, $\neg\mathrm{overlap}(\emptyset)$. It can be determined if a *fud* is *overlapping* between any two *layers* by taking the subset of the *transforms* of the *fud* between the *layers*. If the *fud* is *overlapping* in a *layer*, then it must be *overlapping* in all *layers* below that down to the *substrate*. The function vars in the definition of overlap above could equally be replaced by und.

A *non-overlapping fud* F can be viewed as the union of disjoint *fuds*, or a partition of *depends fud* components. Let $Q = \{\mathrm{depends}(F, \{w\}) : w \in \mathrm{der}(F)\}$ then $F = \bigcup Q$ and $\mathrm{overlap}(F) = (Q \notin \mathrm{B}(F))$.

3.7 Partitions and partition variables

A *partition* is a partition of the *cartesian* set of *states* for some set of *variables*. The *partition* consists of *component* sets of *states*. The *components* are disjoint, but union together to equal the original *cartesian* set.

Define $\mathcal{R} \subset \mathrm{P}(\mathrm{P}(\mathcal{S}) \setminus \{\emptyset\})$ as the set of all *partitions*. Define $\mathrm{vars} \in \mathcal{R} \to \mathrm{P}(\mathcal{V})$

$$\mathrm{vars}(P) := \bigcup\{\mathrm{vars}(S) : S \in \bigcup P\}$$

The *partitions* are constrained

$$\forall P \in \mathcal{R}\ \forall C \in P\ \forall S \in C\ (\mathrm{vars}(S) = \mathrm{vars}(P))$$

and

$$\forall P \in \mathcal{R}\ \forall C, D \in P\ (C \neq D \implies C \cap D = \emptyset)$$

That is, $P \in \mathrm{B}(\bigcup P)$, where B is the partition function (see appendix). The *empty partition*, $\emptyset \in \mathcal{R}$, has no *variables*, $\mathrm{vars}(\emptyset) = \emptyset$. The *scalar partition*, which is the unary partition of the singleton set of *empty state*, $\{\{\emptyset\}\} \in \mathcal{R}$, has no *variables*, $\mathrm{vars}(\{\{\emptyset\}\}) = \emptyset$.

A subset of the parents of a *partition* are those for which the cardinality of the *partition* is decremented. That is, *partition* $Q \in \mathcal{R}$ and parent $P \in \mathrm{parents}(Q)$ such that $|P| = |Q| - 1$. Define $\mathrm{decrements} \in \mathcal{R} \to \mathrm{P}(\mathcal{R})$ as

$$\mathrm{decrements}(Q) := \{P : P \in \mathrm{parents}(Q),\ |P| = |Q| - 1\}$$

which can be constructed explicitly

$$\mathrm{decrements}(Q) = \{Q \setminus \{C, D\} \cup \{C \cup D\} : C, D \in Q,\ C \neq D\}$$

Thus $\forall Q \in \mathcal{R}\ (\mathrm{decrements}(Q) \subseteq \mathrm{parents}(Q))$. The cardinality of the *decremented parent partition* set is $|\mathrm{decrements}(Q)| = |Q|(|Q| - 1)/2$. The search tree initialised from the *self partition* finds all *partitions* in some *variables* V in a *system* U, $\mathrm{elements}(\mathrm{searchTreer}(\mathcal{R}, \mathrm{decrements}, \{V^{\mathrm{CS}\{\}}\})) = \mathrm{B}(V^{\mathrm{CS}})$. See appendix 'Search and optimisation' for a definition of the tree search.

In a particular *system* U let $\mathcal{R}_U \subset \mathcal{R} \cap \mathrm{P}(\mathrm{P}(\mathcal{S}_U) \setminus \{\emptyset\})$. $\mathcal{R}_U$ is additionally constrained

$$\forall P \in \mathcal{R}_U\ (\bigcup P = \mathrm{cartesian}(U)(\mathrm{vars}(P)))$$

The set of all *partitions* in *system* U can be constructed explicitly

$$\mathcal{R}_U = \bigcup \{\mathrm{B}(\mathrm{cartesian}(U)(W)) : W \in \mathrm{P}(\mathrm{vars}(U))\}$$

The *empty partition* is not in $\mathcal{R}_U$, $\emptyset \notin \mathcal{R}_U$. The *scalar partition* is in $\mathcal{R}_U$, $\{\emptyset^{\mathrm{CS}}\} = \{\{\emptyset\}\} \in \mathcal{R}_U$.

A *partition* can be *expanded* to a superset of its *variables* by crossing with the *cartesian states* of the disjoint set of *variables*. Define $\mathrm{expand}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{R}_U \to \mathcal{R}_U$ as

$$\mathrm{expand}(U)(V, P) := \{\{S \cup R : S \in C,\ R \in (V \setminus W)^{\mathrm{CS}}\} : C \in P\}$$

where $W = \mathrm{vars}(P)$. Define shorthand $P^V := \mathrm{expand}(U)(V, P)$. The *variables* of the *expanded partition* are the union $\mathrm{vars}(P^V) = \mathrm{vars}(P) \cup V$. The cardinality of the *partition* is unchanged by the *expansion*, $|P^V| = |P|$. The *expansion* of a *scalar partition* is the unary *partition*, $\{\emptyset^{\mathrm{CS}}\}^V = \{V^{\mathrm{CS}}\}$.

The converse operation is to *contract* a *partition* to the minimum subset of *variables* by removing any *cartesian variables*. Define $\mathrm{contract}(U) \in \mathcal{R}_U \to \mathcal{R}_U$ as

$$\begin{aligned}&\mathrm{contract}(U)(P) := Q :\\ &\quad \{Q\} = \mathrm{mind}(\{(R, |K|) : K \subseteq V,\ R = \{\{S\%K : S \in C\} : C \in P\},\ R^V = P\})\end{aligned}$$

where $V = \mathrm{vars}(P)$. Define shorthand $P^{\%} := \mathrm{contract}(U)(P)$. The *variables* of the *contracted partition* are a subset $\mathrm{vars}(P^{\%}) \subseteq \mathrm{vars}(P)$. The cardinality of the *partition* is unchanged by the *contraction*, $|P^{\%}| = |P|$. There is always exactly one possible *contraction*, $|\mathrm{min}(\{(R, |K|) : K \subseteq V,\ R = \{\{S\%K : S \in C\} : C \in P\},\ R^V = P\})| = 1$. A unary *partition*, $P = \mathrm{unary}(V^{\mathrm{CS}}) = \{V^{\mathrm{CS}}\} \in \mathrm{R}_U$ which is such that $|P| = 1$, *contracts* to the *scalar partition*, $P^{\%} = \{V^{\mathrm{CS}}\}^{\%} = \{\emptyset^{\mathrm{CS}}\} = \{\{\emptyset\}\}$.

Partition variables in *system* $U \in \mathcal{U}$ are defined such that the *partition variable* $P \in \mathrm{vars}(U)$ is itself a *partition*, $P \in \mathcal{R}_U$, and its *values*, U_P, are the *components* of this *partition*, $U_P = P$. That is, $(P, U_P) = (P, P) \in U$. If *system* U' contains all of its *partition variables*, $\forall P \in \mathcal{R}_{U'}\ ((P, P) \in U')$, then the *system*, U', must be infinite by recursive definition, $|U'| = \infty$. Let function implied $\in \mathcal{U} \to \mathcal{U}$ be defined as

$$\mathrm{implied}(U) := U \cup \mathrm{implied}(U \cup \{(P, P) : P \in \mathcal{R}_U\})$$

The implied infinite *system* is $U' = \mathrm{implied}(U)$.

Partitions can be thought of as partitions of the domains of *histograms*. Thus *one functional transforms* can be constructed having a single *derived variable* which is the *partition variable* of the *partition* of the given *underlying variables* in *system* U. Define transform $\in \bigcup\{\mathcal{R}_U \to \mathcal{T}_{U,\mathrm{f},1} : U \in \mathcal{U}\}$

$$\mathrm{transform}(P) := (\{(S \cup \{(P, C)\}, 1) : C \in P,\ S \in C\}, \{P\})$$

Define shorthand $P^{\mathrm{T}} := \mathrm{transform}(P)$. These *one functional transforms* are called *partition transforms*.

The converse function partition $\in \bigcup\{\mathcal{T}_{U,\mathrm{f},1} \to \mathcal{R}_U : U \in \mathcal{U}\}$ recovers the *partition* from a given *functional transform*

$$\mathrm{partition}(T) := \{\mathrm{states}(A) : A \in \mathrm{ran}(\mathrm{inverse}(T))\}$$

Define shorthand $T^{\mathrm{P}} := \mathrm{partition}(T)$. The converse relationship obeys $\forall P \in \mathcal{R}_U\ (P^{\mathrm{TP}} = P)$ in *system* U. Thus P and P^{T} are isomorphic, $P \cong P^{\mathrm{T}}$.

A *system partition* $P \in \mathcal{R}_U$, which is *non-empty*, $P \neq \emptyset$, may be *expanded* using its *partition transform*, P^{T}, $(\mathrm{his}(P^{\mathrm{T}}) * V^{\mathrm{C}}, \{P\})^{\mathrm{P}} = P^V$, where his = histogram. The application of an *expanded partition transform* $P^{V\mathrm{T}}$ to a *histogram* A, where $V = \mathrm{vars}(A)$, forms a bijective map to the application of the *partition transform* P^{T}, where $K = \mathrm{vars}(P)$ and $K \subset V$, $\exists Q \in A * P^{V\mathrm{T}} :\leftrightarrow: A * P^{\mathrm{T}}\ \forall((S,c),(T,d)) \in Q\ (c = d)$. This is because there is a bijective map between the *components* of P^V and P, and so $A * P^{V\mathrm{T}}$ and $A * P^{\mathrm{T}}$ are *non-literal reframes*, $\mathrm{reframe}(X, A * P^{V\mathrm{T}}) = A * P^{\mathrm{T}}$ where $X = \{(P^V, (P, \{(\{S \cup R : S \in C,\ R \in (V \setminus W)^{\mathrm{CS}}\}, C) : C \in P\}))\}$.

Define a function that returns the set of *partition transforms* for a set of *partitions*, creating a *partition functional definition set*, transforms $\in \bigcup\{\mathrm{P}(\mathcal{R}_U) \to \mathcal{F}_{U,\mathrm{P}} : U \in \mathcal{U}\}$

$$\mathrm{transforms}(Q) := \mathrm{map}(\mathrm{transform}, Q)$$

The *partition transforms* form a subset of *one functional transforms* $\mathcal{T}_{U,\mathrm{P}} \subset \mathcal{T}_{U,\mathrm{f},1}$ which is defined

$$\mathcal{T}_{U,\mathrm{P}} := \{T : T \in \mathcal{T}_{U,\mathrm{f},1},\ T^{\mathrm{PT}} = T\} = \{P^{\mathrm{T}} : P \in \mathcal{R}_U\}$$

Partition transforms are such that $\forall T \in \mathcal{T}_{U,\mathrm{P}}\ (\mathrm{der}(T) = \{T^{\mathrm{P}}\})$.

Similarly the *partition functional definition sets* is the subset of *one functional definition sets* which contain only *partition transforms* $\mathcal{F}_{U,\mathrm{P}} \subset \mathcal{F}_{U,1}$, defined as

$$\mathcal{F}_{U,\mathrm{P}} = \mathrm{P}(\mathcal{T}_{U,\mathrm{P}})$$

The *definition constraint* on *derived variables* of *fuds* is implied by definition because $\{(P, (X, \{P\})) : (X, \{P\}) \in F\} \in \mathcal{V}_U \to \mathcal{T}_{U,\mathrm{P}}$ where $F \in \mathcal{F}_{U,\mathrm{P}}$, just as $\mathrm{definitions}(F) \in (\mathrm{vars}(F) \setminus \mathrm{und}(F)) \to F$. Similarly there are no *fud circularities* in $F \in \mathcal{F}_{U,\mathrm{P}}$ by definition rather than by constraint. If there exists a *top transform* of a *partition fud* $F \in \mathcal{F}_{U,\mathrm{P}}$, $\mathrm{depends}(F, \mathrm{der}(\mathrm{top}(F))) = F$, then there is only one *derived variable*, $|\mathrm{der}(\mathrm{top}(F))| = |\mathrm{der}(F)| = 1$. All *partition fuds* are *mono-variate* in the *derived variables* of the *transforms*, $F = \mathrm{mono}(F)$.

The *multi-partition transforms* form a subset of *one functional transforms* $\mathcal{T}_{U,\mathrm{P}^*} \subset \mathcal{T}_{U,\mathrm{f},1}$ which is defined as the set of *transforms* of *single-layer partition fuds*

$$\begin{aligned}
\mathcal{T}_{U,\mathrm{P}^*} &:= \{F^{\mathrm{T}} : F \in \mathcal{F}_{U,\mathrm{P}},\ \mathrm{layer}(F, \mathrm{der}(F)) = 1\} \\
&= \{F^{\mathrm{T}} : F \in \mathcal{F}_{U,\mathrm{P}},\ (\forall T \in F\ (\mathrm{und}(T) \subseteq \mathrm{und}(F))\} \\
&= \{\{P^{\mathrm{T}} : P \in Q\}^{\mathrm{T}} : Q \subset \mathcal{R}_U,\ (\forall P_1, P_2 \in Q\ (P_2 \notin \mathrm{vars}(P_1))\}
\end{aligned}$$

The *partition transforms* is a subset of the *multi-partition transforms*, $\mathcal{T}_{U,\mathrm{P}} \subset \mathcal{T}_{U,\mathrm{P}^*}$.

The *multi-partition functional definition sets* is the subset of *one functional definition sets* which contain only *multi-partition transforms* $\mathcal{F}_{U,\mathrm{P}^*} \subset \mathcal{F}_{U,1}$, defined as

$$\mathcal{F}_{U,\mathrm{P}^*} = \mathrm{P}(\mathcal{T}_{U,\mathrm{P}^*})$$

The *partition fuds* is a subset of the *multi-partition fuds*, $\mathcal{F}_{U,\mathrm{P}} \subset \mathcal{F}_{U,\mathrm{P}^*}$.

Multi-partition transforms may be *expanded* to *expanded multi-partition transforms*. Define $\mathrm{expand}(U, V) \in \mathcal{T}_{U,\mathrm{P}^*} \to \mathcal{T}_{U,\mathrm{P}^*}$ as $\mathrm{expand}(U, V)(T) := \{P^{V\mathrm{T}} : P \in \mathrm{der}(T)\}^{\mathrm{T}}$. Define $T^V := \mathrm{expand}(U, V)(T)$. A *multi-partition fud* $F \in \mathcal{F}_{U,\mathrm{P}^*}$ is said to be *expanded* if all of its *multi-partition transforms* are *expanded*, $\forall T \in F\ (T = T^V)$.

Multi-partition transforms may be *contracted* to *contracted multi-partition transforms*. Define $\mathrm{contract}(U) \in \mathcal{T}_{U,\mathrm{P}^*} \to \mathcal{T}_{U,\mathrm{P}^*}$ as $\mathrm{contract}(U)(T) := \{P^{\%\mathrm{T}} : P \in \mathrm{der}(T)\}^{\mathrm{T}}$. Define $T^{\%} := \mathrm{contract}(U)(T)$. A *multi-partition fud* $F \in \mathcal{F}_{U,\mathrm{P}^*}$ is said to be *contracted* if all of its *multi-partition transforms* are *contracted*, $\forall T \in F\ (T = T^{\%})$.

Multi-partition transforms may be *exploded* to a *partition fud*. Define $\mathrm{explode} \in \bigcup\{\mathcal{T}_{U,\mathrm{P}^*} \to \mathcal{F}_{U,\mathrm{P}} : U \in \mathcal{U}\}$ as $\mathrm{explode}(T) := \{P^{\mathrm{T}} : P \in \mathrm{der}(T)\}$. A *multi-partition fud* may be *exploded* to a *partition fud*. Define $\mathrm{explode} \in \bigcup\{\mathcal{F}_{U,\mathrm{P}^*} \to \mathcal{F}_{U,\mathrm{P}} : U \in \mathcal{U}\}$ as $\mathrm{explode}(F) := \bigcup\{\mathrm{explode}(T) : T \in F\} = \{P^{\mathrm{T}} : T \in F,\ P \in \mathrm{der}(T)\}$.

3.8 Pointed partitions

The set of *pointed partitions* $\mathcal{R}_* \subset \mathcal{R} \times \mathrm{P}(\mathcal{S})$ are pairs of (i) *partitions*, and (ii) *components* of the *partition*, $\forall (P, C_*) \in \mathcal{R}_*\ (C_* \in P)$. The set of *pointed partitions* is defined

$$\mathcal{R}_* = \{(P, C_*) : P \in \mathcal{R},\ C_* \in P\}$$

Define $\mathrm{vars} \in \mathcal{R}_* \to \mathrm{P}(\mathcal{V})$ as $\mathrm{vars}((P, C_*)) := \mathrm{vars}(P)$. The *partition* of a *pointed partition* cannot be the *empty partition*, $\forall (P, \cdot) \in \mathcal{R}_*\ (P \neq \emptyset)$. The *scalar pointed partition* is $(\{\{\emptyset\}\}, \{\emptyset\}) \in \mathcal{R}_*$. For any unary *partition* $P = \{C\} \in \mathcal{R}$ there is exactly one *pointed partition*, $(P, C) \in \mathcal{R}_*$.

Define $\mathrm{transform} \in \mathcal{R}_* \to \mathcal{T}_{\mathrm{f,U}}$ as $\mathrm{transform}((P, C_*)) := \mathrm{transform}(P)$. Define shorthand $P_*^{\mathrm{T}} := \mathrm{transform}(P_*)$. There is no converse function in $\mathcal{T}_{\mathrm{f}} \to \mathcal{R}_*$ except in the case where the *transform* $T \in \mathcal{T}_{\mathrm{f}}$ maps to a *unary partition*, $|\mathrm{inverse}(T)| = 1$. In this case the *pointed partition* is (P, C_*) where $C_* = C^{\mathrm{S}}$, $\{C\} = \mathrm{ran}(\mathrm{inverse}(T))$ and $P = \{C_*\}$.

The *variables* of a *pointed partition* $P_* \in \mathcal{R}_*$ are also known as the *underlying variables* because they equal the *underlying variables* of the *pointed partition transform*, $\mathrm{vars}(P_*) = \mathrm{und}(P_*^{\mathrm{T}})$. The *partition* P of the *pointed partition*, $(P, \cdot) = P_*$ is also known as the *derived variable*, $\{P\} = \mathrm{der}(P_*^{\mathrm{T}})$.

The set of *incremented pointed partitions* of a *pointed partition* are those for which the cardinality of the *point component* is decremented and either (i) the cardinality of the *partition* is incremented, or (ii) the cardinality of one of the other *components* is incremented. That is, *pointed partition* $(P, C_*) \in \mathcal{R}_*$ and *incremented pointed partition* $(Q, D_*) \in \mathcal{R}_*$ are such that $|D_*| = |C_*| - 1$ and either (i) $|Q| = |P| + 1$, or (ii) $\exists C \in P\ \exists D \in Q\ (|D| = |C| + 1)$. Define $\mathrm{increments} \in \mathcal{R}_* \to \mathrm{P}(\mathcal{R}_*)$ as

$$\begin{aligned}
&\mathrm{increments}((P, C_*)) := \\
&\quad \{(P \setminus \{C_*\} \cup \{D_*, D\}, D_*) : |C_*| > 1,\ S \in C_*, \\
&\qquad\qquad D_* = C_* \setminus \{S\},\ D = \{S\}\} \cup \\
&\quad \{(P \setminus \{C_*, C\} \cup \{D_*, D\}, D_*) : |C_*| > 1,\ |P| > 1,\ S \in C_*,\ C \in P,\ C \neq C_*, \\
&\qquad\qquad D_* = C_* \setminus \{S\},\ D = C \cup \{S\}\}
\end{aligned}$$

The *incremented pointed partitions* cardinality is $|\mathrm{increments}((P, C_*))| = |C_*||P|$ if $|C_*| > 1$ otherwise $|\mathrm{increments}((P, C_*))| = 0$. Let the search tree initialised from the *unary partition* of some *variables* V in a *system* U be

$$Z_+ = \mathrm{tree}(\mathrm{searchTreer}(\mathcal{R}_*, \mathrm{increments}, \{(\{V^{\mathrm{CS}}\}, V^{\mathrm{CS}})\}))$$

Z_+ finds all *partitions* in *variables* V, $\mathrm{dom}(\mathrm{elements}(Z_+)) = \mathrm{B}(V^{\mathrm{CS}})$. Compare to the search tree of *decremented partitions* initialised from the *self partition*,

$$Z_- = \mathrm{tree}(\mathrm{searchTreer}(\mathcal{R}, \mathrm{decrements}, \{V^{\mathrm{CS}\{\}}\}))$$

which is also such that $\mathrm{elements}(Z_-) = \mathrm{B}(V^{\mathrm{CS}})$, but the *increments* and *decrements* are not converses of each other, $|\mathrm{nodes}(Z_-)| < |\mathrm{nodes}(Z_+)|$. The paths of the *decrements* tree are subsets of the paths of the *increments* tree, $\forall L_- \in \mathrm{paths}(Z_-)\ \exists L_+ \in \mathrm{paths}(Z_+)\ (\mathrm{set}(L_-) \subset \mathrm{dom}(\mathrm{set}(L_+)))$.

The subset *singleton pointed partitions* $\mathcal{R}_{*,\mathrm{s}} \subset \mathcal{R}_*$ is defined $\mathcal{R}_{*,\mathrm{s}} = \{(P, C_*) : (P, C_*) \in \mathcal{R}_*,\ |C_*| = 1\}$. A *pointed self partition*, $(X^{\{\}}, \{x\})$ where $x \in X$, is necessarily a *singleton pointed partition*.

The subset *pointed binary partitions* $\mathcal{R}_{*,\mathrm{b}} \subset \mathcal{R}_*$ is defined $\mathcal{R}_{*,\mathrm{b}} = \{(P, C_*) : (P, C_*) \in \mathcal{R}_*,\ |P| = 2\}$. A *pointed binary partition* has a *complement*, $P'_* = (\{A, B\}, B)$ where $P_* = (\{A, B\}, A)$. The *complement* is in the same *variables*, $\mathrm{vars}(P'_*) = \mathrm{vars}(P_*)$.

There are logical operators on *pointed binary partitions* which *derive* a *pointed binary partition* from *underlying pointed binary partitions*. Let *pointed binary partition* $P_* = (P, C_*) \in \mathcal{R}_{*,\mathrm{b}}$ be the pair of the *binary partition variable*, P where $|P| = 2$, and the *point component*, $C_* \in P$. So $P_* \in \mathcal{V} \times \mathcal{W}$ and $\{P_*\}$ is a *state*, $\{P_*\} \in \mathcal{S}$. In a *system* U which contains the *partition variable*, $(P, P) \in U$, the *cartesian states* are $\{P\}^{\mathrm{CS}} = \{\{P_*\}, \{P'_*\}\}$ where P'_* is the *complement* of P_*. The only *binary partition* of the *cartesian states* equals the *self partition*, $\{P\}^{\mathrm{CS}\{\}} = \{\{\{P_*\}\}, \{\{P'_*\}\}\} \in \mathcal{R}$. Define $\mathrm{not} \in \mathcal{R}_{*,\mathrm{b}} \to \mathcal{R}_{*,\mathrm{b}}$ as

$$\mathrm{not}(P_*) := (\{\{\{P_*\}\}, \{\{P'_*\}\}\}, \{\{P'_*\}\})$$

The *underlying variable* of the resultant *pointed partition* is the given *partition variable*, $\mathrm{vars}(\mathrm{not}(P_*)) = \{P\} \subset \mathcal{V} \cap \mathcal{R}$. The *derived variable* of the resultant *pointed partition* is the *binary partition* of the *cartesian states*, $\{\{\{P_*\}\}, \{\{P'_*\}\}\} \in \mathcal{V} \cap \mathcal{R}$.

Given two *pointed binary partitions* $P_*, R_* \in \mathcal{R}_{*,\mathrm{b}}$ the *cartesian states* are $\{P, R\}^{\mathrm{CS}} = \{\{P_*, R_*\}, \{P_*, R'_*\}, \{P'_*, R_*\}, \{P'_*, R'_*\}\}$, where P'_* and R'_* are the *complements* of P_* and R_*. The *and binary partition* of the *cartesian states* is $\{X, Y\} \in \mathcal{R}$, where $X = \{\{P_*, R_*\}\}$ and $Y = \{\{P_*, R'_*\}, \{P'_*, R_*\}, \{P'_*, R'_*\}\}$.

The *and pointed binary partition* is $(\{X,Y\},X) \in \mathcal{R}_{*,\mathrm{b}}$. Define $\mathrm{and} \in \mathcal{R}_{*,\mathrm{b}} \times \mathcal{R}_{*,\mathrm{b}} \to \mathcal{R}_{*,\mathrm{b}}$ as

$$\mathrm{and}(P_*,R_*) := (\{X,Y\},X)$$

The *underlying variables* are the given *partition variables*, $\mathrm{vars}(\mathrm{and}(P_*,R_*)) = \{P,R\} \subset \mathcal{V} \cap \mathcal{R}$. The *derived variable* of the resultant *pointed partition* is the *binary partition* of the *cartesian states*, $\{X,Y\} \in \mathcal{V} \cap \mathcal{R}$.

Similarly, the *or binary partition* of the *cartesian states* is $\{A,B\} \in \mathcal{R}$, where $A = \{\{P_*,R_*\},\{P_*,R'_*\},\{P'_*,R_*\}\}$ and $B = \{\{P'_*,R'_*\}\}$. The *or pointed binary partition* is $(\{A,B\},A) \in \mathcal{R}_{*,\mathrm{b}}$. Define $\mathrm{or} \in \mathcal{R}_{*,\mathrm{b}} \times \mathcal{R}_{*,\mathrm{b}} \to \mathcal{R}_{*,\mathrm{b}}$

$$\mathrm{or}(P_*,R_*) := (\{A,B\},A)$$

The *underlying variables* are the given *partition variables*, $\mathrm{vars}(\mathrm{or}(P_*,R_*)) = \{P,R\} \subset \mathcal{V} \cap \mathcal{R}$. The *derived variable* of the resultant *pointed partition* is the *binary partition* of the *cartesian states*, $\{A,B\} \in \mathcal{V} \cap \mathcal{R}$.

The *and* and *not* operators result in *singleton pointed binary partitions*, $\mathrm{and}(P_*,R_*) \in \mathcal{R}_{*,\mathrm{s}} \cap \mathcal{R}_{*,\mathrm{b}}$ and $\mathrm{not}(P_*) \in \mathcal{R}_{*,\mathrm{s}} \cap \mathcal{R}_{*,\mathrm{b}}$, but the *or* operation does not.

The *and* and *or* binary operations can be extended to sets of *pointed binary partitions*. Define $\mathrm{and} \in \mathrm{P}(\mathcal{R}_{*,\mathrm{b}}) \to \mathcal{R}_{*,\mathrm{b}}$ as

$$\mathrm{and}(V_*) := (\{\{V_*\}, V^{\mathrm{CS}} \setminus \{V_*\}\}, \{V_*\})$$

where $V = \{P : (P,\cdot) \in V_*\}$. The *cartesian states*, $V^{\mathrm{CS}} = \mathrm{cartesian}(U)(V)$, requires a *system* U implied from the given *partitions*, $U = \{(P,P) : P \in V\}$. Note that $V_* \in \mathcal{S}_U$. Define $\mathrm{or} \in \mathrm{P}(\mathcal{R}_{*,\mathrm{b}}) \to \mathcal{R}_{*,\mathrm{b}}$

$$\mathrm{or}(V_*) := (\{V^{\mathrm{CS}} \setminus \{V'_*\}, \{V'_*\}\}, V^{\mathrm{CS}} \setminus \{V'_*\})$$

where $V'_* = \{P'_* : P_* \in V_*\}$.

Given a tree of *pointed binary partitions*, $\mathrm{trees}(\mathcal{R}_{*,\mathrm{b}})$, a tree of inherited *and* operations can be derived. Define $\mathrm{and} \in \mathrm{trees}(\mathcal{R}_{*,\mathrm{b}}) \to \mathrm{trees}(\mathcal{R}_{*,\mathrm{b}})$ as

$$\mathrm{and}(Z) := \{(P_*, \mathrm{and}(P_*,X)) : (P_*,X) \in Z\}$$

Define $\mathrm{and} \in \mathcal{R}_{*,\mathrm{b}} \times \mathrm{trees}(\mathcal{R}_{*,\mathrm{b}}) \to \mathrm{trees}(\mathcal{R}_{*,\mathrm{b}})$ as

$$\mathrm{and}(P_*,Z) := \{(M_*, \mathrm{and}(M_*,X)) : (R_*,X) \in Z,\ M_* = \mathrm{and}(P_*,R_*)\}$$

Let $F \in \mathcal{F}_{U,\mathrm{P}}$ be the *fud* of the given *pointed binary partition tree* $Z \in \mathrm{trees}(\mathcal{R}_{*,\mathrm{b}})$ and its derived *and* tree, $\mathrm{and}(Z) \in \mathrm{trees}(\mathcal{R}_{*,\mathrm{b}})$. That is, $F = \{P_*^{\mathrm{T}} : P_* \in \mathrm{elements}(Z) \cup \mathrm{elements}(\mathrm{and}(Z))\}$. The *derived variables* are the leaf *partitions* of the tree, $\mathrm{der}(F) = \mathrm{dom}(\mathrm{leaves}(\mathrm{and}(Z)))$. The *underlying variables* of the *fud* are the *variables* of the *pointed partitions* of the given tree, $\mathrm{und}(F) = \bigcup\{\mathrm{vars}(P_*) : P_* \in \mathrm{elements}(Z)\}$.

The set of *pointed partitions* in a *system* U is $\mathcal{R}_{*,U}$. *Pointed partitions* can be *expanded*, $\mathrm{expand}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{R}_{*,U} \to \mathcal{R}_{*,U}$, and *contracted*, $\mathrm{contract}(U) \in \mathcal{R}_{*,U} \to \mathcal{R}_{*,U}$, such that $P_*^{\%V} = P_*$ where $V = \mathrm{vars}(P_*)$.

Given a *variable* $w \in \mathcal{V}_U$ in *system* U and a *pointed binary partition* $P_* \in \mathcal{R}_{*,\mathrm{b}}$, a *nullable pointed partition* can be *derived* which has *values* for each of the *values* of w when the *value* of P_* is the *point component* C_*, where $(P, C_*) = P_*$, and has a *null value* for any *value* of w when the *value* of P_* is the *complement point component* C'_*, where $(P, C'_*) = P'_*$. Define $\mathrm{nullable}(U) \in \mathcal{V}_U \times \mathcal{R}_{*,U,\mathrm{b}} \to \mathcal{R}_{*,U}$ as

$$\begin{aligned}\mathrm{nullable}(U)(w, P_*) &:= (Q \cup \{D_*\}, D_*) : \\ &\qquad Q = (\{\{P_*\}\}^{\mathrm{U}} * \{w\}^{\mathrm{C}})^{\mathrm{S}\{\}},\ D_* = (\{\{P'_*\}\}^{\mathrm{U}} * \{w\}^{\mathrm{C}})^{\mathrm{S}}\end{aligned}$$

where (i) P'_* is the *complement pointed partition* of P_*, (ii) the *non-null* set of *components* $Q \subset \mathrm{P}(\{w, P\}^{\mathrm{CS}})$ is a self partition, and (iii) D_* is the *null point component* $D_* \in \mathrm{P}(\{w, P\}^{\mathrm{CS}})$. The *underlying variables* are $\{w, P\} \subset \mathcal{V}$. The *derived nullable variable* is $Q \cup \{D_*\} \in \mathcal{V} \cap \mathcal{R}$. That is, $\mathrm{der}((\mathrm{nullable}(U)(w, P_*))^{\mathrm{T}}) = \{Q \cup \{D_*\}\}$. The *point component*, D_*, is the *null value* of the *nullable variable*, $D_* \in Q \cup \{D_*\}$. The resultant *nullable pointed partition* $(Q \cup \{D_*\}, D_*) \in \mathcal{R}_{*,U}$ is only a *pointed binary partition* if w is *mono-valent*, $|U_w| = 1 \implies (Q \cup \{D_*\}, D_*) \in \mathcal{R}_{*,U,\mathrm{b}}$. The cardinality of the *values* of the *nullable variable* is one greater than that of the given *variable*, $|Q \cup \{D_*\}| = |U_w| + 1$. If the *nullable variable* is in the *system*, $Q \cup \{D_*\} \in \mathrm{vars}(U)$, then the *volume* is also incremented, $|\{Q \cup \{D_*\}\}^{\mathrm{C}}| = |\{w\}^{\mathrm{C}}| + 1$.

3.9 Overlapping transforms

The *derived variables* of a *transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ are *non-overlapping* if there exists an equivalent *transform* of a *fud* $F \in \mathcal{F}_{U,1}$ which is *non-overlapping*, $\exists F \in \mathcal{F}_{U,1}\ ((F^{\mathrm{T}} = T) \wedge \neg\mathrm{overlap}(F))$. If the *transform* is *non-overlapping*, there exists at least one equivalent *fud* F that weakly partitions the *underlying variables*, $\{\mathrm{und}(\mathrm{depends}(F, \{w\})) : w \in W\} \in \mathrm{B}'(V)$, where $W = \mathrm{der}(T)$,

$V = \text{und}(T)$ and B' is the weak partition function, $\text{B}'(V) := \text{B}(V) \cup \{Y \cup \{\emptyset\} : Y \in \text{B}(V)\}$ and $\text{B}'(\emptyset) := \{\{\emptyset\}\}$. If the *transform*, T, has *pluri-valent derived variables*, $\exists w \in W\ (|(X\%\{w\})^{\text{F}}| > 1)$ where $(X, W) = T$, then there exists an equivalent *fud* G having *pluri-valent derived variables*, $\text{der}(G) = W'$, where $W' = \{w : w \in W,\ |(X\%\{w\})^{\text{F}}| > 1\}$, such that $G^{\text{T}} = ((X\%(V \cup W'))^{\text{F}}, W')$, that strongly partitions the *underlying variables*, $\{\text{und}(\text{depends}(G, \{w\})) : w \in W'\} \in \text{B}(V)$. *Mono-valent derived variables* of the *transform* map to *disjoint transform variables* in the corresponding weakly partitioning *fud*, $\text{und}(R) = \emptyset$, where $R \in \text{ran}(M)$ and $M \in W \cdot F$. Define overlap $\in \mathcal{T}_{\text{U}} \to \mathbf{B}$ as

$$\begin{aligned}&\text{overlap}(T) := \neg(W' \neq \emptyset \implies \\ &\quad \exists Q \in \text{B}(V)\ \exists R \in W' \cdot Q\ \forall (w, K) \in R\ (((X\%(K \cup \{w\}))^{\text{F}}, \{w\}) \in \mathcal{T}_{\text{f}}))\end{aligned}$$

where $V = \text{und}(T)$, $(X, W) = T$ and $W' = \{w : w \in W,\ |(X\%\{w\})^{\text{F}}| > 1\}$. The *empty transform* is *non-overlapping*, $\neg\text{overlap}((\emptyset, \emptyset))$. If the *transform* T is *non-overlapping*, the corresponding *non-overlapping fud* is

$$F = \{((X\%(K \cup \{w\}))^{\text{F}}, \{w\}) : (w, K) \in R\} \in \mathcal{F}$$

where $R = \{(w, \text{und}(\text{depends}(F, \{w\}))) : w \in \text{der}(T)\} \in W \to \text{P}(V)$.

A *one functional transform* $T \in \mathcal{T}_{U,\text{f},1}$ is *right total* if and only if the *transform* is *non-overlapping*

$$\neg\text{overlap}(T) \iff (X\%W)^{\text{F}} = W^{\text{C}}$$

where $(X, W) = T$. For example, let $W \subset \text{B}(V^{\text{CS}})$ where $V = \text{und}(T)$ and $W = \text{der}(T)$.

The *possible derived states* are the *effective derived states* of the application of the *transform* to the *cartesian*, $(V^{\text{C}} * T)^{\text{FS}} \subseteq W^{\text{CS}}$. The *possible derived states* is the domain of the *transform inverse*, $(V^{\text{C}} * T)^{\text{FS}} = \text{dom}(T^{-1}) = \text{stateDeriveds}(T)$. The *possible derived volume* $w' = |(V^{\text{C}} * T)^{\text{F}}|$ is the cardinality of the *partition*, $w' = |T^{\text{P}}| = |T^{-1}|$, and so is less than or equal to the *derived volume*, $w' \leq w$, where $w = |W^{\text{C}}|$. The *possible derived volume* equals the *derived volume* if and only if the *transform* is *non-overlapped*, $\neg\text{overlap}(T) \iff w' = w$, because it is only in this case that the *transform* is *right total*, $\text{dom}(T^{-1}) = (X\%W)^{\text{FS}} = W^{\text{CS}}$. If the *transform* is *overlapping* there are necessarily *impossible derived states*, $\text{overlap}(T) \implies W^{\text{CS}} \setminus (V^{\text{C}} * T)^{\text{FS}} \neq \emptyset$. The *possible derived volume* is less than or equal to the *underlying volume*, $w' \leq v$, where $v = |V^{\text{C}}|$.

Derived variables $x, y \in W$ are said to be *tautological* if their *partitions* are equal, $\mathrm{partition}((X\%(V \cup \{x\}), \{x\})) = \mathrm{partition}((X\%(V \cup \{y\}), \{y\}))$, where $(X, W) = T$ and $V = \mathrm{und}(T)$. A *transform* is *tautologically overlapped* if all of its *derived variables* are *tautological*. Define tautology $\in \mathcal{T}_\mathrm{f} \to \mathbf{B}$ as $\mathrm{tautology}(T) := |\{\mathrm{partition}((X\%(V \cup \{w\}), \{w\})) : w \in W\}| = 1$. A *tautology* is always *overlapped*, $\forall T \in \mathcal{T}_\mathrm{f} \cap \mathcal{T}_\mathrm{U}\ (\mathrm{tautology}(T) \implies \mathrm{overlap}(T))$.

A *multi-partition transform* $T \in \mathcal{T}_{U,\mathrm{P}^*}$ is *overlapping* if and only if the *contracted transform* is *overlapping*, $\mathrm{overlap}(T) \iff \mathrm{overlap}(T^{\%})$. A *contracted multi-partition transform* is *overlapping* if and only if its *explode fud* is *overlapping*, $\mathrm{overlap}(T^{\%}) \iff \mathrm{overlap}(\mathrm{explode}(T^{\%}))$.

A *multi-partition transform* $T \in \mathcal{T}_{U,\mathrm{P}^*}$ represents a functional map between the *underlying states* and the *derived states*, $V^{\mathrm{CS}} \to W^{\mathrm{CS}}$, so the *possible derived volume* is at most the product of the cardinalities of the *partitions*, $w' = |T^{\mathrm{P}}| \leq \prod_{P \in W} |P| = |W^{\mathrm{CS}}|$. If the *transform* is *expanded*, $W \subset \mathrm{B}(V^{\mathrm{CS}})$, the *right total* case requires that all of the *components* intersect, $\forall P, Q \in W\ \forall C \in P\ \forall D \in Q\ (C \cap D \neq \emptyset)$. This is true for *non-overlapping transforms* because the *contracted partitions* have disjoint *variables*, $\forall P_1, P_2 \in W\ (P_1 \neq P_2 \implies \mathrm{vars}(P_1^{\%}) \cap \mathrm{vars}(P_2^{\%}) = \emptyset)$.

A *contracted multi-partition fud* $F \in \mathcal{F}_{U,\mathrm{P}^*}$, where $\forall T \in F\ (T = T^{\%})$, is *recursively non-overlapping* if the *dependent exploded fud* of each of its *contracted multi-partition transforms* is *non-overlapping*,

$$\forall T \in F\ (\neg\mathrm{overlap}(\mathrm{depends}(\mathrm{explode}(F), \mathrm{der}(T))))$$

3.10 Decompositions

The set of *decompositions* $\mathcal{D}$ is a subset of the trees of pairs of (i) *states*, $\mathcal{S}$, and (ii) *unit functional transforms*, $\mathcal{T}_\mathrm{f} \cap \mathcal{T}_\mathrm{U}$

$$\mathcal{D} \subset \mathrm{trees}(\mathcal{S} \times \mathcal{T}_{\mathrm{f,U}})$$

The set of *decompositions* is constrained such that the set of *transforms* forms a *functional definition set*,

$$\forall D \in \mathcal{D}\ (\mathrm{ran}(\mathrm{elements}(D)) \in \mathcal{F})$$

Define transforms $\in \mathcal{D} \to \mathcal{F}$ as $\mathrm{transforms}(D) := \mathrm{ran}(\mathrm{elements}(D))$. Define underlying $\in \mathcal{D} \to \mathrm{P}(\mathcal{V})$ as $\mathrm{underlying}(D) := \mathrm{underlying}(\mathrm{transforms}(D))$.

The *transforms* form a *fud*, so the sets of *derived variables* are disjoint, $\forall (A,W),(B,X) \in G\ ((A,W) \neq (B,X) \implies W \cap X = \emptyset)$, where $G = \text{transforms}(D)$.

There are some additional constraints on *decompositions*. First, no *underlying variable* of a *transform* in the *decomposition* can be a *derived variable* in another *transform*

$$\forall D \in \mathcal{D}\ (\bigcup\{\text{und}(T) : T \in \text{transforms}(D)\} = \text{underlying}(D))$$

where und = underlying. That is, the *fud* is single *layer*, $\text{layer}(G, \text{der}(G)) = 1$ or $\forall T \in G\ (\text{und}(T) \subseteq \text{und}(G))$. So the *derived variables* of a *transform* do not intersect with the *variables* of any other, $\forall T_1, T_2 \in G\ (T_1 \neq T_2 \implies \text{der}(T_1) \cap \text{vars}(T_2) = \emptyset)$, where $G = \text{transforms}(D)$.

Second, the *states* of the root pairs are *empty states*

$$\forall D \in \mathcal{D}\ (\text{dom}(\text{roots}(D)) = \{\emptyset\})$$

Third, each of the *states* in child pairs are *states* of the *derived variables* of the parent *transform*

$$\forall D \in \mathcal{D}\ \forall ((\cdot,T),(S,\cdot)) \in \text{steps}(D)\ (S \in \text{std}(T))$$

where std = stateDeriveds = dom ∘ inverse. The set of *states* need not be all of the *derived states*, but only a subset or empty, $\forall ((\cdot,T),E) \in \text{nodes}(D)\ (\text{dom}(\text{dom}(E)) \subseteq \text{std}(T))$.

The *empty decomposition* consists of the *empty transform*, $\{((\emptyset,(\emptyset,\emptyset)),\emptyset)\} \in \mathcal{D}$.

The application of a *decomposition* to a *histogram* is a tree of contingent applications of the *transforms*. Define $\text{apply} \in \mathcal{D} \times \mathcal{A} \to \text{trees}(\mathcal{S} \times \mathcal{A})$ as $\text{apply}(D,A) := \text{apply}(D, \text{vars}(A), A)$ where $\text{apply} \in \mathcal{D} \times \text{P}(\mathcal{V}) \times \mathcal{A} \to \text{trees}(\mathcal{S} \times \mathcal{A})$ is

$$\begin{aligned}
&\text{apply}(D,V,A) := \\
&\quad \{((S, B\%W), \text{apply}(E,V,B)) : ((S,(X,W)),E) \in D, \\
&\qquad\qquad B = A * \{S\}^{\text{U}} * X\ \%\ (V \cup W)\}
\end{aligned}$$

Define shorthand $A * D = \text{apply}(D,A)$.

The application of a *decomposition*, D, to a *histogram*, A, can be *contingently* constrained by a query *histogram* $Q \in \mathcal{A}$ to produce a tree of subsets of the given *histogram*, A. Define $\mathrm{query} \in \mathcal{D} \times \mathcal{A} \times \mathcal{A} \to \mathrm{trees}(\mathcal{S} \times \mathcal{A})$ as $\mathrm{query}(D, A, Q) := \mathrm{query}(D, \mathrm{vars}(A), A, Q)$ where $\mathrm{query} \in \mathcal{D} \times \mathrm{P}(\mathcal{V}) \times \mathcal{A} \times \mathcal{A} \to \mathrm{trees}(\mathcal{S} \times \mathcal{A})$ is

$$\begin{aligned}
&\mathrm{query}(D, V, A, Q) := \\
&\quad \{((S, B\%V), \mathrm{query}(E, V, B, R)) : ((S, (X, W)), E) \in D, \\
&\qquad\qquad R = Q * \{S\}^{\mathrm{U}} * X \ \% \ (V \cup W),\ \mathrm{size}(R) > 0, \\
&\qquad\qquad B = A * \{S\}^{\mathrm{U}} * X * (R\%W) \ \% \ (V \cup W)\}
\end{aligned}$$

In the case where the query *variables* are a superset of the *decomposition's underlying variables*, $\mathrm{vars}(Q) \supseteq \mathrm{und}(D)$, and the query has a single *effective state*, $|Q^{\mathrm{F}}| = 1$, the resultant tree has a single path, $|\mathrm{paths}(\mathrm{query}(D, A, Q))| = 1$.

A *decomposition* $D \in \mathcal{D}$ is *distinct* if the elements are a functional map of *states* to *transforms*, $\mathrm{elements}(D) \in \mathcal{S} \to \mathcal{T}$. In fact, a less strict definition is all that is necessary, $\forall E \in \{D\} \cup \mathrm{ran}(\mathrm{nodes}(D))\ (\mathrm{dom}(E) \in \mathcal{S} \to \mathcal{T})$. The subset of *distinct decompositions*, $\mathcal{D}_{\mathrm{d}} \subset \mathcal{D}$ is defined $\mathcal{D}_{\mathrm{d}} = \{D : D \in \mathcal{D},\ \mathrm{dom}(D) \in \mathcal{S} \to \mathcal{T},\ \mathrm{ran}(D) \subset \mathcal{D}_{\mathrm{d}}\}$. The *distinct decompositions* allow the same *transform* to be in more than one path with different ancestors, $\exists D \in \mathcal{D}_{\mathrm{d}}\ (\mathrm{flip}(\mathrm{elements}(D)) \notin \mathcal{T} \to \mathcal{S})$. The function on trees of pairs, $\mathrm{distinct} \in \mathrm{trees}(\mathcal{X} \times \mathcal{Y}) \to \mathrm{P}(\mathrm{trees}(\mathcal{X} \times \mathcal{Y}))$, described in appendix 'Trees', returns the set of *distinct decomposition* trees as $\mathrm{distinct} \in \mathrm{trees}(\mathcal{S} \times \mathcal{T}) \to \mathrm{P}(\mathrm{trees}(\mathcal{S} \times \mathcal{T}))$

$$\begin{aligned}
&\mathrm{distinct}(D) := \\
&\quad \{H : H \subseteq \{((S, T), G) : ((S, T), E) \in D,\ G \in \mathrm{distinct}(E)\}, \\
&\qquad\qquad \mathrm{dom}(H) \in \mathrm{dom}(\mathrm{dom}(D)) :\to \mathrm{ran}(\mathrm{dom}(D))\}
\end{aligned}$$

where $\mathrm{distinct}(\emptyset) := \{\emptyset\}$. Given a *decomposition*, $D \in \mathcal{D}$, the function returns a set of *distinct decompositions*, $\mathrm{distinct}(D) \in \mathrm{P}(\mathcal{D}_{\mathrm{d}})$, because $(\mathcal{D} \to \mathrm{P}(\mathcal{D}_{\mathrm{d}})) \subset (\mathrm{trees}(\mathcal{S} \times \mathcal{T}) \to \mathrm{P}(\mathrm{trees}(\mathcal{S} \times \mathcal{T})))$. A *distinct decomposition* $D \in \mathcal{D}_{\mathrm{d}}$ has singleton roots, $\mathrm{roots}(D) = \{(\emptyset, \cdot)\}$.

The subset of *decompositions* $\mathcal{D}_U \subset \mathcal{D}$ in a *system* U is a subset of the trees of pairs of (i) *states*, $\mathcal{S}_U$, and (ii) *one functional transforms*, $\mathcal{T}_{U,\mathrm{f},1}$

$$\mathcal{D}_U = \mathcal{D} \cap \mathrm{trees}(\mathcal{S}_U \times \mathcal{T}_{U,\mathrm{f},1})$$

If a *decomposition* is *distinct* there exists an inversion of the *decomposition*, $\mathrm{trees}(\mathcal{S}_U \times \mathcal{T}_{U,\mathrm{f},1}) \to \mathrm{trees}(\mathcal{T}_{U,\mathrm{f},1} \times \mathcal{S}_U)$, from pairs of *transforms* and their parent *transform's derived states* to pairs of *transforms* and their own *derived states*. The inversion is such that the *derived states* of the *transforms* are *completed* where they correspond to non-empty *components* of the *partition* of the *underlying variables* of the *decomposition*. That is, the *decomposition completed states* are *possible states* of the *transform*. Define $\mathrm{application}(U) \in \mathcal{D}_{\mathrm{d},U} \to \mathrm{trees}(\mathcal{T}_{U,\mathrm{f},1} \times \mathcal{S}_U)$ as $\mathrm{application}(U)(D) := \mathrm{app}(U)(D, \emptyset, \emptyset, \emptyset)$ where $\mathrm{app}(U) \in \mathrm{trees}(\mathcal{S}_U \times \mathcal{T}_{U,\mathrm{f},1}) \times \mathcal{S}_U \times \mathcal{S}_U \times \mathcal{F}_{U,1} \to \mathrm{trees}(\mathcal{T}_{U,\mathrm{f},1} \times \mathcal{S}_U)$ is

$$
\begin{aligned}
&\mathrm{app}(U)(D,Q,R,F) := \\
&\quad\quad \{((T,S), \mathrm{app}(U)(E,S,R \cup S, F \cup \{T\})) : ((P,T),E) \in D,\ P = Q, \\
&\quad\quad\ \ S \in \mathrm{dom}(\mathrm{dom}(E)),\ R \cup S \in \mathrm{std}((F \cup \{T\})^{\mathrm{T}})\} \\
&\quad \cup\ \ \{((T,S),\emptyset) : ((P,T),E) \in D,\ P = Q, \\
&\quad\quad\ \ S \in \mathrm{der}(T)^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(E)),\ R \cup S \in \mathrm{std}((F \cup \{T\})^{\mathrm{T}})\}
\end{aligned}
$$

The computation of the set of *derived states*, $\mathrm{std}((F \cup \{T\})^{\mathrm{T}}) \subset \mathcal{S}$, in an implementation of this definition may be impracticable if the *volume* of the intermediate *transform*, $(F \cup \{T\})^{\mathrm{T}}$, is too large. An equivalent definition may be given in terms of the application of the *fud*, constrained to contain the *disjoint transform* of the ancestor *state*, $(\{R\}^{\mathrm{U}}, \mathrm{vars}(R))$, to a *unit scalar*, $\mathrm{states}(\mathrm{apply}(F \cup \{T, (\{R\}^{\mathrm{U}}, \mathrm{vars}(R))\}, \mathrm{scalar}(1))) \subset \mathrm{std}((F \cup \{T\})^{\mathrm{T}})$. That is,

$$
\begin{aligned}
&\mathrm{app}(U)(D,Q,R,F) := \\
&\quad\quad \{((T,S), \mathrm{app}(U)(E,S,R \cup S, F \cup \{T\})) : ((P,T),E) \in D,\ P = Q, \\
&\quad\quad\ \ S \in \mathrm{dom}(\mathrm{dom}(E)),\ W = \mathrm{vars}(R) \cup \mathrm{vars}(S), \\
&\quad\quad\ \ X = \mathrm{his}(F \cup \{T\}) \cup \{\{R\}^{\mathrm{U}}, \{S\}^{\mathrm{U}}\},\ R \cup S \in \mathrm{apply}(V,W,X,Z_1)^{\mathrm{S}}\} \\
&\quad \cup\ \ \{((T,S),\emptyset) : ((P,T),E) \in D,\ P = Q, \\
&\quad\quad\ \ S \in \mathrm{der}(T)^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(E)),\ W = \mathrm{vars}(R) \cup \mathrm{vars}(S), \\
&\quad\quad\ \ X = \mathrm{his}(F \cup \{T\}) \cup \{\{R\}^{\mathrm{U}}, \{S\}^{\mathrm{U}}\},\ R \cup S \in \mathrm{apply}(V,W,X,Z_1)^{\mathrm{S}}\}
\end{aligned}
$$

where $V = \mathrm{und}(D)$ and $Z_1 = \mathrm{scalar}(1)$.

The *application tree* $D^* = \mathrm{application}(U)(D)$ is defined only for *distinct decompositions*, $D \in \mathcal{D}_{\mathrm{d},U}$. The *application tree* is such that only one *transform* appears in the roots,

$$\forall E^* \in \{D^*\} \cup \mathrm{ran}(\mathrm{nodes}(D^*))\ (|\mathrm{dom}(\mathrm{dom}(E^*))| = 1)$$

The elements of the *application tree* are pairs of the *transforms* and their *possible derived states*, $\mathrm{elements}(D^*) \subseteq \bigcup\{\{T\} \times \mathrm{std}(T) : T \in \mathrm{transforms}(D)\}$, where $\text{std}(T) = \text{dom}(T^{-1}) = (V^{\text{C}} * T)^{\text{FS}}$. *Impossible derived states*, $W^{\text{CS}} \setminus \text{std}(T)$, where $W = \text{der}(T)$, that exist if the *transform* is *overlapped*, $\neg\text{overlap}(T)$, are excluded.

The *application tree* is constructed by concatenating the *derived state* S of the *transform* T to the accumulated *derived state* R of the ancestors, $R \cup S$. Similarly the *transform* T is concatenated to the accumulated *functional definition set* F of the ancestors, $G = F \cup \{T\}$. *Application trees* exclude contradictions, $R \cup S \notin \mathcal{S}$, because these unions of *states* are not in the *possible derived states* of the accumulated *functional definition set*, $\text{std}(G^{\text{T}}) \subset \mathcal{S}$. The exclusion of contradictions occurs if the same *derived variable* appears more than once in a path. That is, if the same *transform* appears more than once. Multiple *transforms* in the same path are necessarily redundant because the *application* children have the same *derived state*. If the *transform* of the accumulated *fud* is *overlapping*, $\text{overlap}(G^{\text{T}})$, then it is not *right total*, $X\%W \neq W^{\text{C}}$ where $(X, W) = G^{\text{T}}$, and hence some of the *cartesian derived states* are excluded because they are *impossible*. In this case the set of *possible derived states* is a proper subset, $\text{std}(G^{\text{T}}) \subset W^{\text{CS}}$.

The converse function that restores the *distinct decomposition* given the *application* is defined $\text{decomp} \in \text{trees}(\mathcal{T} \times \mathcal{S}) \to \mathcal{D}$ as $\text{decomp}(D^*) := \text{decomp}(\emptyset, D^*)$ where $\text{decomp} \in \mathcal{S} \times \text{trees}(\mathcal{T} \times \mathcal{S}) \to \mathcal{D}$ is

$$\begin{aligned}&\text{decomp}(R, D^*) := \\ &\quad \{((R,T), \bigcup\{\text{decomp}(S, E^*) : ((\cdot, S), E^*) \in D^*\}) : T \in \text{dom}(\text{dom}(D^*))\}\end{aligned}$$

A *well behaved decomposition* is equal to the converse of the *application*, $\text{decomp}(\text{application}(U)(D)) = D$. A *well behaved decomposition* is such that the set of its *transforms* equals that of its *application*, $\text{transforms}(D) = \text{dom}(\text{elements}(D^*))$. Let the set of *well behaved decompositions* in *system* U be defined $\mathcal{D}_{\text{w},U} = \{D : D \in \mathcal{D}_{\text{d},U},\ \text{decomp}(\text{application}(U)(D)) = D\}$. The *empty decomposition* is not *well behaved*, $\{((\emptyset, (\emptyset, \emptyset)), \emptyset)\} \notin \mathcal{D}_{\text{w},U}$.

There are a couple of functions on paths in the *application tree*. Define $\text{transforms} \in \mathcal{L}(\mathcal{T}_{\text{f}} \times \mathcal{S}) \to \mathcal{F}$ as $\text{transforms}(L) := \text{dom}(\text{set}(L))$. Define $\text{state} \in \mathcal{L}(\mathcal{T}_{\text{f}} \times \mathcal{S}) \to \text{P}(\mathcal{V} \times \mathcal{W})$ as $\text{state}(L) := \bigcup\text{ran}(\text{set}(L))$. The pair is sometimes expressed $(F, S) = (\text{trn}(L), \text{st}(L))$ where trn = transforms and st = state.

The *simple partition* of a *decomposition* is G^{TP} where $G = \text{transforms}(D)$. G is the union of the *transforms* of the *decomposition*. However, the *sim-*

ple partition does not correspond to the purpose of *decompositions* which is to represent contingent *application* of child *transforms*. That is, the *transforms* of paths in a *decomposition application tree* are unioned into *functional definition sets* and *applied* to *histograms* separately, $A * F$ where $F = \text{transforms}(L)$ and $L \in \text{paths}(D^*)$.

The *partition* of a *well behaved distinct decomposition* is derived from the paths of the *decomposition application*. Define $\text{partition}(U) \in \mathcal{D}_{\text{w},U} \to \mathcal{R}_U$ as

$$\text{partition}(U)(D) := \{(\text{his}(G^{\text{T}}) * \{S\}^{\text{U}} \ \% \ V)^{\text{S}} : L \in \text{paths}(D^*),\ S = \text{state}(L)\}$$

where $D^* = \text{application}(U)(D)$, $V = \text{und}(D)$, and $G = \text{transforms}(D)$. An equivalent definition in terms of a more tractable navigated *fud* application is

$$\begin{aligned}&\text{partition}(U)(D) := \\ &\quad \{\text{apply}(V, V, \text{his}(F) \cup \{V^{\text{C}}, \{S\}^{\text{U}}\}, Z_1)^{\text{S}} : \\ &\qquad\qquad L \in \text{paths}(D^*),\ (F, S) = (\text{trn}(L), \text{st}(L))\}\end{aligned}$$

Define shorthand $D^{\text{P}} = \text{partition}(U)(D)$. The *partition variables* are the *underlying variables* of the *decomposition*, $\text{vars}(D^{\text{P}}) = \text{und}(D)$. The union of non-empty components is a *partition* because each of the unions of the *one functional transforms* in the initial sub-paths of the *decomposition* is a *one functional definition set* and therefore a *partition*, $\forall L \in \text{subpaths}(D)\ \Diamond F = \text{transforms}(L)\ (F^{\text{TP}} \in \mathcal{R}_U)$. If the *distinct decomposition* consists only of a root *transform*, $\text{elements}(D) = \text{roots}(D) = \{(\emptyset, T)\}$, then the *partition* is simply that of the *transform*, $D^{\text{P}} = T^{\text{P}}$. The *decomposition partition* is a parent partition of the *simple partition*, $\text{parent}(D^{\text{P}}, G^{\text{TP}})$ where $G = \text{transforms}(D)$. Hence of the cardinality of the *simple partition* is greater than or equal to that of the *partition*, $|G^{\text{TP}}| \geq |D^{\text{P}}|$.

A tree of *components* can be mapped cumulatively from the *application tree*. Define $\text{component}(U, V) \in \mathcal{L}(\mathcal{T}_{U,\text{f},1} \times \mathcal{S}_U) \to \text{P}(V^{\text{CS}})$ for some *variables* $V \subset \mathcal{V}_U$ as

$$\text{component}(U, V)(L) := (\text{inverse}(F^{\text{T}})(S) * V^{\text{C}})^{\text{S}}$$

where $F = \text{transforms}(L)$ and $S = \text{state}(L)$. An equivalent definition in terms of *fud* application is

$$\text{component}(U, V)(L) := \text{apply}(V, V, \text{his}(F) \cup \{V^{\text{C}}, \{S\}^{\text{U}}\}, Z_1)^{\text{S}}$$

Define $\text{components}(U) \in \mathcal{D}_{\text{d},U} \to \text{trees}(\text{P}(V^{\text{CS}}))$ as

$$\text{components}(U)(D) := \text{mapAccum}(\text{component}(U,V), D^*)$$

where $D^* = \text{application}(U)(D)$ and $V = \text{und}(D)$. In this definition, the *inverse component* is *expanded* by *multiplication* with V^{C}.

The *decomposition partition* can also be defined in terms of the accumulated path *fuds*

$$\text{partition}(U)(D) := \{\text{component}(U,V)(L) : L \in \text{paths}(D^*)\}$$

or from the *components tree leaves*

$$\text{partition}(U)(D) := \text{leaves}(\text{components}(U)(D))$$

The child *components* of the *components tree* are subsets of their parent *components*, $\forall (C_1, C_2) \in \text{step}(Y)\ (C_2 \subseteq C_1)$ where $Y = \text{components}(U)(D)$. Each *component* of the *components tree* exists at a unique place, $\{(L_{|L|}, L) : L \in \text{subpaths}(Y)\} \in \text{P}(V^{\text{CS}}) \to \mathcal{L}(\text{P}(V^{\text{CS}}))$.

A *well behaved distinct decomposition* $D \in \mathcal{D}_{\text{w},U}$ in *system* U contains a *variable symmetry* if $\exists (L,T),(M,R) \in Q\ ((L \neq M) \wedge (\text{der}(T) \cap \text{der}(R) \neq \emptyset))$ where $Q = \{(L,T) : (L,E^*) \in \text{places}(D^*),\ T \in \text{dom}(\text{dom}(E^*))\} \in \mathcal{L}(\mathcal{T}_{\text{f}} \times \mathcal{S}) \to \mathcal{T}_{\text{f}}$ and $D^* = \text{application}(U)(D)$. The *transforms* of the *decomposition* form a *fud* which implies that the *derived variables* are uniquely defined. Therefore a *variable symmetry* is also a *transform symmetry*. That is, more strictly, $\exists (L,T),(M,R) \in Q\ ((L \neq M) \wedge (T = R))$.

In the still stricter case of $\exists (L,E^*),(M,G^*) \in \text{places}(D^*)\ ((L \neq M) \wedge (E^* = G^* \neq \emptyset))$ then D contains an *application symmetry*. In this case there is a bijection between the child *components* of L, $\{\text{inverse}(F^{\text{T}})(S) : N \in \text{paths}(E^*),\ P = \text{concat}(L,N),\ (F,S) = (\text{trn}(P), \text{st}(P))\}$ and the child *components* of M.

If the same *non-root transform symmetry*, $T \neq T_{\text{r}}$ where $\{(\emptyset, T_{\text{r}})\} = \text{roots}(D)$, exists in all paths, $\forall L \in \text{paths}(D)\ (T \in \text{dom}(\text{set}(L)))$, then the *decomposition* has a *non-contingent symmetry*. If the *non-contingent symmetry*, T, is also an *application symmetry* everywhere, the *decomposition* could be broken into two *distinct decompositions* at the *non-contingent symmetry*.

If it is the case that each node in a *well behaved distinct decomposition* $D \in$

$\mathcal{D}_{\mathrm{w},U}$ has a single *transform*, $\forall E \in \{D\} \cup \mathrm{ran}(\mathrm{nodes}(D))\ (|\mathrm{ran}(\mathrm{dom}(E))| = 1)$ then D is *completely symmetrical* and there is a unique path of *transforms*, $|\{\{(i,T) : (i,(\cdot,T)) \in L\} : L \in \mathrm{paths}(D)\}| = 1$. In this case the *partition* equals the *simple partition*, $D^{\mathrm{P}} = G^{\mathrm{TP}}$ where $G = \mathrm{transforms}(D)$.

A *decomposition application tree* D^* that has at least two children per node, $\forall E^* \in \{D^*\} \cup \mathrm{ran}(\mathrm{nodes}(D^*))\ (E^* \neq \emptyset \implies |E^*| \geq 2)$, has a depth that is limited by the cardinality of the *self-partition* of the *underlying variables* V which is the *underlying volume*, $|V^{\mathrm{CS}\{\}}| = |V^{\mathrm{C}}|$. That is, $\mathrm{depth}(D^*) \leq \mathrm{ceil}(\log_2(|V^{\mathrm{C}}|))$.

A *contingent* tree of pairs of *components* and *transforms* can be mapped cumulatively from a *decomposition* $D \in \mathcal{D}$. Define contingent $\in \mathcal{L}(\mathcal{S} \times \mathcal{T}_{\mathrm{f}}) \to \mathcal{A} \times \mathcal{T}_{\mathrm{f}}$ as

$$\begin{aligned}&\mathrm{contingent}(L) := (\mathrm{inverse}(F^{\mathrm{T}})(S), T) : \\ &\qquad (\cdot,T) = L_{|L|},\ F = \mathrm{ran}(\mathrm{set}(L_{\{1\ldots|L|-1\}})),\ S = \bigcup \mathrm{dom}(\mathrm{set}(L))\end{aligned}$$

where $\mathrm{contingent}(\{(1,(\cdot,T))\}) := (\mathrm{scalar}(1), T)$. An equivalent definition in terms of *fud* application is

$$\mathrm{contingent}(L) := (\mathrm{apply}(V,V,\mathrm{his}(F) \cup \{\{S\}^{\mathrm{U}}\}, Z_1)^{\mathrm{F}}, T)$$

Define contingents $\in \mathcal{D} \to \mathrm{trees}(\mathcal{A} \times \mathcal{T}_{\mathrm{f}})$ as

$$\mathrm{contingents}(D) := \mathrm{mapAccum}(\mathrm{contingent}, D)$$

In this definition, the *inverse component*, $\mathrm{inverse}(F^{\mathrm{T}})(S)$, is not *expanded*. The places of the *application tree*, D^*, are related to the *contingent* tree,

$$\begin{aligned}&\{(\mathrm{inverse}(F^{\mathrm{T}})(S), T) : \\ &\qquad (L, E^*) \in \mathrm{places}(D^*),\ (F,S) = (\mathrm{trn}(L), \mathrm{st}(L)),\ T \in \mathrm{dom}(\mathrm{dom}(E^*))\} \\ \subseteq\ &\mathrm{elements}(\mathrm{contingents}(D))\end{aligned}$$

The application tree of the *decomposition* D applied to *histogram* A, is related to the *contingent* tree,

$$\{B : (\cdot,B) \in \mathrm{elements}(A * D)\} = \{A * C * T : (C,T) \in \mathrm{cont}(D)\}$$

where $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$.

The definition of the *partition*, D^{P}, of a *well behaved distinct decomposition* $D \in \mathcal{D}_{\mathrm{w},U}$ implies a *transform*, D^{PT}, which is *mono-variate* in the *derived variables*, $|\mathrm{der}(D^{\mathrm{PT}})| = 1$, like any *partition transform*. A *crown transform* can be *derived* from a *well behaved distinct decomposition's application*. First, a *transform slice* tree is constructed which consists of *singleton pointed binary partitions*. Define $\mathrm{sliceTransforms}(U) \in \mathcal{D}_{\mathrm{w},U} \to \mathrm{trees}(\mathcal{R}_{*,U,\mathrm{s,b}})$ as

$$\mathrm{sliceTransforms}(U)(D) := \mathrm{map}(\mathrm{slice}(U), D^*)$$

where $D^* = \mathrm{application}(U)(D)$ and $\mathrm{slice}(U) \in (\mathcal{T}_{U,\mathrm{f},1} \times \mathcal{S}_U) \to \mathcal{R}_{*,U,\mathrm{s,b}}$ is defined as

$$\mathrm{slice}(U)((T,S)) := (\{\{S\}, \mathrm{der}(T)^{\mathrm{CS}} \setminus \{S\}\}, \{S\})$$

The same *transform* may appear more than once in the *decomposition*. Thus there may be fewer *transform slices* than the places in the *application tree*, $|\mathrm{elements}(\mathrm{sliceTransforms}(U)(D))| \leq |\mathrm{places}(D^*)|$.

Second, the *contingent slice* tree is constructed by inheriting a logical *and* operation on the underlying *transform slice* and the parent *contingent slice*. Like the *transform slices*, the *contingent slices* are *singleton pointed binary partitions*. Define $\mathrm{sliceContingents}(U) \in \mathcal{D}_{\mathrm{w},U} \to \mathrm{trees}(\mathcal{R}_{*,U,\mathrm{s,b}})$ as

$$\mathrm{sliceContingents}(U)(D) := \mathrm{and}(\mathrm{sliceTransforms}(U)(D))$$

where the *and* binary operation on *pointed binary partition* trees is defined above as $\mathrm{and} \in \mathrm{trees}(\mathcal{R}_{*,\mathrm{b}}) \to \mathrm{trees}(\mathcal{R}_{*,\mathrm{s,b}})$ as

$$\mathrm{and}(Z) := \{(P_*, \mathrm{and}(P_*, X)) : (P_*, X) \in Z\}$$

Define $\mathrm{and} \in \mathcal{R}_{*,\mathrm{b}} \times \mathrm{trees}(\mathcal{R}_{*,\mathrm{b}}) \to \mathrm{trees}(\mathcal{R}_{*,\mathrm{s,b}})$ as

$$\mathrm{and}(P_*, Z) := \{(M_*, \mathrm{and}(M_*, X)) : (R_*, X) \in Z,\ M_* = \mathrm{and}(P_*, R_*)\}$$

Let the set N of *slice partitions* be $N = \mathrm{elements}(\mathrm{sliceTransforms}(U)(D)) \cup \mathrm{elements}(\mathrm{sliceContingents}(U)(D))$. The *point component* C_* of the *transform* and *contingent slices*, $(\{C_*, C'_*\}, C_*) \in N$, is called the *in-slice component*. The *complement point component* C'_* is called the *out-slice component*.

The *slices fud* is the set of *transforms* of the union of (i) the *transform slices* and (ii) the *contingent slices*. Define $\mathrm{slices}(U) \in \mathcal{D}_{\mathrm{w},U} \to \mathcal{F}_{U,1}$ as

$$\begin{aligned}&\mathrm{slices}(U)(D) := \\ &\quad \{P_*^{\mathrm{T}} : P_* \in \mathrm{elements}(\mathrm{sliceTransforms}(U)(D)) \cup \\ &\quad\quad \mathrm{elements}(\mathrm{sliceContingents}(U)(D))\}\end{aligned}$$

The *underlying variables* of the *slices fud*, $\text{slices}(U)(D)$, equals the *derived variables* of the *well behaved decomposition transforms*, $\text{und}(H) = \text{der}(G)$ where $G = \text{transforms}(D)$ and $H = \text{slices}(U)(D)$.

The *transforms* in the *slices fud* have *underlying volumes* no greater than the largest *derived volume* of the *decomposition transforms*, $\text{maxr}(\{(T, |V^{\text{C}}|) : T \in \text{slices}(U)(D),\ V = \text{und}(T)\} = \text{maxr}(\{(R, |W^{\text{C}}|) : R \in G,\ W = \text{der}(T)\}$ where $G = \text{transforms}(D)$. If the *slice transforms* were instead constructed from the accumulated *fud* F and *state* S of each path L in the *application tree* D^*, $\{\{\{S\}, \text{der}(F^{\text{T}})^{\text{CS}} \setminus \{S\}\}^{\text{T}} : L \in \text{paths}(D^*),\ (F,S) = (\text{transforms}(L), \text{state}(L))\}$, there would be no need for a tree of *contingent slices* but the maximum *underlying volume* would be as large as the *derived volume* of the largest accumulated path *fud*, $\text{maxr}(\{(L, |W^{\text{C}}|) : L \in \text{paths}(D^*),\ F = \text{transforms}(L),\ W = \text{der}(F)\}$.

The *crown transform* is *transform* of the union of (i) the *transforms fud* and (ii) the *slices fud*, of a *well behaved distinct decomposition*. Define $\text{transformCrown}(U) \in \mathcal{D}_{\text{w},U} \to \mathcal{T}_{U,\text{f},1}$ as

$$\text{transformCrown}(U)(D) := (G \cup H)^{\text{T}}$$

where $G = \text{transforms}(D)$ and $H = \text{slices}(U)(D)$. The *underlying variables* of the *crown transform* equals the *underlying variables* of the *decomposition*, $\text{und}(T) = \text{und}(D)$ where $T = \text{transformCrown}(U)(D)$. The *crown transform* is such that $\text{crown}(X\%W)$ is true, where $(X,W) = T$. Another way of stating this is $\text{crown}(V^{\text{C}} * T)$, where $V = \text{und}(D)$. The *partition* of the *crown transform* equals the *decomposition partition*, $T^{\text{P}} = D^{\text{P}}$. However the converse does not hold, $T \neq D^{\text{PT}}$ because the *crown transform* is necessarily *pluri-variate* in its *derived variables*, $|\text{der}(T)| > 1$.

The *transforms* of the *slices fud* $H = \text{slices}(U)(D) \in \mathcal{F}$ are derived from *singleton pointed binary partitions*, $\mathcal{R}_{*,\text{s},\text{b}}$, but lose the knowledge of the *in-slice point component* because there is no converse function for $\text{transform} \in \mathcal{R}_* \to \mathcal{T}_{\text{f},\text{U}}$. However, because both the *transform* and *contingent slices* are *singleton pointed partitions*, the *in-slice derived state* S can be identified by the cardinality of its *component*, $|\text{inverse}(T)(S)| = 1$ where $T \in H$, or $S = \{(P,C)\}$ where $P = T^{\text{P}}$, $C \in P$ and $|C| = 1$. Of course, in the case of a *decomposition* with a root *transform*, $T \in \text{ran}(\text{roots}(D))$, which has only two *derived states*, $|\text{der}(T)^{\text{CS}}| = 2$, both the *in-slice* and *out-slice state* has singleton *components*. An alternative is to create *slice variables* explicitly, for example defined for each sub-path in the *application tree*, $\text{subpaths}(D^*) \subset \mathcal{V}$. Each of these *slice* *variables* would have well-known *values*, for example $\{\text{in}, \text{out}\} \subset \mathcal{W}$. The *in-slice derived states* of the *transforms* containing the explicitly defined *slice variables* could then be easily identified without relying on the cardinality.

A set of *nullable pointed partitions* can be *derived* from the *contingent slice pointed partition* tree, $\text{sliceContingents}(U)(D) \in \text{trees}(\mathcal{R}_{*,U,\text{s},\text{b}})$. First, obtain the *or pointed binary partition* of the set of parent *contingent slice partitions* for each *derived variable* of the *transforms* of the *decomposition*. Define $\text{varsSliceAlternates}(U) \in \mathcal{D}_{\text{w},U} \to \text{P}(\mathcal{V}_U \times \mathcal{R}_{*,U,\text{s},\text{b}})$ as

$$\begin{aligned}&\text{varsSliceAlternates}(U)(D) := \\ &\quad \{(w, \text{or}(N)) : (w, N) \in \text{inverse}(\text{slicesVars}(U)(D))\}\end{aligned}$$

where $\text{slicesVars}(U) \in \mathcal{D}_{\text{w},U} \to \text{P}(\mathcal{R}_{*,U,\text{s},\text{b}} \times \mathcal{V}_U)$ is defined as

$$\begin{aligned}&\text{slicesVars}(U)(D) := \\ &\quad \{(P_*, w) : (P_*, Q_*) \in \text{steps}(\text{sliceContingents}(U)(D)), \\ &\qquad (P, \cdot) = P_*,\ \{P, X\} = \text{vars}(Q_*),\ w \in \text{vars}(X)\}\end{aligned}$$

where the n-ary *or* operation on a set of *pointed binary partitions*, $\text{or} \in \text{P}(\mathcal{R}_{*,\text{b}}) \to \mathcal{R}_{*,\text{b}}$, is defined above. If there are *symmetries* for *variable* $w \in \text{der}(G^{\text{T}})$, where $G = \text{transforms}(D)$, then the cardinality of the *slices* corresponding to w is greater than one, $|\text{inverse}(\text{slicesVars}(U)(D))(w)| > 1$, and the *alternate slice* will have more than one underlying *contingent slice*, $|\text{vars}(\text{varsSliceAlternates}(U)(D)(w))| > 1$.

Then, obtain the *nullable pointed partition* of the *derived variable* and the *or pointed binary partition* of the *alternate slices* of the *decomposition*. Define $\text{nullables}(U) \in \mathcal{D}_{\text{w},U} \to \text{P}(\mathcal{R}_{*,U})$ as

$$\begin{aligned}&\text{nullables}(U)(D) := \\ &\quad \{\text{nullable}(U)(w, P_*) : (w, P_*) \in \text{varsSliceAlternates}(U)(D)\}\end{aligned}$$

where $\text{nullable}(U) \in \mathcal{V}_U \times \mathcal{R}_{*,U,\text{b}} \to \mathcal{R}_{*,U}$ is defined above.

The *nullable fud* is the union of (i) *full functional self partition transforms* of the *variables* of the *root transform*, (ii) the non-leaf *transform slice transforms*, (iii) the non-leaf *contingent slice transforms*, (iv) the *alternate slice transforms*, and (v) the *nullable transforms*, of a *distinct decomposition*. Define $\text{nullable}(U) \in \mathcal{D}_{\text{w},U} \to \mathcal{F}_{U,1}$ as

$$\begin{aligned}
&\text{nullable}(U)(D) := \\
&\quad \{\{w\}^{\text{CS}\{\}\text{T}} : T \in \text{ran}(\text{roots}(D)),\ w \in \text{der}(T)\} \cup \\
&\quad \{P_*^{\text{T}} : P_* \in \text{nonleaves}(\text{sliceTransforms}(U)(D)) \cup \\
&\quad\quad \text{nonleaves}(\text{sliceContingents}(U)(D)) \cup \\
&\quad\quad \text{ran}(\text{varsSliceAlternates}(U)(D)) \cup \\
&\quad\quad \text{nullables}(U)(D)\}
\end{aligned}$$

where $\text{nonleaves}(Z) = \text{elements}(Z) \setminus \text{leaves}(Z)$.

The *nullable transform* of a *well behaved distinct decomposition* is the *transform* of the union of (i) the *transforms fud* and (ii) the *nullable fud*. Define $\text{transform}(U) \in \mathcal{D}_{\text{w},U} \to \mathcal{T}_{U,\text{f},1}$ as

$$\text{transform}(U)(D) := (G \cup H)^{\text{T}}$$

where $G = \text{transforms}(D)$ and $H = \text{nullable}(U)(D)$. Define shorthand $D^{\text{T}} := \text{transform}(U)(D)$. The *underlying variables* of the *nullable transform* equals the *underlying variables* of the *decomposition*, $\text{und}(D^{\text{T}}) = \text{und}(D)$. The *partition* of the *nullable transform* equals the *decomposition partition*, $D^{\text{TP}} = D^{\text{P}}$. However the converse does not hold, $D^{\text{T}} \neq D^{\text{PT}}$ because the *nullable transform* is necessarily *pluri-variate* in its *derived variables*, $|\text{der}(D^{\text{T}})| > 1$, unless it trivially consists of nothing but a *mono-variate root transform*, $|\text{der}(T)| = 1$ where $\{(\emptyset, T)\} = D$.

The function $\text{originals}(U) \in \mathcal{D}_{\text{w},U} \to (\mathcal{V}_U \to \mathcal{V}_U)$ recovers the map from the *nullable derived variables* to the *root* and *transform derived variables*

$$\begin{aligned}
&\text{originals}(U)(D) := \\
&\quad \{(\{w\}^{\text{CS}\{\}}, w) : T \in \text{ran}(\text{roots}(D)),\ w \in \text{der}(T)\} \cup \\
&\quad \{(P, w) : P \in \text{dom}(\text{nullables}(U)(D)),\ \{w\} = \text{vars}(P) \cap \text{der}(G)\}
\end{aligned}$$

where $G = \text{transforms}(D)$. So $\text{originals}(U)(D) \in \text{der}(D^{\text{T}}) \to \text{der}(G)$. If it is the case that none of the *root transform derived variables* of a *well behaved decomposition* $D \in \mathcal{D}_{\text{w},U}$ are *symmetrical*, $\text{der}(T) \cap \text{der}(G \setminus \{T\}) = \emptyset$ where $\{T\} = \text{ran}(\text{roots}(D))$, then the mapping is a bijection, $\text{originals}(U)(D) \in \text{der}(D^{\text{T}}) \leftrightarrow \text{der}(G)$.

The *full functional self partition transforms*, $\{\{w\}^{\text{CS}\{\}\text{T}} : w \in W_{\text{r}}\}$, of the *derived variables* $W_{\text{r}} = \text{der}(T_{\text{r}})$ of the *root transform* $T_{\text{r}} \in \text{ran}(\text{roots}(D))$, are *transforms*, $\{P^{\mathrm{T}} : P \in N_{\mathrm{r}}\}$, of the *root frame variables* $N_{\mathrm{r}} = \{\{w\}^{\mathrm{CS}\{\}} : w \in W_{\mathrm{r}}\}$. The use of *root frame variables*, N_{r}, in the *nullable fud*, $\mathrm{nullable}(U)(D)$, is necessary to ensure that the *derived variables* of the *root transform*, W_{r}, are indirectly represented in the *derived variables* of the *nullable transform*. This is because the *root transform derived variables*, W_{r}, are *underlying variables* of the *slice transforms*,

$$W_{\mathrm{r}} \subseteq \mathrm{und}(\{P_*^{\mathrm{T}} : P_* \in \mathrm{elements}(\mathrm{sliceTransforms}(U)(D))\})$$

and so are hidden in the *nullable fud*, $W_{\mathrm{r}} \cap \mathrm{der}(\mathrm{nullable}(U)(D)) = \emptyset$, and hence cannot be in the *derived variables* of the *nullable transform*, $W_{\mathrm{r}} \cap \mathrm{der}(D^{\mathrm{T}}) = \emptyset$. Instead, the *root frame variables* are in the *derived variables* of the *nullable transform*, $N_{\mathrm{r}} \subseteq \mathrm{der}(D^{\mathrm{T}})$.

If the *decomposition* consists solely of a *root transform* T_{r} where $\{((\emptyset, T_{\mathrm{r}}), \emptyset)\} = D$ then the *nullable transform* is a *value full functional transform* of the *root transform*. That is, the *nullable transform* is a *reframe transform*. In this case, the *partition* of the *nullable transform* equals the *partition* of T_{r}, $D^{\mathrm{TP}} = T_{\mathrm{r}}^{\mathrm{P}}$, and the *volume* of the *derived variables* of the *nullable transform* equals the *volume* of the *derived variables* of T_{r}, $|\mathrm{der}(D^{\mathrm{T}})^{\mathrm{C}}| = |\mathrm{der}(T_{\mathrm{r}})^{\mathrm{C}}|$.

If the *decomposition* contains *nullable variables*, $\mathrm{nullables}(U)(D) \neq \emptyset$, then the *volume* of the *derived variables* of the *nullable transform* is greater than the *volume* of the *derived variables* of the *transforms*, $|\mathrm{der}(D^{\mathrm{T}})^{\mathrm{C}}| > |\mathrm{der}(G)^{\mathrm{C}}|$. This is because the *nullable variables* have an additional *null value* with respect to their corresponding *originating variable*, $\exists (w, v) \in \mathrm{originals}(U)(D)\ (|U_w| = |U_v| + 1)$.

The application of the *transform* of a *well behaved distinct decomposition* D to the *cartesian* of the *underlying decomposition variables* $\mathrm{und}(D) = V \subset \mathcal{V}_U$, followed by the *reduction* to the *root frame variables*, form a *completely effective histogram* if the *root transform* is *non-overlapping*. Let $\{T_{\mathrm{r}}\} = \mathrm{ran}(\mathrm{roots}(D))$, $\neg\mathrm{overlap}(T_{\mathrm{r}})$ and $N_{\mathrm{r}} = \{\{w\}^{\mathrm{CS}\{\}} : w \in \mathrm{der}(T_{\mathrm{r}})\}$ where $D \in \mathcal{D}_{\mathrm{w},U}$ and $\mathrm{und}(D) = V$. Then $(V^{\mathrm{C}} * T_{\mathrm{r}})^{\mathrm{F}} = W_{\mathrm{r}}^{\mathrm{C}}$ where $W_{\mathrm{r}} = \mathrm{der}(T_{\mathrm{r}})$ and

$$(V^{\mathrm{C}} * D^{\mathrm{T}} \ \%\ N_{\mathrm{r}})^{\mathrm{F}} = N_{\mathrm{r}}^{\mathrm{C}}$$

If the *decomposition* contains *nullable variables* then the *decomposition transform* is *overlapping*,

$$\mathrm{nullables}(U)(D) \neq \emptyset \implies \mathrm{overlap}(D^{\mathrm{T}})$$

This is because the *nullable variables* depend via *alternate*, *contingent* and *transform slice variables* on ancestor *transform derived variables*, which themselves have dependent *variables* in the *nullable fud derived variables*. Let $F = \text{nullable}(U)(D)$, then $\forall n \in \text{dom}(\text{nullables}(U)(D))\ \exists u \in \text{der}(F)\ ((u \neq n) \wedge (\text{und}(\text{depends}(F, \{u\})) \cap \text{und}(\text{depends}(F, \{n\})) \neq \emptyset))$. The *transform* is necessarily *overlapping* whatever *symmetries* exist because of the additional *null value* of the *nullable variables*. The *null value* exists even if none of the *possible derived states* are *null*. In particular, all *nullable variables* depend on the *root transform derived variables*, $\forall n \in \text{dom}(\text{nullables}(U)(D))\ (\text{der}(T_{\text{r}}) \subset \text{und}(\text{depends}(F, \{n\})))$.

In the case where there are *nullable variables*, $\text{nullables}(U)(D) \neq \emptyset$, and the *transforms* are also all *mono-derived-variate transforms*, $\forall T \in G\ (|\text{der}(T)| = 1)$, the application to the *cartesian* is *skeletal*, $\text{skeleton}(V^{\text{C}} * D^{\text{T}})$, and any non-singleton subset of the *decomposition derived variables* is *contingent* and *overlapping*

$$\forall J \subseteq \text{der}(D^{\text{T}})\ (|J| > 1 \implies (V^{\text{C}} * D^{\text{T}}\ \%\ J)^{\text{F}} \neq J^{\text{C}})$$

The application of an *overlapping decomposition transform*, $\text{overlap}(D^{\text{T}})$, to the *cartesian* is *incompletely effective*, $(V^{\text{C}} * D^{\text{T}})^{\text{F}} < N^{\text{C}}$ where $N = \text{der}(D^{\text{T}})$. That is, the *possible derived volume* $w' = |(V^{\text{C}} * D^{\text{T}})^{\text{F}}| = |D^{\text{P}}| = |(D^{\text{T}})^{-1}|$ is less the *derived volume*, $w' < w$, where $w = |N^{\text{C}}|$. The *possible derived volume*, w', may be calculated from the *contingent possible derived volumes* of the *decomposition's transforms*,

$$\begin{aligned} w' &= |(V^{\text{C}} * T_{\text{r}})^{\text{F}}| + \sum \big(|(C * T)^{\text{F}}| - 1 : (C,T) \in \text{cont}(D),\ (C,T) \neq (V^{\text{C}}, T_{\text{r}})\big) \\ &= \sum \big(|(C * T)^{\text{F}}| : (C,T) \in \text{cont}(D)\big) + 1 - |\text{cont}(D)| \end{aligned}$$

where $\text{cont} = \text{elements} \circ \text{contingents}$. The *possible derived volume*, w', is bounded by the *possible derived volumes* of the individual *transforms*,

$$\begin{aligned} w' &\leq |T_{\text{r}}^{-1}| + \sum_{T \in G,\ T \neq T_{\text{r}}} \big(|T^{-1}| - 1\big) \\ &= \sum_{T \in G} \big(|T^{-1}|\big) + 1 - |G| \end{aligned}$$

where there are no *transform symmetries*, $|\text{nodes}(D)| = |G|$. In the case where all *transforms* are *non-overlapping*, $\forall T \in G\ (\neg\text{overlap}(T))$, the *possible* *derived volume* is bounded

$$\begin{aligned} w' &\leq |W_{\mathrm{r}}^{\mathrm{C}}| + \sum_{T \in G,\ T \neq T_{\mathrm{r}}} \big(|W_T^{\mathrm{C}}| - 1\big) \\ &= \sum_{T \in G} \big(|W_T^{\mathrm{C}}|\big) + 1 - |G| \end{aligned}$$

This calculation is only an upper bound for the *possible derived volume*, w', because it ignores *overlaps* between *transforms* in the same path, whether the *transforms* themselves are *overlapping* or not. Let $(C_1, T_1), (C_2, T_2) \in \mathrm{cont}(D)$ such that $T_1 \neq T_2$, $C_1 \supset C_2$, $\neg\mathrm{overlap}(T_1)$ and $\neg\mathrm{overlap}(T_2)$. Then $\mathrm{overlap}(\{T_1, T_2\}) \implies |(\{T_1, T_2\})^{-1}| < |T_1^{-1}||T_2^{-1}| = |\mathrm{der}(\{T_1, T_2\})^{\mathrm{C}}|$, and $\mathrm{overlap}(\{T_1, T_2\}) \implies |(C_1 * T_1)^{\mathrm{F}}| + |(C_2 * T_2)^{\mathrm{F}}| \leq |(C_1 * T_1)^{\mathrm{F}}| + |(C_1 * T_2)^{\mathrm{F}}|$. This is more obvious when viewed in terms of the accumulated *states* and *fuds* of subpaths of the *decomposition*. Let $L \in \mathrm{paths}(D)$, $L_2 \in \mathrm{subpaths}(L)$, $L_1 \in \mathrm{subpaths}(L_2)$, such that (i) $(F_1, R_1) = (\mathrm{trn}(L_1), \mathrm{st}(L_1))$ and $\mathrm{his}(F_1) * \{R_1\}^{\mathrm{U}} \ \%\ V = C_1$, and (ii) $(F_2, R_2) = (\mathrm{trn}(L_2), \mathrm{st}(L_2))$ and $\mathrm{his}(F_2) * \{R_2\}^{\mathrm{U}} \ \%\ V = C_2$. Then $\{T_1, T_2\} \subseteq F_2$ and $\mathrm{overlap}(F_2)$.

The *reduced* application of the *nullables* of a *non-root transform* T form a *pivot histogram*. Let $Y = \{P_* : P_* \in \mathrm{nullables}(U)(D),\ \mathrm{vars}(P_*) \cap \mathrm{der}(T) \neq \emptyset\}$ where $T \in G \setminus \{T_{\mathrm{r}}\}$ and $G = \mathrm{transforms}(D)$

$$\mathrm{pivot}(V^{\mathrm{C}} * D^{\mathrm{T}} \ \%\ \mathrm{dom}(Y))$$

The *nullable variables* are $\mathrm{dom}(Y) = \mathrm{ran}(\mathrm{filter}(\mathrm{der}(T), \mathrm{flip}(\mathrm{originals}(U)(D))))$. The *pivot state* is the *out-slice* or *null state*, $Y \in \mathcal{S}_U$.

A pair of *nullable variables* taken from different *non-root transforms* form an *axial histogram* when applied to the *cartesian* V^{C}. Let $P_{1*}, P_{2*} \in \mathrm{nullables}(U)(D)$ such that $\mathrm{originals}(U)(D)(P_1) \in \mathrm{der}(T_1)$, $\mathrm{originals}(U)(D)(P_2) \in \mathrm{der}(T_2)$, $T_1, T_2 \in G \setminus \{T_{\mathrm{r}}\}$ and $T_1 \neq T_2$, then

$$\mathrm{axial}(V^{\mathrm{C}} * D^{\mathrm{T}} \ \%\ \{P_1, P_2\})$$

If the *transforms* are on different paths, $M_{T_1} \neq M_{T_2}$ where $M = \{(T, \mathrm{first} \circ L) : L \in \mathrm{paths}(D^*),\ T \in \mathrm{dom}(\mathrm{set}(L))\}$, where $\mathrm{first}((x, \cdot)) = x$, then the *pivot* corresponds to the *null state*, $Y = \{P_{1*}, P_{2*}\} \in \mathcal{S}_U$, where both *nullable variables* are in their *out-slices*. If the *transforms* are the only children of the *root transform*, $M_{T_1}(2) = T_1$, $M_{T_2}(2) = T_2$ and $|\{L_{\{1\ldots 2\}} : L \in \mathrm{paths}(D^*)\}| = 2$, then the *pivot state* is zero, $(V^{\mathrm{C}} * D^{\mathrm{T}} * \{Y\}^{\mathrm{U}})^{\mathrm{F}} = \emptyset$, because the *out-slices* are exclusive. Consider the case where the *transforms* each appear once on the same path $L = M_{T_1} = M_{T_2}$. Let the ancestor *transform* be T_1 and the descendant *transform* be T_2, where $\mathrm{flip}(L)(T_1) < \mathrm{flip}(L)(T_2)$. The *pivot state* corresponds to an *in-slice non-null value* of the ancestor *variable* P_1 and an *out-slice null value* of the descendant *variable* P_2. That is, the *pivot state* is one of $\{\{(P_1, C), P_{2*}\} : C \in P_1,\ (P_1, C) \neq P_{1*}\} \subset \mathcal{S}_U$.

Thus the *reduced transformed cartesian* of a selection of the *derived variables* $K \subset N$ originating bijectively, $K \leftrightarrow G$, forms a *skeletal histogram*. Let $\mathrm{trans}(D) = \{(w, T) : T \in G,\ w \in \mathrm{der}(T)\}$, then $\mathrm{trans}(D) \circ \mathrm{originals}(U)(D) \in N \to G$ and

$$\begin{aligned}
&\forall Q \subseteq \mathrm{trans}(D) \circ \mathrm{originals}(U)(D) \\
&\quad (Q \in N \leftrightarrow G \implies \mathrm{skeletal}(V^{\mathrm{C}} * D^{\mathrm{T}} \ \%\ \mathrm{dom}(Q))
\end{aligned}$$

A *functional transform* $T \in \mathcal{T}_{\mathrm{f}}$ has a set of *reduced transforms* with respect to a *histogram* $A \in \mathcal{A}$, where $\mathrm{vars}(A) = \mathrm{underlying}(T)$. Section 'Transform, Action and Function histograms', above, defines the function $\mathrm{reductions} \in \mathcal{A} \times \mathcal{T}_{\mathrm{f}} \to \mathrm{P}(\mathcal{T}_{\mathrm{f}})$ such that $\forall T' \in \mathrm{reductions}(A, T)\ (\{\mathrm{trim}(A * C) : (\cdot, C) \in \mathrm{inverse}(T')\} = \{\mathrm{trim}(A * C) : (\cdot, C) \in \mathrm{inverse}(T)\})$. The *nullable transform* of a *decomposition* has *reductions*, $\mathrm{reductions}(A, D^{\mathrm{T}})$, but note that only the *nullable variables* of the *nullable transform* are *reduced*. The underlying *transforms*, $\mathrm{transforms}(D)$, are not *reduced*. However, a *decomposition* can also be *contingently reduced* to a set of *distinct decompositions*. Define $\mathrm{reductions} \in \mathcal{A} \times \mathcal{D} \to \mathrm{P}(\mathcal{D})$ as $\mathrm{reductions}(A, D) := \mathrm{distinct}(\mathrm{reductions}(A, \emptyset, \emptyset, D))$ where $\mathrm{reductions} \in \mathcal{A} \times \mathcal{F} \times \mathcal{S} \times \mathcal{D} \to \mathcal{D}$ is

$$\begin{aligned}
&\mathrm{reductions}(A, F, R, D) := \\
&\quad \{((Q, X), G) : ((S, T), E) \in D,\ Q = S \% \mathrm{der}(F), \\
&\quad\ X \in \mathrm{reductions}(A * \mathrm{inverse}(F^{\mathrm{T}})(R \cup Q), T), \\
&\quad\ G = \mathrm{reductions}(A, F \cup \{X\}, R \cup Q, E)\}
\end{aligned}$$

The intersection with the *nullable transform reductions* contains the *nullable transform*, $D^{\mathrm{T}} \in \{E^{\mathrm{T}} : E \in \mathrm{reductions}(A, D)\} \cap \mathrm{reductions}(A, D^{\mathrm{T}})$. The cardinality of the *nullable transform reductions* is greater than or equal to the cardinality of the *contingent reduction decomposition nullable transforms*, $|\mathrm{reductions}(A, D^{\mathrm{T}})| \geq |\{E^{\mathrm{T}} : E \in \mathrm{reductions}(A, D)\}|$.

If it is the case that all of the *transforms* of a *decomposition* are *contingently diagonalised* with respect to the *histogram*,

$$\forall B \in \mathrm{ran}(\mathrm{elements}(A * D))\ (\mathrm{diagonal}(B))$$

or

$$\forall (C,T) \in \text{elements}(\text{contingents}(D))\ (\text{diagonal}(A * C * T))$$

then there must exist a *contingent reduction decomposition nullable transform* that is *skeletal*,

$$\exists D' \in \text{reductions}(A,D)\ (\text{skeletal}(A * D^{'\text{T}}))$$

That is, the *reduction* to any two *derived variables* of the *derived histogram* of the *nullable transform* of the *contingent reduction*, D', is *linear* or *axial*

$$\forall P_1, P_2 \in \text{der}(D^{'\text{T}})\ (\text{line}(A * D^{'\text{T}}\ \%\ \{P_1,P_2\}) \lor \text{axial}(A * D^{'\text{T}}\ \%\ \{P_1,P_2\}))$$

Also, there must exist a *reduction nullable transform* that is *skeletal*,

$$\exists T \in \text{reductions}(A, D^{\text{T}})\ (\text{skeletal}(A * T))$$

A *skeletal contingent reduction decomposition*, D', of a *contingently diagonalised decomposition*, D, implies a tree of *derived variables* $Z = \text{map}(\text{vr}, D') \in \text{trees}(\text{der}(D^{'\text{T}}))$ where $\text{vr}((\cdot,T)) := x$ and $\{x\} = \text{der}(T)$. The *reduction* to any two of these *derived variables*, $B = A * D^{'\text{T}}\ \%\ \{P_1,P_2\}$ where $\{P_1,P_2\} \subseteq \text{elements}(Z)$, is *linear* or *axial*, $\text{line}(B) \lor \text{axial}(B)$. If *axial*, the *pivot state* is $X = \bigcup\{S \cap T : S,T \in B^{\text{FS}}\} \in \{P_1,P_2\}^{\text{CS}}$. If one of the *variables* is a descendant of the other, $\exists L \in \text{paths}(Z)\ (P_1 \neq P_2 \land \{P_1,P_2\} \subseteq \text{set}(L))$, then the descendant, P_2, has the *null value* in the *pivot state*, $P_{2*} \in X$, where $P_{2*} \in \text{nullables}(U)(D')$ and $\text{flip}(L)(P_2) > \text{flip}(L)(P_1)$. In this case a non-root ancestor, $P_1 \notin \text{roots}(Z)$, has a *non-null value*, $P_{1*} \notin X$, where $P_{1*} \in \text{nullables}(U)(D')$. The *pivot state*, X, is therefore not the ancestor *out-slice state*, $X \neq Y$, where the *out-slice* or *null state* is $Y = \{P_{1*}, P_{2*}\}$. If the *variables* are not ancestor-descendant, $P_1 \neq P_2 \land P_1 \in \text{set}(L) \land P_2 \in \text{set}(M) \implies L \neq M$ where $L,M \in \text{paths}(Z)$, then both are necessarily *null-valued* in the *pivot state*. That is, the *pivot state* is the *out-slice/null state*, $X = Y$. The *count* of the *pivot state*, B_X, depends on how closely related the pair of *variables* is. If P_1 is an immediate parent of P_2, then the *pivot state* is zero, $(P_1,P_2) \in \text{steps}(Z) \implies B_X = 0$. Note, however, that the *out-slice count*, B_Y, is not necessarily zero. If P_1 and P_2 are the only siblings of the root *variable* then the *pivot state*/*out-slice* is zero, $\{(\cdot,\{(P_1,\cdot),(P_2,\cdot)\})\} = Z \implies B_X = B_Y = 0$. If the pair are, for example, in different leaves, $P_1 \neq P_2 \land \{P_1,P_2\} \subseteq \text{leaves}(Z)$, then the *pivot/out-slice count*, $B_X = B_Y$, will vary as the depth of the tree, $\text{depth}(Z)$. The more distant the pair, the more the *axial* resembles a *singleton*.

Given a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ and a *histogram* $A \in \mathcal{A}$, where $\mathrm{vars}(A) = \mathrm{underlying}(T)$, a set of *well behaved decompositions* $H \subset \mathcal{D}_{\mathrm{w},U}$, where $\forall D \in H\ (\mathrm{und}(D) = \mathrm{vars}(A))$, may be inferred such that the *nullable transforms* of the *decompositions* correspond to the given *transform*. That is, the application of the *nullable transforms* of the *decompositions* to the given *histogram* are *non-literal reframes* of the application of the given *transform*, $\forall D \in H\ \exists X \in \mathrm{der}(T) \leftrightarrow (\mathrm{der}(D^{\mathrm{T}}) \times (\mathcal{W} \leftrightarrow \mathcal{W}))\ (A * D^{\mathrm{T}} = \mathrm{reframe}(X, A * T))$. The inferred *decompositions* contains at least a *decomposition* consisting of a *reframe transform* of the given *transform* as the *root transform*, $\{((\emptyset, \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in \mathrm{der}(T)\}^{\mathrm{T}}), \emptyset)\} \in H$.

Let the subset of the given *derived variables* $K \subset \mathrm{der}(T)$ correspond to the *non-root transforms* of an inferred *decomposition* $D \in H$. Let the *null state* be $R \in K^{\mathrm{CS}}$. Consider the case where a partition of the *null state* $M \in \mathrm{B}(R)$ exists such that the *reduction* of the application of the given *transform* to the *variables* of a component $P \in M \subset \mathcal{S}$ forms a *pivoted histogram*, $\forall P \in M\ (\mathrm{pivot}(A * T\%\mathrm{vars}(P)))$, such that the component, P, is the *pivot state*, $\forall P \in M\ \forall S \in (A * T\%\mathrm{vars}(P))^{\mathrm{FS}}\ (S \neq P \implies S \cap P = \emptyset)$. Then a *decomposition* $D \in H$ exists such that the *reframe* mapping $X \in \mathrm{der}(T) \leftrightarrow (\mathrm{der}(D^{\mathrm{T}}) \times (\mathcal{W} \leftrightarrow \mathcal{W}))$ is constrained such that the point *components* of the *nullable pointed partition variables* correspond to the *null values* in the *values* map, $\forall (v, ((Q, C), W)) \in X\ (v \in K \implies ((Q, C) \in \mathrm{nullables}(U)(D)) \wedge ((M_v, C) \in W))$.

In the case where two components of the *null state* partition $P_1, P_2 \in M$, where $P_1 \neq P_2$, are such that the *reduction* to the union of their *variables* is also a *pivoted histogram*, $\mathrm{pivot}(A * T\%(\mathrm{vars}(P_1) \cup \mathrm{vars}(P_2)))$, such that the *pivot state* is $P_1 \cup P_2$, then the corresponding *non-root transforms*, $T_1, T_2 \in \mathrm{transforms}(D)$, of an inferred *decomposition*, $D \in H$, must be on different paths. That is, there exists no path in which the *transforms* are in an ancestor-descendant relation, $\forall L \in \mathrm{paths}(D)\ (T_1 \notin \mathrm{ran}(\mathrm{set}(L)) \vee T_2 \notin \mathrm{ran}(\mathrm{set}(L)))$.

A non-empty *sub-decomposition* $E \in \mathcal{D}$ of non-empty *decomposition* $D \in \mathcal{D}$ is a subtree $E \in \mathrm{subtrees}(D) \setminus \{\emptyset\}$. The *underlying variables* of E are a subset of the *underlying variables* of D, $\mathrm{und}(E) \subseteq \mathrm{und}(D)$. If D is in *system* U, $D \in \mathcal{D}_U$, then E is in *system* U, $E \in \mathcal{D}_U$.

If D is *distinct*, $D \in \mathcal{D}_{\mathrm{d}}$, then E is *distinct*, $E \in \mathcal{D}_{\mathrm{d}}$. Both *decompositions* share the same *root transform*, $\mathrm{roots}(E) = \mathrm{roots}(D) = \{(\emptyset, T_{\mathrm{r}})\}$. The *expanded partition* of E is a *parent partition* of D, $\mathrm{parent}(E^{\mathrm{P}V}, D^{\mathrm{P}})$ where $V = \mathrm{und}(D)$. The *components* tree of E is a subtree of the *components* tree of D, $\mathrm{components}(U)(E) \in \mathrm{subtrees}(\mathrm{components}(U)(D))$. If there are no *symmetries* between the *transforms* of E, $\mathrm{transforms}(E)$, and the disjoint *transforms*, $\mathrm{transforms}(D) \setminus \mathrm{transforms}(E)$, then the *nullable fud* of E is a subset of the *nullable fud* of D, $\mathrm{nullable}(U)(E) \subseteq \mathrm{nullable}(U)(D)$. In this case, the application of the *decomposition* D to *histogram* $A \in \mathcal{A}$, having *variables* V, *reduces* to the application of the *decomposition* E, $A*D^{\mathrm{T}} \ \% \ \mathrm{der}(E^{\mathrm{T}}) = A*E^{\mathrm{T}}$.

Let *non-zero* sample *histogram* $A \in \mathcal{A}_U$ have non-empty *variables* $V = \mathrm{vars}(A) \neq \emptyset$. The normalisation is a *probability histogram*, $\hat{A} \in \mathcal{A} \cap \mathcal{P}$. Let *non-zero* query *histogram* $Q \in \mathcal{A}_U$ have *variables* $K = \mathrm{vars}(Q)$ that are a subset of the sample *variables*, $K \subseteq V$. The normalisation of the query *histogram* is a *probability histogram*, $\hat{Q} \in \mathcal{A} \cap \mathcal{P}$. The difference between the sample *variables* and the query *variables*, $V \setminus K$, is called the set of label *variables*. As discussed above in section 'Transforms', given a *one functional transform* $T = (M, W) \in \mathcal{T}_{U,\mathrm{f},1}$, having *underlying variables* $J = \mathrm{und}(T)$, the *model* analog of the *transformed conditional product*, $\hat{Q} * T'_A = \hat{Q} * (A/(A\%K), (V \setminus K)) \in \mathcal{A} \cap \mathcal{P}'$, is the renormalisation of the application of the normalised *sample action*, $(T, (\hat{A} * M, V))$, to the expanded query *probability histogram*, $\hat{Q}_J = \hat{Q} * (J \setminus K)^{\mathrm{C}\wedge} \in \mathcal{A} \cap \mathcal{P}$,

$$(\hat{Q}_J * T * (\hat{A} * M, V))^{\wedge} \ \% \ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}$$

or

$$\hat{Q}_J * T * T^{\odot A} \ \% \ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}$$

if the intersection of *derived effective states* is not empty, $(Q*T)^{\mathrm{F}} \cap (A*T)^{\mathrm{F}} \neq \emptyset$.

The *modelled transformed conditional product* can be computed for a *well behaved decomposition* $D \in \mathcal{D}_{\mathrm{w},U}$ by constructing the *nullable transform*, D^{T},

$$(\hat{Q}_J * D^{\mathrm{T}} * (\hat{A} * \mathrm{his}(D^{\mathrm{T}}), V))^{\wedge} \ \% \ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}$$

where his = histogram. However, in some cases the computation of the *nullable transform*, D^{T}, may be impracticable. In these cases the query function, $\mathrm{query} \in \mathcal{D} \times \mathcal{A} \times \mathcal{A} \to \mathrm{trees}(\mathcal{S} \times \mathcal{A})$, may sometimes be used instead. If (i) the set of query *variables* is a superset of the set of *decomposition underlying variables*, $K \supseteq J$, where $J = \mathrm{und}(D)$, (ii) the query *histogram* is an *effective singleton*, $|Q^{\mathrm{F}}| = 1$, and (iii) the intersection of *derived effective states* is not empty, $R \in (A * D^{\mathrm{T}})^{\mathrm{FS}}$, where $\{R\} = (Q * D^{\mathrm{T}})^{\mathrm{F}}$, then

$$(\hat{Q}_J * D^{\mathrm{T}} * (\hat{A} * \mathrm{his}(D^{\mathrm{T}}), V))^{\wedge} \ \% \ (V \setminus K) = \hat{N} \ \% \ (V \setminus K)$$

where

$$\{N\} = \text{leaves}(\text{query}(D, A, Q))$$

In this case of *effective singleton* query, $|Q^{\text{F}}| = 1$, the query tree has a single path, $|\text{paths}(\text{query}(D, A, Q))| = 1$, and hence a single leaf element, $|\text{leaves}(\text{query}(D, A, Q))| = 1$. If the intersection of *derived effective states* is not empty, $R \in (A * D^{\text{T}})^{\text{FS}}$, then the last *histogram* in the path is not *empty*, $\text{size}(L_{|L|}) > 0$, where $\{L\} = \text{paths}(\text{query}(D, A, Q))$, and so the leaf *histogram*, $N = L_{|L|}$, is not *empty*, $\text{size}(N) > 0$.

If, however, the intersection of *derived effective states* is empty, $R \notin (A * D^{\text{T}})^{\text{FS}}$, because of *over-fitting*, it may be that some ancestor *slice* is not *empty*. That is, sometimes there exists $(C_1, T_1), (C_2, T_2) \in \text{cont}(D)$ such that $C_1 \supset C_2$, $\text{size}(Q*C_2) > 0$, $\text{size}(A*C_2) = 0$, but $\text{size}(A*C_1) > 0$. The ancestor *slice* is the last of the *non-empty* filtered path. Let $L' = \text{filter}(\{(P, \text{size}(P) > 0) : P \in \text{set}(L)\}, L)$, then $N = L'_{|L'|}$, if $|L'| \neq 0$.

If the filtered path is empty, $|L'| = 0$, then the query is an *ineffective derived state* of the *root transform*, $R \notin (A * T_{\text{r}})^{\text{FS}}$, where $\{R\} = (Q * T_{\text{r}})^{\text{FS}}$ and $\{((\emptyset, T_{\text{r}}), \emptyset)\} = D$. In this case the best best guess is simply $\hat{A}$ % $(V \setminus K)$.

The set of *functional definition set decompositions* $\mathcal{D}_{\text{F}}$ is a subset of the trees of pairs of (i) *states*, $\mathcal{S}$, and (ii) *functional definition sets*, $\mathcal{F}$

$$\mathcal{D}_{\text{F}} \subset \text{trees}(\mathcal{S} \times \mathcal{F})$$

The set of *fud decompositions* is constrained such that the *derived variables* of the *transforms* in the *fuds* are each uniquely *defined*

$$\forall D \in \mathcal{D}_{\text{F}}\ (\bigcup\{\text{def}(F) : F \in \text{fuds}(D)\} \in \mathcal{V} \to \mathcal{T})$$

where def = definitions and $\text{fuds} \in \mathcal{D}_{\text{F}} \to \text{P}(\mathcal{F})$ is defined as $\text{fuds}(D) := \text{ran}(\text{elements}(D))$. In other words, if in *fud decomposition* D *fuds* $F, G \in \text{fuds}(D)$ share a *defined variable* $w \in \text{dom}(\text{def}(F)) \cap \text{dom}(\text{def}(G))$ then the *definition* is the same in each, $\text{def}(F)(w) = \text{def}(G)(w)$. This also allows the same *fud* to appear more than once in a *fud decomposition* tree. The union of the *fuds* is therefore a *fud*,

$$\forall D \in \mathcal{D}_{\text{F}}\ (\bigcup \text{ran}(\text{elements}(D)) \in \mathcal{F})$$

Define $\text{fud} \in \mathcal{D}_{\text{F}} \to \mathcal{F}$ as $\text{fud}(D) := \bigcup \text{fuds}(D)$. Define $\text{underlying} \in \mathcal{D}_{\text{F}} \to \text{P}(\mathcal{V})$ as $\text{underlying}(D) := \text{underlying}(\text{fud}(D))$.

There are some additional constraints on *fud decompositions*. First, no *underlying variable* of a *fud* in the *decomposition* can be a *defined variable* in another *fud*

$$\forall D \in \mathcal{D}_{\mathrm{F}}\ (\bigcup\{\mathrm{und}(F) : F \in \mathrm{fuds}(D)\} = \mathrm{underlying}(D))$$

where und = underlying. So only the *underlying variables* of a *fud* can intersect with the *underlying variables* of any other, $\forall F_1, F_2 \in \mathrm{fuds}(D)\ (\mathrm{vars}(F_1) \setminus \mathrm{und}(F_1) \cap \mathrm{und}(F_2) = \emptyset)$.

Second, the *states* of the root pairs are *empty states*

$$\forall D \in \mathcal{D}_{\mathrm{F}}\ (\mathrm{dom}(\mathrm{roots}(D)) = \{\emptyset\})$$

Third, each of the *states* in child pairs are *states* of the *derived variables* of the parent *fud*

$$\forall D \in \mathcal{D}_{\mathrm{F}}\ \forall ((\cdot, F), (S, \cdot)) \in \mathrm{steps}(D)\ (S \in \mathrm{std}(F^{\mathrm{T}}))$$

where std = stateDeriveds.

The *empty fud decomposition* consists of the *empty fud*, $\{((\emptyset, \emptyset), \emptyset)\} \in \mathcal{D}_{\mathrm{F}}$.

A *fud decomposition* $D \in \mathcal{D}_{\mathrm{F}}$ may be converted to a *transform decomposition* if the sets of *derived variables* of the *fuds* are disjoint, $(\forall F_1, F_2 \in \mathrm{fuds}(D)\ (F_1 \neq F_2 \implies \mathrm{der}(F_1) \cap \mathrm{der}(F_2) = \emptyset)) \implies \mathrm{map}(\mathrm{transform}, D) \in \mathcal{D}$, where $\mathrm{transform} \in (\mathcal{S} \times \mathcal{F}) \to (\mathcal{S} \times \mathcal{T}_{\mathrm{f,U}})$ is defined $\mathrm{transform}((S, F)) := (S, F^{\mathrm{T}})$. Define shorthand $D^{\mathrm{D}} := \mathrm{map}(\mathrm{transform}, D)$.

There is no constraint on converting a *transform decomposition* to a *fud decomposition*, $\forall D \in \mathcal{D}\ (\mathrm{map}(\mathrm{fud}, D) \in \mathcal{D}_{\mathrm{F}})$, where $\mathrm{fud} \in (\mathcal{S} \times \mathcal{T}_{\mathrm{f,U}}) \to (\mathcal{S} \times \mathcal{F})$ is defined $\mathrm{fud}((S, T)) := (S, \{T\})$.

The application of a *fud decomposition* to a *histogram* is a tree of contingent applications of the *fuds*. Define $\mathrm{apply} \in \mathcal{D}_{\mathrm{F}} \times \mathcal{A} \to \mathrm{trees}(\mathcal{S} \times \mathcal{A})$ as $\mathrm{apply}(D, A) := \mathrm{apply}(D, \mathrm{vars}(A), A)$ where $\mathrm{apply} \in \mathcal{D}_{\mathrm{F}} \times \mathrm{P}(\mathcal{V}) \times \mathcal{A} \to \mathrm{trees}(\mathcal{S} \times \mathcal{A})$ is

$$\begin{aligned}
&\mathrm{apply}(D, V, A) := \\
&\quad \{((S, B\%W), \mathrm{apply}(E, V, B)) : ((S, F), E) \in D,\ W = \mathrm{der}(F), \\
&\qquad\qquad B = \mathrm{apply}(V, V \cup W, \mathrm{his}(F) \cup \{\{S\}^{\mathrm{U}}\}, A)\}
\end{aligned}$$

where his = histograms. Define shorthand $A * D = \text{apply}(D, A)$. The application of a *fud decomposition* to a *histogram* equals the application of the corresponding *transform decomposition* to a *histogram*, $A * D = A * D^{\text{D}}$.

The subset of *fud decompositions* $\mathcal{D}_{\text{F},U} \subset \mathcal{D}_{\text{F}}$ in a *system* U is a subset of the trees of pairs of (i) *states*, $\mathcal{S}_U$, and (ii) *one functional definition sets*, $\mathcal{F}_{U,1}$

$$\mathcal{D}_{\text{F},U} = \mathcal{D}_{\text{F}} \cap \text{trees}(\mathcal{S}_U \times \mathcal{F}_{U,1})$$

The *fuds* of a *system fud decomposition* are *non-circular*, $\forall D \in \mathcal{D}_{\text{F},U}\ \forall F \in \text{fuds}(D)\ (\neg\text{circular}(F))$.

A further subset is the set of *system fud decompositions* where the *fuds* are *partition fuds*. Define the *partition fud decompositions*

$$\mathcal{D}_{\text{F},U,\text{P}} = \mathcal{D}_{\text{F}} \cap \text{trees}(\mathcal{S}_{U'} \times \mathcal{F}_{U,\text{P}})$$

where the finite *system* U' is defined $U' = \{(P, P) : F \in \mathcal{F}_{U,\text{P}},\ P \in \text{dom}(\text{def}(F))\} \cup U$.

A *contingent* tree of pairs of *components* and *fuds* can be mapped cumulatively from a *fud decomposition* $D \in \mathcal{D}_{\text{F}}$. Define contingent $\in \mathcal{L}(\mathcal{S} \times \mathcal{F}) \to \mathcal{A} \times \mathcal{F}$ as

$$\begin{aligned}&\text{contingent}(L) := (\text{inverse}(G^{\text{T}})(S), F) :\\&\qquad (\cdot, F) = L_{|L|},\ G = \bigcup \text{ran}(\text{set}(L_{\{1\ldots|L|-1\}})),\ S = \bigcup \text{dom}(\text{set}(L))\end{aligned}$$

$\text{contingent}(\{(1, (\cdot, F))\}) := (\text{scalar}(1), F)$. Define contingents $\in \mathcal{D}_{\text{F}} \to \text{trees}(\mathcal{A} \times \mathcal{F})$ as

$$\text{contingents}(D) := \text{mapAccum}(\text{contingent}, D)$$

The subset *distinct fud decompositions* $\mathcal{D}_{\text{F,d}} \subset \mathcal{D}_{\text{F}}$ is defined $\mathcal{D}_{\text{F,d}} = \mathcal{D}_{\text{F}} \cap ((\mathcal{S} \to \mathcal{F}) \to \mathcal{D}_{\text{F,d}})$. Define $\text{partition}(U) \in \mathcal{D}_{\text{F,d},U} \to \mathcal{R}_U$ as

$$\text{partition}(U)(D) := \text{partition}(U)(D^{\text{D}})$$

The *well behaved distinct fud decompositions* is a subset of the *distinct fud decompositions*, $\mathcal{D}_{\text{F,w},U} \subset \mathcal{D}_{\text{F,d},U}$. A *fud decomposition* is *well behaved* if its *transform decomposition* is *well behaved*, $\forall D \in \mathcal{D}_{\text{F,w},U}\ (D^{\text{D}} \in \mathcal{D}_{\text{w},U})$. Define $\text{transform}(U) \in \mathcal{D}_{\text{F,w},U} \to \mathcal{T}_{U,\text{f},1}$ as

$$\text{transform}(U)(D) := (G \cup H)^{\text{T}}$$

where $G = \text{fud}(D)$ and $H = \text{nullable}(U)(D^{\text{D}})$.

A *well behaved distinct fud decomposition* $D \in \mathcal{D}_{\mathrm{F,w},U}$ in *system* U contains a *variable symmetry* if $\exists (L,T),(M,R) \in Q\ ((L \neq M) \wedge (\mathrm{der}(T) \cap \mathrm{der}(R) \neq \emptyset))$ where $Q = \{(L,T) : (L,E) \in \mathrm{places}(D^{\mathrm{D}*}),\ T \in \mathrm{dom}(\mathrm{dom}(E))\} \in \mathcal{L}(\mathcal{T}_{\mathrm{f}} \times \mathcal{S}) \to \mathcal{T}_{\mathrm{f}}$ and $D^{\mathrm{D}*} = \mathrm{application}(U)(D^{\mathrm{D}})$. If all of the *derived variables* are equal the *variable symmetry* is also a *fud symmetry*. That is, more strictly, $\exists (L,T),(M,R) \in Q\ ((L \neq M) \wedge (T = R))$.

3.11 Substrate structures

Consider the *partition functional definition set* $F_{U,V} \in \mathcal{F}_{U,\mathrm{P}}$ of all possible *partition transforms* of the *cartesian* set of *states* of some *substrate* set of *underlying variables* V in *system* U,

$$F_{U,V} = \mathrm{transforms}(\mathrm{B}(\mathrm{cartesian}(U)(V)))$$

$F_{U,V}$ is called the *base functional definition set* of V in *system* U. It may be more concisely defined $F_{U,V} = \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\}$. The *base fud*, $F_{U,V}$, contains all its *partition variables* in one *layer*, $|\mathrm{der}(F_{U,V})| = |F_{U,V}|$. All of the *partition variables* are *expanded*, $\forall P \in \mathrm{der}(F_{U,V})\ (\mathrm{vars}(P) = V)$ and $\forall T \in F_{U,V}\ (\mathrm{und}(T) = V)$. $F_{U,V}$ has cardinality

$$|F_{U,V}| = |\mathrm{B}(\mathrm{cartesian}(U)(V))| = |\mathrm{B}(\{1 \ldots y\})| = \mathrm{bell}(y)$$

where the *volume* $y = |V^{\mathrm{C}}|$. The function $\mathrm{bell} \in \mathbf{N}_{>0} \to \mathbf{N}_{>0}$ is Bell's number which is factorial, $\mathrm{bell} \in \mathrm{O}(\{(n, n^n) : n \in \mathbf{R}_{>0}\})$. The cardinality of the *base functional definition set* is finite, though because it is of factorial complexity in the *volume*, it is impracticable in some cases. For example, if the *variables* V have the same *valency* d having a *regular volume* $y = d^n$, where *dimension* $n = |V|$, then for four *bi-valent variables* the Bell's number is $\mathrm{bell}(2^4) = 10480142147$. Similarly for two *quad-valent variables*, $\mathrm{bell}(4^2) = 10480142147$. For three *tri-valent variables*, $\mathrm{bell}(3^3) = 545717047936059989389$.

The *base functional definition set*, $F_{U,\emptyset}$, of *empty substrate*, $\emptyset$, is a singleton containing only the *unary scalar partition transform*, $F_{U,\emptyset} = \{P^{\mathrm{T}} : P \in \mathrm{B}(\emptyset^{\mathrm{CS}})\} = \{(\{\{(R, \{\emptyset\})\}\}^{\mathrm{U}}, \{R\})\}$ where $R = \mathrm{unary}(\emptyset^{\mathrm{CS}}) = \{\emptyset^{\mathrm{CS}}\} = \{\{\emptyset\}\}$.

The *base fud* contains the *unary partition transform*, $R^{\mathrm{T}} = (\{S \cup \{(R, V^{\mathrm{CS}})\} : S \in V^{\mathrm{CS}}\}^{\mathrm{U}}, \{R\}) \in F_{U,V} = \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\}$, where R is the *unary partition*, $R = \mathrm{unary}(V^{\mathrm{CS}}) = \{V^{\mathrm{CS}}\} \in \mathrm{B}(V^{\mathrm{CS}})$. The corresponding *contracted unary partition transform* is the *unary scalar partition transform*, $R^{\%\mathrm{T}} = (\{\{(R^{\%}, \{\emptyset\})\}\}^{\mathrm{U}}, \{R^{\%}\})$ where $R^{\%} = \{\emptyset^{\mathrm{CS}}\} = \{\{\emptyset\}\}$.

The *base fud* contains the *self partition transform*, $Q^{\mathrm{T}} = (\{S \cup \{(Q, \{S\})\} : S \in V^{\mathrm{CS}}\}^{\mathrm{U}}, \{Q\}) \in F_{U,V} = \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\}$, which corresponds to the *self partition* $Q = \mathrm{self}(V^{\mathrm{CS}}) = V^{\mathrm{CS}\{\}} = \{\{S\} : S \in V^{\mathrm{CS}}\} \in \mathrm{B}(V^{\mathrm{CS}})$.

Given a set of *substrate variables* $V \subset \mathcal{V}_U$ in *system* U, let the set $\mathcal{T}_{U,V} \subset \mathcal{T}_{U,\mathrm{P}*}$ be the set of *multi-partition transforms* that are the *equivalent transforms* of *fud* subsets of the *base functional definition set*, $F_{U,V}$, of *substrate* V

$$\mathcal{T}_{U,V} = \{\mathrm{transform}(F) : F \subseteq \mathrm{transforms}(\mathrm{B}(\mathrm{cartesian}(U)(V)))\}$$

or more succinctly $\mathcal{T}_{U,V} = \{F^{\mathrm{T}} : F \subseteq F_{U,V}\}$. This finite set is called the *substrate transforms set* on *variables* V. Constrain the *system* U such that it contains all of the *partition variables* in the *transforms*, $\bigcup\{\mathrm{der}(T) : T \in \mathcal{T}_{U,V}\} \subset \mathrm{vars}(U)$. $\mathcal{T}_{U,V}$ can also be rewritten in various ways

$$\begin{aligned}
\mathcal{T}_{U,V} &= \{F^{\mathrm{T}} : F \subseteq \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\}\} \\
&= \{\{P^{\mathrm{T}} : P \in X\}^{\mathrm{T}} : X \subseteq \mathrm{B}(V^{\mathrm{CS}})\} \\
&= \{\{\{A^{\mathrm{S}} : A \in M\}^{\mathrm{T}} : M \in Y\}^{\mathrm{T}} : Y \subseteq \mathrm{B}(V^{\mathrm{C}})\}
\end{aligned}$$

The *substrate transforms set* can also be defined in terms of *expanded partitions*, $\mathcal{T}_{U,V} = \{F^{\mathrm{T}} : F \subseteq \{P^{V\mathrm{T}} : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}})\}\}$.

Let y be the *volume* of the *substrate*, $y = |V^{\mathrm{C}}|$. The cardinality of the *substrate transforms set* is $|\mathcal{T}_{U,V}| = 2^{\mathrm{bell}(y)}$. A large practicable *volume* is the *bi-valent bi-variate* case $y = 2^2$ where $|\mathcal{T}_{U,V}| = 2^{15} = 32768$. The next *volume* is the *5-valent mono-variate* case, $y = 5^1$, where $|\mathcal{T}_{U,V}| = 2^{52} = 4503599627370496$.

The *substrate transforms set* contains the *empty transform*, $(\emptyset, \emptyset) \in \mathcal{T}_{U,V}$. The *empty transform* is the *equivalent transform* of the *empty fud*, $(\emptyset, \emptyset) = \emptyset^{\mathrm{T}}$. It is the only element of the *substrate transforms set* which has no *derived variables*, $|\mathrm{der}((\emptyset, \emptyset))| = 0$. It is the only element which has *underlying variables* not equal to V, $\mathrm{und}((\emptyset, \emptyset)) = \emptyset \neq V$.

The *substrate transforms set* contains the *unary partition transform*, $\{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$, and the *self partition transform*, $V^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U,V}$.

The *substrate transforms set* contains the *value full functional transform*, $\{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$. This *transform* is the only *value full functional transform* of the *substrate variables* in the *substrate transforms set.*

The *derived states*, $(X\%W)^{\mathrm{S}} = \{\{v\}^{\mathrm{CS}\{\}V} : v \in V\}^{\mathrm{CS}}$, are *reframed underlying states*, $(X\%V)^{\mathrm{S}} = V^{\mathrm{CS}}$, where $(X, W) = \{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}}$. That is, $\exists Q \in V \cdot W\ \forall (v, w) \in Q\ (\mathrm{split}(\{v\}, (X\%\{v, w\})^{\mathrm{S}}) \in \{v\}^{\mathrm{CS}} \leftrightarrow \{w\}^{\mathrm{CS}})$.

The *substrate transforms set* contains the *base fud transform*, $F^{\mathrm{T}}_{U,V} \in \mathcal{T}_{U,V}$.

The *substrate transforms set* is a superset of the *base fud*, $F_{U,V} \subset \mathcal{T}_{U,V}$. The subset of *substrate transforms*, having non-empty *substrate variables*, $V \neq \emptyset$, that are *mono-variate* in the *derived variables* equals the *base fud*, $\{T : T \in \mathcal{T}_{U,V},\ |\mathrm{der}(T)| = 1\} = F_{U,V}$. The *substrate transforms set* contains the *unary partition transform*, $R^{\mathrm{T}} \in \mathcal{T}_{U,V}$ where $R = \{V^{\mathrm{CS}}\}$. The subset of the *substrate transforms set* which contains the *unary partition*, $\{T : T \in \mathcal{T}_{U,V},\ \{V^{\mathrm{CS}}\} \in \mathrm{der}(T)\}$, has a complement of the same cardinality, $\{T : T \in \mathcal{T}_{U,V},\ \{V^{\mathrm{CS}}\} \in \mathrm{der}(T)\} \leftrightarrow \{T : T \in \mathcal{T}_{U,V},\ \{V^{\mathrm{CS}}\} \notin \mathrm{der}(T)\}$. The *empty transform*, $(\emptyset, \emptyset)$, complements the *unary partition transform*, R^{T}.

A *substrate partition-set* $N \in \mathrm{P}(\mathcal{R}_U)$ in *variables* V is constrained to be such that each of the *partitions* in the *partition-set* has *variables* which are a subset of V, $\forall P \in N\ (\mathrm{vars}(P) \subseteq V)$. Define the *substrate partition-sets set*

$$\mathcal{N}_{U,V} = \mathrm{P}(\{P : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}})\})$$

The cardinality of the *substrate partition-sets set* is

$$|\mathcal{N}_{U,V}| = 2^c : c = \sum_{K \subseteq V} \mathrm{bell}(|K^{\mathrm{CS}}|)$$

This is bounded

$$2^{\mathrm{bell}(y)} \leq |\mathcal{N}_{U,V}| \leq 2^{2^n \mathrm{bell}(y)}$$

where $y = |V^{\mathrm{CS}}|$. In the case of *regular variables* V, having *valency* $\{d\} = \{|U_w| : w \in V\}$ and *dimension* $n = |V|$, the cardinality is

$$|\mathcal{N}_{U,V}| = 2^c : c = \sum_{k \in \{0 \ldots n\}} \binom{n}{k} \mathrm{bell}(d^k)$$

A *substrate partition-set* $N \in \mathcal{N}_{U,V}$ maps to a *multi-partition transform*, $\{P^{\mathrm{T}} : P \in N\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{P}*}$. A *substrate partition-set* $N \in \mathcal{N}_{U,V}$ maps to a *substrate transforms set* by *expanding* the *partitions* to V, $\{P^{V\mathrm{T}} : P \in N\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$. The *empty partition-set*, $\emptyset \in \mathcal{N}_{U,V}$, *expands* to the *empty transform*, $(\emptyset, \emptyset) \in \mathcal{T}_{U,V}$. A converse function maps a *substrate transform* $T \in \mathcal{T}_{U,V}$

to a *contracted partition-set*, $\{P^{\%} : P \in \mathrm{der}(T)\} \in \mathcal{N}_{U,V}$. The *substrate transforms set* can therefore be defined in terms of *substrate partition-sets*,

$$\mathcal{T}_{U,V} = \{\{P^{V\mathrm{T}} : P \in N\}^{\mathrm{T}} : N \in \mathcal{N}_{U,V}\} = \{\{P^{\mathrm{T}} : P \in N\}^{\mathrm{T}V} : N \in \mathcal{N}_{U,V}\}$$

Define transform $\in \bigcup\{\mathrm{P}(\mathcal{R}_U) \to \mathcal{T}_{U,\mathrm{P}^*} : U \in \mathcal{U}\}$ as

$$\mathrm{transform}(N) := \{P^{\mathrm{T}} : P \in N\}^{\mathrm{T}}$$

and shorthand $N^{\mathrm{T}} := \mathrm{transform}(N)$. Then the *substrate transforms set* is defined $\mathcal{T}_{U,V} = \{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V}\}$.

Define the converse function that constructs a *substrate partition-set* given a *one functional transform*, partitionset $\in \bigcup\{\mathcal{T}_{U,\mathrm{f},1} \to \mathrm{P}(\mathcal{R}_U) : U \in \mathcal{U}\}$

$$\mathrm{partitionset}(T) := \{(X\%(V \cup \{w\}), \{w\})^{\mathrm{P}} : w \in W\}$$

where $W = \mathrm{der}(T)$, $V = \mathrm{und}(T)$ and $X = \mathrm{his}(T)$.

Similarly a *substrate pointed partition-set* $N \in \mathrm{P}(\mathcal{R}_{*,U})$ in *variables* V is constrained to be such that each of the *pointed partitions* has *variables* which are a subset of V, $\forall P_* \in N\ (\mathrm{vars}(P_*) \subseteq V)$. A *substrate pointed partition-set* is also constrained such that the *partitions* are unique, $|\mathrm{dom}(N)| = |N|$. Define the *substrate pointed partition-sets set*

$$\mathcal{N}_{*,U,V} = \mathrm{P}(\{(P, C_*) : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}}),\ C_* \in P\}) \cap (\mathcal{R}_U \to \mathrm{P}(\mathcal{S}_U))$$

A *substrate pointed partition-set* $N \in \mathcal{N}_{*,U,V}$ maps to a *substrate transform* by *expanding* the *partitions* to V, $\{P^{V\mathrm{T}} : (P, \cdot) \in N\} \in \mathcal{T}_{U,V}$. The *point component* is forgotten. This is no converse function to map a *substrate transform* $T \in \mathcal{T}_{U,V}$ to a *substrate pointed partition-set* except in the case where the *contractions* are *unary partitions*, $\forall P \in \mathrm{der}(T)\ (|P^{\%}| = 1)$. In this case the *substrate pointed partition-set* is $\{(P^{\%}, C_*) : P \in \mathrm{der}(T),\ \{C_*\} = P^{\%}\} \in \mathcal{N}_{U,V}$.

Consider the set of all the *flattened expanded partition transforms* generated from a *one functional definition set*. Define flatten $\in \mathcal{F} \to \mathcal{F}$

$$\mathrm{flatten}(F) := \{(X\%(V \cup W), W)^{\mathrm{PT}} : T \in F,\ W = \mathrm{der}(T)\} : V = \mathrm{und}(F),\ X = \prod \mathrm{his}(F)$$

where his = histograms, und = underlying and der = derived. If $F \in \mathcal{F}_{U,1}$ then flatten(F) is a subset of the *base fud*, $\mathrm{flatten}(F) \subseteq F_{U,V} \in \mathcal{F}_{U,\mathrm{P}}$. All of the *partition transforms* are *expanded*, $\forall T \in \text{flatten}(F)\ (T^{\text{P}} = T^{\text{P}V})$, where $V = \text{und}(F)$. The *flatten* function uses a similar method to that used to create the *equivalent transform* by taking the product, X, of all the *histograms* in the *fud*. The *flattened fud* can also be defined in terms of the *equivalent transform* of the *depends functional definition set* where the *system* U is given. The resultant *partitions* are *expanded* to the *substrate* of *underlying variables* of the *fud*

$$\begin{aligned}\text{flatten}(F) &= \{G^{\text{TP}V\text{T}} : T \in F,\ G = \text{depends}(F, \text{der}(T))\} \\ &= \{G^{\text{TP}V\text{T}} : G \subseteq F,\ \text{und}(G) \subseteq V\}\end{aligned}$$

where $V = \text{und}(F)$. The *flattened fud* is not equal to the given *fud* unless the *fud* is a subset of the *base fud*, $F \cap F_{U,V} \neq F \implies \text{flatten}(F) \neq F$.

Another special case of a *partition functional definition set*, $F \in \mathcal{F}_{U,\text{P}}$, that consists only of *partition transforms* is the *power functional definition set*. This set is constructed from a *substrate* V in a *system* U by *partitioning* the *cartesian* of all subsets of V, and recursing on the union of the set of newly *derived variables* and the *underlying variables*. The *power functional definition set* excludes *partition circularities* because each of its *transforms* is constrained such that none is equivalent to an element of the *flattened fud* of its underlying *transforms* in the *depends functional definition set*. Define $\text{power}(U) \in \text{P}(\mathcal{V}_U) \to \mathcal{F}_{U,\text{P}}$

$$\text{power}(U)(V) := \text{power}(U)(V, \emptyset)$$

Define $\text{power}(U) \in \text{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\text{P}} \to \mathcal{F}_{U,\text{P}}$

$$\begin{aligned}&\text{power}(U)(V, F) := \\ &\quad \text{if}(G = \emptyset,\ F,\ \text{power}(U)(V, F \cup G)) : \\ &\quad\ G = \{T : K \subseteq \text{vars}(F) \cup V,\ H = \text{depends}(F, K), \\ &\quad\quad T \in F_{U,K} \setminus F,\ (H \cup \{T\})^{\text{TPT}} \notin \text{flatten}(H)\}\end{aligned}$$

The *power fud* contains as subsets all possible non-circular *layered partition fuds* on the given *substrate* V or subset of the *substrate*. Let $G_{U,V} = \text{power}(U)(V)$. The *power fud* is a superset of the *power fuds* of subsets of the *substrate*, $\forall K \subseteq V\ (\text{power}(U)(K) \subseteq G_{U,V})$. The *power fud* is a superset of the *base fud*, $F_{U,V} \subset G_{U,V}$. Also the *base fud* is equal to *flattened fud* of the *power fud*, $F_{U,V} = \text{flatten}(G_{U,V}) = \{T : T \in G_{U,V},\ \text{und}(T) = V\} \subset G_{U,V}$. The *power fud* of a non-empty set of *substrate variables*, $V \neq \emptyset$, excludes *partition circularities* and so has a finite depth or number of *layers*,

$\text{layer}(G_{U,V}, \text{der}(G_{U,V})) = |\text{B}(V^{\text{CS}})| + 1 = \text{bell}(y) + 1$ where *substrate volume* $y = |V^{\text{CS}}|$. The depth of the *power fud* is therefore greater than the cardinality of its *base fud* and *flattened fud*. The *base fud* is a subset of the *first layer* subset, $F_{U,V} = \{P^{\text{T}} : P \in \text{B}(V^{\text{CS}})\} \subset \{T : T \in G_{U,V},\ \text{und}(T) \subseteq V\} \subset G_{U,V}$, where the *first layer* subset is $\{T : T \in G_{U,V},\ \text{und}(T) \subseteq V\} = \{M : M \subset G_{U,V},\ \text{und}(M) \subseteq V,\ \text{layer}(M, \text{der}(M)) = 1\} = \{P^{\text{T}} : K \subseteq V,\ P \in \text{B}(K^{\text{CS}})\}$. The cardinality of the *first layer* subset is the width of the bottom *layer*, therefore the cardinality of the *power fud* of a non-empty set of *substrate variables* is greater than twice the cardinality of non-powerset *base fud*, $|G_{U,V}| > 2\ \text{bell}(y)$. Thus the *power fud* is less practicable than the *base fud*. See appendix 'Cardinality of the power functional definition set'.

Given a set of *substrate variables* $V \subset \mathcal{V}_U$ in *system* U, let the set $\mathcal{F}_{U,V} \subset \mathcal{F}_{U,\text{P}}$ be the set of all subsets of the *power fud* having *underlying variables* equal to or a subset of V

$$\mathcal{F}_{U,V} = \{F : F \subseteq \text{power}(U)(V),\ \text{und}(F) \subseteq V\}$$

This finite set is called the *substrate fuds set* on *variables* V. Constrain the *system* U such that it contains all of the *partition variables* in the *transforms* of the *fuds*, $\bigcup\{\text{der}(T) : F \in \mathcal{F}_{U,V},\ T \in F\} \subset \text{vars}(U)$. The *substrate fuds set* can also be defined in terms of the *variables* in the *power fud*, $\mathcal{F}_{U,V} = \{\text{depends}(G_{U,V}, X) : X \subseteq \text{vars}(G_{U,V})\}$ where $G_{U,V} = \text{power}(U)(V)$. The *substrate fuds set* can also be defined in terms of the *transforms* in the *power fud*, $\mathcal{F}_{U,V} = \{\bigcup\{\text{depends}(G_{U,V}, \text{der}(T)) : T \in F\} : F \subseteq G_{U,V}\}$. Thus the cardinality of the *substrate fuds set* is such that $|\mathcal{F}_{U,V}| \leq 2^{|G_{U,V}|}$. The *substrate fuds set* contains the *empty fud*, $\emptyset \in \mathcal{F}_{U,V}$. The *substrate fuds set* is a superset of the powerset of the *base fud*, $\text{P}(F_{U,V}) \subset \mathcal{F}_{U,V}$.

The *consistent one functional definition sets* $\mathcal{F}_{U,1,\text{x}} \subset \mathcal{F}_{U,1}$ is the subset of *one functional definition sets* that do not contain circularities or contradictions or duplicates. All *consistent one functional definition sets*, $F \in \mathcal{F}_{U,1,\text{x}}$ having *underlying variables* V, $\text{und}(F) = V$, have an equivalence class defined by a member of the *substrate fuds set*, $E \in \mathcal{F}_{U,V}$. That is, *fud* F and *substrate fud* E form a pair in a surjective functional map Z, $(F, E) \in Z \in \mathcal{F}_{U,1,\text{x},V} \to \mathcal{F}_{U,V}$ where $\mathcal{F}_{U,1,\text{x},V} = \{F : F \in \mathcal{F}_{U,1,\text{x}},\ \text{und}(F) = V\}$. Each F implies $G = \text{mono}(F) \in \mathcal{F}_{U,1}$ which is the *one functional definition set* that is *monovariate* in its *derived variables* in each *transform*, $\forall T \in G\ (|\text{der}(G)| = 1)$, such that $G^{\text{T}} = F^{\text{T}}$. There exists exactly one *non-literal frame variables* mapping X in *system* U such that $\exists X \subset \{(v, (w, W)) : Q \in (\text{vars}(F) \setminus V) \cdot (\text{vars}(E) \setminus V),\ (v, w) \in Q,\ W \in U_v \cdot U_w\}\ (X \in \mathcal{V}_U \leftrightarrow (\mathcal{V}_U \times (\mathcal{W}_U \leftrightarrow \mathcal{W}_U)))$

such that $\text{reframe}(X, G)$ is defined and $\text{reframe}(X, G) = E$. The existence of a mapping between the *transforms*, $G \leftrightarrow E$, implies that the cardinalities are equal, $|G| = |E|$. The cardinalities of the *underlying variables* are equal, $\exists M \in G \leftrightarrow E\ \forall (T, T') \in M\ (|\text{und}(T)| = |\text{und}(T')|)$. The number of *layers* are equal $\text{layer}(F, \text{der}(F)) = \text{layer}(E, \text{der}(E))$. Thus the possibly infinite set $\mathcal{F}_{U,1,\text{x}}$ is partitioned into sets of *fuds*, $\mathcal{F}_{U,1,\text{x},V}$, having the same *underlying variables*, V, $\bigcup\{\mathcal{F}_{U,1,\text{x},V} : F \in \mathcal{F}_{U,1,\text{x}},\ V = \text{und}(F)\} = \mathcal{F}_{U,1,\text{x}}$. Then these possibly infinite sets are partitioned into equivalence classes by finite *substrate fuds sets*, $\mathcal{F}_{U,1,\text{x},V} \to \mathcal{F}_{U,V}$.

The *substrate transforms set*, $\mathcal{T}_{U,V}$, and *substrate fuds set*, $\mathcal{F}_{U,V}$, in *substrate variables* V in *system* U are closely related. Each *substrate transform* $T \in \mathcal{T}_{U,V}$ can be *exploded* to a *substrate fud*, $\text{explode}(T) = \{P^{\text{T}} : P \in \text{der}(T)\} = \{(\text{his}(T)\%(V \cup \{P\}), \{P\}) : P \in \text{der}(T)\} \in \mathcal{F}_{U,V}$. A similar *explode* exists for *contracted partitions*, $\text{explode}(T^{\%}) = \{P^{\%\text{T}} : P \in \text{der}(T)\} \in \mathcal{F}_{U,V}$. The subset of *substrate transforms* that are *mono-variate* in the *derived variables* form singleton *substrate fuds*, $\{\{T\} : T \in \mathcal{T}_{U,V},\ |\text{der}(T)| = 1\} \subset \mathcal{F}_{U,V}$. The *substrate transforms set* is equal to the set of *expanded equivalent transforms* of the *substrate fud set*

$$\mathcal{T}_{U,V} \quad = \quad \{\{G^{\text{TPVT}} : w \in \text{der}(F),\ G = \text{depends}(F, \{w\})\}^{\text{T}} : F \in \mathcal{F}_{U,V}\}$$

Thus there exists a functional surjection $\mathcal{F}_{U,V} \to \mathcal{T}_{U,V}$. Each *partition* $P \in \text{B}(V^{\text{CS}})$ forms an equivalence class for *non-empty substrate transforms*. Each *substrate transform* $T \in \mathcal{T}_{U,V}$ forms an equivalence class for a *substrate fud* $E \in \mathcal{F}_{U,V}$ which in turn forms an equivalence class for a *consistent one fud* $F \in \mathcal{F}_{U,1,\text{x}}$ where $\text{und}(F) = V$

$$\mathcal{F}_{U,1,\text{x},V} \to \mathcal{F}_{U,V} \to \mathcal{T}_{U,V} \to (F_{U,V} \cup \{(\emptyset, \emptyset)\}) \leftrightarrow (\text{B}(V^{\text{CS}}) \cup \{\emptyset\})$$

The *empty consistent fud*, $\emptyset \in \mathcal{F}_{U,1,\text{x},V}$, maps to the *empty fud*, $\emptyset \in \mathcal{F}_{U,V}$, thence to the *empty transform*, $(\emptyset, \emptyset) \in \mathcal{T}_{U,V}$, and thence to the *empty partition* $\emptyset \in (\text{B}(V^{\text{CS}}) \cup \{\emptyset\})$.

The *substrate decompositions set* $\mathcal{D}_{U,V} \subset \text{trees}(\mathcal{S} \times \mathcal{T})$ is the subset of *distinct decompositions*, $\mathcal{D}_{\text{d}}$, such that the *transforms* are in the *substrate transforms set*, $\mathcal{T}_{U,V}$, and no *transform* appears more than once in any path

$$\begin{aligned} \mathcal{D}_{U,V} = \{D : D \in \mathcal{D}_{\text{d}},\ \text{transforms}(D) \subseteq \mathcal{T}_{U,V}, \\ \forall L \in \text{paths}(D)\ (\text{maxr}(\text{count}(\{(T, i) : (i, (\cdot, T)) \in L\})) = 1)\} \end{aligned}$$

The *substrate decompositions set*, $\mathcal{D}_{U,V}$, is finite because $\mathcal{T}_{U,V}$ is finite. The depth of the tree of a *substrate decomposition* $D \in \mathcal{D}_{U,V}$ is less than or equal to the cardinality of the *substrate transforms set*, $\text{depth}(D) \leq |\mathcal{T}_{U,V}|$. Note that if the constraint on the paths were relaxed to be $\text{maxr}(\text{count}(\text{flip}(L))) = 1$, the maximum depth would be much greater but the depth of the *decomposition application tree*, D^*, would still be limited, $\text{depth}(D^*) \leq |\mathcal{T}_{U,V}|$.

The *substrate decompositions set*, $\mathcal{D}_{U,V}$, can be created explicitly from the *power decomposition*. Define $\text{power}(U) \in \text{P}(\mathcal{V}_U) \to \text{trees}(\mathcal{S}_U \times \mathcal{T}_U)$ as

$$\text{power}(U)(V) := \{((\emptyset, T), \text{power}(U)(V, T, \{T\})) : T \in \mathcal{T}_{U,V}\}$$

where $\text{power}(U) \in \text{P}(\mathcal{V}_U) \times \mathcal{T}_{U,\text{f},1} \times \mathcal{F}_{U,1} \to \text{trees}(\mathcal{S}_U \times \mathcal{T}_U)$ is defined as

$$\begin{aligned}
&\text{power}(U)(V, T, F) := \\
&\quad \{((S, R), \text{power}(U)(V, R, F \cup \{R\})) : S \in \text{std}(T),\ R \in \mathcal{T}_{U,V},\ R \notin F\} \cup \\
&\quad \{((S, R), \emptyset) : S \in \text{std}(T),\ R \in \mathcal{T}_{U,V},\ R \notin F\}
\end{aligned}$$

where std = stateDeriveds. Then the *substrate decompositions set* is the set of *distinct decompositions* in the *power decompositions* such that the *transforms* form a *fud*,

$$\begin{aligned}
&\mathcal{D}_{U,V} = \{D : D \in \text{distinct}(\text{power}(U)(V)), \\
&\quad \forall (A, W), (B, X) \in \text{transforms}(D)\ ((A, W) \neq (B, X) \implies W \cap X = \emptyset)\}
\end{aligned}$$

The *substrate decompositions set*, $\mathcal{D}_{U,V}$, maps to $\mathcal{T}_{U,V}$ in several ways. First, a *substrate decomposition* $D \in \mathcal{D}_{U,V}$ has a *partition*, $D^{\text{P}} \in \text{B}(V^{\text{CS}})$, which is already *expanded* to V, $D^{\text{PT}} \in \mathcal{T}_{U,V}$. Second, if the *substrate decomposition* is *well behaved*, $D \in \mathcal{D}_{\text{w},U}$, the *crown transform*, $\text{transformCrown}(U)(D)$, constructed from the *transforms fud*, $\text{transforms}(D)$ and the *slices fud*, $\text{slices}(U)(D)$, can be *expanded*, $\{(X\%(V \cup \{w\}), \{w\})^{\text{PVT}} : w \in W\}^{\text{T}} \in \mathcal{T}_{U,V}$ where $(X, W) = \text{transformCrown}(U)(D)$. Third, if the *substrate decomposition* is *well behaved*, $D \in \mathcal{D}_{\text{w},U}$, the *nullable transform*, D^{T}, constructed from the *transforms fud*, $\text{transforms}(D)$ and the *nullable fud*, $\text{nullable}(U)(D)$, can also be *expanded*, $\{(X\%(V \cup \{w\}), \{w\})^{\text{PVT}} : w \in W\}^{\text{T}} \in \mathcal{T}_{U,V}$ where $(X, W) = D^{\text{T}}$.

The *substrate functional definition set decompositions* $\mathcal{D}_{\text{F},U,V} \subset \text{trees}(\mathcal{S} \times \mathcal{F})$ is a subset of the *distinct fud decompositions*, $\mathcal{D}_{\text{F,d}}$, and can be defined similarly to *substrate transform decompositions*, $\mathcal{D}_{U,V} \subset \text{trees}(\mathcal{S} \times \mathcal{T})$. That is, all of the *fuds* are *substrate fuds* and none can appear more than once in a path

$$\begin{aligned}
&\mathcal{D}_{\text{F},U,V} = \{D : D \in \mathcal{D}_{\text{F,d}},\ \text{fuds}(D) \subseteq \mathcal{F}_{U,V}, \\
&\quad \forall L \in \text{paths}(D)\ (\text{maxr}(\text{count}(\{(F, i) : (i, (\cdot, F)) \in L\})) = 1)\}
\end{aligned}$$

The set of *substrate fud decompositions*, $\mathcal{D}_{\mathrm{F},U,V}$, is also finite. The *substrate fuds* are a subset of the *partition fuds*, $\mathcal{F}_{U,V} \subset \mathcal{F}_{U,\mathrm{P}}$, so the *substrate fud decompositions* are a subset of the *partition fud decompositions*, $\mathcal{D}_{\mathrm{F},U,V} \subset \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$. The depth of the tree of a *substrate fud decomposition* $D \in \mathcal{D}_{\mathrm{F},U,V}$ is less than or equal to the cardinality of the *substrate fuds set*, $\mathrm{depth}(D) \leq |\mathcal{F}_{U,V}|$. The accumulated *fud* along any path $L \in \mathrm{paths}(D)$ is a subset of the *power fud*, $\bigcup \mathrm{ran}(\mathrm{set}(L)) \subseteq \mathrm{power}(U)(V) \in \mathcal{F}_{U,\mathrm{P}}$. Note that the map of a *substrate fud decomposition* $D \in \mathcal{D}_{\mathrm{F},U,V}$ to a *transform decomposition*, $D^{\mathrm{D}} \in \mathcal{D}_U$, is not generally a *substrate transform decomposition*, $D^{\mathrm{D}} \notin \mathcal{D}_{U,V}$, unless it so happens that the *fuds* are singletons of *substrate transforms*, $\mathrm{fuds}(D) \subset \{\{T\} : T \in \mathcal{T}_{U,V}\}$.

The *substrate fud decompositions set*, $\mathcal{D}_{\mathrm{F},U,V}$, maps to the *substrate transforms set* $\mathcal{T}_{U,V}$. First, a *substrate fud decomposition* $D \in \mathcal{D}_{\mathrm{F},U,V}$ has a *partition*, D^{DP}, which requires *expanding* to V, $D^{\mathrm{DPVT}} \in \mathcal{T}_{U,V}$. Second, the *nullable transform*, D^{T}, constructed from the union of the *fuds*, $\bigcup \mathrm{fuds}(D)$, and the *nullable fud*, $\mathrm{nullable}(U)(D^{\mathrm{D}})$, can be *expanded*, $\{(X\%(V \cup \{w\}), \{w\})^{\mathrm{PVT}} : w \in W\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$ where $(X, W) = D^{\mathrm{T}}$.

Given *substrate variables* V, the *non-overlapping* subset of the *substrate transforms set* $\mathcal{T}_{U,V,\mathrm{n}} \subset \mathcal{T}_{U,V}$ is defined

$$\mathcal{T}_{U,V,\mathrm{n}} \quad = \quad \{T : T \in \mathcal{T}_{U,V},\ \neg\mathrm{overlap}(T)\}$$

The *non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n}}$, can also be defined in terms of *substrate partition-sets*. Let the *non-overlapping substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{n}} \subset \mathcal{N}_{U,V}$ be defined in terms of the strong partition of the *substrate variables*

$$\begin{aligned}&\mathcal{N}_{U,V,\mathrm{n}} = \\ &\qquad \{N : N \in \mathcal{N}_{U,V},\ M = \{\mathrm{vars}(P) : P \in N\},\ |M| = |N|,\ M \in \mathrm{B}(V)\}\end{aligned}$$

The *non-overlapping substrate partition-sets set* is empty if the *substrate variables* is empty, $\mathcal{N}_{U,\emptyset,\mathrm{n}} = \emptyset$. The *non-overlapping substrate partition-sets set* excludes the *empty partition-set*, $\emptyset \notin \mathcal{N}_{U,V,\mathrm{n}}$.

Let the *weak non-overlapping substrate partition-sets set* $\mathcal{N}'_{U,V,\mathrm{n}} \subset \mathcal{N}_{U,V}$ be defined in terms of the weak partition of the *substrate variables* including the *empty partition-set*

$$\begin{aligned}&\mathcal{N}'_{U,V,\mathrm{n}} = \\ &\qquad \{N : N \in \mathcal{N}_{U,V},\ M = \{\mathrm{vars}(P) : P \in N\},\ |M| = |N|,\ M \in \mathrm{B}'(V)\} \cup \{\emptyset\}\end{aligned}$$

where the weak partition function is $\mathrm{B}'(V) := \mathrm{B}(V) \cup \{Y \cup \{\emptyset\} : Y \in \mathrm{B}(V)\}$ and $\mathrm{B}'(\emptyset) := \{\{\emptyset\}\}$. The *weak non-overlapping substrate partition-sets set* includes the *empty partition-set*, $\emptyset \in \mathcal{N}'_{U,V,\mathrm{n}}$. If the *substrate variables* set is empty, the *weak non-overlapping substrate partition-sets set* is a set of the *empty partition-set* and a singleton of the *contracted unary partition*, $\mathcal{N}'_{U,\emptyset,\mathrm{n}} = \{\emptyset, \{\{\emptyset^{\mathrm{CS}}\}\}\}$, where the *unary partition* is $\{\emptyset^{\mathrm{CS}}\}^V = \{V^{\mathrm{CS}}\}$. The *non-overlapping substrate partition-sets set* is a subset of the *weak non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n}} \subset \mathcal{N}'_{U,V,\mathrm{n}}$,

$$V \neq \emptyset \implies \mathcal{N}'_{U,V,\mathrm{n}} = \mathcal{N}_{U,V,\mathrm{n}} \cup \{N \cup \{\{\emptyset^{\mathrm{CS}}\}\} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} \cup \{\emptyset\}$$

The *non-overlapping substrate transforms set* is defined in terms of the *weak non-overlapping substrate partition-sets set*

$$\mathcal{T}_{U,V,\mathrm{n}} = \{\{P^{V\mathrm{T}} : P \in N\}^{\mathrm{T}} : N \in \mathcal{N}'_{U,V,\mathrm{n}}\} = \{N^{\mathrm{T}V} : N \in \mathcal{N}'_{U,V,\mathrm{n}}\}$$

where $P^V := \mathrm{expand}(U)(V,P)$. The *non-overlapping substrate transforms set* is therefore such that

$$\forall T \in \mathcal{T}_{U,V,\mathrm{n}}\ \forall P_1, P_2 \in \mathrm{der}(T^{\%})\ (P_1 \neq P_2 \implies \mathrm{vars}(P_1) \cap \mathrm{vars}(P_2) = \emptyset)$$

The *non-overlapping substrate partition-sets set* can be defined explicitly

$$\mathcal{N}_{U,V,\mathrm{n}} = \{N : Y \in \mathrm{B}(V),\ N \in \prod_{K \in Y} \mathrm{B}(K^{\mathrm{CS}})\}$$

and, similarly, the *weak non-overlapping substrate partition-sets set* can be defined explicitly

$$\mathcal{N}'_{U,V,\mathrm{n}} = \{N : Y \in \mathrm{B}'(V),\ N \in \prod_{K \in Y} \mathrm{B}(K^{\mathrm{CS}})\} \cup \{\emptyset\}$$

and so the *non-overlapping substrate transforms set* can be defined explicitly

$$\mathcal{T}_{U,V,\mathrm{n}} = \{N^{\mathrm{T}V} : Y \in \mathrm{B}'(V),\ N \in \prod_{K \in Y} \mathrm{B}(K^{\mathrm{CS}})\} \cup \{(\emptyset, \emptyset)\}$$

The *non-overlapping substrate transforms set* includes the *empty transform*, $(\emptyset, \emptyset) \in \mathcal{T}_{U,V,\mathrm{n}}$, the *unary partition transform*, $\{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{n}}$, the *self partition transform*, $V^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{n}}$, and the *value full functional transform*, $\{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{n}}$. The *base functional definition set* is a subset, $F_{U,V} \subseteq \mathcal{T}_{U,V,\mathrm{n}}$.

The *strong non-overlapping substrate transforms set* is the set of *transforms* of the *non-overlapping substrate partition-sets set*, $\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\}$. The *non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n}}$, is constructed from strong partitions of the *substrate variables*, $\mathrm{B}(V)$. The *strong non-overlapping substrate transforms set* is a subset of the *non-overlapping substrate transforms set*,

$$\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} \subseteq \mathcal{T}_{U,V,\mathrm{n}}$$

In the case of non-empty *substrate variables*, $V \neq \emptyset$, the cardinality of the *non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n}}$, is

$$|\mathcal{N}_{U,V,\mathrm{n}}| = \sum_{Y \in \mathrm{B}(V)} \prod_{K \in Y} |\mathrm{B}(K^{\mathrm{CS}})|$$

The cardinality of the *strong non-overlapping substrate transforms set* is therefore bounded

$$|\mathrm{B}(V^{\mathrm{CS}})| \leq |\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\}| \leq \sum_{Y \in \mathrm{B}(V)} \prod_{K \in Y} |\mathrm{B}(K^{\mathrm{CS}})|$$

If the *underlying variables* are *regular*, having *dimension* $n = |V|$ and common *valency* d, $\{d\} = \{|U_x| : x \in V\}$, then the cardinality of the *non-overlapping substrate partition-sets set* is

$$|\mathcal{N}_{U,V,\mathrm{n}}| = \sum_{Y \in \mathrm{B}(V)} \prod_{K \in Y} \mathrm{bell}(d^{|K|}) = \sum_{(L,c) \in \mathrm{bcd}(n)} \left(c \prod_{(k,m) \in L} \mathrm{bell}(d^k)^m \right)$$

where $\mathrm{bcd} = \mathrm{bellcd}$ and the partition function cardinality function $\mathrm{bellcd} \in \mathbf{N}_{>0} \to (\mathcal{L}(\mathbf{N}) \to \mathbf{N})$, defined in appendix 'Partitions', below, computes the histogram of the histograms of the component cardinalities.

The cardinality of the *strong non-overlapping substrate transforms set* is bounded

$$\mathrm{bell}(d^n) \leq |\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\}| \leq \mathrm{bell}(n) \times \mathrm{bell}(d^n)$$

Generalising to *irregular*,

$$\mathrm{bell}(y) \leq |\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\}| \leq \mathrm{bell}(n) \times \mathrm{bell}(y)$$

where $y = \mathrm{volume}(U)(V)$.

The cardinality of the *weak non-overlapping substrate partition-sets set* is twice that of the *non-overlapping substrate partition-sets set* plus one, $|\mathcal{N}'_{U,V,\mathrm{n}}| = 2 \times |\mathcal{N}_{U,V,\mathrm{n}}| + 1$. The cardinality of the *non-overlapping substrate transforms set* is therefore bounded

$$2 \times |\mathrm{B}(V^{\mathrm{CS}})| \leq |\mathcal{T}_{U,V,\mathrm{n}}| \leq 2 \times \sum_{Y \in \mathrm{B}(V)} \prod_{K \in Y} |\mathrm{B}(K^{\mathrm{CS}})| + 1$$

If the *underlying variables* are *regular*, the cardinality of the *non-overlapping substrate transforms set* is bounded $2 \times \mathrm{bell}(d^n) \leq |\mathcal{T}_{U,V,\mathrm{n}}| \leq 2 \times \mathrm{bell}(n) \times \mathrm{bell}(d^n) + 1$. Generalising to *irregular*,

$$2 \times \mathrm{bell}(y) \leq |\mathcal{T}_{U,V,\mathrm{n}}| \leq 2 \times \mathrm{bell}(n) \times \mathrm{bell}(y) + 1$$

where $y = \mathrm{volume}(U)(V)$. Compare the cardinality of the *substrate transforms set* itself,

$$|\mathcal{T}_{U,V}| = 2^{\mathrm{bell}(y)}$$

The cardinality of the *non-overlapping* subset, $|\mathcal{T}_{U,V,\mathrm{n}}|$, may be compared to the cardinality of the subset of the *substrate transforms set* which limits the cardinality of the *derived variables* to the *dimension*, the *fixed dimension substrate transforms set*,

$$|\{T : T \in \mathcal{T}_{U,V},\ |\mathrm{der}(T)| \leq n\}| = \sum_{k \in \{0 \ldots n\}} \frac{\mathrm{bell}(y)!}{k!\ (\mathrm{bell}(y) - k)!} > \frac{1}{n!}(\mathrm{bell}(y))^{\underline{n}}$$

where $x^{\underline{n}}$ is the falling factorial.

A practicable *volume* of *regular variables* V of the *non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n}}$, is $y = 8$, for example, the *bi-valent tri-variate* case $y = 2^3$.

Consider the *binary non-overlapping substrate transforms set* $\mathcal{T}_{U,V,\mathrm{n,b}}$ which is a subset of the *non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,b}} \subseteq \mathcal{T}_{U,V,\mathrm{n}}$. The *binary non-overlapping substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{n,b}} \subset \mathcal{N}_{U,V,\mathrm{n}}$ constrains the partition of the *substrate variables* to have a cardinality of two

$$\mathcal{N}_{U,V,\mathrm{n,b}} = \{N : N \in \mathcal{N}_{U,V,\mathrm{n}},\ |N| = 2\}$$

The *binary non-overlapping substrate transforms set* is defined in terms of the *binary non-overlapping substrate partition-sets set*

$$\mathcal{T}_{U,V,\mathrm{n,b}} = \{N^{\mathrm{TV}} : N \in \mathcal{N}_{U,V,\mathrm{n,b}}\}$$

The *binary non-overlapping substrate transforms set* is such that

$$\forall T \in \mathcal{T}_{U,V,\mathrm{n,b}}\ (T \neq \{V^{\mathrm{CS}}\}^{\mathrm{T}} \implies |\mathrm{der}(T)| = 2)$$

In the case of empty or singleton *substrate variables* the *binary non-overlapping substrate transforms set* is empty, $|V| < 2 \implies \mathcal{T}_{U,V,\mathrm{n,b}} = \emptyset$. In the case of *pluri-variate substrate* the *binary non-overlapping substrate transforms set* includes the *unary partition transform*, $|V| \geq 2 \implies \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{n,b}}$.

The *binary non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n,b}}$, can be defined explicitly

$$\begin{aligned}\mathcal{N}_{U,V,\mathrm{n,b}} = \{\{P,Q\} :\ & K \subset V,\ K \neq \emptyset,\ K \neq V, \\ & P \in \mathrm{B}(K^{\mathrm{CS}}),\ Q \in \mathrm{B}((V \setminus K)^{\mathrm{CS}})\}\end{aligned}$$

The cardinality is

$$|\mathcal{N}_{U,V,\mathrm{n,b}}| = 1/2 \times \sum_{K \in \mathrm{P}(V) \setminus \{\emptyset,V\}} |\mathrm{B}(K^{\mathrm{CS}})| \times |\mathrm{B}((V \setminus K)^{\mathrm{CS}})|$$

In the case of *regular variables* of *valency* d and *dimension* n, the cardinality is

$$|\mathcal{N}_{U,V,\mathrm{n,b}}| = 1/2 \times \sum_{k \in \{1 \ldots n-1\}} \binom{n}{k} \mathrm{bell}(d^k) \times \mathrm{bell}(d^{n-k})$$

The *binary non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,b}}$, can be defined explicitly

$$\begin{aligned}\mathcal{T}_{U,V,\mathrm{n,b}} = \{\{P^{V\mathrm{T}}, Q^{V\mathrm{T}}\}^{\mathrm{T}} :\ & K \subset V,\ K \neq \emptyset,\ K \neq V, \\ & P \in \mathrm{B}(K^{\mathrm{CS}}),\ Q \in \mathrm{B}((V \setminus K)^{\mathrm{CS}})\}\end{aligned}$$

The cardinality of $\mathcal{T}_{U,V,\mathrm{n,b}}$ is less than $2^{n-2} \times \mathrm{bell}(y)$ where *volume* $y = |V^{\mathrm{C}}|$.

Consider the *self non-overlapping substrate transforms set* $\mathcal{T}_{U,V,\mathrm{n,s}}$ which is a subset of the *non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}} \subseteq \mathcal{T}_{U,V,\mathrm{n}}$. The *self non-overlapping substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{n,s}} \subset \mathcal{N}_{U,V,\mathrm{n}}$ constrains the partition of the *substrate variables* to have a cardinality equal to that of the *substrate variables*

$$\mathcal{N}_{U,V,\mathrm{n,s}} = \{N : N \in \mathcal{N}_{U,V,\mathrm{n}},\ |N| = |V|\}$$

The *self non-overlapping substrate transforms set* is defined in terms of the *self non-overlapping substrate partition-sets set*

$$\mathcal{T}_{U,V,\mathrm{n,s}} = \{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n,s}}\}$$

The *self non-overlapping substrate transforms set* is such that

$$\forall T \in \mathcal{T}_{U,V,\mathrm{n,s}}\ \forall P \in \mathrm{der}(T^{\%})\ (|\mathrm{vars}(P)| \leq 1)$$

In the case of empty *substrate variables* the *self non-overlapping substrate transforms set* is empty, $\mathcal{T}_{U,\emptyset,\mathrm{n,s}} = \emptyset$. In the case of *multi-variate substrate* the *self non-overlapping substrate transforms set* includes the *unary partition transform*, $V \neq \emptyset \implies \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{n,s}}$, and the *value full functional transform*, $V \neq \emptyset \implies \{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{n,s}}$.

The *self non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n,s}}$, can be defined explicitly

$$\mathcal{N}_{U,V,\mathrm{n,s}} = \prod_{w \in V} \mathrm{B}(\{w\}^{\mathrm{CS}})$$

The cardinality is

$$|\mathcal{N}_{U,V,\mathrm{n,s}}| = \prod_{w \in V} |\mathrm{B}(\{w\}^{\mathrm{CS}})|$$

If V is *regular* having *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$, the cardinality is

$$|\mathcal{N}_{U,V,\mathrm{n,s}}| = \mathrm{bell}(d)^n$$

The *self non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}}$, can be defined explicitly

$$\mathcal{T}_{U,V,\mathrm{n,s}} = \{N^{\mathrm{T}V} : N \in \prod_{w \in V} \mathrm{B}(\{w\}^{\mathrm{CS}})\}$$

If V is *regular* the cardinality of $\mathcal{T}_{U,V,\mathrm{n,s}}$ is less than $\mathrm{bell}(d)^n$.

In the case of *pluri-variate substrate* the intersection of the *binary non-overlapping substrate transforms set* and the *self non-overlapping substrate transforms set* is non-empty, including at least the *unary partition transform*, $|V| \geq 2 \implies \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{n,b}} \cap \mathcal{T}_{U,V,\mathrm{n,s}}$. The *binary non-overlapping substrate transforms set* equals the *self non-overlapping substrate transforms set* if the *substrate* is *bi-variate*, $|V| = 2 \implies \mathcal{T}_{U,V,\mathrm{n,b}} = \mathcal{T}_{U,V,\mathrm{n,s}}$.

Having considered the *self non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}}$, now consider the less constrained *self overlapping substrate transforms set* $\mathcal{T}_{U,V,\mathrm{o,s}} \subseteq \mathcal{T}_{U,V}$. Here the *substrate* is still self-partitioned, $V^{\{\}}$, but more than one *derived partition variable* for each *substrate variable* is allowed. The *self overlapping substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{o,s}} \subset \mathcal{N}_{U,V}$ is explicitly defined

$$\mathcal{N}_{U,V,\mathrm{o,s}} = \{\bigcup H : H \in \prod_{w \in V} \mathrm{P}(\mathrm{B}(\{w\}^{\mathrm{CS}}))\}$$

The *self overlapping substrate transforms set* is defined in terms of the *self overlapping substrate partition-sets set*

$$\mathcal{T}_{U,V,\mathrm{o,s}} = \{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{o,s}}\}$$

The *self overlapping substrate transforms set* is such that

$$\forall T \in \mathcal{T}_{U,V,\mathrm{o,s}}\ \forall P \in \mathrm{der}(T^{\%})\ (|\mathrm{vars}(P)| \leq 1)$$

The *self non-overlapping substrate transforms set* is a subset of the *self overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}} \subseteq \mathcal{T}_{U,V,\mathrm{o,s}}$. In the case of empty *substrate variables* the *self overlapping substrate transforms set* is empty, $\mathcal{T}_{U,\emptyset,\mathrm{o,s}} = \emptyset$. In the case of *multi-variate substrate* the *self overlapping substrate transforms set* includes the *empty transform*, $V \neq \emptyset \implies (\emptyset, \emptyset) \in \mathcal{T}_{U,V,\mathrm{o,s}}$, the *unary partition transform*, $V \neq \emptyset \implies \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{o,s}}$, and the *value full functional transform*, $V \neq \emptyset \implies \{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{o,s}}$.

The *self overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{o,s}}$, has cardinality

$$|\mathcal{N}_{U,V,\mathrm{o,s}}| = \prod(2^m : w \in V,\ m = |\mathrm{B}(\{w\}^{\mathrm{CS}})|)$$

If V is *regular*, having *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$, the cardinality of $\mathcal{T}_{U,V,\mathrm{o,s}}$ equals $2^{n \times \mathrm{bell}(d)}$.

In the case of non-empty *substrate variables*, $V \neq \emptyset$, the subset of the *non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n}}$, which are unary partitions of the *substrate*, $\mathrm{unary}(V) = \{V\}$, is simply the *base fud*, $|V| \geq 1 \implies \{T : T \in \mathcal{T}_{U,V,\mathrm{n}},\ |\mathrm{der}(T)| = 1\} = F_{U,V}$, which has cardinality $\mathrm{bell}(y)$ where $y = |V^{\mathrm{CS}}|$. In the case of *pluri-variate substrate* the intersection of the *unary substrate transforms set*, the *binary non-overlapping substrate transforms set* and the *self non-overlapping substrate transforms set* is non-empty, including at least the *unary partition transform*, $|V| \geq 2 \implies \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \bigcap\{F_{U,V}, \mathcal{T}_{U,V,\mathrm{n,b}}, \mathcal{T}_{U,V,\mathrm{n,s}}\}$.

In contrast to subsets of the *substrate transforms set*, $\mathcal{T}_{U,V}$, that partition the *substrate variables* in various ways, consider subsets that impose constraints on the partitions of the *cartesian states* of subsets of the *substrate*. Consider the *substrate self-cartesian transforms set* $\mathcal{T}_{U,V,\mathrm{c}}$ which is a subset of the *substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{c}} \subseteq \mathcal{T}_{U,V}$. Let the *substrate self-cartesian partition-sets set* $\mathcal{N}_{U,V,\mathrm{c}} \subset \mathcal{N}_{U,V}$ be defined such that the *partition sets* consist only of *self partitions* of the *cartesian states* of subsets of the *substrate*, $\mathrm{self}(K^{\mathrm{CS}})^V$ where $K \subseteq V$

$$\begin{aligned}\mathcal{N}_{U,V,\mathrm{c}} &= \{N : N \in \mathcal{N}_{U,V},\ \forall P \in N\ (P = (\bigcup P)^{\{\}})\} \\ &= \{N : N \in \mathcal{N}_{U,V},\ \forall P \in N\ (|P| = |\bigcup P|)\}\end{aligned}$$

using the shorthand $X^{\{\}} = \mathrm{self}(X)$. The *substrate self-cartesian partition-sets set* includes the *empty partition-set*, $\emptyset \in \mathcal{N}_{U,V,\mathrm{c}}$. If the *substrate variables* is empty, the *substrate self-cartesian partition-sets set* is a set of the *empty partition-set* and a singleton of the *contracted unary partition*, $\mathcal{N}_{U,\emptyset,\mathrm{c}} = \mathcal{N}'_{U,\emptyset,\mathrm{n}} = \{\emptyset, \{\{\emptyset^{\mathrm{CS}}\}\}\}$.

The *substrate self-cartesian transforms set* is defined in terms of *substrate self-cartesian partition-sets set*

$$\mathcal{T}_{U,V,\mathrm{c}} = \{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{c}}\}$$

The *substrate self-cartesian transforms set* is such that

$$\forall T \in \mathcal{T}_{U,V,\mathrm{c}}\ \forall P \in \mathrm{der}(T^{\%})\ \lozenge K = \mathrm{vars}(P)\ (P = K^{\mathrm{CS}\{\}}))$$

The *substrate self-cartesian partition-sets set*, $\mathcal{N}_{U,V,\mathrm{c}}$, can be defined explicitly

$$\mathcal{N}_{U,V,\mathrm{c}} = \{\{K^{\mathrm{CS}\{\}} : K \in X\} : X \subseteq \mathrm{P}(V)\}$$

The *substrate self-cartesian transforms set*, $\mathcal{T}_{U,V,\mathrm{c}}$, can be defined explicitly

$$\mathcal{T}_{U,V,\mathrm{c}} = \{\{P^{V\mathrm{T}} : K \in X,\ P = K^{\mathrm{CS}\{\}}\}^{\mathrm{T}} : X \subseteq \mathrm{P}(V)\}$$

The *substrate self-cartesian transforms set* includes the *empty transform*, $(\emptyset, \emptyset) \in \mathcal{T}_{U,V,\mathrm{c}}$, the *unary partition transform*, $\{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{c}}$, the *self partition transform*, $V^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{c}}$, and the *value full functional transform*, $\{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{c}}$.

The cardinality of the *substrate self-cartesian partition-sets set* is $|\mathcal{N}_{U,V,\mathrm{c}}| = 2^{2^n}$ where *dimension* $n = |V|$. Therefore the cardinality of the *substrate self-cartesian transforms set* is bounded $|\mathcal{T}_{U,V,\mathrm{c}}| \leq 2^{2^n}$.

The intersection between the *substrate self-cartesian partition-sets set*, $\mathcal{N}_{U,V,\mathrm{c}}$, and the *non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n}}$, is

$$\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n}} = \{\{K^{\mathrm{CS}\{\}} : K \in Y\} : Y \in \mathrm{B}(V)\}$$

In the case of non-empty *substrate variables*, the cardinality of the intersection is $V \neq \emptyset \implies |\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n}}| = \mathrm{bell}(n)$ where *dimension* $n = |V|$.

The intersection between the *substrate self-cartesian partition-sets set*, $\mathcal{N}_{U,V,\mathrm{c}}$, and the *weak non-overlapping substrate partition-sets set*, $\mathcal{N}'_{U,V,\mathrm{n}}$, is

$$\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}'_{U,V,\mathrm{n}} = \{\{K^{\mathrm{CS}\{\}} : K \in Y\} : Y \in \mathrm{B}'(V)\} \cup \{\emptyset\}$$

In the case of non-empty *substrate variables*, the cardinality of the intersection is $V \neq \emptyset \implies |\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}'_{U,V,\mathrm{n}}| = 2 \times \mathrm{bell}(n) + 1$ where *dimension* $n = |V|$. The cardinality of the *substrate transforms* of the intersection between the *substrate self-cartesian partition-sets set*, $\mathcal{N}_{U,V,\mathrm{c}}$, and the *weak non-overlapping substrate partition-sets set*, $\mathcal{N}'_{U,V,\mathrm{n}}$, is bounded

$$|\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}'_{U,V,\mathrm{n}}\}| \leq 2 \times \mathrm{bell}(n) + 1$$

The *transforms* of the intersection are bijective between the *underlying states* and *derived states*, $\mathrm{split}(V, X^{\mathrm{S}}) \in V^{\mathrm{CS}} \leftrightarrow W^{\mathrm{CS}}$, where $T \in \{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}'_{U,V,\mathrm{n}}\}$ and $(X, W) = T$. The *substrate transforms* of the intersection form a subset of the intersection between the *substrate self-cartesian transforms set*, $\mathcal{T}_{U,V,\mathrm{c}}$, and the *non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n}}$,

$$\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}'_{U,V,\mathrm{n}}\} \subseteq \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n}}$$

The *unary transform* is a member of the intersection between the *substrate self-cartesian transforms set* and the *non-overlapping substrate transforms set*, $\{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n}}$, but it is only a member of the *substrate transforms* of the intersection between the *substrate self-cartesian partition-sets set* and the *weak non-overlapping substrate partition-sets set* if the *substrate* consists only of *mono-valent variables*, $\exists v \in V\ (|U_v| > 1)) \implies \{V^{\mathrm{CS}}\}^{\mathrm{T}} \notin \{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}'_{U,V,\mathrm{n}}\}$.

The intersection between the *substrate self-cartesian partition-sets set*, $\mathcal{N}_{U,V,\mathrm{c}}$, and the *binary non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n,b}}$, is

$$\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,b}} = \{\{K^{\mathrm{CS}\{\}}, (V \setminus K)^{\mathrm{CS}\{\}}\} : K \subset V,\ K \neq \emptyset,\ K \neq V\}$$

In the case of *pluri-variate substrate variables*, the cardinality of the intersection is $|\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,b}}| = 2^{n-1} - 1$ where *dimension* $n = |V|$. The cardinality of the *substrate transforms* of the intersection between the *substrate self-cartesian partition-sets set*, $\mathcal{N}_{U,V,\mathrm{c}}$, and the *binary substrate partition-sets set*, $\mathcal{N}'_{U,V,\mathrm{n,b}}$, is bounded

$$|\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,b}}\}| \leq 2^{n-1} - 1$$

The *substrate transforms* of the intersection form a subset of the intersection between the *substrate self-cartesian transforms set*, $\mathcal{T}_{U,V,\mathrm{c}}$, and the *binary non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,b}}$,

$$\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,b}}\} \subseteq \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n,b}}$$

The intersection between the *substrate self-cartesian transforms set*, $\mathcal{N}_{U,V,\mathrm{c}}$, and the *self non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n,s}}$, is a singleton

$$\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,s}} = \{\{\{v\}^{\mathrm{CS}\{\}} : v \in V\}\}$$

The corresponding *substrate transform* is the *value full functional transform*, $\{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}}$. The *value full functional transform* is a member of the intersection between the *substrate self-cartesian transforms set*, $\mathcal{T}_{U,V,\mathrm{c}}$, and the *self non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}}$,

$$\{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n,s}}$$

Consider the *substrate binary-cartesian partition transforms set* $\mathcal{T}_{U,V,2}$ which is a subset of the *substrate transforms set*, $\mathcal{T}_{U,V,2} \subseteq \mathcal{T}_{U,V}$. Let the *substrate binary-cartesian partition-sets set* $\mathcal{N}_{U,V,2} \subset \mathcal{N}_{U,V}$ be defined such that the *partition sets* consist only of *binary partitions* of the *cartesian states* of subsets of the *substrate*, $|P^V| = |P| = 2$ where $P \in \mathrm{B}(K^{\mathrm{CS}})$ and $K \subseteq V$,

$$\mathcal{N}_{U,V,2} = \{N : N \in \mathcal{N}_{U,V},\ \forall P \in N\ (|P| = 2)\}$$

The *substrate binary-cartesian partition-sets set* includes the *empty partition-set*, $\emptyset \in \mathcal{N}_{U,V,2}$.

The *substrate binary-cartesian transforms set* is defined in terms of *substrate binary-cartesian partition-sets set*

$$\mathcal{T}_{U,V,2} = \{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,2}\}$$

The *substrate binary-cartesian transforms set* is such that

$$\forall T \in \mathcal{T}_{U,V,2}\ \forall P \in \mathrm{der}(T)\ (|P| = 2)$$

The *substrate binary-cartesian partition-sets set*, $\mathcal{N}_{U,V,2}$, can be defined explicitly

$$\begin{aligned}\mathcal{N}_{U,V,2} &= \mathrm{P}(\{P : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}}),\ |P| = 2\}) \\ &= \mathrm{P}(\{\{C, K^{\mathrm{CS}} \setminus C\} : K \subseteq V,\ C \in \mathrm{P}(K^{\mathrm{CS}}) \setminus \{\emptyset, K^{\mathrm{CS}}\}\})\end{aligned}$$

The cardinality of the *substrate binary-cartesian partition partition-sets set* is $|\mathcal{N}_{U,V,2}| = 2^m$ where

$$\begin{aligned}m &= \sum |\mathrm{P}(K^{\mathrm{CS}})|/2 - 1 : K \subseteq V,\ |K^{\mathrm{CS}}| \geq 2 \\ &= \sum 2^{x-1} - 1 : K \subseteq V,\ x = |K^{\mathrm{CS}}|,\ x \geq 2\end{aligned}$$

In the case where the *substrate volume* is at least two, $y \geq 2$, where $y = |V^{\mathrm{CS}}|$, the cardinality of the *substrate binary-cartesian partition partition-sets set* is bounded, $2^{2^{y-1}-1} \leq |\mathcal{N}_{U,V,2}| \leq 2^{2^{n+y-1}}$, where $n = |V|$.

In the case of *regular variables* of *valency* $d \geq 2$ and *dimension* n, the cardinality of the *substrate binary-cartesian partition partition-sets set* is

$$|\mathcal{N}_{U,V,2}| = 2^m : m = \sum_{k \in \{1 \ldots n\}} \binom{n}{k} (2^{d^k - 1} - 1)$$

The *substrate binary-cartesian partition transforms set* includes the *empty transform*, $(\emptyset, \emptyset) \in \mathcal{T}_{U,V,2}$, but excludes the *self partition transform*, $V^{\mathrm{CS}\{\}\mathrm{T}} \notin \mathcal{T}_{U,V,2}$, unless the *volume* is two, $y = 2$. The *substrate binary-cartesian partition transforms set* excludes the *unary partition transform*, $\{V^{\mathrm{CS}}\}^{\mathrm{T}} \notin \mathcal{T}_{U,V,2}$.

The *crown transform*, transformCrown(D), of a *substrate decomposition* $D \in \mathcal{D}_{U,V}$ maps to a *substrate transform* in the *substrate binary-cartesian partition transforms set*, $\{(X\%(V \cup \{w\}), \{w\})^{\mathrm{PVT}} : w \in W\}^{\mathrm{T}} \in \mathcal{T}_{U,V,2}$ where $(X, W) = \mathrm{transformCrown}(D)$. This is because *contingent slice binary partitions* form the *derived variables* of the *crown transform*, $\forall P \in W\ (|P| = 2)$.

The intersection between the *substrate binary-cartesian partition partition-sets set* and the *non-overlapping substrate partition-sets set* is

$$\begin{aligned}&\mathcal{N}_{U,V,2} \cap \mathcal{N}_{U,V,\mathrm{n}} \\ &= \{N : Y \in \mathrm{B}(V),\ N \in \prod_{K \in Y} \{P : P \in \mathrm{B}(K^{\mathrm{CS}}),\ |P| = 2\}\} \\ &= \{N : Y \in \mathrm{B}(V),\ N \in \prod_{K \in Y} \{\{C, K^{\mathrm{CS}} \setminus C\} : C \in \mathrm{P}(K^{\mathrm{CS}}) \setminus \{\emptyset, K^{\mathrm{CS}}\}\}\}\end{aligned}$$

In the case of non-empty *substrate variables*, $V \neq \emptyset$, the intersection has cardinality

$$|\mathcal{N}_{U,V,2} \cap \mathcal{N}_{U,V,\mathrm{n}}| = \sum \left(\prod_{K \in Y} 2^{|K^{\mathrm{CS}}|-1} - 1 \right) : Y \in \mathrm{B}(V),\ \forall K \in Y\ (|K^{\mathrm{CS}}| \geq 2)$$

The cardinality is bounded $|\mathcal{N}_{U,V,2} \cap \mathcal{N}_{U,V,\mathrm{n}}| \leq \mathrm{bell}(n) \times (2^y - 1)$, where *volume* $y = |V^{\mathrm{CS}}|$ and *dimension* $n = |V|$.

In the case of *regular variables* of *valency* $d \geq 2$ and *dimension* n, the cardinality of the intersection is

$$|\mathcal{N}_{U,V,2} \cap \mathcal{N}_{U,V,\mathrm{n}}| = \sum_{(L,c) \in \mathrm{bcd}(n)} \left(c \prod_{(k,m) \in L} (2^{d^k - 1} - 1)^m \right)$$

The subset of the *substrate transforms set*, $\mathcal{T}_{U,V}$, which consists of *transforms* having *unary partitions* of the *cartesian states* is simply the *empty transform* and the *unary partition transform*. Let the *substrate unary-cartesian partition-sets set* $\mathcal{N}_{U,V,1} \subset \mathcal{N}_{U,V}$ be defined such that the *partition sets* consist only of *unary partitions* of the *cartesian states* of subsets of the *substrate*, $|P^V| = |P| = 1$ where $P \in \mathrm{B}(K^{\mathrm{CS}})$ and $K \subseteq V$,

$$\begin{aligned} \mathcal{N}_{U,V,1} &= \{N : N \in \mathcal{N}_{U,V},\ \forall P \in N\ (|P| = 1)\} \\ &= \mathrm{P}(\{\{K^{\mathrm{CS}}\} : K \subseteq V\}) \end{aligned}$$

The cardinality of the *substrate unary-cartesian partition-sets set* is $|\mathcal{N}_{U,V,1}| = 2^{2^n}$. The set of *substrate transforms* is

$$\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,1}\} = \{(\emptyset, \emptyset), \{V^{\mathrm{CS}}\}^{\mathrm{T}}\} \subset \mathcal{T}_{U,V}$$

Consider the *substrate decremented transforms set* $\mathcal{T}_{U,V,-}$ which is a subset of the *substrate transforms set*, $\mathcal{T}_{U,V,-} \subseteq \mathcal{T}_{U,V}$. Let the *substrate decremented partition-sets set* $\mathcal{N}_{U,V,-} \subset \mathcal{N}_{U,V}$ be defined such that the *partition sets* contain exactly one *decremented self partition*, $\mathrm{decs}(\mathrm{self}(J^{\mathrm{CS}}))$ where $J \subseteq V$ and $\mathrm{decs} = \mathrm{decrements} \in \mathcal{R}_U \to \mathrm{P}(\mathcal{R}_U)$, with the remainder being *self partitions* of the *cartesian states*, $\mathrm{self}(K^{\mathrm{CS}})$ where $K \subseteq V$

$$\begin{aligned} \mathcal{N}_{U,V,-} = \{N : N \in \mathcal{N}_{U,V},\ \exists Q \in N\ ((|Q| = |\bigcup Q| - 1) \wedge \\ (\forall P \in N\ (P \neq Q \implies |P| = |\bigcup P|)))\} \end{aligned}$$

The *substrate decremented partition-sets set* excludes the *empty partition-set*, $\emptyset \notin \mathcal{N}_{U,V,-}$. Note that this definition does not prevent the *partition-set* from containing both the *decremented partition*, $Q \in \mathrm{decs}(\mathrm{self}(J^{\mathrm{CS}}))$ and the *self-partition*, $P = \mathrm{self}(J^{\mathrm{CS}})$, on the same subset, $J \subseteq V$, of the *substrate variables*.

The *substrate decremented transforms set* is defined in terms of *substrate decremented partition-sets set*

$$\mathcal{T}_{U,V,-} = \{N^{\mathrm{TV}} : N \in \mathcal{N}_{U,V,-}\}$$

The *substrate decremented transforms set* is such that

$$\begin{aligned}&\forall T \in \mathcal{T}_{U,V,-}\ \exists Q \in \mathrm{der}(T^{\%})\ \Diamond J = \mathrm{vars}(Q)\ ((Q \in \mathrm{decs}(J^{\mathrm{CS}\{\}}))\wedge \\ &\quad (\forall P \in \mathrm{der}(T^{\%})\ \Diamond K = \mathrm{vars}(P)\ (P \neq Q \implies P = K^{\mathrm{CS}\{\}})))\end{aligned}$$

The *substrate decremented partition-sets set*, $\mathcal{N}_{U,V,-}$, can be defined explicitly

$$\begin{aligned}\mathcal{N}_{U,V,-} &= \{\{Q\} \cup N : J \subseteq V,\ Q \in \mathrm{decs}(J^{\mathrm{CS}\{\}}),\ N \in \mathcal{N}_{U,V,\mathrm{c}}\} \\ &= \{\{Q\} \cup \{K^{\mathrm{CS}\{\}} : K \in X\} : J \subseteq V,\ Q \in \mathrm{decs}(J^{\mathrm{CS}\{\}}),\ X \subseteq \mathrm{P}(V)\} \\ &= \{\{J^{\mathrm{CS}\{\}} \setminus \{\{S_1\},\{S_2\}\} \cup \{\{S_1,S_2\}\}\} \cup \{K^{\mathrm{CS}\{\}} : K \in X\} : \\ &\qquad\qquad J \subseteq V,\ S_1,S_2 \in J^{\mathrm{CS}},\ S_1 \neq S_2,\ X \subseteq \mathrm{P}(V)\}\end{aligned}$$

The *substrate decremented transforms set* excludes the *empty transform*, $(\emptyset,\emptyset) \notin \mathcal{T}_{U,V,-}$. If the *substrate variables* is a singleton of a *bi-valent variable*, the *substrate decremented transforms set* includes the *self partition transform*, $(|V| = 1) \wedge (|V^{\mathrm{CS}}| = 2) \implies V^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U,V,-}$. The *substrate decremented transforms set* includes the *decremented self partition transforms*, $\{Q^{\mathrm{T}} : Q \in \mathrm{decs}(V^{\mathrm{CS}\{\}})\} \subset \mathcal{T}_{U,V,-}$. The *substrate decremented transforms set* excludes the *unary partition transform*, $\{V^{\mathrm{CS}}\}^{\mathrm{T}} \notin \mathcal{T}_{U,V,-}$, if none of the *substrate variables* are *bi-valent*, $\forall w \in V\ (|U_w| \neq 2)$.

The cardinality of the *substrate decremented partition-sets set* is

$$|\mathcal{N}_{U,V,-}| = 2^{2^n} \sum_{J \in \mathrm{P}(V)\setminus\{\emptyset\}} |J^{\mathrm{CS}}|(|J^{\mathrm{CS}}|-1)/2$$

where *dimension* $n = |V|$. Thus $|\mathcal{N}_{U,V,-}| < 2^{2^n+n-1}y^2$ where *substrate volume* $y = |V^{\mathrm{CS}}|$. If the *substrate* V is *regular* having *valency* d, the cardinality is

$$|\mathcal{N}_{U,V,-}| = 2^{2^n} \sum_{k \in \{1 \ldots n\}} \binom{n}{k} d^k(d^k-1)/2$$

which is bounded $|\mathcal{N}_{U,V,-}| < 2^{2^n+n-1} d^{2n}$.

The intersection of the *substrate decremented partition-sets set* and the *weak non-overlapping substrate partition-sets set* can be defined explicitly

$$\begin{aligned}&\mathcal{N}_{U,V,-} \cap \mathcal{N}'_{U,V,\mathrm{n}} = \\ &\quad \{\{Q\} \cup \{K^{\mathrm{CS}\{\}} : K \in Y,\ K \neq J\} : Y \in \mathrm{B}'(V),\ J \in Y,\ Q \in \mathrm{decs}(J^{\mathrm{CS}\{\}})\}\end{aligned}$$

The cardinality of the intersection is

$$|\mathcal{N}_{U,V,-} \cap \mathcal{N}'_{U,V,\mathrm{n}}| = \sum_{Y \in \mathrm{B}'(V)} \sum_{J \in Y \setminus \{\emptyset\}} |J^{\mathrm{CS}}|(|J^{\mathrm{CS}}|-1)/2$$

So $|\mathcal{N}_{U,V,-} \cap \mathcal{N}'_{U,V,\mathrm{n}}| \leq \mathrm{bell}(n) \times ny^2$, where *dimension* $n = |V|$ and *volume* $y = |V^{\mathrm{CS}}|$.

Similarly, the intersection of the *substrate decremented partition-sets set* and the *non-overlapping substrate partition-sets set* can be defined explicitly

$$\begin{aligned}&\mathcal{N}_{U,V,-} \cap \mathcal{N}_{U,V,\mathrm{n}} = \\ &\quad \{\{Q\} \cup \{K^{\mathrm{CS}\{\}} : K \in Y,\ K \neq J\} : Y \in \mathrm{B}(V),\ J \in Y,\ Q \in \mathrm{decs}(J^{\mathrm{CS}\{\}})\}\end{aligned}$$

The cardinality of the intersection is

$$|\mathcal{N}_{U,V,-} \cap \mathcal{N}_{U,V,\mathrm{n}}| = \sum_{Y \in \mathrm{B}(V)} \sum_{J \in Y \setminus \{\emptyset\}} |J^{\mathrm{CS}}|(|J^{\mathrm{CS}}|-1)/2$$

So $|\mathcal{N}_{U,V,-} \cap \mathcal{N}_{U,V,\mathrm{n}}| \leq \mathrm{bell}(n) \times ny^2/2$, where *dimension* $n = |V|$ and *volume* $y = |V^{\mathrm{CS}}|$. In the case of *regular variables* of *valency* d and *dimension* n, the cardinality is

$$|\mathcal{N}_{U,V,-} \cap \mathcal{N}_{U,V,\mathrm{n}}| = \sum_{(L,c) \in \mathrm{bcd}(n)} \left(c \sum_{(k,m) \in L} md^k(d^k-1)/2 \right)$$

The intersection of the *substrate decremented partition-sets set* and the *self non-overlapping substrate partition-sets set* can be defined explicitly

$$\begin{aligned}&\mathcal{N}_{U,V,-} \cap \mathcal{N}_{U,V,\mathrm{n,s}} \\ &= \{\{Q\} \cup \{\{u\}^{\mathrm{CS}\{\}} : u \in V \setminus \{w\}\} : w \in V,\ Q \in \mathrm{decs}(\{w\}^{\mathrm{CS}\{\}})\} \\ &= \{\{\{w\}^{\mathrm{CS}\{\}} \setminus \{\{S_1\},\{S_2\}\} \cup \{\{S_1,S_2\}\}\} \cup \{\{u\}^{\mathrm{CS}\{\}} : u \in V \setminus \{w\}\} : \\ &\qquad w \in V,\ s,t \in U_w,\ s \neq t,\ S_1 = \{(w,s)\},\ S_2 = \{(w,t)\}\}\end{aligned}$$

The cardinality of the intersection is

$$|\mathcal{N}_{U,V,-} \cap \mathcal{N}_{U,V,\mathrm{n,s}}| = \sum_{w \in V} |\{w\}^{\mathrm{CS}}|(|\{w\}^{\mathrm{CS}}| - 1)/2$$

So $|\mathcal{N}_{U,V,-} \cap \mathcal{N}_{U,V,\mathrm{n,s}}| \leq 1/2 \times ny^2$, where *dimension* $n = |V|$ and *volume* $y = |V^{\mathrm{CS}}|$. If the *substrate* V is *regular* having *valency* d, then

$$|\mathcal{N}_{U,V,-} \cap \mathcal{N}_{U,V,\mathrm{n,s}}| = nd(d-1)/2$$

By contrast, consider the *substrate incremented transforms set* $\mathcal{T}_{U,V,+}$ which is a subset of the *substrate transforms set*, $\mathcal{T}_{U,V,+} \subseteq \mathcal{T}_{U,V}$. The *substrate incremented transforms set* consists of the *transforms* of the increments of a *pointed partition*, increments $\in \mathcal{R}_* \to \mathrm{P}(\mathcal{R}_*)$. The only *pointed partition* that can be constructed from a *substrate transform*, without specifically defining a point *component*, is that of the *unary partition transform*, $\{V^{\mathrm{CS}}\}^{\mathrm{T}}$, which has only one *component*. Let the *substrate incremented partition-sets set* $\mathcal{N}_{U,V,+} \subset \mathcal{N}_{U,V}$ be defined such that the *partition sets* are singletons of the *incremented self partitions* which are *singleton pointed binary partitions incremented* from the *unary partition*, $\mathrm{incs}((\{J^{\mathrm{CS}}\}, J^{\mathrm{CS}}))$ where $J \subseteq V$ and $\mathrm{incs} = \mathrm{increments} \in \mathcal{R}_{*,U} \to \mathrm{P}(\mathcal{R}_{*,U})$,

$$\mathcal{N}_{U,V,+} \quad = \quad \{N : N \in \mathcal{N}_{U,V},\ N = \{Q\},\ |Q| = 2,\ (\exists C \in Q\ (|C| = 1))\}$$

The *substrate incremented partition-sets set* excludes the *empty partition-set*, $\emptyset \notin \mathcal{N}_{U,V,+}$.

The *substrate incremented transforms set* is defined in terms of *substrate incremented partition-sets set*

$$\mathcal{T}_{U,V,+} = \{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,+}\}$$

The *substrate incremented transforms set* is such that

$$\begin{aligned}&\forall T \in \mathcal{T}_{U,V,+}\ (|\mathrm{der}(T^{\%})| = 1 \wedge \\ &\qquad (\forall Q \in \mathrm{der}(T^{\%})\ \Diamond J = \mathrm{vars}(Q)\ (Q \in \mathrm{incs}((\{J^{\mathrm{CS}}\}, J^{\mathrm{CS}})))))\end{aligned}$$

The *substrate incremented partition-sets set*, $\mathcal{N}_{U,V,+}$, can be defined explicitly

$$\mathcal{N}_{U,V,+} \quad = \quad \{\{\{\{S\}, J^{\mathrm{CS}} \setminus \{S\}\}\} : J \subseteq V,\ |J^{\mathrm{CS}}| > 1,\ S \in J^{\mathrm{CS}}\}$$

The *substrate incremented transforms set* is a subset of the *substrate binary-cartesian partition transforms set*, $\mathcal{T}_{U,V,+} \subset \mathcal{T}_{U,V,2}$. The *substrate incremented* *transforms set* excludes the *empty transform*, $(\emptyset, \emptyset) \notin \mathcal{T}_{U,V,+}$, and the *unary partition transform*, $\{V^{\mathrm{CS}}\}^{\mathrm{T}} \notin \mathcal{T}_{U,V,+}$. The *substrate incremented transforms set* excludes the *self partition transform*, $V^{\mathrm{CS}\{\}\mathrm{T}} \notin \mathcal{T}_{U,V,+}$, unless the *volume* is two, $|V^{\mathrm{CS}}| = 2$.

The cardinality of the *substrate incremented partition-sets set* is

$$|\mathcal{N}_{U,V,+}| = \sum(|J^{\mathrm{CS}}|/2 : J \subseteq V,\ |J^{\mathrm{CS}}| = 2) + \sum(|J^{\mathrm{CS}}| : J \subseteq V,\ |J^{\mathrm{CS}}| > 2)$$

The cardinality of the *substrate incremented partition-sets set* is bounded $|\mathcal{N}_{U,V,+}| \leq 2^n y$ where *dimension* $n = |V|$ and $y = |V^{\mathrm{CS}}|$ if $y > 2$.

Now consider subsets of the *substrate partition-sets set* which are defined by parameter. The definition of the *substrate partition-sets set* is

$$\mathcal{N}_{U,V} = \mathrm{P}(\{P : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}})\})$$

The corresponding *substrate transforms set* is $\mathcal{T}_{U,V} = \{N^{\mathrm{TV}} : N \in \mathcal{N}_{U,V}\}$.

The cardinality of the *substrate partition-sets set* is

$$|\mathcal{N}_{U,V}| = 2^c : c = \sum_{K \subseteq V} \mathrm{bell}(|K^{\mathrm{CS}}|)$$

This is bounded

$$2^{\mathrm{bell}(y)} \leq |\mathcal{N}_{U,V}| \leq 2^{2^n \mathrm{bell}(y)}$$

where $y = |V^{\mathrm{CS}}|$. In the case of *regular variables* V, having *valency* $\{d\} = \{|U_w| : w \in V\}$ and *dimension* $n = |V|$, the cardinality is

$$|\mathcal{N}_{U,V}| = 2^c : c = \sum_{k \in \{0 \ldots n\}} \binom{n}{k} \mathrm{bell}(d^k)$$

First consider the *limited-underlying-dimension substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{kmax}}$ which is parameterised by $\mathrm{kmax} \in \mathbf{N}$ such that the cardinality of the *variables* of each of the *partitions* is *limited*,

$$\mathcal{N}_{U,V,\mathrm{kmax}} = \mathrm{P}(\{P : K \subseteq V,\ |K| \leq \mathrm{kmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\})$$

The cardinality of the *limited-underlying-dimension substrate partition-sets set* is

$$|\mathcal{N}_{U,V,\mathrm{kmax}}| = 2^c\ :\ c = \sum(\mathrm{bell}(|K^{\mathrm{CS}}|) : K \subseteq V,\ |K| \leq \mathrm{kmax})$$

In the case of *regular variables* V, having *valency* d and *dimension* n, such that $\mathrm{kmax} \leq n$, the cardinality is

$$|\mathcal{N}_{U,V,\mathrm{kmax}}| = 2^c \ : \ c = \sum_{k \in \{0 \ldots \mathrm{kmax}\}} \binom{n}{k} \mathrm{bell}(d^k)$$

This is bounded

$$2^{\mathrm{bell}(d^{\mathrm{kmax}})} \leq |\mathcal{N}_{U,V,\mathrm{kmax}}| \leq 2^{2^n \mathrm{bell}(d^{\mathrm{kmax}})}$$

where $d^{\mathrm{kmax}} \leq |V^{\mathrm{CS}}|$.

Similarly, consider the *limited-underlying-volume substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{xmax}}$ which is parameterised by $\mathrm{xmax} \in \mathbf{N}_{>0}$ such that the *underlying volume* of each of the *partitions* is *limited*,

$$\mathcal{N}_{U,V,\mathrm{xmax}} = \mathrm{P}(\{P : K \subseteq V,\ |K^{\mathrm{CS}}| \leq \mathrm{xmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\})$$

The cardinality of the *limited-underlying-volume substrate partition-sets set* is

$$|\mathcal{N}_{U,V,\mathrm{xmax}}| = 2^c \ : \ c = \sum (\mathrm{bell}(|K^{\mathrm{CS}}|) : K \subseteq V,\ |K^{\mathrm{CS}}| \leq \mathrm{xmax})$$

This is bounded

$$|\mathcal{N}_{U,V,\mathrm{xmax}}| \leq 2^{2^n \mathrm{bell}(\mathrm{xmax})}$$

where $\mathrm{xmax} \leq |V^{\mathrm{CS}}|$.

In the case of *pluri-valent regular variables* V, having *valency* $d > 1$ and *dimension* n, the cardinality is defined in terms of the implied *underlying-dimension limit* $\mathrm{kmax} = \ln \mathrm{xmax} \ / \ln d$, where $\ln \mathrm{xmax} \ / \ln d \in \mathbf{N}$,

$$|\mathcal{N}_{U,V,\mathrm{xmax}}| = 2^c \ : \ c = \sum_{k \in \{0 \ldots \mathrm{kmax}\}} \binom{n}{k} \mathrm{bell}(d^k)$$

The *limited-valency substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{umax}}$ is parameterised by $\mathrm{umax} \in \mathbf{N}_{>0}$ such that the *valency* of the *partition variables* is *limited*,

$$\mathcal{N}_{U,V,\mathrm{umax}} = \mathrm{P}(\{P : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}}),\ |P| \leq \mathrm{umax}\})$$

The cardinality of the *limited-valency substrate partition-sets set* is

$$\begin{aligned}&|\mathcal{N}_{U,V,\mathrm{umax}}| = 2^c : \\ &\quad c = \sum (\mathrm{stir}(|K^{\mathrm{CS}}|, u) : K \subseteq V,\ u \in \{1 \ldots \mathrm{umax}\},\ u \leq |K^{\mathrm{CS}}|)\end{aligned}$$

where $\mathrm{stir} \in \mathbf{N}_{>0} \times \mathbf{N} \to \mathbf{N}_{>0}$ is the Stirling number of the second kind.

In the case of *regular variables* V, having *valency* d and *dimension* n, the cardinality is

$$\begin{aligned}
&|\mathcal{N}_{U,V,\mathrm{umax}}| = 2^c : \\
&\quad c = \sum \left(\binom{n}{k} \mathrm{stir}(d^k, u) : k \in \{0 \ldots n\},\ u \in \{1 \ldots \mathrm{umax}\},\ u \leq d^k \right)
\end{aligned}$$

Similarly, the *lower-limited-valency substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{umin}}$ is parameterised by $\mathrm{umin} \in \mathbf{N}_{>0}$ such that the *valency* of the *partition variables* is *lower-limited*,

$$\mathcal{N}_{U,V,\mathrm{umin}} = \mathrm{P}(\{P : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}}),\ |P| \geq \mathrm{umin}\})$$

The cardinality of the *lower-limited-valency substrate partition-sets set* is

$$|\mathcal{N}_{U,V,\mathrm{umin}}| = 2^c \ :\ c = \sum (\mathrm{stir}(|K^{\mathrm{CS}}|, u) : K \subseteq V,\ u \in \{\mathrm{umin} \ldots |K^{\mathrm{CS}}|\})$$

In the case of *regular variables* V, having *valency* d and *dimension* n, the cardinality is

$$|\mathcal{N}_{U,V,\mathrm{umin}}| = 2^c \ :\ c = \sum \left(\binom{n}{k} \mathrm{stir}(d^k, u) : k \in \{0 \ldots n\},\ u \in \{\mathrm{umin} \ldots d^k\} \right)$$

In the special case where $\mathrm{umin} = 2$ the cardinality is

$$|\mathcal{N}_{U,V,\mathrm{umin}}| = 2^c \ :\ c = \sum \left(\binom{n}{k} (\mathrm{bell}(d^k) - 1) : k \in \{0 \ldots n\} \right)$$

The *limited-component substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{cmin}}$ is parameterised by $\mathrm{cmin} \in \mathbf{N}_{>0}$ such that the cardinality of the *components* of the *partition variables* is *limited*,

$$\mathcal{N}_{U,V,\mathrm{cmin}} = \mathrm{P}(\{P : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}}),\ \forall C \in P\ (|C| \geq \mathrm{cmin})\})$$

In the case of *regular variables* V, having *valency* d and *dimension* n, the cardinality is

$$\begin{aligned}
&|\mathcal{N}_{U,V,\mathrm{cmin}}| = 2^c : \\
&\quad c = \sum \left(\binom{n}{k} b : k \in \{0 \ldots n\},\ (L, b) \in \mathrm{bcd}(d^k), \right. \\
&\qquad\qquad \left. \forall (j, m) \in L\ (m \neq 0 \implies j \geq \mathrm{cmin}) \right)
\end{aligned}$$

The *intersecting substrate partition-sets set* $\mathcal{N}_{U,V,X}$ is parameterised by a set of *variables* $X \subseteq \mathrm{vars}(U)$ such that the *variables* of the *partitions* intersect with the given set,

$$\mathcal{N}_{U,V,X} = \mathrm{P}(\{P : K \subseteq V,\ K \cap X \neq \emptyset,\ P \in \mathrm{B}(K^{\mathrm{CS}})\})$$

The cardinality of the *intersecting substrate partition-sets set* is

$$|\mathcal{N}_{U,V,X}| = 2^c \ :\ c = \sum(\mathrm{bell}(|K^{\mathrm{CS}}|) : K \subseteq V,\ K \cap X \neq \emptyset)$$

In the case where the intersection with the *substrate variables* is not empty, $V \cap X \neq \emptyset$, the cardinality is bounded

$$2^{\mathrm{bell}(y)} \leq |\mathcal{N}_{U,V,X}| \leq 2^{x2^{n-1}\mathrm{bell}(y)}$$

where $x = |V \cap X|$ and $y = |V^{\mathrm{CS}}|$. In the case of *regular variables* V, having *valency* d and *dimension* n, the cardinality is

$$\begin{aligned} |\mathcal{N}_{U,V,X}| = 2^c \ :\ c \ &= \ \left(\sum_{k \in \{1 \ldots n-x\}} \left(\binom{n}{k} - \binom{n-x}{k}\right) \mathrm{bell}(d^k)\right) + \\ &\quad \left(\sum_{k \in \{n-x+1 \ldots n\}} \binom{n}{k} \mathrm{bell}(d^k)\right) \end{aligned}$$

This may be written more succinctly if the binomial coefficient is defined $\forall a, b \in \mathbf{N}\ (b > a \implies \binom{a}{b} = 0)$,

$$|\mathcal{N}_{U,V,X}| = 2^c \ :\ c = \sum_{k \in \{1 \ldots n\}} \left(\binom{n}{k} - \binom{n-x}{k}\right) \mathrm{bell}(d^k)$$

The *limited-breadth substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{bmax}}$ is parameterised by $\mathrm{bmax} \in \mathbf{N}$ such that the cardinalities of the *partition-sets* are *limited*,

$$\mathcal{N}_{U,V,\mathrm{bmax}} = \{N : N \in \mathcal{N}_{U,V},\ |N| \leq \mathrm{bmax}\}$$

The cardinality of the *limited-breadth substrate partition-sets set* is

$$|\mathcal{N}_{U,V,\mathrm{bmax}}| = \left(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\right) \ :\ c = \sum_{K \subseteq V} \mathrm{bell}(|K^{\mathrm{CS}}|)$$

This is bounded

$$|\mathcal{N}_{U,V,\mathrm{bmax}}| \leq \sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{2^n \mathrm{bell}(y)}{b}$$

where $y = |V^{\mathrm{CS}}|$. In the case of *regular variables* V, having *valency* d and *dimension* n, the cardinality is

$$|\mathcal{N}_{U,V,\mathrm{bmax}}| = \Big(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\Big) \ : \ c = \sum_{k \in \{0 \ldots n\}} \binom{n}{k} \mathrm{bell}(d^k)$$

The intersection of the *limited-underlying-dimension substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{kmax}}$, and the *limited-breadth substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{bmax}}$, is

$$\begin{aligned}&\mathcal{N}_{U,V,\mathrm{kmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}} = \\ &\qquad \{N : N \subseteq \{P : K \subseteq V,\ |K| \leq \mathrm{kmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\},\ |N| \leq \mathrm{bmax}\}\end{aligned}$$

The cardinality of the intersection is

$$\begin{aligned}&|\mathcal{N}_{U,V,\mathrm{kmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}}| = \\ &\qquad \Big(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\Big) \ : \ c = \sum(\mathrm{bell}(|K^{\mathrm{CS}}|) : K \subseteq V,\ |K| \leq \mathrm{kmax})\end{aligned}$$

In the case of *regular variables* V, having *valency* d and *dimension* n, the cardinality of the intersection is

$$\begin{aligned}&|\mathcal{N}_{U,V,\mathrm{kmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}}| = \\ &\qquad\qquad \Big(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\Big) \ : \ c = \sum_{k \in \{0 \ldots \mathrm{kmax}\}} \binom{n}{k} \mathrm{bell}(d^k)\end{aligned}$$

The intersection of the *intersecting substrate partition-sets set* $\mathcal{N}_{U,V,X}$, the *limited-underlying-dimension substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{kmax}}$, and the *limited-breadth substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{bmax}}$, is

$$\begin{aligned}&\mathcal{N}_{U,V,X} \cap \mathcal{N}_{U,V,\mathrm{kmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}} = \\ &\qquad \{N : N \subseteq \{P : K \subseteq V,\ K \cap X \neq \emptyset,\ |K| \leq \mathrm{kmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\}, \\ &\qquad\qquad |N| \leq \mathrm{bmax}\}\end{aligned}$$

The cardinality of the intersection is

$$\begin{aligned}&|\mathcal{N}_{U,V,X} \cap \mathcal{N}_{U,V,\mathrm{kmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}}| = \\ &\qquad \Big(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\Big) : \\ &\qquad\qquad c = \sum(\mathrm{bell}(|K^{\mathrm{CS}}|) : K \subseteq V,\ K \cap X \neq \emptyset,\ |K| \leq \mathrm{kmax})\end{aligned}$$

In the case of *regular variables* V, having *valency* d, *dimension* n and *intersecting dimension* $x = |X|$, the cardinality of the intersection is

$$|\mathcal{N}_{U,V,X} \cap \mathcal{N}_{U,V,\mathrm{kmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}}| = \\ \Big(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\Big) \ :\ c = \sum_{k \in \{0 \ldots \mathrm{kmax}\}} \Big(\binom{n}{k} - \binom{n-x}{k}\Big)\,\mathrm{bell}(d^k)$$

The *range-limited-breadth substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{bran}}$ is parameterised by $\mathrm{bran} = (\mathrm{bmin}, \mathrm{bmax}) \in \mathbf{N^2}$ such that the cardinalities of the *partition-sets* are *limited*,

$$\mathcal{N}_{U,V,\mathrm{bran}} = \{N : N \in \mathcal{N}_{U,V},\ \mathrm{bmin} \leq |N| \leq \mathrm{bmax}\}$$

The cardinality of the *range-limited-breadth substrate partition-sets set* is

$$|\mathcal{N}_{U,V,\mathrm{bran}}| = \Big(\sum_{b \in \{\mathrm{bmin} \ldots \mathrm{bmax}\}} \binom{c}{b}\Big) \ :\ c = \sum_{K \subseteq V} \mathrm{bell}(|K^{\mathrm{CS}}|)$$

In the case of *regular variables* V, having *valency* d and *dimension* n, the cardinality is

$$|\mathcal{N}_{U,V,\mathrm{bran}}| = \Big(\sum_{b \in \{\mathrm{bmin} \ldots \mathrm{bmax}\}} \binom{c}{b}\Big) \ :\ c = \sum_{k \in \{0 \ldots n\}} \binom{n}{k}\mathrm{bell}(d^k)$$

The intersection of the *non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n}}$, and the *range-limited-breadth substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{bran}}$, is

$$\begin{aligned} &\mathcal{N}_{U,V,\mathrm{n},\mathrm{bran}} \\ &\quad = \{N : Y \in \mathrm{B}(V),\ N \in \prod_{K \in Y} \mathrm{B}(K^{\mathrm{CS}}),\ \mathrm{bmin} \leq |N| \leq \mathrm{bmax}\} \\ &\quad = \{N : Y \in \mathrm{B}(V),\ \mathrm{bmin} \leq |Y| \leq \mathrm{bmax},\ N \in \prod_{K \in Y} \mathrm{B}(K^{\mathrm{CS}})\} \end{aligned}$$

In the case where $\mathrm{bmax} \leq n$,

$$\mathcal{N}_{U,V,\mathrm{n},\mathrm{bran}} = \{N : b \in \{\mathrm{bmin} \ldots \mathrm{bmax}\},\ Y \in \mathrm{S}(V,b),\ N \in \prod_{K \in Y} \mathrm{B}(K^{\mathrm{CS}})\}$$

where the fixed cardinality partition function is $\mathrm{S} \in \mathrm{P}(\mathcal{X}) \times \mathbf{N}_{>0} \to \mathrm{P}(\mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}))$.

In the case of *regular variables* V, having *valency* d and *dimension* n, the cardinality of the *range-limited-breadth non-overlapping substrate partition-sets set* is

$$\begin{aligned}
&|\mathcal{N}_{U,V,\mathrm{n,bran}}| \\
&\quad= \sum \Big(\prod_{K \in Y} \mathrm{bell}(d^{|K|}) \Big) : b \in \{\mathrm{bmin} \ldots \mathrm{bmax}\},\ Y \in \mathrm{S}(V,b) \\
&\quad= \sum \Big(c \prod_{(k,m) \in L} \mathrm{bell}(d^k)^m \Big) : b \in \{\mathrm{bmin} \ldots \mathrm{bmax}\},\ (L,c) \in \mathrm{sscd}(n,b)
\end{aligned}$$

where sscd = stircd and the fixed cardinality partition function cardinality function is $\mathrm{stircd} \in \mathbf{N}_{>0} \times \mathbf{N}_{>0} \to (\mathcal{L}(\mathbf{N}) \to \mathbf{N})$, defined in appendix 'Partitions', below.

The special case of the *fixed-breadth non-overlapping substrate partition-sets set* given cardinality $b \in \mathbf{N}_{>0}$ is defined

$$\mathcal{N}_{U,V,\mathrm{n},b} = \{N : Y \in \mathrm{B}(V),\ N \in \prod_{K \in Y} \mathrm{B}(K^{\mathrm{CS}}),\ |N| = b\}$$

In the case of *regular variables* V, having *valency* d and *dimension* n, the cardinality is

$$|\mathcal{N}_{U,V,\mathrm{n},b}| = \sum_{(L,c) \in \mathrm{sscd}(n,b)} \Big(c \prod_{(k,m) \in L} \mathrm{bell}(d^k)^m \Big)$$

In the case where the *fixed-breadth* is two, $b = 2$, the *fixed-breadth non-overlapping substrate partition-sets set* equals the *binary non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n},2} = \mathcal{N}_{U,V,\mathrm{n,b}}$. In the case where the *fixed-breadth* equals the *dimension*, $b = n$, the *fixed-breadth non-overlapping substrate partition-sets set* equals the *self non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n},n} = \mathcal{N}_{U,V,\mathrm{n,s}}$.

The intersection of the *substrate self-cartesian partition-sets set* and the *limited-breadth non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,bmax}}$, is

$$\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,bmax}} = \{\{K^{\mathrm{CS}\{\}} : K \in Y\} : Y \in \mathrm{B}(V),\ |Y| \leq \mathrm{bmax}\}$$

In the case where $\mathrm{bmax} \leq n$, the cardinality of the intersection is

$$|\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,bmax}}| = \sum_{b \in \{1 \ldots \mathrm{bmax}\}} \mathrm{stir}(n,b)$$

where $\mathrm{stir} \in \mathbf{N}_{>0} \times \mathbf{N} \to \mathbf{N}_{>0}$ is the Stirling number of the second kind.

Similarly to the *limited-breadth substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{bmax}}$, the *limited-derived-volume substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{wmax}}$ is parameterised by wmax $\in \mathbf{N}_{>0}$ such that the *volumes* of the *partition-sets* are *limited*,

$$\mathcal{N}_{U,V,\mathrm{wmax}} = \{N : N \in \mathcal{N}_{U,V},\ |N^{\mathrm{C}}| \leq \mathrm{wmax}\}$$

Having considered the analysis of the *substrate transforms set*, $\mathcal{T}_{U,V}$, into various subsets, now consider its synthesis and the synthesis of its subsets. That is, consider possible sequences of construction from smaller subsets.

Substrate structures may be constructed by means of *linear fuds*. A *linear fud* $F \in \mathcal{F}_{U,1}$ is a *non-circular fud* such that the *underlying variables* of the *transforms* in each *layer fud* are the *derived variables* of the *layer fud* immediately below, $\forall i \in \{2 \ldots |L|\}\ (\mathrm{und}(L_i) \subseteq \mathrm{der}(L_{i-1}))$ where $L \in \mathcal{L}(\mathrm{P}(F))$ and $\mathrm{set}(L) \in \mathrm{B}(F)$. Thus a *linear fud* may be represented as a list of *transforms*, $\{(i, G^{\mathrm{T}}) : (i, G) \in L\} \in \mathcal{L}(\mathcal{T}_{\mathrm{f,U}})$.

Let *system* U contain all of the *partition variables* of the *substrate fuds set*, $\mathcal{F}_{U,V}$, on *variables* V, $\bigcup\{\mathrm{vars}(T) : T \in \mathcal{F}_{U,V}\} \subseteq \mathrm{vars}(U)$. The *substrate transforms set*, $\mathcal{T}_{U,V}$, can be constructed from *partition-sets* of *linear fuds* of pairs of *multi-partition transforms*, $T_1, T_2 \in \mathcal{T}_{U,\mathrm{P}^*}$. The first *transform* is a *substrate self-cartesian transform*, $T_1 \in \mathcal{T}_{U,V,\mathrm{c}}$. The second *transform* is a *non-empty self overlapping substrate transform*, $T_2 \in \mathcal{T}_{U,W,\mathrm{o,s}}$, $T_2 \neq (\emptyset, \emptyset)$, in the *derived variables* of the first, $W = \mathrm{der}(T_1)$,

$$\mathcal{T}_{U,V} = \{\{T_1, T_2\}^{\mathrm{TNT}} : T_1 \in \mathcal{T}_{U,V,\mathrm{c}},\ W = \mathrm{der}(T_1),\ T_2 \in \mathcal{T}_{U,W,\mathrm{o,s}} \setminus \{(\emptyset, \emptyset)\}\} \cup \{(\emptyset, \emptyset)\}$$

where shorthand $T^{\mathrm{N}} := \mathrm{partitionset}(T)$. Note that the second *transform*, T_2, is *non-empty* so that the *fud* is *one functional*, $\{T_1, T_2\} \in \mathcal{F}_{U,1}$.

The *substrate self-cartesian transform set*, $\mathcal{T}_{U,V,\mathrm{c}}$, is constructed explicitly as

$$\mathcal{T}_{U,V,\mathrm{c}} = \{\{K^{\mathrm{CS}\{\}V\mathrm{T}} : K \in X\}^{\mathrm{T}} : X \subseteq \mathrm{P}(V)\}$$

The *self overlapping substrate transforms set*, $\mathcal{T}_{U,W,\mathrm{n,s}}$, is constructed explicitly as

$$\mathcal{T}_{U,W,\mathrm{o,s}} = \{(\bigcup H)^{\mathrm{T}W} : H \in \prod_{w \in W} \mathrm{P}(\mathrm{B}(\{w\}^{\mathrm{CS}}))\}$$

So the *substrate transforms set*, $\mathcal{T}_{U,V}$, can be constructed explicitly as

$$\begin{aligned}\mathcal{T}_{U,V} = \{(W \cup \bigcup H)^{\mathrm{TNT}V} : X \subseteq \mathrm{P}(V),\\ W = \{K^{\mathrm{CS}\{\}} : K \in X\},\ H \in \prod_{w \in W} \mathrm{P}(\mathrm{B}(\{w\}^{\mathrm{CS}}))\} \cup \{(\emptyset,\emptyset)\}\end{aligned}$$

which has cardinality of construction

$$\begin{aligned}|\{(X,H) : X \subseteq \mathrm{P}(V),\ H \in \prod_{K \in X} \mathrm{P}(\mathrm{B}(\{K^{\mathrm{CS}\{\}}\}^{\mathrm{CS}}))\}| &= \sum_{X \subseteq \mathrm{P}(V)} \prod_{K \in X} 2^{|\mathrm{B}(K^{\mathrm{CS}})|}\\ &\leq 2^{2^n(1+\mathrm{bell}(y))}\end{aligned}$$

where *volume* $y = |V^{\mathrm{CS}}|$, and *dimension* $n = |V|$. This may be compared to the explicit constructions from (i) subsets of the *base substrate partitions*, $\mathrm{B}(V^{\mathrm{CS}})$,

$$\mathcal{T}_{U,V} = \{N^{\mathrm{T}} : N \subseteq \mathrm{B}(V^{\mathrm{CS}})\}$$

which has cardinality of construction

$$|\mathrm{P}(\mathrm{B}(V^{\mathrm{CS}}))| = 2^{\mathrm{bell}(y)}$$

and (ii) the *substrate partition-sets set*, $\mathcal{N}_{U,V}$,

$$\mathcal{T}_{U,V} = \{N^{\mathrm{T}V} : N \subseteq \{P : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}})\}\}$$

which has cardinality of construction

$$|\mathcal{N}_{U,V}| = |\mathrm{P}(\{P : K \subseteq V,\ P \in \mathrm{B}(K^{\mathrm{CS}})\})| = \prod_{K \subseteq V} 2^{|\mathrm{B}(K^{\mathrm{CS}})|}$$

and which is bounded $2^{\mathrm{bell}(y)} \leq |\mathcal{N}_{U,V}| \leq 2^{2^n \times \mathrm{bell}(y)}$.

The *non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n}} \subset \mathcal{T}_{U,V}$, can also be constructed from *partition-sets* of *linear fuds* of pairs of *multi-partition transforms*, $T_1, T_2 \in \mathcal{T}_{U,\mathrm{P}^*}$. In this case the first *transform* is in the *non-overlapping* subset of the *substrate self-cartesian transform set*, $T_1 \in \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n}}$. The second *transform* is a *self non-overlapping substrate transform*, $T_2 \in \mathcal{T}_{U,W,\mathrm{n,s}}$, in the *derived variables* of the first, $W = \mathrm{der}(T_1)$,

$$\begin{aligned}&\mathcal{T}_{U,V,\mathrm{n}} =\\ &\quad \{\{T_1,T_2\}^{\mathrm{TNT}} : T_1 \in \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n}},\ W = \mathrm{der}(T_1),\ T_2 \in \mathcal{T}_{U,W,\mathrm{n,s}}\} \cup \{(\emptyset,\emptyset)\}\end{aligned}$$

The *strong non-overlapping substrate transforms set*, $\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} \subseteq \mathcal{T}_{U,V,\mathrm{n}}$ can be constructed from the *transforms* of the *non-overlapping* subset of the *substrate self-cartesian partition-sets set*, $T_1 \in \{M^{\mathrm{T}} : M \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n}}\}$, followed by the *transforms* of the *self non-overlapping substrate partition-sets set*, $T_2 \in \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,W,\mathrm{n,s}}\}$, where $W = \mathrm{der}(T_1)$,

$$\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} = \{(M \cup N)^{\mathrm{TNT}} : M \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n}},\ N \in \mathcal{N}_{U,M,\mathrm{n,s}}\}$$

The intersection between the *substrate self-cartesian partition-sets set*, $\mathcal{N}_{U,V,\mathrm{c}}$, and the *non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n}}$, is constructed explicitly as

$$\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n}} = \{\{K^{\mathrm{CS}\{\}} : K \in Y\} : Y \in \mathrm{B}(V)\}$$

The *self non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n,s}}$, is constructed explicitly as

$$\mathcal{N}_{U,M,\mathrm{n,s}} = \prod_{w \in M} \mathrm{B}(\{w\}^{\mathrm{CS}})$$

So the *strong non-overlapping substrate transforms set*, $\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\}$, can be constructed explicitly as

$$\begin{aligned}\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} = \{&(M \cup N)^{\mathrm{TNT}} : Y \in \mathrm{B}(V), \\ &M = \{K^{\mathrm{CS}\{\}} : K \in Y\},\ N \in \prod_{w \in M} \mathrm{B}(\{w\}^{\mathrm{CS}})\}\end{aligned}$$

which has cardinality of construction equal to the cardinality of construction of the *non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n}}$,

$$\begin{aligned}|\{(Y,N) : Y \in \mathrm{B}(V),\ N \in \prod_{K \in Y} \mathrm{B}(\{K^{\mathrm{CS}\{\}}\}^{\mathrm{CS}})\}| &= \sum_{Y \in \mathrm{B}(V)} \prod_{K \in Y} |\mathrm{B}(K^{\mathrm{CS}})| \\ &\leq \mathrm{bell}(n) \times \mathrm{bell}(y)\end{aligned}$$

where *volume* $y = |V^{\mathrm{CS}}|$, and *dimension* $n = |V|$.

The *self non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}}$, can be constructed from *linear fuds* of *multi-partition transforms*, $L \in \mathcal{L}(\mathcal{T}_{U,\mathrm{P}^*})$. The starting *transform* in the sequence is the singleton *strong self non-overlapping substrate self-cartesian transforms set*, $L_1 \in \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,s}}\}$. That is, the first *transform* is the *value full functional transform*, $L_1 =$

$\{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}}$. The subsequent *transforms* are *strong self non-overlapping substrate decremented transforms*, $L_2 \in \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,W_1,-} \cap \mathcal{N}_{U,W_1,\mathrm{n,s}}\}$, $L_3 \in \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,W_2,-} \cap \mathcal{N}_{U,W_2,\mathrm{n,s}}\}$, and so on,

$$\begin{aligned}
&\mathcal{T}_{U,V,\mathrm{n,s}} = \\
&\quad \{\mathrm{set}(L)^{\mathrm{TNT}} : L \in \mathcal{L}(\mathcal{T}_{U,\mathrm{f},1}),\ \{L_1\} = \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,s}}\}, \\
&\qquad (\forall i \in \{2 \ldots |L|\}\ \Diamond W = \mathrm{der}(L_{i-1})\ (L_i \in \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,W,-} \cap \mathcal{N}_{U,W,\mathrm{n,s}}\}))\}
\end{aligned}$$

The intersection of the *substrate decremented partition-sets set* and the *self non-overlapping substrate partition-sets set* is constructed explicitly

$$\begin{aligned}
&\mathcal{N}_{U,W,-} \cap \mathcal{N}_{U,W,\mathrm{n,s}} = \\
&\quad \{\{Q\} \cup \{\{u\}^{\mathrm{CS}\{\}} : u \in W \setminus \{w\}\} : w \in W,\ Q \in \mathrm{decs}(\{w\}^{\mathrm{CS}\{\}})\}
\end{aligned}$$

where decs = decrements $\in \mathcal{R}_U \to \mathrm{P}(\mathcal{R}_U)$.

So the *self non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}}$, can be constructed explicitly as

$$\begin{aligned}
&\mathcal{T}_{U,V,\mathrm{n,s}} = \\
&\quad \{(\bigcup \mathrm{set}(L))^{\mathrm{TNT}} : M \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,s}}, \\
&\qquad\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdec}(U)(M))\})\}
\end{aligned}$$

and, in the case of non-empty *substrate*, $V \neq \emptyset$,

$$\begin{aligned}
&V \neq \emptyset \implies \mathcal{T}_{U,V,\mathrm{n,s}} = \\
&\quad \{(\bigcup \mathrm{set}(L))^{\mathrm{TNT}} : M = \{\{v\}^{\mathrm{CS}\{\}} : v \in V\}, \\
&\qquad\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdec}(U)(M))\})\}
\end{aligned}$$

where the tree of *self non-overlapping substrate decremented partition-sets* is defined $\mathrm{tdec}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathrm{P}(\mathcal{R}_U))$ as

$$\mathrm{tdec}(U)(M) := \{(N, \mathrm{tdec}(U)(N)) : N \in \mathcal{N}_{U,M,-} \cap \mathcal{N}_{U,M,\mathrm{n,s}}\}$$

and $\mathrm{tdec}(U)(\emptyset) := \emptyset$. Explicitly this is

$$\begin{aligned}
&\mathrm{tdec}(U)(M) := \{(N, \mathrm{tdec}(U)(N)) : \\
&\quad w \in M,\ Q \in \mathrm{decs}(\{w\}^{\mathrm{CS}\{\}}),\ N = \{Q\} \cup \{\{u\}^{\mathrm{CS}\{\}} : u \in M,\ u \neq w\}\}
\end{aligned}$$

The cardinality of the *self non-overlapping substrate decremented partition-sets tree* may be computed by defining $\mathrm{tdeccd}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}))$ as

$$\mathrm{tdeccd}(U)(V) := \{((1, L), \mathrm{tdeccd}(1, L)) : L = \{(i, |U_v|) : (v, i) \in \mathrm{order}(D_{\mathrm{V}}, V)\}\}$$

where order $D_{\rm V}$ is such that $\mathrm{order}(D_{\rm V}, V) \in \mathrm{enums}(V)$, and $\mathrm{tdeccd} \in \mathbf{N} \times \mathcal{L}(\mathbf{N}) \to \mathrm{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}))$ as

$$\begin{aligned}&\mathrm{tdeccd}(k, L) := \{((m, M), \mathrm{tdeccd}(m, M)) :\\ &\quad i \in \{1 \ldots |L|\},\ L_i > 1,\ m = kL_i(L_i - 1),\ M = L \setminus \{(i, L_i)\} \cup \{(i, L_i - 1)\}\}\end{aligned}$$

In the case of *regular substrate variables* of *valency* d and *dimension* n, the *self non-overlapping substrate decremented partition-sets tree* is defined $\mathrm{tdeccd} \in \mathbf{N} \times \mathbf{N} \to \mathrm{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}))$ as

$$\mathrm{tdeccd}(d, n) := \{((1, L), \mathrm{tdeccd}(1, L)) : L = \{1 \ldots n\} \times \{d\}\}$$

In the case of non-empty *substrate variables*, $V \neq \emptyset$, the depth is

$$\mathrm{depth}(\mathrm{tdec}(U)(V)) = \mathrm{depth}(\mathrm{tdeccd}(U)(V)) - 1 = \sum_{v \in V} (|U_v| - 1)$$

and the cardinalities are

$$|\mathrm{paths}(\mathrm{tdec}(U)(V))| = \sum (m : L \in \mathrm{paths}(\mathrm{tdeccd}(U)(V)),\ (m, \cdot) = L_{|L|})$$

and

$$|\mathrm{nodes}(\mathrm{tdec}(U)(V))| = \sum (m : L \in \mathrm{subpaths}(\mathrm{tdeccd}(U)(V)),\ (m, \cdot) = L_{|L|}) - 1$$

If the *substrate*, V, is non-empty, $n > 0$, and *regular* having *valency* $d > 1$, then the depth is

$$\mathrm{depth}(\mathrm{tdec}(U)(V)) = n(d - 1)$$

the initial cardinality of the *decrements* is

$$|\mathcal{N}_{U,V,-} \cap \mathcal{N}_{U,V,\mathrm{n,s}}| = nd(d - 1)/2$$

and the cardinalities are bounded

$$|\mathrm{paths}(\mathrm{tdec}(U)(V))| \leq (nd(d - 1)/2)^{n(d-1)}$$

and

$$|\mathrm{nodes}(\mathrm{tdec}(U)(V))| \leq n(d - 1)(nd(d - 1)/2)^{n(d-1)} \leq (nd^2)^{nd}$$

This may be compared to the explicit construction of the *self non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n,s}}$,

$$\mathcal{N}_{U,V,\mathrm{n,s}} = \prod_{w \in V} \mathrm{B}(\{w\}^{\mathrm{CS}})$$

which has cardinality of construction

$$|\mathcal{N}_{U,V,\mathrm{n,s}}| = \prod_{w \in V} |\mathrm{B}(\{w\}^{\mathrm{CS}})|$$

that is bounded $|\mathcal{N}_{U,V,\mathrm{n,s}}| \leq \mathrm{bell}(d)^n \leq d^{nd}$.

In the special case of *mono-variate substrate*, $n = 1$, the cardinality of construction is

$$|\mathrm{paths}(\mathrm{tdec}(U)(V))| = \frac{d!(d-1)!}{2^{d-1}} \leq d^{2d}$$

and

$$|\mathrm{nodes}(\mathrm{tdec}(U)(V))| = \frac{d!(d-1)!}{2^{d-1}} \sum_{j \in \{1 \ldots d-1\}} \frac{2^j}{j!(j-1)!} \leq d^{2d+1}$$

Similarly the *strong non-overlapping substrate transforms set*, $\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} \subseteq \mathcal{T}_{U,V,\mathrm{n}}$, can be constructed explicitly in terms of *strong self substrate decremented transforms* as

$$\begin{aligned}
&\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} \\
&= \{(\bigcup \mathrm{set}(L))^{\mathrm{TNT}} : M \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n}}, \\
&\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdec}(U)(M))\})\} \\
&= \{(\bigcup \mathrm{set}(L))^{\mathrm{TNT}} : Y \in \mathrm{B}(V),\ M = \{K^{\mathrm{CS}\{\}} : K \in Y\}, \\
&\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdec}(U)(M))\})\}
\end{aligned}$$

which has cardinality of construction

$$\begin{aligned}
&|\{(Y, L) : Y \in \mathrm{B}(V), L \in \mathrm{subpaths}(\mathrm{tdec}(U)(\{K^{\mathrm{CS}\{\}} : K \in Y\})) \cup \{\emptyset\}\}| \\
&= \sum_{Y \in \mathrm{B}(V)} (|\mathrm{nodes}(\mathrm{tdec}(U)(\{K^{\mathrm{CS}\{\}} : K \in Y\}))| + 1) \\
&= \sum_{Y \in \mathrm{B}(V)} \sum (m : L \in \mathrm{subpaths}(\mathrm{tdeccd}(U)(\{K^{\mathrm{CS}\{\}} : K \in Y\})),\ (m, \cdot) = L_{|L|}) \\
&\leq \mathrm{bell}(n) \times y^{2y}
\end{aligned}$$

where *volume* $y = |V^{\mathrm{CS}}|$, and *dimension* $n = |V|$. This may be compared to the cardinality of construction by *transform* pair *linear fud* which equals the cardinality of construction of the *non-overlapping substrate partition-sets* *set*, $\mathcal{N}_{U,V,\mathrm{n}}$,

$$\begin{aligned}
&|\{(Y,N) : Y \in \mathrm{B}(V),\ N \in \prod_{K \in Y} \mathrm{B}(\{K^{\mathrm{CS}\{\}}\}^{\mathrm{CS}})\}| \\
&\qquad = |\{N : Y \in \mathrm{B}(V),\ N \in \prod_{K \in Y} \mathrm{B}(K^{\mathrm{CS}})\}| \\
&\qquad = \sum_{Y \in \mathrm{B}(V)} \prod_{K \in Y} |\mathrm{B}(K^{\mathrm{CS}})| \\
&\qquad \leq \mathrm{bell}(n) \times \mathrm{bell}(y)
\end{aligned}$$

where *volume* $y = |V^{\mathrm{CS}}|$, and *dimension* $n = |V|$.

The *self non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}}$, can also be constructed more directly by means of a tree of *decremented partitions* as

$$\begin{aligned}
&\mathcal{T}_{U,V,\mathrm{n,s}} = \\
&\quad \{N^{\mathrm{TNT}} : M \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,s}},\ N \in \mathrm{elements}(\{(M, \mathrm{tpdec}(M))\})\}
\end{aligned}$$

and, in the case of non-empty *substrate*, $V \neq \emptyset$,

$$\begin{aligned}
&V \neq \emptyset \implies \mathcal{T}_{U,V,\mathrm{n,s}} = \\
&\quad \{N^{\mathrm{TNT}} : M = \{\{v\}^{\mathrm{CS}\{\}} : v \in V\},\ N \in \mathrm{elements}(\{(M, \mathrm{tpdec}(M))\})\}
\end{aligned}$$

where the *partition* tree of *self non-overlapping substrate decremented partition-sets* is defined $\mathrm{tpdec} \in \mathrm{P}(\mathcal{R}) \to \mathrm{trees}(\mathrm{P}(\mathcal{R}))$ as

$$\mathrm{tpdec}(M) := \{(N, \mathrm{tpdec}(N)) : P \in M,\ Q \in \mathrm{decs}(P),\ N = M \setminus \{P\} \cup \{Q\}\}$$

Here the subpaths of the tree of *decremented partitions* do not form *multi-layer linear fuds*. Instead the node *partition-sets* form the second *transform* of a *transform* pair *linear fud*. The *self non-overlapping substrate decremented partition-sets partition tree* maps bijectively to the *self non-overlapping substrate decremented partition-sets tree*, $\mathrm{places}(\mathrm{tpdec}(M)) :\leftrightarrow: \mathrm{places}(\mathrm{tdec}(U)(M))$ where the *value full functional partition-set* is $M = \{\{v\}^{\mathrm{CS}\{\}} : v \in V\}$, so the cardinalities of construction are equal.

The *strong non-overlapping substrate transforms set*, $\{N^{\mathrm{TV}} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} \subseteq \mathcal{T}_{U,V,\mathrm{n}}$, can be constructed using the *self non-overlapping substrate decre-*

mented partition-sets partition tree

$$\begin{aligned}
&\{N^{\mathrm{TV}} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} \\
&\quad = \{N^{\mathrm{TNT}} : M \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n}}, \\
&\qquad\qquad N \in \mathrm{elements}(\{(M, \mathrm{tpdec}(M))\})\} \\
&\quad = \{N^{\mathrm{TNT}} : Y \in \mathrm{B}(V),\ M = \{K^{\mathrm{CS}\{\}} : K \in Y\}, \\
&\qquad\qquad N \in \mathrm{elements}(\{(M, \mathrm{tpdec}(M))\})\}
\end{aligned}$$

which has cardinality of construction

$$\begin{aligned}
&|\{(Y,L) : Y \in \mathrm{B}(V), L \in \mathrm{subpaths}(\mathrm{tpdec}(\{K^{\mathrm{CS}\{\}} : K \in Y\})) \cup \{\emptyset\}\}| \\
&\quad = \sum_{Y \in \mathrm{B}(V)} \sum (m : L \in \mathrm{subpaths}(\mathrm{tdeccd}(U)(\{K^{\mathrm{CS}\{\}} : K \in Y\})),\ (m,\cdot) = L_{|L|}) \\
&\quad \leq \mathrm{bell}(n) \times y^{2y}
\end{aligned}$$

where *volume* $y = |V^{\mathrm{CS}}|$, and *dimension* $n = |V|$.

The *self non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}}$, can be constructed by means of a tree of *incremented pointed partitions* as

$$\begin{aligned}
&\mathcal{T}_{U,V,\mathrm{n,s}} = \\
&\quad \{N_*^{\mathrm{TNT}} : M \in \mathcal{N}_{U,V,1} \cap \mathcal{N}_{U,V,\mathrm{n,s}},\ M_* = \{(\{C_*\}, C_*) : \{C_*\} \in M\}, \\
&\qquad\qquad N_* \in \mathrm{elements}(\{(M_*, \mathrm{tinc}(M_*))\})\}
\end{aligned}$$

and, in the case of non-empty *substrate*, $V \neq \emptyset$,

$$\begin{aligned}
&V \neq \emptyset \implies \mathcal{T}_{U,V,\mathrm{n,s}} = \\
&\quad \{N_*^{\mathrm{TNT}} : M_* = \{(\{\{v\}^{\mathrm{CS}}\}, \{v\}^{\mathrm{CS}}) : v \in V\}, \\
&\qquad\qquad N_* \in \mathrm{elements}(\{(M_*, \mathrm{tinc}(M_*))\})\}
\end{aligned}$$

where the tree of *self non-overlapping substrate incremented pointed partition-sets* is defined $\mathrm{tinc} \in \mathrm{P}(\mathcal{R}_*) \to \mathrm{trees}(\mathrm{P}(\mathcal{R}_*))$ as

$$\mathrm{tinc}(M_*) := \{(N_*, \mathrm{tinc}(N_*)) : P_* \in M_*,\ Q_* \in \mathrm{incs}(P_*),\ N_* = M_* \setminus \{P_*\} \cup \{Q_*\}\}$$

where $\mathrm{incs} = \mathrm{increments} \in \mathcal{R}_* \to \mathrm{P}(\mathcal{R}_*)$. Again, the subpaths of the tree of *incremented pointed partitions* do not form *multi-layer linear fuds* because the *transforms* would forget the *point component*. Instead the node *pointed partition-sets* form the second *transform* of a *transform* pair *linear fud*.

The cardinality of the *self non-overlapping substrate incremented pointed* *partition-sets tree* may be computed by defining $\text{tinccd}(U) \in \text{P}(\mathcal{V}_U) \to \text{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}^2))$ as

$$\begin{aligned}&\text{tinccd}(U)(V) := \\ &\quad \{((1,L),\text{tinccd}(1,L)) : L = \{(i,(|U_v|,0)) : (v,i) \in \text{order}(D_{\text{V}},V)\}\}\end{aligned}$$

where order D_{V} is such that $\text{order}(D_{\text{V}},V) \in \text{enums}(V)$, and $\text{tinccd} \in \mathbf{N} \times \mathcal{L}(\mathbf{N}^2) \to \text{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}^2))$ is

$$\begin{aligned}&\text{tinccd}(k,L) := \\ &\quad \{((m,M),\text{tinccd}(m,M)) : i \in \{1 \ldots |L|\},\ (d,c) = L_i,\ d > 1,\ m = kd, \\ &\qquad M = L \setminus \{(i,(d,c))\} \cup \{(i,(d-1,c+1))\}\} \cup \\ &\quad \{((m,M),\text{tinccd}(m,M)) : i \in \{1 \ldots |L|\},\ (d,c) = L_i,\ d > 1,\ m = kdc, \\ &\qquad M = L \setminus \{(i,(d,c))\} \cup \{(i,(d-1,c))\}\}\end{aligned}$$

Then in the case of non-empty *substrate variables*, $V \neq \emptyset$, the depth is

$$\text{depth}(\text{tinc}(M_*)) = \text{depth}(\text{tinccd}(U)(V)) - 1 = \sum_{v \in V} (|U_v| - 1)$$

where $M_* = \{(\{\{v\}^{\text{CS}}\},\{v\}^{\text{CS}}) : v \in V\}$. The cardinalities are

$$|\text{paths}(\text{tinc}(M_*))| = \sum(m : L \in \text{paths}(\text{tinccd}(U)(V)),\ (m,\cdot) = L_{|L|})$$

and

$$|\text{places}(\text{tinc}(M_*))| = \sum(m : L \in \text{subpaths}(\text{tinccd}(U)(V)),\ (m,\cdot) = L_{|L|}) - 1$$

If the *substrate*, V, is non-empty, $n > 0$, and *regular* having *valency* $d > 1$, then the depth is

$$\text{depth}(\text{tinc}(M_*)) = n(d-1)$$

the initial cardinality of the *increments* is nd. The cardinalities are bounded

$$|\text{paths}(\text{tinc}(U)(V))| \leq (2nd^2)^{n(d-1)}$$

and

$$|\text{nodes}(\text{tinc}(U)(V))| \leq n(d-1)(2nd^2)^{n(d-1)} \leq (2nd^2)^{nd}$$

The *strong non-overlapping substrate transforms set*, $\{N^{\text{TV}} : N \in \mathcal{N}_{U,V,\text{n}}\} \subseteq \mathcal{T}_{U,V,\text{n}}$, can be constructed using the *self non-overlapping substrate incremented pointed partition-sets tree*

$$\begin{aligned}&\{N^{\text{TV}} : N \in \mathcal{N}_{U,V,\text{n}}\} \\ &\quad = \{N_*^{\text{TNT}} : M \in \mathcal{N}_{U,V,1} \cap \mathcal{N}_{U,V,\text{n}},\ M_* = \{(\{C_*\},C_*) : \{C_*\} \in M\}, \\ &\qquad\qquad N_* \in \text{elements}(\{(M_*,\text{tinc}(M_*))\})\} \\ &\quad = \{N_*^{\text{TNT}} : Y \in \text{B}(V),\ M_* = \{(\{K^{\text{CS}}\},K^{\text{CS}}) : K \in Y\}, \\ &\qquad\qquad N_* \in \text{elements}(\{(M_*,\text{tinc}(M_*))\})\}\end{aligned}$$

which has cardinality of construction

$$\begin{aligned}
&|\{(Y,L) : Y \in \mathrm{B}(V), L \in \mathrm{subpaths}(\mathrm{tinc}(U)(\{(\{K^{\mathrm{CS}}\}, K^{\mathrm{CS}}) : K \in Y\})) \cup \{\emptyset\}\}| \\
&\quad = \sum_{Y \in \mathrm{B}(V)} \sum (m : L \in \mathrm{subpaths}(\mathrm{tinccd}(U)(\{K^{\mathrm{CS}\{\}} : K \in Y\})),\ (m,\cdot) = L_{|L|}) \\
&\quad \leq \mathrm{bell}(n) \times (2y)^{2y}
\end{aligned}$$

where *volume* $y = |V^{\mathrm{CS}}|$, and *dimension* $n = |V|$.

The *substrate models set* $\mathcal{M}_{U,V}$ of *system* U and *variables* V is the set of *substrate* structures that correspond to a *transform* in the *substrate transforms set*, $\mathcal{T}_{U,V}$. Define $\mathrm{transform}(U,V) \in \mathcal{M}_{U,V} \to \mathcal{T}_{U,V}$. Thus

$$F_{U,V}, \mathcal{T}_{U,V}, \mathcal{N}_{U,V}, \mathcal{N}_{*,U,V}, \mathcal{F}_{U,V}, \mathcal{D}_{U,V}, \mathcal{D}_{\mathrm{F},U,V} \subset \mathcal{M}_{U,V}$$

Define $\mathrm{transform}(U,V)(F) := F^{\mathrm{TV}}$ where $F \in \mathcal{F}_{U,V}$. Define $\mathrm{transform}(U,V)(D) := D^{\mathrm{TV}}$ where $D \in \mathcal{D}_{U,V}$.

3.12 Independent histograms

A *histogram* A of *variables* $V = \mathrm{vars}(A)$ is said to be *partially independent* in a subset K of its *variables* $K \subseteq V$ if

$$A = \frac{1}{Z_A} * (A \ \%\ K) * (A \ \%\ (V \setminus K))$$

where $Z_A = \mathrm{scalar}(\mathrm{size}(A))$. The scaling factor ensures that the *histogram expression* has the same *size* as A, because the *size* of the product of two *histograms* of disjoint *variables* is the product of their *sizes*. In the trivial cases of $K = \emptyset$ and $K = V$ the *histogram expression* always evaluates to A.

Having considered the partitioning of *states* in the discussion of *derived variables*, consider the special case of the partitioning of *variables* which is the *partially independent set*. Given an argument *histogram* A of *variables* $V = \mathrm{vars}(A)$, construct a set of *histogram expressions* in a single free variable, each element of which evaluates to a *partially independent histogram*. The *partially independent set* R_A is

$$R_A = \{Z_A * \prod\{\frac{A}{Z_A} \ \%\ C : C \in P\} : P \in \mathrm{B}(V)\}$$

where B is the partition function.

A *histogram* A of *variables* V that is *partially independent* in a subset K is the special case of a *partially independent histogram* where the partition is binary, $P = \{K,\ V \setminus K\}$,

$$\frac{1}{Z_A} * (A \ \% \ K) * (A \ \% \ (V \setminus K)) \in R_A$$

The special case of a *partially independent histogram* where the cardinality of each of the partition components is one, i.e. the self partition, $P = V^{\{\}}$, is the *independent histogram*. Define function independent $\in \mathcal{A} \to \mathcal{A}$

$$\text{independent}(A) := Z_A * \prod\{\frac{A}{Z_A} \ \% \ \{w\} : w \in V\}$$

where $V = \text{vars}(A)$ and $Z_A = \text{scalar}(\text{size}(A))$. Define independent$(\emptyset) := \emptyset$. If A is a *zero histogram*, size$(A) = 0$, define independent$(A) := \prod\{A \ \% \ \{w\} : w \in V\}$. Define notation $A^{\text{X}} = \text{independent}(A)$.

The *independent* function can also be equivalently defined

$$\text{independent}(A) := \text{scalar}(z^{-(n-1)}) * \prod_{w \in V} A \ \% \ \{w\}$$

where $V = \text{vars}(A)$, $n = |V|$ and $z = \text{size}(A) > 0$.

The *independent histogram* is in the *partially independent set*, $A^{\text{X}} \in R_A$. The *independent histogram* has the greatest degree of *independence* of any of the *partially independent set*. See later for the definition of degree of *dependence* with respect to the *partially independent set* in the discussion of 'Minimum alignment'. The *independent histogram* is contained recursively in all the *partially independent sets* of elements of *partially independent sets*, $\forall B \in R_A \ (A^{\text{X}} \in R_B)$.

Independent histograms are such that states$(A^{\text{X}}) \supseteq$ states(A), size$(A^{\text{X}}) =$ size(A), vars$(A^{\text{X}}) =$ vars(A) and so A and A^{X} are *congruent*, congruent(A, A^{X}). Also volume$(U)(A^{\text{X}}) =$ volume$(U)(A)$ where A is in *system* U. Note, however, that independent(A) can be calculated without reference to a *system*.

A *histogram* is said to be *independent* if it is *equivalent* to its own *independent*, $A \equiv A^{\text{X}}$. An *independent histogram* is its own *independent histogram*, independent$(A^{\text{X}}) = A^{\text{X}}$. *Empty histograms* and *scalars*, $V = \emptyset$, are defined as *independent*. If the *histogram* is *mono-variate* $|V| = 1$ then it is *independent* $A = A\%\{w\} = A^{\text{X}}$ where $\{w\} = V$. *Uniform-cartesian histograms*, which are *scalar* multiples of the *cartesian*, $A = \text{scalar}(z/v) * A^{\text{C}}$ where $z = \text{size}(A)$ and $v = |A^{\text{C}}|$, including *zero histograms*, are *independent*. *Singleton histograms* $|A^{\text{F}}| = 1$ are *independent*. The *effective states* of an *independent histogram* form a *cartesian sub-volume*, that is, $A^{\text{XF}} = \prod\{(A\%\{w\})^{\text{F}} : w \in V\}$. Putting it the other way around, a *histogram* A is a *cartesian sub-volume* if $A^{\text{F}} = A^{\text{XF}}$.

The *perimeter* of a *histogram* is the set of its *reductions* indexed by *variable*. *Histogram* A in *variables* V has *perimeter* $Q_A = \{(w, A\%\{w\}) : w \in V\}$. The *histograms* of the *perimeter* have the same *size* as the given *histogram*, $\forall B \in \text{ran}(Q_A)\ (\text{size}(B) = \text{size}(A))$. The *histograms* of the *perimeter* are *integral* if the given *histogram* is *integral*, $A \in \mathcal{A}_{\text{i}} \implies \forall B \in \text{ran}(Q_A)\ (B \in \mathcal{A}_{\text{i}})$. The *independent* is constructed from the *perimeter* $A^{\text{X}} = Z_A * \prod_{w \in V}(Q_A(w)/Z_A)$ where $Z_A = \text{scalar}(\text{size}(A))$.

A *completely effective pluri-variate independent histogram*, $A^{\text{XF}} = A^{\text{C}}$, for which all of the *variables* are *pluri-valent*, $\forall w \in V\ (|A^{\text{XF}}\%\{w\}| > 1)$, must be *non-causal*, $\neg\text{causal}(A^{\text{X}})$. Thus it cannot be the *histogram* of a *functional transform*, $\forall T \in \mathcal{T}_{\text{f}}\ (A \neq \text{his}(T))$.

Given a *partially independent histogram* $A \in R_A$ of *variables* V having a partition of the *variables* P, where $A = Z_A * \prod_{K \in P}(A/Z_A)\%K$, subsets of the *variables* can be chosen such that the *reduction* is *independent*. Let $M \in P \leftrightarrow V$ be such that $\forall (K, w) \in M\ (w \in K)$, then $A\%J = (A\%J)^{\text{X}}$ where $J = \text{ran}(M)$.

Consider the *regular integral histogram* $A \in \mathcal{A}_{\text{i}}$ of *variables* $V = \text{vars}(A)$, *dimension* $n = |V|$, *valency* $\{d\} = \{|U_w| : w \in V\}$, *size* $z = \text{size}(A)$, and such that the *independent* is *completely effective*, $A^{\text{XF}} = V^{\text{C}}$. The fraction of *integral histograms congruent* to A that are *independent* may be estimated. In the case where the *histogram* is *binary*, $\{w_1, w_2\} = V$, consider *perimeter states* $\{(\cdot, x)\} \in A\%\{w_1\}$ and $\{(\cdot, y)\} \in A\%\{w_2\}$, which are such that $x, y \in \{1 \ldots z\}$. If the *histogram* is *independent*, $A = A^{\text{X}}$, then $xy = kz$, where $k \in \mathbf{N}_{>0}$. If z is prime, the fraction of the z^2 pairs $(x, y) \in \{1 \ldots z\}^2$ for which $xy = kz$ is $2z/z^2 = 2/z$. If z is the product of two primes, then the fraction is $(2z + 2p_2)/z^2$, where $p_1 p_2 = z$ and $p_1 \leq p_2$. For arbitrary *size*, z, numerical analysis suggests that the fraction is of the order of $(\ln z)/z$. Generalising to arbitrary *dimension*, n, gives $(\ln z)^{n-1}/z$. The fraction of *congruent integral histograms* that are *independent* for a given set of *perimeters* is therefore estimated as $((\ln z)^{n-1}/z)^{d^n}$. The cardinality of *perimeters* is $((z+d-1)!/(z!(d-1)!))^n$, so the cardinality of *congruent integral independent* *histograms* is estimated as

$$\left(\frac{(z+d-1)!}{z!(d-1)!} \right)^n \quad \times \quad \left(\frac{(\ln z)^{n-1}}{z} \right)^{d^n}$$

The logarithm of this cardinality approximates to

$$nd \ln \frac{z}{d} \quad - \quad d^n \ln \frac{z}{(\ln z)^{n-1}}$$

So, in the case where the logarithm of the *size* is of the order of the *valency*, $\ln z \approx d$, the logarithm of the cardinality of *congruent integral independent histograms* varies against the *volume* $v = d^n$,

$$\ln |\{A : A \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{XF}} = V^{\mathrm{C}},\ \mathrm{size}(A) = z,\ A = A^{\mathrm{X}}\}| \quad \sim \quad -v$$

That is, for a given *size*, the smaller the *volume*, the more probable an *integral histogram* is *independent*.

3.12.1 Transforms and Independent

The application $A * T$ of a *functional transform* $T \in \mathcal{T}_{\mathrm{f}}$ to a *histogram* $A \in \mathcal{A}$, such that the *underlying variables* are a subset of the *histogram variables*, $\mathrm{und}(T) \subseteq \mathrm{vars}(A)$, is called the *derived histogram*, $A * T \in \mathcal{A}$. In this context, A is called the *underlying histogram*.

A *functional transform* T is said to be *abstract* with respect to a *histogram* A if the *derived histogram*, $A * T$, is *independent*, $A * T \equiv (A * T)^{\mathrm{X}}$. Define $\mathrm{abstract} \in \mathcal{A} \times \mathcal{T}_{\mathrm{f}} \to \mathbf{B}$

$$\mathrm{abstract}(A, T) := A * T \equiv (A * T)^{\mathrm{X}}$$

where $\mathrm{size}(A) > 0$ and $\mathrm{und}(T) \subseteq \mathrm{vars}(A)$. The *independent* of the *derived histogram*, $(A * T)^{\mathrm{X}} \in \mathcal{A}$, is called the *abstract histogram*.

A *functional transform* T is said to be *formal* with respect to a *histogram* A if the *derived histogram*, $A * T$, is equivalent to the *transformed independent*, $A^{\mathrm{X}} * T$. Define $\mathrm{formal} \in \mathcal{A} \times \mathcal{T}_{\mathrm{f}} \to \mathbf{B}$

$$\mathrm{formal}(A, T) := A * T \equiv A^{\mathrm{X}} * T$$

where $\mathrm{size}(A) > 0$ and $\mathrm{vars}(A) \subseteq \mathrm{und}(T)$. The *derived* of the *independent histogram*, $A^{\mathrm{X}} * T \in \mathcal{A}$, is called the *formal histogram*. The *formal independent histogram*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} \in \mathcal{A}$, is sometimes called the *independent abstract* *histogram*.

If the *derived histogram* is equivalent to the *formal independent histogram* then the *transform* is *abstract*. That is, the *derived histogram* is equivalent to the *abstract histogram*, because the *derived histogram* is *independent*

$$A * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}} \implies A * T \equiv (A * T)^{\mathrm{X}}$$

If the *formal histogram* is equivalent to the *abstract histogram* then the *formal histogram* is *independent*

$$A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$$

and the *abstract histogram* is equivalent to the *formal independent histogram*

$$A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}} \implies (A * T)^{\mathrm{X}} \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$$

Therefore the *formal histogram* is equivalent to the *abstract histogram* only if the *abstract histogram* is equivalent to the *formal independent histogram* and the *formal histogram* is *independent*

$$((A * T)^{\mathrm{X}} \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}) \wedge (A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}) \iff A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}}$$

The *formal histogram* is equivalent to the *abstract histogram* in another stricter case if the *formal histogram* is equivalent to the *derived histogram* and the *derived histogram* is *independent*.

$$(A^{\mathrm{X}} * T \equiv A * T) \wedge (A * T \equiv (A * T)^{\mathrm{X}}) \implies A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}}$$

In this case the *transform* is *formal*, $\mathrm{formal}(A, T)$, and *abstract*, $\mathrm{abstract}(A, T)$.

At first sight, it would appear that the *derived histogram* of a *functional transform* $T \in \mathcal{T}_{\mathrm{f}}$, which has more than one *derived variable* $|\mathrm{der}(T)| \geq 2$, cannot be non-trivially *independent* if A is not *independent*, $A \neq A^{\mathrm{X}} \wedge (\forall w \in \mathrm{der}(T)\ (|(A * T\ \%\ \{w\})^{\mathrm{F}}| > 1)) \implies A * T \neq (A * T)^{\mathrm{X}}$. This is because the *transform* can only do one *reduction* to *derived variables* that are functionally synchronised, whereas the *independent* operator requires a *reduction* for each of the *derived variables*. However, there are cases where a single *reduction* is sufficient to *reduce* all of the *derived variables* so that the *derived histogram* is *independent*. A *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ in *system* U that is *non-overlapping*, $\neg\mathrm{overlap}(T)$, and such that the *derived variables* partition the *underlying variables* of a *partially independent underlying histogram*, A, must be *abstract*, $A * T = (A * T)^{\mathrm{X}}$. Let $Q \in \mathrm{B}(\mathrm{und}(T))$ be the partition of the *underlying variables* such that $\text{resize}(z, \prod\{A\%K : K \in Q\}) \equiv A$. Let $F \in \mathcal{F}_{U,1}$ be the *non-overlapping fud* having *equivalent transform*, $F^{\mathrm{T}} = T$, such that $\{\text{und}(R) : R \in F\} = Q$. Then

$$\begin{aligned}
A * T &= A * F^{\mathrm{T}} \\
&= Z_n * \prod_{K \in Q} A\%K * \prod_{R \in F} \text{his}(R) \ \% \bigcup_{R \in F} \text{der}(R) \\
&= Z_n * \prod\{A\%K * \text{his}(R) \ \% \ \text{der}(R) : K \in Q,\ R \in F,\ \text{und}(R) = K\} \\
&= (A * F^{\mathrm{T}})^{\mathrm{X}} \\
&= (A * T)^{\mathrm{X}}
\end{aligned}$$

where $Z_n = \text{scalar}(z/z^n)$ and $n = |\text{und}(T)|$.

The *independent underlying histogram* A^{X} is a *partially independent histogram* by definition and so it follows that for *non-overlapping transforms* the *formal histogram*, $A^{\mathrm{X}} * T$, is *independent*, $\neg\text{overlap}(T) \implies A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$.

However, note that it not always the case that the converse, $A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}} \implies \neg\text{overlap}(T)$, is true. That is, the *formal histogram* is *independent* for some *overlapping transforms*, $\exists A \in \mathcal{A}_U\ \exists T \in \mathcal{T}_{U,\mathrm{f},1}\ (\text{overlap}(T) \wedge (A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}))$. For example, a *tautology* having *singleton derived*, $\text{tautology}(T) \wedge |W^{\mathrm{C}}| = 1$. Or for example, let $\{P_1, P_2\} = W \subset \mathrm{B}(V^{\mathrm{CS}})$ where $V = \text{und}(T)$ and $W = \text{der}(T)$. Then the *formal* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, if

$$\forall C_1 \in P_1\ \forall C_2 \in P_2\ (\text{size}(A^{\mathrm{X}} * C_1^{\mathrm{U}} * C_2^{\mathrm{U}}) = \frac{1}{z} * \text{size}(A^{\mathrm{X}} * C_1^{\mathrm{U}}) * \text{size}(A^{\mathrm{X}} * C_2^{\mathrm{U}}))$$

where $z = \text{size}(A)$. This condition may be satisfied by some *overlapping transforms*. In these cases the *components* intersect, $\forall C_1 \in P_1\ \forall C_2 \in P_2\ (C_1 \cap C_2 \neq \emptyset)$, and the *transform* is *right total*, $(X\%W)^{\mathrm{F}} = W^{\mathrm{C}}$ where $X = \text{his}(T)$.

The *formal histogram* is equivalent to the *abstract histogram* if the *transform* is *non-overlapping*, $\neg\text{overlap}(T) \implies A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$, and the *abstract histogram* is equivalent to the *formal independent histogram*

$$\neg\text{overlap}(T) \wedge ((A * T)^{\mathrm{X}} \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}) \implies A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}}$$

Given *congruent histograms* $A, B \in \mathcal{A}$, in *variables* $V = \text{vars}(A) = \text{vars}(B)$, and *substrate transform* $T \in \mathcal{T}_{U,V}$, having *derived variables* $W = \text{der}(T)$, the *abstracts* are equal if all of the *derived reductions* to *partition variable* are equal,

$$(B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \iff \forall P \in W\ (B * T \ \%\ \{P\} = A * T \ \%\ \{P\})$$

or

$$(B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \iff \forall P \in W\ (B * P^{\mathrm{T}} = A * P^{\mathrm{T}})$$

In the special case where $B = A^{\mathrm{X}}$, the *formal independent*, or *independent abstract*, equals the *abstract* if and only if each *partition derived* equals its *partition independent derived*,

$$\begin{aligned}
&(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \iff \\
&\qquad \forall P \in W\ (A * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}}) \\
&= \quad \forall P \in W\ \forall (R, \cdot) \in (P^{\mathrm{T}})^{-1}\ ((A * P^{\mathrm{T}})_R = (A^{\mathrm{X}} * P^{\mathrm{T}})_R) \\
&= \quad \forall P \in W\ \forall C \in P\ (\text{size}(A * C^{\mathrm{U}}) = \text{size}(A^{\mathrm{X}} * C^{\mathrm{U}}))
\end{aligned}$$

In the case where the *formal* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, the *formals* are equal if and only if all of the *independent derived reductions* to *partition variable* are equal,

$$\begin{aligned}
&A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}} \implies \\
&\qquad B^{\mathrm{X}} * T = A^{\mathrm{X}} * T \iff \forall P \in W\ (B^{\mathrm{X}} * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}})
\end{aligned}$$

This is the case where the *transform* is *non-overlapping*, $\neg\text{overlap}(T) \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$. In this case the equality can be *reduced* to the *underlying variables* of the *partition*,

$$\begin{aligned}
&\neg\text{overlap}(T) \implies \\
&\qquad B^{\mathrm{X}} * T = A^{\mathrm{X}} * T \iff \forall P \in W\ (B^{\mathrm{X}}\%V_P * P^{\%\mathrm{T}} = A^{\mathrm{X}}\%V_P * P^{\%\mathrm{T}})
\end{aligned}$$

where $V_P = \text{vars}(P^{\%}) \subseteq V$.

If the *transform* is both *non-overlapping* and the *formal independent* equals the *abstract*, then the constraint on the *partition transforms* can be expressed in terms of the *contraction* of the *partition variable*,

$$\begin{aligned}
&\neg\text{overlap}(T)\ \wedge\ (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \iff \\
&\qquad\qquad \forall P \in W\ (A\%V_P * P^{\%\mathrm{T}} = A^{\mathrm{X}}\%V_P * P^{\%\mathrm{T}})
\end{aligned}$$

The *application* of an *action*, $\text{action}(C, A)$ where $C \in \text{actions} \subset \mathcal{T} \times \mathcal{T}$ and $A \in \mathcal{A}$, may be *independent* even if A is not. This is because the second *transform* R of the pair $(L, R) = C$ need not be *functional*. A trivial example is where L is a *null transform* and thus A is *reduced* to a *scalar*.

3.12.2 Independent converse

The *simple converse* $\text{converseSimple} \in \mathcal{T} \to \mathcal{T}$ and *natural converse* $\text{converseNatural} \in \mathcal{T} \to \mathcal{T}$ for a *transform* $T \in \mathcal{T}$ have already been defined. The *independent converse* is defined $\text{converseIndependent} \in \mathcal{A} \times \mathcal{T}_{\text{f}} \to \mathcal{T}$

$$\begin{aligned}&\text{converseIndependent}(B, T) :=\\ &\quad (\sum \frac{(B * C)^{\text{X}}}{(B * C)\%\emptyset} * \{R\}^{\text{U}} : (R, C) \in \text{inverse}(T),\ V)\end{aligned}$$

where $\text{size}(B) > 0$ and $\text{vars}(B) = V = \text{underlying}(T)$. Define notation

$$T^{\dagger B} = \text{converseIndependent}(B, T)$$

Unlike the other *converses*, the argument *transform* must be *functional* $T \in \mathcal{T}_{\text{f}}$. This is the case if T is the *transform* of a *partition* $P \in \mathcal{R}_U$, $T = \text{transform}(U)(P)$, because then $T \in \mathcal{T}_{U,\text{f},1}$.

Also the *independent converse* requires an extra argument $B \in \mathcal{A}$ to provide the *independent histogram* for each *component* of the *functional transform*. The *histogram* B must be *non-zero*, $\text{size}(B) > 0$, and have the same *variables* as the *underlying variables* of the *transform*, $\text{vars}(B) = \text{underlying}(T)$.

The *action* of a *one functional transform* $T \in \mathcal{T}_{U,\text{f},1}$ and its *independent converse*, $(T, T^{\dagger B})$, is *size* conserving if all of the *components* of T are *non-zero* when applied to B. Thus

$$\text{size}(A * T * T^{\dagger B}) = \text{size}(A)$$

if $\forall C \in \text{ran}(\text{inverse}(T))\ (\text{size}(B * C) > 0)$ or $(A * T)^{\text{F}} \leq (B * T)^{\text{F}}$. Thus $\text{size}((A * B^{\text{F}}) * T * T^{\dagger B}) = \text{size}(A * B^{\text{F}})$. When $B = A$, then the *action* $(T, T^{\dagger A})$ is always *size* conserving $\text{size}(A * T * T^{\dagger A}) = \text{size}(A)$ irrespective of whether *zero components* exist.

If any of the *components* of the *one functional transform* $T \in \mathcal{T}_{U,\text{f},1}$ are not *cartesian sub-volumes* $\exists C \in \text{ran}(\text{inverse}(T))\ ((B * C)^{\text{F}} < (B * C)^{\text{XF}})$ then the *independent converse* may be more *effective*

$$(A * T * T^{\dagger B})^{\text{F}} \geq A^{\text{F}}$$

If the given *histogram* is *unit cartesian* $B = V^{\mathrm{C}}$ and all of the *components* of *one funtional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ are *cartesian sub-volumes* $\forall C \in \mathrm{ran}(\mathrm{inverse}(T))\ (C^{\mathrm{F}} = C^{\mathrm{XF}})$ then

$$\mathrm{converseIndependent}(V^{\mathrm{C}}, T) = \mathrm{converseNatural}(T)$$

or $T^{\dagger V^{\mathrm{C}}} = T^{\dagger}$. Let $(X, W) = T$

$$\begin{aligned}
\frac{X^{\mathrm{F}}}{X\%W} &= \sum\{\frac{C^{\mathrm{F}}}{C\%\emptyset} * \{R\}^{\mathrm{U}} : (R, C) \in \mathrm{inverse}(T)\} \\
&= \sum\{\frac{(V^{\mathrm{C}} * C)^{\mathrm{F}}}{(V^{\mathrm{C}} * C)\%\emptyset} * \{R\}^{\mathrm{U}} : (R, C) \in \mathrm{inverse}(T)\} \\
&= \sum\{\frac{(V^{\mathrm{C}} * C)^{\mathrm{X}}}{(V^{\mathrm{C}} * C)\%\emptyset} * \{R\}^{\mathrm{U}} : (R, C) \in \mathrm{inverse}(T)\}
\end{aligned}$$

where the last step holds only if all *components* of T, the partition of V^{C}, are *cartesian sub-volumes*.

The application to A of the *independent converse action* $(T, T^{\dagger A})$ of a *unit functional transform* $T \in \mathcal{T}_{\mathrm{f,U}}$ with respect to A is called the *idealisation*, $A * T * T^{\dagger A}$. If the *idealisation* of A is equivalent to A then the *transform* is *ideal*. Define $\mathrm{ideal} \in \mathcal{A} \times \mathcal{T}_{\mathrm{f,U}} \to \mathbf{B}$

$$\mathrm{ideal}(A, T) := A * T * T^{\dagger A} \equiv A$$

where $\mathrm{size}(A) > 0$ and $\mathrm{vars}(A) = \mathrm{und}(T)$. An *idealisation* is *size-conserving*, $\mathrm{size}(A * T * T^{\dagger A}) = \mathrm{size}(A)$. An *idealisation* is *effective* if the *transform* is *effective* with respect to the *histogram*, $(X\%V)^{\mathrm{F}} \geq A^{\mathrm{F}}$ where $X = \mathrm{his}(T)$ and $V = \mathrm{und}(T)$. An *ideal transform* must be *effective*. A *one functional transform*, $T \in \mathcal{T}_{U,\mathrm{f},1}$, is always *effective*.

An *idealisation* can be defined as the *summation* of its *independent components*

$$A * T * T^{\dagger A} \equiv \sum_{C \in T^{\mathrm{P}}} (A * C^{\mathrm{U}})^{\mathrm{X}}$$

or,

$$A * T * T^{\dagger A} \equiv \sum_{(R,C) \in T^{-1}} (A * C)^{\mathrm{X}}$$

In some cases, but not all, the *components* of an *effective ideal transform* T with respect to *histogram* A, $\mathrm{ideal}(A, T)$, are *cartesian sub-volumes*, $\exists A \in \mathcal{A}\ \exists T \in \mathcal{T}_{\mathrm{f,U}}\ (\mathrm{ideal}(A, T) \wedge \forall C \in T^{\mathrm{P}}\ ((A * C^{\mathrm{U}})^{\mathrm{F}} = (A * C^{\mathrm{U}})^{\mathrm{XF}}))$.

The *independent* of an *effective idealisation* equals the *independent* of the given *histogram*

$$(A * T * T^{\dagger A})^{\mathrm{X}} \equiv A^{\mathrm{X}}$$

This is true because the sum of the *perimeters* of the *components* are equal, $\forall w \in V$

$$\begin{aligned} A * T * T^{\dagger A} \ \% \ \{w\} &\equiv \sum_{C \in T^{\mathrm{P}}} A * T * T^{\dagger A} * C^{\mathrm{U}} \ \% \ \{w\} \\ &\equiv \sum_{C \in T^{\mathrm{P}}} (A * C^{\mathrm{U}})^{\mathrm{X}} \ \% \ \{w\} \\ &\equiv \sum_{C \in T^{\mathrm{P}}} A * C^{\mathrm{U}} \ \% \ \{w\} \\ &\equiv A\%\{w\} \\ &\equiv A^{\mathrm{X}}\%\{w\} \end{aligned}$$

where $V = \mathrm{und}(T) = \mathrm{vars}(A)$. Thus

$$(A * T * T^{\dagger A})^{\mathrm{X}} \equiv \left(\sum_{C \in T^{\mathrm{P}}} (A * C^{\mathrm{U}})^{\mathrm{X}} \right)^{\mathrm{X}} \equiv A^{\mathrm{X}}$$

A special case of an *ideal transform* is where the *transform* is a self-partition. In this case, the *transform* is *ideal* with respect to any *histogram* A in its underlying *variables* $\mathrm{vars}(A) = V = \mathrm{und}(T)$, so long as the *transform* T is as *effective* as A

$$((X\%V)^{\mathrm{F}} \geq A^{\mathrm{F}}) \wedge (\forall C \in \mathrm{ran}(\mathrm{inverse}(T)) \ (|C| = 1)) \implies \mathrm{ideal}(A, T)$$

where $X = \mathrm{histogram}(T)$. In this case, each *component* is a *singleton* and hence is *independent*. An example is the *self partition transform* $V^{\mathrm{CS\{\}T}} \in \mathcal{T}_{U,V} \subset \mathcal{T}_{U,\mathrm{f},1}$ which is always *ideal*, $\mathrm{ideal}(A, V^{\mathrm{CS\{\}T}})$. Another example is the *value full functional transform* $\{\{w\}^{\mathrm{CS\{\}T}} : w \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V} \subset \mathcal{T}_{U,\mathrm{f},1}$ which is always *ideal*, $\mathrm{ideal}(A, \{\{w\}^{\mathrm{CS\{\}T}} : w \in V\}^{\mathrm{T}})$.

If a *transform* T is a unary-partition, $|X\%W| = 1$ where $(X, W) = T$, then T is an *ideal transform* of A^{X} if T is as *effective* as A^{X}, $(X\%V)^{\mathrm{F}} \geq A^{\mathrm{XF}}$ where $\mathrm{vars}(A) = V = \mathrm{und}(T)$ and A^{X} is *non-zero*, $\mathrm{size}(A^{\mathrm{X}}) > 0$,

$$((X\%V)^{\mathrm{F}} \geq A^{\mathrm{XF}}) \wedge (|\mathrm{inverse}(T)| = 1) \implies \mathrm{ideal}(A^{\mathrm{X}}, T)$$

It is also the case that the *idealisation*, $A * T * T^{\dagger A}$, equals the *independent*, A^{X}, $A * T * T^{\dagger A} = A^{\mathrm{X}}$. An example is the *unary partition transform*

$\{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V} \subset \mathcal{T}_{U,\mathrm{f},1}$ which is always an *ideal transform* of the *independent histogram* A^{X}, $\mathrm{ideal}(A^{\mathrm{X}}, \{V^{\mathrm{CS}}\}^{\mathrm{T}})$ or $A * \{V^{\mathrm{CS}}\}^{\mathrm{T}} * \{V^{\mathrm{CS}}\}^{\mathrm{T}\dagger A} = A^{\mathrm{X}}$.

A *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ must be *ideal* with respect to *histogram* A if each of the *effective states* of A is in a separate *component*, $A^{\mathrm{F}} \leftrightarrow T^{\mathrm{P}}$, because each *component* is a *singleton* and therefore *independent*, $\forall C \in T^{\mathrm{P}}\ (|(A * C^{\mathrm{U}})^{\mathrm{F}}| \leq 1)$. Thus, in the case where the cardinality of *effective states* is less than the *volume*, $b < v$, there must exist at least $\mathrm{stir}(v - b, b)$ *ideal partition transforms*, where $v = |V^{\mathrm{C}}|$, $b = |A^{\mathrm{F}}|$ and $\mathrm{stir} \in \mathbf{N}_{>0} \times \mathbf{N} \to \mathbf{N}_{>0}$ is the Stirling number of the second kind. This is the case if the *size* is less than the *volume*, $z < v$, where $z = \mathrm{size}(A)$. Thus the cardinality of the set of *ideal substrate transforms* is bounded $|\{T : T \in \mathcal{T}_{U,V},\ A * T * T^{\dagger A} \equiv A\}| \geq \mathrm{stir}(v - b, b)$. This lower bound may be compared to the cardinality of the *substrate partition transforms* $|F_{U,V}| = |\{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\}| = |\mathrm{B}(V^{\mathrm{CS}})| = \mathrm{bell}(v) = \sum_{k \in 0 \ldots v} \mathrm{stir}(v, k)$. Note that the Stirling number of the second kind, $\mathrm{stir}(n, k)$ is maximised where $k \approx n/\ln n$ for large n. So conjecture that the maximisation of the fraction of the cardinality of the *idealisations* per cardinality of the *substrate transforms*, $|\{T : T \in \mathcal{T}_{U,V},\ A * T * T^{\dagger A} \equiv A\}|/|\mathcal{T}_{U,V}|$, occurs approximately where $b \approx v/\ln v$.

The *nullable transform* D^{T} of a *well behaved distinct decomposition* $D \in \mathcal{D}_{\mathrm{w},U}$ is *ideal* with respect to *histogram* A if each of the *effective states* of A is in a separate *component* or *slice*, $A^{\mathrm{F}} \leftrightarrow D^{\mathrm{TP}}$. That is, $\forall C \in D^{\mathrm{TP}}\ (|(A * C^{\mathrm{U}})^{\mathrm{F}}| \leq 1) \implies \mathrm{ideal}(A, D^{\mathrm{T}})$. This is also true of the *partition transform*, D^{PT}, because both have the same *partition*, $D^{\mathrm{TP}} = D^{\mathrm{PTP}} = D^{\mathrm{P}}$. The *decomposition* is said to be *effectively sliced* with respect to the *histogram*, A. Note that a *decomposition* may be *ideal* even when not *effectively sliced*. If a *sub-decomposition* $E \in \mathrm{subtrees}(D)$ is *effectively sliced* with respect to A, $\forall C \in E^{\mathrm{P}}\ (|(A * C^{\mathrm{U}})^{\mathrm{F}}| \leq 1)$, then D must also be *effectively sliced*, because the *expanded partition* of E is a *parent partition*, $\mathrm{parent}(E^{\mathrm{P}V}, D^{\mathrm{P}})$.

A special case of an *ideal transform* is a *naturally ideal transform*

$$A * T * T^{\dagger} = A$$

If it is the case that all of the *components* of $T \in \mathcal{T}_{U,\mathrm{f},1}$ are *cartesian sub-volumes*, $\forall C \in \mathrm{ran}(\mathrm{inverse}(T))\ (C^{\mathrm{F}} = C^{\mathrm{XF}})$, then $T^{\dagger V^{\mathrm{C}}} = T^{\dagger}$ and so T is a *naturally ideal transform* with respect to V^{C}, $V^{\mathrm{C}} * T * T^{\dagger} = V^{\mathrm{C}}$, where $V = \mathrm{und}(T)$.

If T is a *naturally ideal transform* of *histogram* A then each of the *components* must be *uniform*

$$\forall C \in \text{ran}(\text{inverse}(T))\ (|\text{ran}(A * C)| = 1)$$

Given the *naturally ideal transform* T and the *sizes* of each of the *components* $Q = \{(R, \text{size}(A * C)) : (R, C) \in \text{inverse}(T)\}$, then A can be reconstructed

$$A = \sum\{\text{scalar}(Q_R) * \frac{C}{C\%\emptyset} : (R, C) \in \text{inverse}(T)\}$$

Similarly, a *histogram* A can be reconstructed from an *ideal transform* T, where the *components* of the *effective transform* are *cartesian sub-volumes*, given the perimeters of each of these *components*

$$Q = \{(R, \{B\%\emptyset\} \cup \{\frac{B\%\{w\}}{B\%\emptyset} : w \in V\}) : (R, C) \in \text{inv}(\text{eff}(A, T)),\ B = A * C\}$$

where inv = inverse and eff = effective. So

$$A = \sum\{\prod Q_R : R \in \text{dom}(Q)\}$$

Let *non-zero* sample *histogram* $A \in \mathcal{A}_U$ have non-empty *variables* $V = \text{vars}(A) \neq \emptyset$. The normalisation is a *probability histogram*, $\hat{A} \in \mathcal{A} \cap \mathcal{P}$. Let *non-zero* query *histogram* $Q \in \mathcal{A}_U$ have *variables* $K = \text{vars}(Q)$ that are a subset of the sample *variables*, $K \subseteq V$. The normalisation of the query *histogram* is a *probability histogram*, $\hat{Q} \in \mathcal{A} \cap \mathcal{P}$. The difference between the sample *variables* and the query *variables*, $V \setminus K$, is called the set of label *variables*. As discussed above in section 'Transforms', given a *one functional transform* $T = (M, W) \in \mathcal{T}_{U,\text{f},1}$, having *underlying variables* $J = \text{und}(T)$, the *model* analog of the *transformed conditional product*, $\hat{Q} * T'_A = \hat{Q} * (A/(A\%K), (V \setminus K)) \in \mathcal{A} \cap \mathcal{P}'$, is the renormalisation of the application of the normalised *sample action*, $(T, (\hat{A} * M, V))$, to the expanded query *probability histogram*, $\hat{Q}_J = \hat{Q} * (J \setminus K)^{\text{C}\wedge} \in \mathcal{A} \cap \mathcal{P}$,

$$(\hat{Q}_J * T * (\hat{A} * M, V))^\wedge\ \%\ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}$$

if the intersection of *derived effective states* is not empty, $(Q * T)^{\text{F}} \cap (A * T)^{\text{F}} \neq \emptyset$. The *modelled transformed conditional product* may be expressed in terms of the *actual converse transform*,

$$\hat{Q}_J * T * T^{\odot A}\ \%\ (V \setminus K) \in \mathcal{A} \cap \mathcal{P}$$

In the case where the *transform* is *ideal* with respect to the sample *histogram*, ideal(A,T), the *actual converse* equals the *ideal converse*, $T^{\odot A} = T^{\dagger A}$, and so

$$\begin{aligned}
& \hat{Q}_J * T * T^{\odot A} \ \% \ (V \setminus K) \\
= \ & \hat{Q}_J * T * T^{\dagger A} \ \% \ (V \setminus K) \\
= \ & \hat{Q}_J * T * \sum_{(R,C) \in T^{-1}} \{R\}^{\mathrm{U}} * (A * C)^{\mathrm{X}\wedge} \ \% \ (V \setminus K)
\end{aligned}$$

3.12.3 Actual converse and Independent

The *actual converse* is very similar to the *independent converse* except that the literal application of the *component* to the argument *histogram* is used, rather than the *independent* of the applied *component*. The *actual converse* is defined above as converseActual $\in \mathcal{A} \times \mathcal{T}_{\mathrm{f}} \to \mathcal{T}$

$$\begin{aligned}
& \text{converseActual}(B,T) := \\
& \quad (\sum \frac{B * C}{(B * C)\%\emptyset} * \{R\}^{\mathrm{U}} : (R,C) \in \text{inverse}(T), \ V)
\end{aligned}$$

where size$(B) > 0$ and vars$(B) = V =$ underlying(T). Define notation

$$T^{\odot B} = \text{converseActual}(B,T)$$

Like the *independent converse*, the argument *transform* must be *functional* $T \in \mathcal{T}_{\mathrm{f}}$. The *actual converse* may be expressed more concisely,

$$T^{\odot B} \ := \ (\sum_{(R,C) \in T^{-1}} \{R\}^{\mathrm{U}} * (B * C)^{\wedge}, \ V) \tag{4}$$

where the normalisation is defined $\hat{A} = A/(A\%\emptyset)$ so that normalised *zero histograms* are *empty*, $(V^{\mathrm{CZ}})^{\wedge} = \emptyset$.

The *actual converse* of *unit functional transform* $T \in \mathcal{T}_{\mathrm{f,U}}$ of a *histogram* A applied to the *abstract histogram* $(A * T)^{\mathrm{X}}$ is called the *surrealisation*, $(A * T)^{\mathrm{X}} * T^{\odot A}$. The *surrealisation* is *equivalent* to the *histogram* only if the *transform* is *abstract*, abstract(A,T),

$$A * T \equiv (A * T)^{\mathrm{X}} \iff (A * T)^{\mathrm{X}} * T^{\odot A} \equiv A$$

An example of an *abstract transform* is the *unary partition transform* $\{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$ which is always *abstract*, abstract$(A, \{V^{\mathrm{CS}}\}^{\mathrm{T}})$. Another example is the *self partition transform* $V^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$. In this case there is only one *derived variable* and so the *derived histogram* is *independent*, $\mathrm{abstract}(A, V^{\mathrm{CS}\{\}\mathrm{T}})$. A *value full functional transform* is *abstract* if the *histogram* is *independent*, $\mathrm{abstract}(A^{\mathrm{X}}, \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}})$.

A *surrealisation* is *size-conserving* if the *derived histogram* is as *effective* as the *abstract histogram*, $(A * T)^{\mathrm{F}} = (A * T)^{\mathrm{XF}} \implies \mathrm{size}((A * T)^{\mathrm{X}} * T^{\odot A}) = \mathrm{size}(A)$. Otherwise the *size* of the *surrealisation* is less than the *size* of the *histogram*, $(A * T)^{\mathrm{F}} < (A * T)^{\mathrm{XF}} \implies \mathrm{size}((A * T)^{\mathrm{X}} * T^{\odot A}) < \mathrm{size}(A)$. An *abstract transform* is necessarily *size-conserving*.

The *independent* of an *effective surrealisation* equals the *independent* of the given *histogram* if the *transform* is *abstract*, $\mathrm{abstract}(A, T)$

$$A * T \equiv (A * T)^{\mathrm{X}} \implies ((A * T)^{\mathrm{X}} * T^{\odot A})^{\mathrm{X}} \equiv A^{\mathrm{X}}$$

because the *surrealisation* is *equivalent* to the *histogam*, $(A * T)^{\mathrm{X}} * T^{\odot A} \equiv A$.

In the case where the *formal histogram* is *independent*, $A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$, then the *surrealisation* of the *independent*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\odot A^{\mathrm{X}}}$, equals the *independent*, A^{X}

$$A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}} \implies (A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\odot A^{\mathrm{X}}} \equiv A^{\mathrm{X}}$$

This is the case where the *transform* is *non-overlapping* because the *formal histogram* is *independent*, $\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$.

A *histogram* A can be reconstructed from an *abstract transform* T if the *components* of the *transform* are *diagonalised*, $\forall C \in T^{\mathrm{P}}\ (\mathrm{diagonal}((A * C^{\mathrm{U}})^{\mathrm{F}}))$, given the choice of *diagonal* and the *counts* along it. In the case where the applied *component* is a *fully diagonalised regular cartesian volume* the cardinality of *diagonals* is $(d!)^{n-1}$ where $n = |\mathrm{vars}(A)|$ and $d = |(B * C)^{\mathrm{F}}|$. Then the choice of *diagonal* is in $\{1 \ldots (d!)^{n-1}\}$.

In the case where the *transform* is a *substrate transform* $T \in \mathcal{T}_{U,V}$, having *derived variables* $W = \mathrm{der}(T)$, the *derived* of the *partition transform* is *independent*, $\forall P \in W\ (A * P^{\mathrm{T}} = (A * P^{\mathrm{T}})^{\mathrm{X}})$, because the *partition transforms* are *mono-derived-variate*, $|\mathrm{der}(P^{\mathrm{T}})| = 1$. So the *surrealisations* of the *partition transforms* equal the *histogram*,

$$\forall P \in W\ ((A * P^{\mathrm{T}})^{\mathrm{X}} * P^{\mathrm{T}\odot A} = A)$$

The *actual converse* of *unit functional transform* $T \in \mathcal{T}_{\mathrm{f,U}}$ of a *histogram* A applied to the *formal histogram* $A^{\mathrm{X}} * T$ is called the *contentisation*, $A^{\mathrm{X}} * T * T^{\odot A}$. A *transform* T *formal* with respect to a *histogram* A if the *derived histogram*, $A * T$, equals the *formal histogram*, $A^{\mathrm{X}} * T$. This is the case if the *contentisation* of A is *equivalent* to A. The *contentisation* is *equivalent* to the *histogram* only if the *transform* is *formal*, $\mathrm{formal}(A, T)$

$$A * T \equiv A^{\mathrm{X}} * T \iff A^{\mathrm{X}} * T * T^{\odot A} \equiv A$$

An example of a *formal transform* is the *unary partition transform* $\{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$ which is always *formal*, $\mathrm{formal}(A, \{V^{\mathrm{CS}}\}^{\mathrm{T}})$. All *effective unit functional transforms* are *formal* if the *histogram* is *independent*, $A = A^{\mathrm{X}} \implies A * T = A^{\mathrm{X}} * T$. A *value full functional transform* is *formal* only if the *histogram* is *independent*, $\mathrm{formal}(A^{\mathrm{X}}, \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}})$.

A *contentisation* is *size-conserving* if the *derived histogram* is as *effective* as the *formal histogram*, $(A*T)^{\mathrm{F}} \geq (A^{\mathrm{X}}*T)^{\mathrm{F}} \implies \mathrm{size}(A^{\mathrm{X}}*T*T^{\odot A}) = \mathrm{size}(A)$. Otherwise the *size* of the *contentisation* is less than the *size* of the *histogram*, $(A*T)^{\mathrm{F}} < (A^{\mathrm{X}}*T)^{\mathrm{F}} \implies \mathrm{size}(A^{\mathrm{X}}*T*T^{\odot A}) < \mathrm{size}(A)$. A *formal transform* is necessarily *size-conserving*.

The *effective contentisation* of an *independent* equals the *independent*

$$A^{\mathrm{X}} * T * T^{\odot A^{\mathrm{X}}} \equiv A^{\mathrm{X}}$$

The *contentisation* equals the *histogram* if the *transform* is *formal*, $\mathrm{formal}(A, T)$, by definition

$$A * T \equiv A^{\mathrm{X}} * T \implies A^{\mathrm{X}} * T * T^{\odot A} \equiv A$$

The *independent* of a *contentisation* equals the *independent* of the given *histogram* if the *transform* is *formal*, $\mathrm{formal}(A, T)$

$$A * T \equiv A^{\mathrm{X}} * T \implies (A^{\mathrm{X}} * T * T^{\odot A})^{\mathrm{X}} \equiv A^{\mathrm{X}}$$

because the *contentisation* is *equivalent* to the *histogam*, $A^{\mathrm{X}} * T * T^{\odot A^{\mathrm{X}}} \equiv A$.

The *contentisation* equals the *surrealisation* if the *formal histogram* is equivalent to the *abstract histogram*, $A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}}$

$$A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} * T * T^{\odot A} \equiv (A * T)^{\mathrm{X}} * T^{\odot A}$$

In the case where the *formal histogram* is *independent*, $A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$, then the *surrealisation* of the *independent*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\odot A^{\mathrm{X}}}$, equals *contentisation* of the *independent*, $A^{\mathrm{X}} * T * T^{\odot A^{\mathrm{X}}}$, which equals the *independent*,

$A^{\rm X}$

$$A^{\rm X} * T \equiv (A^{\rm X} * T)^{\rm X} \implies (A^{\rm X} * T)^{\rm X} * T^{\odot A^{\rm X}} \equiv A^{\rm X} * T * T^{\odot A^{\rm X}} \equiv A^{\rm X}$$

This is the case where the *transform* is *non-overlapping*, $\neg\mathrm{overlap}(T) \implies A^{\rm X} * T \equiv (A^{\rm X} * T)^{\rm X}$.

The *actual converse* of *unit functional transform* $T \in \mathcal{T}_{\rm f,U}$ of the *independent histogram* $A^{\rm X}$ applied to the *derived histogram* $A*T$ is called the *neutralisation*, $A*T*T^{\odot A^{\rm X}}$. A *neutralisation* is *size-conserving*, $\mathrm{size}(A*T*T^{\odot A^{\rm X}}) = \mathrm{size}(A)$. If the *transform* is a *unary partition transform* $\{V^{\rm CS}\}^{\rm T} \in \mathcal{T}_{U,\rm f,1}$ then the *neutralisation* equals the *independent*, $A * \{V^{\rm CS}\}^{\rm T} * \{V^{\rm CS}\}^{{\rm T}\odot A^{\rm X}} = A^{\rm X}$. If the *transform* is a *full functional transform*, for example a *value full functional transform* $\{\{w\}^{\rm CS\{\}T} : w \in V\}^{\rm T}$, then the *neutralisation* equals the *histogram* only if the *histogram* is *independent*, $A^{\rm X} * \{\{w\}^{\rm CS\{\}T} : w \in V\}^{\rm T} * \{\{w\}^{\rm CS\{\}T} : w \in V\}^{{\rm T}\odot A^{\rm X}} = A^{\rm X}$.

The *effective neutralisation* of an *independent* equals the *independent*

$$A^{\rm X} * T * T^{\odot A^{\rm X}} \equiv A^{\rm X}$$

The *effective neutralisation* equals the *independent* if the *transform* is *formal*, $\mathrm{formal}(A,T)$,

$$A * T \equiv A^{\rm X} * T \implies A * T * T^{\odot A^{\rm X}} \equiv A^{\rm X}$$

So the *independent* of a *neutralisation* equals the *independent* if the *transform* is *formal*

$$A * T \equiv A^{\rm X} * T \implies (A * T * T^{\odot A^{\rm X}})^{\rm X} \equiv A^{\rm X}$$

If a *transform* T is *formal*, $\mathrm{formal}(A,T)$, then the *contentisation* equals the *histogram*, $A^{\rm X} * T * T^{\odot A} \equiv A$, and the *neutralisation* equals the *independent*, $A * T * T^{\odot A^{\rm X}} \equiv A^{\rm X}$

$$A^{\rm X} * T * T^{\odot A} \equiv A \iff A * T * T^{\odot A^{\rm X}} \equiv A^{\rm X}$$

The *neutralisation* equals the *idealisation* when each of the *components* of the *independent* equals the *independent component* of the *histogram*

$$\forall C \in T^{\rm P}\ (A^{\rm X} * C^{\rm U} = (A * C^{\rm U})^{\rm X}) \iff A * T * T^{\odot A^{\rm X}} = A * T * T^{\dagger A}$$

In the case where the *transform* is a *substrate transform* $T \in \mathcal{T}_{U,V}$ and the *formal independent* equals the *abstract*, $(A^{\rm X} * T)^{\rm X} = (A * T)^{\rm X}$, then the *neutralisations* of the *partition transforms* equal the *independent*,

$$\begin{aligned} & (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \iff \\ & \quad \forall P \in W\ (A * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}}) \\ = & \quad \forall P \in W\ (A * P^{\mathrm{T}} * P^{\mathrm{T}\odot A^{\mathrm{X}}} = A^{\mathrm{X}}) \end{aligned}$$

and the *contentisations* of the *partition transforms* equal the *histogram*,

$$\begin{aligned} & (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \iff \\ & \quad \forall P \in W\ (A * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}}) \\ = & \quad \forall P \in W\ (A = A^{\mathrm{X}} * P^{\mathrm{T}} * P^{\mathrm{T}\odot A}) \end{aligned}$$

Of the *idealisation* and *actualisations*, the *idealisation*, $A * T * T^{\dagger A}$, and the *surrealisation*, $(A * T)^{\mathrm{X}} * T^{\odot A}$, may be grouped together as *abstract converse actions* which depend on the *derived histogram*, $A * T$, and the *independent* of the *derived histogram*, or *abstract histogram*, $(A * T)^{\mathrm{X}}$. The *neutralisation*, $A * T * T^{\odot A^{\mathrm{X}}}$, and the *contentisation*, $A^{\mathrm{X}} * T * T^{\odot A}$, may be grouped together as *formal converse actions* which depend on the *histogram*, A, and the *independent* of the *histogram*, A^{X}.

3.12.4 Converse action entropy

Consider the *histogram-transform* pair $(A, T) \in \mathcal{A} \times \mathcal{T}_{U,\mathrm{f},1}$ in *variables* $V = \mathrm{vars}(A)$ where (i) the *independent histogram* is *completely effective*, $A^{\mathrm{XF}} = V^{\mathrm{C}}$, (ii) the *one functional transform* has *underlying variables* equal to the *histogram variables*, $\mathrm{und}(T) = V$, and (iii) the *derived histogram* is as *effective* as the *formal histogram*, $(A * T)^{\mathrm{F}} \geq (A^{\mathrm{X}} * T)^{\mathrm{F}}$, so that the *contentisation* is *size-conserving*, $\mathrm{size}(A^{\mathrm{X}} * T * T^{\odot A}) = \mathrm{size}(A)$.

The *formal converse actions*, which depend on the *histogram*, A, and the *independent* of the *histogram*, A^{X}, are related. Conjecture that the *sum* of the *contentisation* and the *neutralisation* is approximately equal to the *sum* of the *histogram* and the *independent*

$$A^{\mathrm{X}} * T * T^{\odot A} + A * T * T^{\odot A^{\mathrm{X}}} \cong A + A^{\mathrm{X}}$$

The *sizes* sum exactly, $\mathrm{size}(A^{\mathrm{X}} * T * T^{\odot A}) + \mathrm{size}(A * T * T^{\odot A^{\mathrm{X}}}) = \mathrm{size}(A) + \mathrm{size}(A^{\mathrm{X}})$. In the case when the *transform* is *formal* with respect to the *histogram*, $\mathrm{formal}(A, T) := A * T \equiv A^{\mathrm{X}} * T$, then (i) the *contentisation* is *equivalent* to the *histogram*, $A^{\mathrm{X}} * T * T^{\odot A} \equiv A$, and (ii) the *neutralisation* is *equivalent* to the *independent*, $A * T * T^{\odot A^{\mathrm{X}}} \equiv A^{\mathrm{X}}$, and so the *sum* of the *contentisation* and the *neutralisation* is exactly equal to the *sum* of the *histogram* and the *independent*, $A * T * T^{\odot A} + A^{\mathrm{X}} * T * T^{\odot A^{\mathrm{X}}} \equiv A + A^{\mathrm{X}}$.

If the *transform* is a *unary partition transform* $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$ then (i) the *neutralisation* equals the *independent*, $A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\odot A^{\mathrm{X}}} \equiv A^{\mathrm{X}}$, and (ii) the *contentisation* equals the *histogram*, $A^{\mathrm{X}} * T_{\mathrm{u}} * T_{\mathrm{u}}^{\odot A} \equiv A$, and so in this case the sums are exactly equal, $A^{\mathrm{X}} * T_{\mathrm{u}} * T_{\mathrm{u}}^{\odot A} + A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\odot A^{\mathrm{X}}} \equiv A + A^{\mathrm{X}}$.

If the *transform* is a *full functional transform*, for example a *value full functional transform* $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}}$, then (i) the *neutralisation* equals the *histogram*, $A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\odot A^{\mathrm{X}}} \equiv A$, and (ii) the *contentisation* equals the *independent* because the *histogram* is as *effective* as the *independent*, $(A * T_{\mathrm{s}})^{\mathrm{F}} \geq (A^{\mathrm{X}} * T_{\mathrm{s}})^{\mathrm{F}} \implies A^{\mathrm{F}} = A^{\mathrm{XF}} \implies A^{\mathrm{X}} * T_{\mathrm{s}} * T_{\mathrm{s}}^{\odot A} \equiv A^{\mathrm{X}}$, and so in this case the sums are exactly equal, $A^{\mathrm{X}} * T_{\mathrm{s}} * T_{\mathrm{s}}^{\odot A} + A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\odot A^{\mathrm{X}}} \equiv A^{\mathrm{X}} + A$.

In the special case where the *independent* equals the *scaled cartesian*, $A^{\mathrm{X}} = V_z^{\mathrm{C}}$, where $z = \mathrm{size}(A)$, $v = |V^{\mathrm{C}}|$ and $V_z^{\mathrm{C}} = \mathrm{scalar}(z/v) * V^{\mathrm{C}}$, then the sum of the *unnaturalisation* and the *naturalisation* is approximately equal to the *sum* of the *histogram* and the *scaled cartesian*

$$V_z^{\mathrm{C}} * T * T^{\odot A} + A * T * T^{\dagger} \cong A + V_z^{\mathrm{C}}$$

Conjecture that the sum of the *entropies* of the *contentisation* and the *neutralisation* varies as the sum of the *entropies* of the *histogram* and the *independent*

$$\mathrm{entropy}(A^{\mathrm{X}} * T * T^{\odot A}) + \mathrm{entropy}(A * T * T^{\odot A^{\mathrm{X}}}) \sim \mathrm{entropy}(A) + \mathrm{entropy}(A^{\mathrm{X}})$$

In the special case where the *independent* equals the *scaled cartesian*, $A^{\mathrm{X}} = V_z^{\mathrm{C}}$, then the *entropies* are such that

$$\mathrm{entropy}(V_z^{\mathrm{C}} * T * T^{\odot A}) + \mathrm{entropy}(A * T * T^{\dagger}) \sim \mathrm{entropy}(A) + \mathrm{entropy}(V_z^{\mathrm{C}})$$

The relationship between the *entropies* of the *formal converse actions* can be lifted to the *abstract converse actions*. In the case where the *formal histogram* is equivalent to the *abstract histogram*, $A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}}$, then insofar as the *idealisation* approximates to the *neutralisation*, $A * T * T^{\dagger A} \cong A * T * T^{\odot A^{\mathrm{X}}}$, conjecture that the sum of the *entropies* of the *surrealisation* and the *idealisation* varies as the sum of the *entropies* of the *histogram* and the *independent*

$$\mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\odot A}) + \mathrm{entropy}(A * T * T^{\dagger A}) \sim \mathrm{entropy}(A) + \mathrm{entropy}(A^{\mathrm{X}})$$

In the case where the *formal* equals the *abstract*, $A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}}$, conjecture that the *entropies* of the *converse action histograms* are subject to the following inequalities. First via the *actual converse* of the *independent*,

$$\begin{aligned}
& \mathrm{entropy}(V_z^{\mathrm{C}} * T * T^{\dagger}) = \mathrm{entropy}(V_z^{\mathrm{C}}) \\
\geq\ & \mathrm{entropy}(V_z^{\mathrm{C}} * T * T^{\odot A^{\mathrm{X}}}) \\
\geq\ & \mathrm{entropy}(A^{\mathrm{X}} * T * T^{\odot A^{\mathrm{X}}}) = \mathrm{entropy}(A^{\mathrm{X}}) \\
\geq\ & \mathrm{entropy}(A^{\mathrm{X}} * T * T^{\odot A}) = \mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\odot A}) \\
\geq\ & \mathrm{entropy}(A) \\
\geq\ & \mathrm{entropy}(Z_A) = 0
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{entropy}(V_z^{\mathrm{C}} * T * T^{\dagger}) = \mathrm{entropy}(V_z^{\mathrm{C}}) \\
\geq\ & \mathrm{entropy}(V_z^{\mathrm{C}} * T * T^{\odot A^{\mathrm{X}}}) \\
\geq\ & \mathrm{entropy}(A^{\mathrm{X}} * T * T^{\odot A^{\mathrm{X}}}) = \mathrm{entropy}(A^{\mathrm{X}}) \\
\geq\ & \mathrm{entropy}(A * T * T^{\odot A^{\mathrm{X}}}) \\
\geq\ & \mathrm{entropy}(A * T * T^{\dagger A}) \\
\geq\ & \mathrm{entropy}(A) \\
\geq\ & \mathrm{entropy}(Z_A) = 0
\end{aligned}$$

where $Z_A = A\%\emptyset = \mathrm{scalar}(\mathrm{size}(A))$.

Now via the *natural converse*,

$$\begin{aligned}
& \mathrm{entropy}(V_z^{\mathrm{C}} * T * T^{\dagger}) = \mathrm{entropy}(V_z^{\mathrm{C}}) \\
\geq\ & \mathrm{entropy}(A^{\mathrm{X}} * T * T^{\dagger}) = \mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\dagger}) \\
\geq\ & \mathrm{entropy}(A * T * T^{\dagger}) \\
\geq\ & \mathrm{entropy}(A * T * T^{\odot A^{\mathrm{X}}}) \\
\geq\ & \mathrm{entropy}(A * T * T^{\dagger A}) \\
\geq\ & \mathrm{entropy}(A) \\
\geq\ & \mathrm{entropy}(Z_A) = 0
\end{aligned}$$

Now via the *actual converse*,

$$\begin{aligned}
& \text{entropy}(V_z^{\text{C}} * T * T^{\dagger}) = \text{entropy}(V_z^{\text{C}}) \\
\geq\ & \text{entropy}((V_z^{\text{C}} * T)^{\text{X}} * T^{\dagger}) \\
\geq\ & \text{entropy}((V_z^{\text{C}} * T)^{\text{X}} * T^{\odot A}) \\
\geq\ & \text{entropy}((A^{\text{X}} * T)^{\text{X}} * T^{\odot A}) \\
\geq\ & \text{entropy}(A^{\text{X}} * T * T^{\odot A}) = \text{entropy}((A * T)^{\text{X}} * T^{\odot A}) \\
\geq\ & \text{entropy}(A) \\
\geq\ & \text{entropy}(Z_A) = 0
\end{aligned}$$

In section 'Minimum alignment', below, it is shown that the *relative entropy* of the *independent* with respect to the *histogram* equals the difference between the *independent entropy* and the *histogram entropy*,

$$\text{entropyRelative}(A, A^{\text{X}}) \quad = \quad \text{entropy}(A^{\text{X}}) - \text{entropy}(A)$$

and so the *independent entropy* is greater than or equal to the *histogram entropy*, $\text{entropy}(A^{\text{X}}) \geq \text{entropy}(A)$. Therefore the *surrealisation derived entropy* is greater than or equal to the *idealisation derived entropy*,

$$\begin{aligned}
& \text{entropy}((A * T)^{\text{X}} * T^{\odot A} * T) = \text{entropy}((A * T)^{\text{X}}) \\
\geq\ & \text{entropy}(A * T * T^{\dagger} * T) = \text{entropy}(A * T)
\end{aligned}$$

The *idealisation independent* equals the *histogram independent*, so the *idealisation entropy* is less than or equal to the *independent entropy*, $\text{entropy}(A^{\text{X}}) = \text{entropy}((A * T * T^{\dagger A})^{\text{X}}) \geq \text{entropy}(A * T * T^{\dagger A})$. The *idealisation entropy* is greater than or equal to the *histogram entropy*, $\text{entropy}(A * T * T^{\dagger A}) \geq \text{entropy}(A)$. Therefore the *idealisation entropy* is between the *independent entropy* and the *histogram entropy*,

$$\text{entropy}(A^{\text{X}}) \ \geq\ \text{entropy}(A * T * T^{\dagger A}) \ \geq\ \text{entropy}(A)$$

3.12.5 Iso-sets

A *histogram* A and its *independent* A^{X} are *congruent*. Let the set of *complete congruent histograms* in *system* U, of *variables* V and *size* z be

$$\mathcal{A}_{U,V,z} = \{A : A \in \mathcal{A}_U,\ A^{\text{U}} = V^{\text{C}},\ \text{size}(A) = z\}$$

The set of *complete congruent histograms* is also known as the set of *substrate histograms*. The set of *complete congruent histograms*, $\mathcal{A}_{U,V,z}$, is infinite if the *volume* is greater than one, $|V^{\mathrm{C}}| > 1$. Let *histogram* A be a *complete congruent histogram*, $A \in \mathcal{A}_{U,V,z}$. Then the *independent histogram*, A^{X}, is a *complete congruent histogram*, $A^{\mathrm{X}} \in \mathcal{A}_{U,V,z}$. The *independent* function partitions $\mathcal{A}_{U,V,z}$ into equivalence classes of *iso-independents*. Let $Y_{U,V,z} \in \mathcal{A}_{U,V,z} \to \mathcal{A}_{U,V,z}$ be the subset of the *independent* function,

$$Y_{U,V,z} = \{(A, A^{\mathrm{X}}) : A \in \mathcal{A}_{U,V,z}\} \subset \text{independent}$$

Then $\text{inverse}(Y_{U,V,z}) \in \mathcal{A}_{U,V,z} \to \mathrm{P}(\mathcal{A}_{U,V,z})$ and $\text{ran}(\text{inverse}(Y_{U,V,z})) \in \mathrm{B}(\mathcal{A}_{U,V,z})$. Thus the set of *iso-independents* of *histogram* $A \in \mathcal{A}_{U,V,z}$ is

$$\text{inverse}(Y_{U,V,z})(A^{\mathrm{X}}) = \{B : B \in \mathcal{A}_{U,V,z},\ B^{\mathrm{X}} = A^{\mathrm{X}}\}$$

The set of *iso-independents* is infinite if more than one of its *perimeter histograms* is not an *effective singleton*, $|\{w : w \in V,\ |(A^{\mathrm{X}}\%\{w\})^{\mathrm{F}}| > 1\}| > 1 \iff |Y_{U,V,z}^{-1}(A^{\mathrm{X}})| = \infty$. Otherwise the set of *iso-independents* is a singleton, $|\{w : w \in V,\ |(A^{\mathrm{X}}\%\{w\})^{\mathrm{F}}| > 1\}| \leq 1 \iff Y_{U,V,z}^{-1}(A^{\mathrm{X}}) = \{A^{\mathrm{X}}\}$.

Both the *histogram*, A, and the *independent histogram*, A^{X}, are *iso-independents*, $A, A^{\mathrm{X}} \in \text{inverse}(Y_{U,V,z})(A^{\mathrm{X}})$.

The *idealisation* of a *histogram* given an *effective transform*, $A*T*T^{\dagger A}$, is also in the *iso-independents*, $A * T * T^{\dagger A} \in Y_{U,V,z}^{-1}(A^{\mathrm{X}})$, because the *independent* of the *idealisation* equals the *independent histogram*, $(A * T * T^{\dagger A})^{\mathrm{X}} = A^{\mathrm{X}}$.

Let the set of *complete congruent integral histograms*, also called the *integral congruent support*, in *system* U, of *variables* V and *size* z be

$$\mathcal{A}_{U,\mathrm{i},V,z} = \{A : A \in \mathcal{A}_{U,\mathrm{i}},\ A^{\mathrm{U}} = V^{\mathrm{C}},\ \text{size}(A) = z\}$$

The set of *complete congruent integral histograms* is also known as the set of *integral substrate histograms*. The *integral congruent support* is a finite subset of the *complete congruent histograms*, $\mathcal{A}_{U,\mathrm{i},V,z} \subset \mathcal{A}_{U,V,z}$. Its cardinality is the cardinality of weak compositions $|\mathrm{C}'(V^{\mathrm{C}}, z)|$

$$|\mathcal{A}_{U,\mathrm{i},V,z}| = \frac{(z + v - 1)!}{z!\ (v - 1)!}$$

where $v = |V^{\mathrm{C}}|$. The *independent* function also partitions the *integral congruent support*, $\mathcal{A}_{U,\mathrm{i},V,z}$, into equivalence classes of *integral iso-independents*. Let $Y_{U,\mathrm{i},V,z} \in \mathcal{A}_{U,\mathrm{i},V,z} \to \mathcal{A}_{U,V,z}$ be such that $Y_{U,\mathrm{i},V,z} = \{(A, A^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\} \subset$

$Y_{U,V,z} \subset$ independent. Thus the finite set of *integral iso-independents* of *histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ is

$$\mathrm{inverse}(Y_{U,\mathrm{i},V,z})(A^{\mathrm{X}}) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} = A^{\mathrm{X}}\}$$

The *integral histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ is an *integral iso-independent*, $A \in \mathrm{inverse}(Y_{U,\mathrm{i},V,z})(A^{\mathrm{X}})$. The *independent histogram*, A^{X}, is not necessarily *integral*. If it is *integral* then it is a member of the *integral iso-independents*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \iff A^{\mathrm{X}} \in \mathrm{inverse}(Y_{U,\mathrm{i},V,z})(A^{\mathrm{X}})$.

The range of the *integral congruent independent function*, $\mathrm{ran}(Y_{U,\mathrm{i},V,z})$, consists of the set of all of the *independent histograms* in the *complete congruent histograms* which have *integral perimeter*, $\mathrm{ran}(Y_{U,\mathrm{i},V,z}) = \{A^{\mathrm{X}} : A \in \mathcal{A}_{U,V,z},\ \forall w \in V\ (A^{\mathrm{X}}\%\{w\} \in \mathcal{A}_{U,\mathrm{i},\{w\},z})\}$, because (i) each of the *histograms* of the domain, $\mathrm{dom}(Y_{U,\mathrm{i},V,z})$, has *integral perimeter*, $\forall A \in \mathcal{A}_{U,\mathrm{i},V,z}\ \forall w \in V\ (A\%\{w\} \in \mathcal{A}_{U,\mathrm{i},\{w\},z})$, (ii) the *perimeter* of an *independent histogram* equals the *perimeter* of its *histogram*, $\forall A \in \mathcal{A}_{U,\mathrm{i},V,z}\ \forall w \in V\ (A^{\mathrm{X}}\%\{w\} = A\%\{w\})$, and (iii) all *integral perimeters* imply the existence of at least one *integral histogram* having that *perimeter*, $\forall Q \in V \to \mathcal{A}_U\ ((|Q| = |V| \wedge (\forall (w,B) \in Q\ (B \in \mathcal{A}_{U,\mathrm{i},\{w\},z}))) \implies (\exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ (\{(w, A\%\{w\}) : w \in V\} = Q)))$. This can be shown by constructing the set of *integral iso-independents* explicitly given an *integral perimeter*. Define $\mathrm{iiso} \in (\mathcal{V} \to \mathcal{A}_{\mathrm{i}}) \to \mathrm{P}(\mathcal{A}_{\mathrm{i}})$ as $\mathrm{iiso}(Q) := \mathrm{iiso}(Q, \emptyset)$, and $\mathrm{iiso} \in (\mathcal{V} \to \mathcal{A}_{\mathrm{i}}) \times \mathcal{A}_{\mathrm{i}} \to \mathrm{P}(\mathcal{A}_{\mathrm{i}})$ as

$$\begin{aligned}
&\mathrm{iiso}(Q, A) := \\
&\quad \bigcup\ \{\mathrm{iiso}(Q', A') : X \in \prod \mathrm{ran}(Q),\ \mathrm{minr}(X) > 0,\ S = \bigcup \mathrm{dom}(X), \\
&\qquad\quad A' = A + \{S\}^{\mathrm{U}},\ Q' = \{(w, B - \{S\%\{w\}\}^{\mathrm{U}}) : (w,B) \in Q\}\} \\
&\quad \cup\ \ \{A : \mathrm{maxr}(\mathrm{ran}(Q)) = 0\}
\end{aligned}$$

where $(\prod)$ is the monoidal product of a set of sets. The function, iiso, is such that $\mathrm{iiso}(Q) \subset \mathcal{A}_{U,\mathrm{i},V,z}$ where $\forall (w,B) \in Q\ (B \in \mathcal{A}_{U,\mathrm{i},\{w\},z})$. It is also the case that $|\mathrm{iiso}(Q)| > 0$. Thus there always exists at least one *integral iso-independent histogram* having the given *integral perimeter*. Therefore the range of the *integral congruent independent function* is bijective with the *integral congruent perimeters* and the cardinality is

$$|\mathrm{ran}(Y_{U,\mathrm{i},V,z})| = \prod_{w \in V} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!}$$

Thus, $\mathrm{iiso}(\{(w, A^{\mathrm{X}}\%\{w\}) : w \in V\}) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$.

The cardinality of the *integral congruent perimeters* must be less than or equal to the cardinality of the *integral congruent support*

$$|\mathrm{ran}(Y_{U,\mathrm{i},V,z})| \leq |\mathrm{dom}(Y_{U,\mathrm{i},V,z})| = |\mathcal{A}_{U,\mathrm{i},V,z}|$$

and so

$$\prod_{w \in V} \frac{(z+|U_w|-1)!}{z!\ (|U_w|-1)!} \leq \frac{(z+v-1)!}{z!\ (v-1)!}$$

In the case of *regular variables* V having *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$ the cardinality is

$$|\mathrm{ran}(Y_{U,\mathrm{i},V,z})| = \left(\frac{(z+d-1)!}{z!\ (d-1)!}\right)^n$$

The average cardinality of the *integral iso-independents* is

$$\frac{|\mathcal{A}_{U,\mathrm{i},V,z}|}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|} = \frac{(z+v-1)!}{z!\ (v-1)!} / \prod_{w \in V} \frac{(z+|U_w|-1)!}{z!\ (|U_w|-1)!}$$

The average cardinality of the *integral iso-independents* varies with both *size*, z, and *volume*, v. For a given *volume*, v, the average cardinality of the *integral iso-independents* varies with the entropy of the *valencies*, $\mathrm{entropy}(\{(w,|U_w|) : w \in V\})$. Thus the average cardinality of the *integral iso-independents* tends to increase with *dimension*, $n = |V|$. *Regular histograms* tend to have higher average cardinality than *irregular*.

Integral iso-independents sets cannot be singletons of a *non-independent histogram*, $A \neq A^{\mathrm{X}} \implies Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \neq \{A\}$. Thus $A \neq A^{\mathrm{X}} \implies |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| > 1$. *Integral iso-independents* sets are singletons only if no more than one of its *perimeter histograms* is not an *effective singleton*, $|\{w : w \in V,\ |(A^{\mathrm{X}}\%\{w\})^{\mathrm{F}}| > 1\}| \leq 1 \iff Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) = \{A^{\mathrm{X}}\}$.

Given any *integral* subset of the *substrate histograms* $I \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$ that contains the *histogram* $A \in I$, the degree to which the subset is said to be *aligned-like* is called the *iso-independence*. The *iso-independence* is defined as the ratio of (i) the cardinality of the intersection between the *integral substrate histograms* subset and the set of *integral iso-independents*, and (ii) the cardinality of the union,

$$\frac{1}{|\mathcal{A}_{U,\mathrm{i},V,z}|} \ \leq\ \frac{|I\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|I\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} \ \leq\ 1$$

If the *iso-independence* is low, for example in the case of the *integral substrate histograms*, $I = \mathcal{A}_{U,\mathrm{i},V,z}$, the subset is said to be *classical-like*. If the *iso-independence* is high, for example in the case of the *integral iso-independents*, $I = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, the subset is said to be *aligned-like*.

Let *histogram* $A \in \mathcal{A}_{U,V,z}$ be in *system* U and have *variables* V and *size* z. Consider the *iso-independents* given some *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ where $\mathrm{und}(T) = V$. Let $W = \mathrm{der}(T)$. Let the *formal*-valued function of the *substrate histograms* $Y_{U,T,\mathrm{V},z} \in \mathcal{A}_{U,V,z} \to \mathcal{A}_{U,W,z}$ be defined

$$Y_{U,T,\mathrm{V},z} = \{(A, A^{\mathrm{X}} * T) : A \in \mathcal{A}_{U,V,z}\}$$

Note that $Y_{U,T,\mathrm{V},z}$ is not a subset of the *independent* function, $Y_{U,T,\mathrm{V},z} \cap \mathrm{independent} = \emptyset$. The infinite set of *iso-formals* of $A^{\mathrm{X}} * T$ is

$$\mathrm{inverse}(Y_{U,T,\mathrm{V},z})(A^{\mathrm{X}} * T) = \{B : B \in \mathcal{A}_{U,V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T\}$$

Iso-formals have the same *formal histogram*, $\forall B \in Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T)\ (B^{\mathrm{X}} * T = A^{\mathrm{X}} * T)$. The equivalence classes implied by $Y_{U,T,\mathrm{V},z}$ partition the *complete congruent histograms*, $\mathrm{ran}(Y^{-1}_{U,T,\mathrm{V},z}) \in \mathrm{B}(\mathcal{A}_{U,V,z})$. The equivalence classes implied by $Y_{U,T,\mathrm{V},z}$ form a parent partition of equivalence classes implied by the subset of the *independent* function $Y_{U,V,z}$, $\mathrm{parent}(\mathrm{ran}(Y^{-1}_{U,T,\mathrm{V},z}), \mathrm{ran}(Y^{-1}_{U,V,z}))$, so that $Y^{-1}_{U,V,z}(A^{\mathrm{X}}) \subseteq Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T)$. Thus both the *histogram* and its *independent* are *iso-formals*, $A, A^{\mathrm{X}} \in Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T)$. The *iso-formals* is a superset of the *iso-independents*, $Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \supseteq Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, so the *iso-independence*, or degree of *aligned-likeness*, is

$$\frac{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)|}$$

In the case where the *transform* is a *substrate transform*, $T \in \mathcal{T}_{U,V}$, and the *formal* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, the *iso-formals* can be written in terms of the *partition variables*,

$$\begin{aligned} &A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}} \implies \\ &\quad Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) = \{B : B \in \mathcal{A}_{U,V,z},\ \forall P \in W\ (B^{\mathrm{X}} * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}})\} \end{aligned}$$

This is the case if the *transform* is *non-overlapping*, $\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$.

Let the *formal independent*-valued function of the *substrate histograms* $Y_{U,T,\mathrm{V},\mathrm{x},z} \in \mathcal{A}_{U,V,z} \to \mathcal{A}_{U,W,z}$ be defined

$$Y_{U,T,\mathrm{V},\mathrm{x},z} = \{(A, (A^{\mathrm{X}} * T)^{\mathrm{X}}) : A \in \mathcal{A}_{U,V,z}\}$$

The infinite set of *iso-formal-independents* of $(A^{\mathrm{X}} * T)^{\mathrm{X}}$ is

$$\mathrm{inverse}(Y_{U,T,\mathrm{V},\mathrm{x},z})((A^{\mathrm{X}} * T)^{\mathrm{X}}) = \{B : B \in \mathcal{A}_{U,V,z},\ (B^{\mathrm{X}} * T)^{\mathrm{X}} = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}$$

Iso-formal-independents have the same *formal independent*, $\forall B \in Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ ((B^{\mathrm{X}} * T)^{\mathrm{X}} = (A^{\mathrm{X}} * T)^{\mathrm{X}})$. The equivalence classes implied by $Y_{U,T,\mathrm{V},\mathrm{x},z}$ partition the *complete congruent histograms*, $\mathrm{ran}(Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}) \in \mathrm{B}(\mathcal{A}_{U,V,z})$. The equivalence classes implied by $Y_{U,T,\mathrm{V},\mathrm{x},z}$ form a parent partition of equivalence classes implied by the subset of the *formal* function $Y_{U,T,\mathrm{V},z}$,

$$\mathrm{parent}(\mathrm{ran}(Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}), \mathrm{ran}(Y^{-1}_{U,T,\mathrm{V},z}))$$

so that $Y^{-1}_{U,V,z}(A^{\mathrm{X}}) \subseteq Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \subseteq Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})$. Thus both the *histogram* and its *independent* are *iso-formal-independents*, $A, A^{\mathrm{X}} \in Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})$. The *iso-independence* of the *integral iso-formal independents* is less than or equal to the *iso-independence* of the *integral iso-formals*, $|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|/|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})| \leq |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|/|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)|$.

In the case where the *transform* is a *substrate transform*, $T \in \mathcal{T}_{U,V}$, the *iso-formal-independents* equals the subset of the *substrate histograms* having the same set of *partition formals*,

$$Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ = \ \{B : B \in \mathcal{A}_{U,V,z},\ \forall P \in W\ (B^{\mathrm{X}} * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}})\}$$

In fact the *iso-partition-formal* sets are bijective to the *iso-formal-independents* sets,

$$\{\{A^{\mathrm{X}} * P^{\mathrm{T}} : P \in W\} : A \in \mathcal{A}_{U,V,z}\} \quad :\leftrightarrow: \quad \mathrm{ran}(Y_{U,T,\mathrm{V},\mathrm{x},z})$$

Similarly to the definition of the *iso-formals*, let the *abstract*-valued function of the *substrate histograms* $Y_{U,T,\mathrm{W},z} \in \mathcal{A}_{U,V,z} \to \mathcal{A}_{U,W,z}$ be defined

$$Y_{U,T,\mathrm{W},z} = \{(A, (A * T)^{\mathrm{X}}) : A \in \mathcal{A}_{U,V,z}\}$$

The infinite set of *iso-abstracts* of $(A * T)^{\mathrm{X}}$ is

$$\mathrm{inverse}(Y_{U,T,\mathrm{W},z})((A * T)^{\mathrm{X}}) = \{B : B \in \mathcal{A}_{U,V,z},\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}$$

Iso-abstracts have the same *abstract histogram*, $\forall B \in Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})\ ((B*T)^{\mathrm{X}} = (A*T)^{\mathrm{X}})$. The equivalence classes implied by $Y_{U,T,\mathrm{W},z}$ partition the *complete congruent histograms*, $\mathrm{ran}(Y^{-1}_{U,T,\mathrm{W},z}) \in \mathrm{B}(\mathcal{A}_{U,V,z})$. The *lifted iso-abstracts* is a subset of the *iso-independents* of the *derived*,

$$\{B*T : B \in Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})\} \subseteq Y^{-1}_{U,W,z}((A*T)^{\mathrm{X}})$$

If the *transform* is *non-overlapping* then the lifted *iso-abstracts* equals the *derived iso-independents*

$$\neg\mathrm{overlap}(T) \implies \{B*T : B \in Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})\} = Y^{-1}_{U,W,z}((A*T)^{\mathrm{X}})$$

because the *transform* is *right total* when *non-overlapping*, $(X\%W)^{\mathrm{F}} = W^{\mathrm{C}}$ where $(X,W) = T$.

The *derived iso-independence* of the *integral lifted iso-abstracts* is

$$\frac{|\{B*T : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})|}$$

The *histogram* is an *iso-abstract*, $A \in Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})$. If the *formal independent histogram* equals the *abstract histogram*, $(A^{\mathrm{X}}*T)^{\mathrm{X}} = (A*T)^{\mathrm{X}}$, then the *independent* is an *iso-abstract*,

$$(A^{\mathrm{X}}*T)^{\mathrm{X}} = (A*T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})$$

This is also the case if the *formal histogram* equals the *abstract histogram*

$$(A^{\mathrm{X}}*T) = (A*T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})$$

In the case where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and an *integral iso-abstract*, $A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})$, the *iso-independence* of the *iso-abstracts*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

is greater than would otherwise be the case because the *independent* is in the intersection, $A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}}) \cap Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$.

In the case where the *transform* is a *substrate transform*, $T \in \mathcal{T}_{U,V}$, the set of *iso-abstracts* can be written in terms of the *partition variables*,

$$\begin{aligned} Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}}) &= \{B : B \in \mathcal{A}_{U,V,z},\ \forall P \in W\ (B*P^{\mathrm{T}} = A*P^{\mathrm{T}})\} \\ &= \bigcap_{P \in W} D^{-1}_{U,P^{\mathrm{T}},z}(A*P^{\mathrm{T}}) \end{aligned}$$

where $D^{-1}_{U,T,z}(A * T)$ is the set of *iso-deriveds*, defined below.

In fact the *iso-partition-derived* sets are bijective to the *iso-abstract* sets,

$$\{\{A * P^{\mathrm{T}} : P \in W\} : A \in \mathcal{A}_{U,V,z}\} \quad :\leftrightarrow: \quad \mathrm{ran}(Y_{U,T,\mathrm{W},z})$$

For this reason, all subsets of the *iso-abstracts* that include the *histogram*,

$$\{I : I \subseteq Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}),\ A \in I\}$$

are called *entity-like iso-sets* of the *histogram*, A.

The *lifted iso-sets* of all *entity-like integral iso-sets* are subsets of the *integral iso-independents* of the *derived*,

$$\{B * T : B \in I\} \subseteq Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})$$

where $I \subseteq Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$ and $A \in I$. So the *derived iso-independence* is

$$\frac{|\{B * T : B \in I\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

In some cases the *derived iso-independence* of *entity-like integral iso-sets* may be greater than the *iso-independence*,

$$\frac{|\{B * T : B \in I\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|} \ \geq\ \frac{|I\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|I\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

because the *lifted iso-sets* of *entity-like integral iso-sets* are subsets of the *derived iso-independents*, $\{B * T : B \in I\} \subseteq Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})$, and not just intersections.

The degree to which an *integral iso-set* $I \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$ that contains the *histogram*, $A \in I$, is said to be *entity-like* is called the *iso-abstractence*. The *iso-abstractence* is defined as the ratio of (i) the cardinality of the intersection between the *integral iso-set* and the set of *integral iso-abstracts*, and (ii) the cardinality of the union,

$$\frac{1}{|\mathcal{A}_{U,\mathrm{i},V,z}|} \ \leq\ \frac{|I\ \cap\ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}{|I\ \cup\ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \ \leq\ 1$$

Consider the *iso-deriveds* subset of the *iso-abstracts*. Let the *derived*-valued function of the *substrate histograms* $D_{U,T,z} \in \mathcal{A}_{U,V,z} \to \mathcal{A}_{U,W,z}$ be defined

$$D_{U,T,z} = \{(A, A * T) : A \in \mathcal{A}_{U,V,z}\}$$

The infinite set of *iso-deriveds* of $A * T$ is

$$\mathrm{inverse}(D_{U,T,z})(A * T) = \{B : B \in \mathcal{A}_{U,V,z},\ B * T = A * T\}$$

The set of equivalence classes implied by $D_{U,T,z}$ is a child partition of the set of equivalence classes implied by $Y_{U,T,\mathrm{W},z}$, $\mathrm{parent}(\mathrm{ran}(Y^{-1}_{U,T,\mathrm{W},z}), \mathrm{ran}(D^{-1}_{U,T,z}))$. The *lifted iso-deriveds* is a singleton, $\{B * T : B \in D^{-1}_{U,T,z}(A * T)\} = \{A * T\}$. The *naturalisation* is in the *iso-deriveds*, $A * T * T^{\dagger} \in D^{-1}_{U,T,z}(A * T)$, because the *derived* of the *naturalisation* equals the *derived*, $(A * T * T^{\dagger}) * T = A * T$. The *iso-deriveds* might equally well be called the *iso-naturalisations*, since all have the same *naturalisation*, $\forall B \in D^{-1}_{U,T,z}(A*T)\ (B*T*T^{\dagger} = A*T*T^{\dagger})$.

The *iso-abstractence* or degree of *entity-likeness* is

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}\ \leq\ 1$$

All subsets of the *iso-derived* that include the *histogram*,

$$\{I : I \subseteq D^{-1}_{U,T,z}(A * T),\ A \in I\}$$

are called *law-like iso-sets* of the *histogram*, A. All *law-like iso-sets* are *entity-like iso-sets*, $D^{-1}_{U,T,z}(A * T) \subseteq Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$.

The *iso-independence* of the *integral iso-derived* is

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|D^{-1}_{U,\mathrm{i},T,z}(A * T)\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

The *lifted iso-set* of any *law-like iso-set* is a singleton of the *derived*, $\{A*T\}$, so the *derived iso-independence* of the *integral lifted iso-derived* is

$$\frac{1}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

The degree to which an *integral iso-set* $I \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$ that contains the *histogram*, $A \in I$, is said to be *law-like* is called the *iso-derivedence*. The *iso-derivedence* is defined as the ratio of (i) the cardinality of the intersection between the *integral iso-set* and the set of *integral iso-deriveds*, and (ii) the cardinality of the union,

$$\frac{1}{|\mathcal{A}_{U,\mathrm{i},V,z}|} \leq \frac{|I \cap D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|I \cup D^{-1}_{U,\mathrm{i},T,z}(A * T)|} \leq 1$$

The *iso-derivedence* of the *iso-abstracts* equals the *iso-abstractence* of the *iso-deriveds*,

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}$$

That is, the *iso-deriveds* is as *entity-like* as the *iso-abstracts* is *law-like*.

In the case where (i) the *transform* is a *substrate transform*, $T \in \mathcal{T}_{U,V}$, and (ii) the *derived* is *independent*, and so equals the *abstract*, $A * T = (A * T)^{\mathrm{X}}$, the set of *iso-deriveds* equals the set of *iso-partition-deriveds* and so equals the set of *iso-abstracts*,

$$\begin{aligned}
D^{-1}_{U,T,z}(A * T) &= D^{-1}_{U,T,z}((A * T)^{\mathrm{X}}) \\
&= \bigcap_{P \in W} D^{-1}_{U,P^{\mathrm{T}},z}(A * P^{\mathrm{T}}) \\
&= Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})
\end{aligned}$$

So in this case the *iso-abstractence*, or degree of *entity-likeness*, of the *integral iso-deriveds* is maximal.

In fact the *iso-deriveds* equals the *iso-abstracts* if and only if the *derived* is *independent*,

$$A * T = (A * T)^{\mathrm{X}} \iff D^{-1}_{U,T,z}(A * T) = Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

So if the *derived* is not *independent*, the *iso-deriveds* is a proper subset of the *iso-abstracts*,

$$A * T \neq (A * T)^{\mathrm{X}} \implies D^{-1}_{U,T,z}(A * T) \subset Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

and the *iso-abstractence* of the *integral iso-deriveds* is sub-maximal,

$$A * T \neq (A * T)^{\mathrm{X}} \implies \frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} < 1$$

The cardinality of the set of *integral iso-deriveds* is the product of the weak compositions of the *components*,

$$|D^{-1}_{U,\mathrm{i},T,z}(A * T)| \;\; = \prod_{(R,C) \in T^{-1}} \frac{((A * T)_R + |C| - 1)!}{(A * T)_R! \; (|C| - 1)!}$$

The cardinality of the set of *integral iso-abstracts* is constrained,

$$\forall P \in W \; \left(|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| \; \leq \prod_{(R,C) \in (P^{\mathrm{T}})^{-1}} \frac{((A * P^{\mathrm{T}})_R + |C| - 1)!}{(A * P^{\mathrm{T}})_R! \; (|C| - 1)!} \right)$$

In the case where the *derived* is *independent*, $A*T = (A*T)^{\mathrm{X}}$, the cardinality of the set of *integral iso-abstracts* can also be stated explicitly,

$$|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| \;\; = \prod_{(R,C) \in T^{-1}} \frac{((A * T)^{\mathrm{X}}_R + |C| - 1)!}{(A * T)^{\mathrm{X}}_R! \; (|C| - 1)!}$$

Corresponding to the *derived* valued function of the *substrate histograms*, $D_{U,T,z}$, the *normalised components* valued function of the *substrate histograms* is $C_{U,T,z} \in \mathcal{A}_{U,V,z} \to \mathrm{P}(\mathcal{A}_{U,V,z})$, defined

$$C_{U,T,z} = \{(A, \{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) : A \in \mathcal{A}_{U,V,z}\}$$

where $()^{\wedge} \in \mathcal{A} \to \mathcal{A}$ is defined $\hat{A} := \mathrm{normalise}(A)$ if $\mathrm{size}(A) > 0$ otherwise $\hat{A} := A$, and $T^{-1} := \mathrm{inverse}(T)$. The infinite set of *iso-components* of $\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}$ is

$$\begin{aligned} &\mathrm{inverse}(C_{U,T,z})(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) = \\ &\quad \{B : B \in \mathcal{A}_{U,V,z},\ \forall C \in T^{\mathrm{P}} \ ((B * C^{\mathrm{U}})^{\wedge} = (A * C^{\mathrm{U}})^{\wedge})\} \end{aligned}$$

The *unnaturalisation* is in the *iso-components*, $V^{\mathrm{C}}_z * T * T^{\odot A} \in C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\})$. The *iso-components* might equally well be called the *iso-unnaturalisations*, since all have the same *unnaturalisation*, $\forall B \in C^{-1}_{U,T,z}(\{(A* C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) \ (V^{\mathrm{C}}_z * T * T^{\odot B} = V^{\mathrm{C}}_z * T * T^{\odot A})$.

The *normalised components* of the *independent* valued function of the *substrate histograms* is $C_{U,\mathrm{x},T,z} \in \mathcal{A}_{U,V,z} \to \mathrm{P}(\mathcal{A}_{U,V,z})$, defined

$$C_{U,\mathrm{x},T,z} = \{(A, \{(A^{\mathrm{X}} * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) : A \in \mathcal{A}_{U,V,z}\}$$

The infinite set of *iso-independent-components* of $\{(A^{\mathrm{X}} * C^{\mathrm{U}})^\wedge : C \in T^{\mathrm{P}}\}$ is

$$\begin{aligned}&\mathrm{inverse}(C_{U,\mathrm{x},T,z})(\{(A^{\mathrm{X}} * C^{\mathrm{U}})^\wedge : C \in T^{\mathrm{P}}\}) = \\ &\quad \{B : B \in \mathcal{A}_{U,V,z},\ \forall C \in T^{\mathrm{P}}\ ((B^{\mathrm{X}} * C^{\mathrm{U}})^\wedge = (A^{\mathrm{X}} * C^{\mathrm{U}})^\wedge)\}\end{aligned}$$

The *normalised independent components* valued function of the *substrate histograms* is $C_{U,T,\mathrm{x},z} \in \mathcal{A}_{U,V,z} \to \mathrm{P}(\mathcal{A}_{U,V,z})$, defined

$$C_{U,T,\mathrm{x},z} = \{(A, \{(A * C^{\mathrm{U}})^{\mathrm{X}\wedge} : C \in T^{\mathrm{P}}\}) : A \in \mathcal{A}_{U,V,z}\}$$

The infinite set of *iso-component-independents* of $\{(A * C^{\mathrm{U}})^{\mathrm{X}\wedge} : C \in T^{\mathrm{P}}\}$ is

$$\begin{aligned}&\mathrm{inverse}(C_{U,T,\mathrm{x},z})(\{(A * C^{\mathrm{U}})^{\mathrm{X}\wedge} : C \in T^{\mathrm{P}}\}) = \\ &\quad \{B : B \in \mathcal{A}_{U,V,z},\ \forall C \in T^{\mathrm{P}}\ ((B * C^{\mathrm{U}})^{\mathrm{X}\wedge} = (A * C^{\mathrm{U}})^{\mathrm{X}\wedge})\}\end{aligned}$$

Now consider the *iso-formals* and the *iso-abstracts* together. Let the *formal-abstract* pair valued function of the *substrate histograms* $Y_{U,T,z} \in \mathcal{A}_{U,V,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})$ be defined

$$Y_{U,T,z} = \{(A, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,V,z}\}$$

The infinite set of *iso-transform-independents* of $((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})$ is

$$\begin{aligned}&\mathrm{inverse}(Y_{U,T,z})(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) = \\ &\quad \{B : B \in \mathcal{A}_{U,V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}\end{aligned}$$

Iso-transform-independents have (i) the same *formal histogram* and (ii) the same *abstract histogram*, $\forall B \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\ ((B^{\mathrm{X}} * T = A^{\mathrm{X}} * T) \wedge ((B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}))$. The *formal histogram*, $A^{\mathrm{X}} * T$, is not necessarily equal to the *abstract histogram*, $(A * T)^{\mathrm{X}}$. The equivalence classes implied by $Y_{U,T,z}$ partition the *complete congruent histograms*, $\mathrm{ran}(Y^{-1}_{U,T,z}) \in \mathrm{B}(\mathcal{A}_{U,V,z})$. The equivalence classes implied by both $Y_{U,T,\mathrm{V},z}$ and $Y_{U,T,\mathrm{W},z}$ form parent partitions of the partition implied by $Y_{U,T,z}$, $\mathrm{parent}(\mathrm{ran}(Y^{-1}_{U,T,\mathrm{V},z}), \mathrm{ran}(Y^{-1}_{U,T,z}))$ and $\mathrm{parent}(\mathrm{ran}(Y^{-1}_{U,T,\mathrm{W},z}), \mathrm{ran}(Y^{-1}_{U,T,z}))$. So the set of *iso-transform-independents* is the intersection of the *iso-formals* and *iso-abstracts*

$$Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) = Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

The *histogram* is an *iso-transform-independent*, $A \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$.

The *idealisation* of the *histogram* with the given *transform*, $A*T*T^{\dagger A}$, is also in the *iso-transform-independents*, $A*T*T^{\dagger A} \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}}))$. This is because (i) the *idealisation* is in the *iso-abstracts*, $A*T*T^{\dagger A}*T = A*T \implies (A*T*T^{\dagger A}*T)^{\mathrm{X}} = (A*T)^{\mathrm{X}}$, and (ii) *idealisation* is in the *iso-formals*, $(A*T*T^{\dagger A})^{\mathrm{X}} = A^{\mathrm{X}} \implies (A*T*T^{\dagger A})^{\mathrm{X}}*T = A^{\mathrm{X}}*T$.

The set of *iso-transform-independents* is a subset of the *iso-abstracts*, so it is an *entity-like iso-set* of the *histogram*, A,

$$Y^{-1}_{U,T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}})) \ \subseteq \ Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})$$

The *iso-abstractence* or degree of *entity-likeness* is

$$\frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}}))|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})|} \ \leq \ 1$$

The set of *iso-transform-independents* is not necessarily more *entity-like* than the *iso-deriveds*, which has an *iso-abstractence* of

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A*T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})|}$$

because the *iso-transform-independents* is not necessarily a superset of the *iso-deriveds*, $|D^{-1}_{U,\mathrm{i},T,z}(A*T) \setminus Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}}))| \geq 0$.

The *iso-derivedence* or degree of *law-likeness* is

$$\frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}})) \ \cap \ D^{-1}_{U,\mathrm{i},T,z}(A*T)|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}})) \ \cup \ D^{-1}_{U,\mathrm{i},T,z}(A*T)|} =$$

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}}*T) \ \cap \ D^{-1}_{U,\mathrm{i},T,z}(A*T)|}{|(Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}}*T) \ \cup \ D^{-1}_{U,\mathrm{i},T,z}(A*T)) \ \cap \ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})|}$$

So the set of *iso-transform-independents* is not necessarily more *law-like* than the *iso-abstracts*, which has an *iso-derivedence* of

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A*T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})|}$$

The set of *iso-transform-independents* is *entity-like* so the *lifted iso-transform-independents* is a subset of the *iso-independents* of the *derived*,

$$\begin{aligned}\{B*T : B \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}}))\}& \\ &\subseteq \{B*T : B \in Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})\} \\ &\subseteq Y^{-1}_{U,W,z}((A*T)^{\mathrm{X}})\end{aligned}$$

So the *derived iso-independence* of the *integral lifted iso-transform-independents* is

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

The *derived iso-independence* of the *integral lifted iso-transform-independents* is less than or equal to the *derived iso-independence* of the *integral lifted iso-abstracts*,

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|} \leq \frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

The *iso-independence* of the *iso-transform-independents* is

$$\frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \cap Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \cup Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} = \frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \cap Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|(Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \cup Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})) \cap Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)|}$$

So the set of *iso-transform-independents* is not necessarily more *aligned-like* than the *iso-formals*, which has an *iso-independence* of

$$\frac{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)|}$$

depending on the relative intersection cardinality.

If the *formal independent histogram* equals the *abstract histogram* then the *independent* is an *iso-abstract*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$, and hence is an *iso-transform-independent*,

$$(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))$$

This is also the case where the *formal histogram* equals the *abstract histogram*,

$$A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))$$

In the case where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and an *integral iso-transform-independent*, $A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}}))$, the *iso-independence* of the *iso-transform-independents*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \cap Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \cup Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

is greater than would otherwise be the case because the *independent* is in the intersection, $A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \cap Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$.

If the *formal histogram* equals the *abstract histogram* then the *lifted iso-transform-independents* contains the *abstract histogram*

$$(A * T)^{\mathrm{X}} = A^{\mathrm{X}} * T \in \{B * T : B \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))\}$$

In this case, if the *abstract* is also *integral*, $(A * T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, the *derived iso-independence* of the *iso-transform-independents*,

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

is greater than would otherwise be the case because the *abstract* is in the intersection, $(A*T)^{\mathrm{X}} \in \{B*T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}}))\} \cap Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})$.

Note that it is only in the subset where the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}}*T = (A*T)^{\mathrm{X}}$, that the *lifted iso-transform-independent* relation is functional

$$\begin{aligned}&\{(A * T,((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\}\\ \in\ &\mathcal{A}_{U,W,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})\end{aligned}$$

That is, the *lifted iso-transform-independent* sets do not partition the *complete congruent histograms* in the *derived variables*

$$\mathrm{ran}(\{(A*T,((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}})) : (A,((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}})) \in Y_{U,T,z}\}^{-1}) \notin \mathrm{B}(\mathcal{A}_{U,W,z})$$

except where the *formal histogram* equals the *abstract histogram*

$$\begin{aligned}&\mathrm{ran}(\{(A * T,((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) :\\ &\quad (A,((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \in Y_{U,T,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\}^{-1})\\ \in\ &\mathrm{B}(\{B * T : B \in \mathcal{A}_{U,V,z},\ B^{\mathrm{X}} * T = (B * T)^{\mathrm{X}}\})\end{aligned}$$

Similarly it is only in the subset where the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, that the *formal* domained relation of the *iso-transform-independents* is functional

$$\begin{aligned}&\{(A^{\mathrm{X}} * T,((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\}\\ \in\ &\mathcal{A}_{U,W,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})\end{aligned}$$

If the *transform* is a *self partition transform*, $T = V^{\mathrm{CS}\{\}\mathrm{T}}$, or it is *value full functional*, $T = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}}$, then the set of *iso-transform-independents* equals the set of *iso-independents* in the *underlying variables*, $Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) = Y^{-1}_{U,V,z}(A^{\mathrm{X}})$. In this case the *iso-independence* is maximised and the *iso-transform-independents* is *aligned-like*.

If the *transform* is a *unary partition*, $T^{\mathrm{P}} = \{V^{\mathrm{CS}}\}$, then the set of *iso-transform-independents* equals the set of *complete congruent histograms* in the *underlying variables*, $Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) = \mathcal{A}_{U,V,z}$. In this case the *iso-independence* is minimised and the *iso-transform-independents* is *classical-like*.

In the case where the *transform* is a *substrate transform*, $T \in \mathcal{T}_{U,V}$, and the *formal* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, the *iso-transform-independents* can be written in terms of the *partition variables*,

$$\begin{aligned}
&A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}} \implies \\
&\quad Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) = \{B : B \in \mathcal{A}_{U,V,z}, \\
&\qquad \forall P \in W\ (B^{\mathrm{X}} * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}}\ \wedge\ B * P^{\mathrm{T}} = A * P^{\mathrm{T}})\}
\end{aligned}$$

In the stricter case where the *formal* equals the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, this is,

$$\begin{aligned}
&A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies \\
&\quad Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) = \{B : B \in \mathcal{A}_{U,V,z}, \\
&\qquad \forall P \in W\ (B^{\mathrm{X}} * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}} = B * P^{\mathrm{T}} = A * P^{\mathrm{T}})\}
\end{aligned}$$

because $\forall P \in W\ (A * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}})$.

The set of *iso-partition-independents* is the intersection of the *iso-formal-independents* and *iso-abstracts*

$$Y^{-1}_{U,T,\mathrm{V,x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

In the case where the *transform* is a *substrate transform*, $T \in \mathcal{T}_{U,V}$, each *iso-partition-independent* has the same set of *partition formals* and *partition deriveds*,

$$\begin{aligned}
&\forall B \in Y^{-1}_{U,T,\mathrm{V,x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \\
&\qquad \forall P \in W\ (B^{\mathrm{X}} * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}}\ \wedge\ B * P^{\mathrm{T}} = A * P^{\mathrm{T}})
\end{aligned}$$

In the case where the *formal independent* equals the *abstract* each of the *partition transforms* is *formal*, $(A^{\mathrm{X}}*T)^{\mathrm{X}} = (A*T)^{\mathrm{X}} \implies \forall P \in W\ (A*P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}})$, so

$$\begin{aligned}&(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies \\ &\quad \forall B \in Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \\ &\qquad \forall P \in W\ (B^{\mathrm{X}} * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}} = B * P^{\mathrm{T}} = A * P^{\mathrm{T}})\end{aligned}$$

In this case the *independent* is an *iso-partition-independent* too,

$$(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

The set of *iso-partition-independents* is a subset of the *iso-abstracts*, so it is an *entity-like iso-set* of the *histogram*, A,

$$Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})\ \subseteq\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

The *iso-transform-independents* is a subset of the *iso-partition-independents*,

$$\begin{aligned}&Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \subseteq \\ &\qquad Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})\end{aligned}$$

The *iso-abstractence* of the set of *iso-partition-independents* is greater than or equal to the *iso-abstractence* of the set of *iso-transform-independents*

$$\begin{aligned}&\frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \geq \\ &\qquad \frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)\ \cap\ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}\end{aligned}$$

So the *iso-partition-independents* is more *entity-like* than the *iso-transform-independents*.

The set of *iso-neutralisations* is the intersection of the *iso-independent-components* and *iso-deriveds*

$$C^{-1}_{U,\mathrm{x},T,z}(\{(A^{\mathrm{X}} * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\})\ \cap\ D^{-1}_{U,T,z}(A * T)$$

Each *iso-neutralisation* has the same *neutralisation*,

$$\forall B \in C^{-1}_{U,\mathrm{x},T,z}(\{(A^{\mathrm{X}}*C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) \cap D^{-1}_{U,T,z}(A*T)\ (B*T*T^{\odot B^{\mathrm{X}}} = A*T*T^{\odot A^{\mathrm{X}}})$$

The *neutralisation* is necessarily in the *iso-deriveds*, $A*T*T^{\odot A^{\mathrm{X}}}*T = A*T$, but not necessarily in the *iso-neutralisations*. In the case where the *transform* is *formal*, $\mathrm{formal}(A,T)$, the *neutralisation* is in the *iso-neutralisations* because it is in the *iso-independent-components*

$$A*T = A^{\mathrm{X}}*T \implies \{((A*T*T^{\odot A^{\mathrm{X}}})^{\mathrm{X}}*C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\} = \{(A^{\mathrm{X}}*C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}$$

The set of *iso-neutralisations* is a subset of the *iso-deriveds*, so it is a *law-like iso-set* of the *histogram*, A,

$$C^{-1}_{U,\mathrm{x},T,z}(\{(A^{\mathrm{X}} * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) \ \cap \ D^{-1}_{U,T,z}(A*T) \ \subseteq \ D^{-1}_{U,T,z}(A*T)$$

The set of *iso-contentisations* is the intersection of the *iso-components* and *iso-formals*

$$C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) \ \cap \ Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T)$$

Each *iso-contentisation* has the same *contentisation*,

$$\forall B \in C^{-1}_{U,T,z}(\{(A*C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) \cap Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}}*T)\ (B^{\mathrm{X}}*T*T^{\odot B} = A^{\mathrm{X}}*T*T^{\odot A})$$

The *contentisation* is necessarily in the *iso-components*, $(A^{\mathrm{X}} * T * T^{\odot A} * C^{\mathrm{U}})^{\wedge} = (A * C^{\mathrm{U}})^{\wedge}$, but not necessarily in the *iso-contentisations*. In the case where the *transform* is *formal*, $\mathrm{formal}(A,T)$, the *contentisation* is in the *iso-contentisations* because it is in the *iso-formals*,

$$A^{\mathrm{X}} * T = A * T \implies (A^{\mathrm{X}} * T * T^{\odot A})^{\mathrm{X}} * T = A^{\mathrm{X}} * T$$

The set of *iso-liftisations* is the intersection of the *iso-formals* and *iso-deriveds*

$$Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \ \cap \ D^{-1}_{U,T,z}(A*T)$$

The *iso-liftisations* is a subset of the *iso-transform-independents* which is the intersection of the *iso-formals* and *iso-abstracts*

$$\begin{aligned} Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \ \cap \ D^{-1}_{U,T,z}(A*T) \ &\subseteq \ Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T),(A*T)^{\mathrm{X}})) \\ &= \ Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}}) \end{aligned}$$

In the case of *formal histogram*, $\mathrm{formal}(A,T)$, the *naturalisation* is in the *iso-liftisations*,

$$A * T = A^{\mathrm{X}} * T \implies A * T * T^{\dagger} \in Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \ \cap \ D^{-1}_{U,T,z}(A*T)$$

The set of *iso-liftisations* is a subset of the *iso-deriveds*, so it is a *law-like iso-set* of the *histogram*, A,

$$Y_{U,T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \;\cap\; D_{U,T,z}^{-1}(A * T) \;\;\subseteq\;\; D_{U,T,z}^{-1}(A * T)$$

The *iso-derivedence* of the set of *iso-liftisations* is greater than or equal to the *iso-derivedence* of the set of *iso-transform-independents*

$$\frac{|Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \;\cap\; D_{U,\mathrm{i},T,z}^{-1}(A * T)|}{|D_{U,\mathrm{i},T,z}^{-1}(A * T)|} \geq \frac{|Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \;\cap\; D_{U,\mathrm{i},T,z}^{-1}(A * T)|}{|(Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \;\cap\; Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})) \;\cup\; D_{U,\mathrm{i},T,z}^{-1}(A * T)|}$$

So the *iso-liftisations* is more *law-like* than the *iso-transform-independents*.

Let the *iso-idealisation* function $Y_{U,T,\dagger,z} \in \mathcal{A}_{U,V,z} \to \mathrm{P}(\mathcal{A}_{U,V,z})$ be defined $Y_{U,T,\dagger,z} = \{(A, A * T * T^{\dagger A}) : A \in \mathcal{A}_{U,V,z}\}$. The infinite set of *iso-idealisations* of $A * T * T^{\dagger A}$ is

$$\begin{aligned}&\mathrm{inverse}(Y_{U,T,\dagger,z})(A * T * T^{\dagger A}) = \\ &\qquad \{B : B \in \mathcal{A}_{U,V,z},\; B * T * T^{\dagger B} = A * T * T^{\dagger A}\}\end{aligned}$$

Each *iso-idealisation* has the same set of *component independents* as the given histogram A,

$$\forall B \in Y_{U,T,\dagger,z}^{-1}(A * T * T^{\dagger A})\ \forall C \in T^{\mathrm{P}}\ \ ((B * C^{\mathrm{U}})^{\mathrm{X}} = (A * C^{\mathrm{U}})^{\mathrm{X}})$$

The *iso-idealisations* equals the intersection of the *iso-component-independents* and the *iso-derived*,

$$Y_{U,T,\dagger,z}^{-1}(A * T * T^{\dagger A}) \;\;=\;\; C_{U,T,\mathrm{x},z}^{-1}(\{(A * C^{\mathrm{U}})^{\mathrm{X}\wedge} : C \in T^{\mathrm{P}}\}) \;\cap\; D_{U,T,z}^{-1}(A * T)$$

The set of *iso-idealisations* is a subset of the *iso-deriveds*, so it is a *law-like iso-set* of the *histogram*, A,

$$Y_{U,T,\dagger,z}^{-1}(A * T * T^{\dagger A}) \;\;\subseteq\;\; D_{U,T,z}^{-1}(A * T)$$

The *iso-derivedence* or degree of *law-likeness* is

$$\frac{|Y_{U,\mathrm{i},T,\dagger,z}^{-1}(A * T * T^{\dagger A})|}{|D_{U,\mathrm{i},T,z}^{-1}(A * T)|} \;\leq\; 1$$

The set of *iso-idealisations* is a subset of the *iso-abstracts*, so it is an *entity-like iso-set* of the *histogram*, A,

$$Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \ \subseteq \ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

The *iso-abstractence* or degree of *entity-likeness* is less than or equal to the *iso-derivedence*

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \ \leq \ \frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|}$$

so the set of *iso-idealisations* is more *law-like* than *entity-like*.

The equivalence classes implied by $Y_{U,T,\dagger,z}$ partition the *complete congruent histograms*, $\mathrm{ran}(Y^{-1}_{U,T,\dagger,z}) \in \mathrm{B}(\mathcal{A}_{U,V,z})$. The set of *iso-idealisations* is (i) the intersection of the *iso-component-independents* and the *iso-derived* which is (ii) a subset of the intersection of the *iso-independents* and *iso-deriveds* which is (iii) a subset of the *iso-liftisations* which is (iv) a subset of the *iso-transform-independents* which is (v) a subset of the *iso-partition-independents* which is (vi) a subset of the *iso-abstracts*,

$$\begin{aligned}
Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \ &= \ C^{-1}_{U,T,\mathrm{x},z}(\{(A * C^{\mathrm{U}})^{\mathrm{X}\wedge} : C \in T^{\mathrm{P}}\}) \ \cap \ D^{-1}_{U,T,z}(A * T) \\
&\subseteq \ Y^{-1}_{U,V,z}(A^{\mathrm{X}}) \ \cap \ D^{-1}_{U,T,z}(A * T) \\
&\subseteq \ Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \ \cap \ D^{-1}_{U,T,z}(A * T) \\
&\subseteq \ Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \ \cap \ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \\
&\subseteq \ Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \\
&\subseteq \ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})
\end{aligned}$$

The *lifted iso-idealisations* is a singleton, $\{B*T : B \in Y^{-1}_{U,T,\dagger,z}(A*T*T^{\dagger A})\} = \{A * T\}$.

The *histogram* is an *iso-idealisation*, $A \in Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A})$. The *histogram idealisation* is an *iso-idealisation*, $A * T * T^{\dagger A} \in Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A})$.

There is a bijection between the sets of *component independents* and the *idealisations* of the *histograms* of the *complete congruent histograms*, $\{\{(A*C^{\mathrm{U}})^{\mathrm{X}} : C \in T^{\mathrm{P}}\} : A \in \mathcal{A}_{U,V,z}\} :\leftrightarrow: \mathrm{ran}(Y^{-1}_{U,T,\dagger,z})$.

The set of *iso-idealisations* is a subset of the *iso-independents*, $Y^{-1}_{U,T,\dagger,z}(A*T*T^{\dagger A}) \subseteq Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, so the degree to which the *iso-idealisations* is *aligned-like*, or the *iso-independence*, is $|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|/|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|$.

The *iso-independence* of the intersection of the *iso-derived* and the *iso-independents* is greater than or equal to the *iso-independence* of the *iso-derived*,

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}\ \geq\ \frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|D^{-1}_{U,\mathrm{i},T,z}(A * T)\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

The set of *iso-idealisations* is a subset of the intersection of the *iso-derived* and the *iso-independents*, $Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \subseteq D^{-1}_{U,T,z}(A * T)\ \cap\ Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, so in some cases the *iso-independence* of the *iso-idealisations* is greater than or equal to the *iso-independence* of the *iso-derived*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}\ \geq\ \frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|D^{-1}_{U,\mathrm{i},T,z}(A * T)\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

The set of *iso-idealisations* is also a subset of the intersection of the *iso-abstracts* and the *iso-independents*, $Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \subseteq Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \cap Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, so in some cases the *iso-independence* of the *iso-idealisations* is greater than or equal to the *iso-independence* of the *iso-abstract*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}\ \geq\ \frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

If the *transform* is a *self partition transform*, $T = V^{\mathrm{CS\{\}T}}$, or it is *value full functional*, $T = \{\{w\}^{\mathrm{CS\{\}T}} : w \in V\}^{\mathrm{T}}$, then the set of *iso-idealisations* equals a singleton of the *histogram*, $Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) = \{A\}$. In this case the *iso-idealisations* is neither *aligned-like* nor *classical-like*.

If the *transform* is a *unary partition*, $T^{\mathrm{P}} = \{V^{\mathrm{CS}}\}$, then the set of *iso-idealisations* equals the set of *iso-independents* in the *underlying variables*, $Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) = Y^{-1}_{U,V,z}(A^{\mathrm{X}})$. In this case the *iso-independence* is maximised and the *iso-idealisations* is *aligned-like*.

The set of *iso-surrealisations* is the intersection of the *iso-abstracts* and *iso-components*

$$Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})\ \cap\ C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\})$$

Each *iso-surrealisation* has the same *surrealisation*,

$$\forall B \in Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \cap C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\})$$
$$((B * T)^{\mathrm{X}} * T^{\odot B} = (A * T)^{\mathrm{X}} * T^{\odot A})$$

The set of *iso-surrealisations* is a subset of the *iso-abstracts*, so it is an *entity-like iso-set* of the *histogram*, A,

$$Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \;\cap\; C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) \;\;\subseteq\;\; Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

The *iso-abstractence* or degree of *entity-likeness* is

$$\frac{|C^{-1}_{U,\mathrm{i},T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) \;\cap\; Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \;\leq\; 1$$

Define the *iso-extremes* as the union of the *iso-liftisations* and the *iso-surrealisations*,

$$\begin{gathered}(Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \;\;\cap\;\; D^{-1}_{U,T,z}(A * T)) \;\cup \\ (Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \;\;\cap\;\; C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}))\end{gathered}$$

The *iso-extremes* are required to be members of the *iso-liftisations* or the *iso-surrealisations*, but not necessarily both. So the set of *iso-extremes* is not, strictly speaking, an *iso-set*, because there is no function of the *substrate histograms* that implies a partition for which the set of *iso-extremes* is a component. The *iso-extremes* can be re-arranged as the intersection of (i) the union of the *iso-formals* and the *iso-abstracts*, and (ii) the union of the *iso-deriveds* and the *iso-components*,

$$\begin{gathered}(Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \;\;\cup\;\; Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})) \;\cap \\ (D^{-1}_{U,T,z}(A * T) \;\;\cup\;\; C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}))\end{gathered}$$

So, although the *iso-transform-independents* is the intersection of the *iso-formals* and the *iso-abstracts*, the *iso-extreme* set is not a superset of the set of *iso-transform-independents*. That is, the *iso-transform-independents* that are neither *iso-deriveds* nor *iso-components* are not *iso-extreme*,

$$\begin{gathered}(Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \;\cap\; Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})) \;\setminus \\ (D^{-1}_{U,T,z}(A * T) \;\cup\; C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}))\end{gathered}$$

The *iso-idealisations*, which is a subset of the *iso-liftisations*, is a subset of the *iso-extremes*.

The set of *iso-extremes* is a subset of the *iso-abstracts*, so it is an *entity-like iso-set* of the *histogram*, A,

$$\begin{aligned}&(Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \;\;\cap\;\; D^{-1}_{U,T,z}(A * T)) \;\cup \\ &(Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \;\;\cap\;\; C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\})) \\ &\qquad\qquad\qquad\qquad\subseteq Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})\end{aligned}$$

The *integral* subset of the *iso-transform-independents* is formally defined as follows. Let $Y_{U,\mathrm{i},T,z} \in \mathcal{A}_{U,\mathrm{i},V,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})$ be defined, $Y_{U,\mathrm{i},T,z} = \{(A, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$. The finite set of *integral iso-transform-independents* of $((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})$ is

$$\begin{aligned}&\mathrm{inverse}(Y_{U,\mathrm{i},T,z})(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) = \\ &\qquad \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}\end{aligned}$$

The equivalence classes implied by $Y_{U,\mathrm{i},T,z}$ partition the *integral congruent support*, $\mathrm{ran}(Y^{-1}_{U,\mathrm{i},T,z}) \in \mathrm{B}(\mathcal{A}_{U,\mathrm{i},V,z})$. The *histogram* is an *integral iso-transform-independent*, $A \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$. If the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and the *formal independent histogram* equals the *abstract histogram*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$, then the *independent* is an *integral iso-abstract* and hence an *integral iso-transform-independent*,

$$(A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}) \wedge ((A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}) \implies A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$$

The cardinality of the set of *integral iso-formal* sets is such that

$$|\mathrm{ran}(Y_{U,\mathrm{i},T,\mathrm{V},z})| \leq \prod_{w \in V} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!}$$

The cardinality of the set of *integral iso-abstract* sets is

$$|\mathrm{ran}(Y_{U,\mathrm{i},T,\mathrm{W},z})| \leq \prod_{w \in W} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!}$$

Therefore the cardinality of the set of *integral iso-transform-independent* sets is such that

$$|\mathrm{ran}(Y_{U,\mathrm{i},T,z})| \leq \prod_{w \in V} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!} \times \prod_{w \in W} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!}$$

The cardinality of the set of *integral iso-transform-independent* sets is also such that

$$|\mathrm{ran}(Y_{U,\mathrm{i},T,z})| \leq |\mathrm{dom}(Y_{U,\mathrm{i},T,z})| = \frac{(z + v - 1)!}{z!\ (v - 1)!}$$

In the *derived*-valued function of the *substrate histograms*, $D_{U,T,z}$, the *model* is a *transform*. Now consider extending the *model* to (i) *fuds*, (ii) *decompositions*, and (iii) *fud decompositions*.

Let *substrate histogram* $A \in \mathcal{A}_{U,V,z}$ be in *system* U and have *variables* V

and *size* z. Given the *one functional definition set* $F \in \mathcal{F}_{U,1}$, such that $\mathrm{und}(F) \subseteq V$, let the *derived* set valued function of the *substrate histograms* $D_{U,F,z} \in \mathcal{A}_{U,V,z} \to \mathrm{P}(\mathcal{A}_U)$ be defined

$$D_{U,F,z} = \{(A, \{A * T_F : T \in F\}) : A \in \mathcal{A}_{U,V,z}\}$$

where $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$.

The set of *iso-fuds* is the intersection of the *iso-deriveds* of each *transform*

$$D_{U,F,z}^{-1}(\{A * T_F : T \in F\}) \quad = \quad \bigcap_{T \in F} D_{U,T_F,z}^{-1}(A * T_F)$$

If the *top transform* exists the set of *iso-fuds* is a subset of the *iso-deriveds*,

$$\exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F)) \implies D_{U,F,z}^{-1}(\{A * T_F : T \in F\}) \subseteq D_{U,F^{\mathrm{T}},z}^{-1}(A * F^{\mathrm{T}})$$

In this case the set of *iso-fuds* is *law-like* with an *iso-derivedence* of

$$\frac{|D_{U,\mathrm{i},F,z}^{-1}(\{A * T_F : T \in F\})|}{|D_{U,\mathrm{i},F^{\mathrm{T}},z}^{-1}(A * F^{\mathrm{T}})|}$$

In the case where the *fud* is a singleton, $F = \{T\}$, the *iso-fuds* equals the *iso-deriveds*,

$$D_{U,\{T\},z}^{-1}(\{A * T\}) \quad = \quad D_{U,T,z}^{-1}(A * T)$$

and the *iso-fuds* is maximally *law-like*.

The set of *iso-fuds* is a subset of the *iso-abstracts*,

$$D_{U,F,z}^{-1}(\{A * T_F : T \in F\}) \subseteq Y_{U,F^{\mathrm{T}},\mathrm{W},z}^{-1}((A * F^{\mathrm{T}})^{\mathrm{X}})$$

so the set of *iso-fuds* is *entity-like* with an *iso-abstractence* of

$$\frac{|D_{U,\mathrm{i},F,z}^{-1}(\{A * T_F : T \in F\})|}{|Y_{U,\mathrm{i},F^{\mathrm{T}},\mathrm{W},z}^{-1}((A * F^{\mathrm{T}})^{\mathrm{X}})|}$$

In the case where the *fud* consists of a single *layer* of *partition transforms*, $F \in \mathcal{F}_{U,\mathrm{P}}$ such that $|\mathrm{der}(F)| = |F|$, the *iso-fuds* equals the *iso-abstracts*, which is the intersection of the *iso-deriveds* of the *partition transforms*,

$$\begin{aligned} D_{U,F,z}^{-1}(D_{U,F,z}(A)) \quad &= \quad Y_{U,F^{\mathrm{T}},\mathrm{W},z}^{-1}((A * F^{\mathrm{T}})^{\mathrm{X}}) \\ &= \quad \bigcap_{T \in F} D_{U,T,z}^{-1}(A * T) \end{aligned}$$

In this case the *iso-fuds* is maximally *entity-like*.

Given the *decomposition* of *one functional transforms* $D \in \mathcal{D}_U = \mathcal{D} \cap \mathrm{trees}(\mathcal{S}_U \times \mathcal{T}_{U,\mathrm{f},1})$, such that $\mathrm{und}(D) \subseteq V$, let the *component-derived* function valued function of the *substrate histograms* $D_{U,D,z} \in \mathcal{A}_{U,V,z} \to (\mathcal{A}_U \to \mathcal{A}_U)$ be defined

$$D_{U,D,z} = \{(A, \{(C, A * C * T) : (C,T) \in \mathrm{cont}(D)\}) : A \in \mathcal{A}_{U,V,z}\}$$

where $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$.

The set of *iso-decompositions* can be related to the set *iso-deriveds* for each *slice*,

$$\forall B \in D_{U,D,z}^{-1}(D_{U,D,z}(A))\ \forall (C,T) \in \mathrm{cont}(D)\ (B * C \in D_{U,T,z_C}^{-1}(A * C * T))$$

where $z_C = \mathrm{size}(A * C)$. So each set of *slice iso-decompositions* is *law-like*,

$$\forall (C,T) \in \mathrm{cont}(D)\ (D_{U,D,z_C}^{-1}(D_{U,D,z_C}(A * C)) \subseteq D_{U,T,z_C}^{-1}(A * C * T))$$

The set of *iso-decompositions* is a subset of the *iso-deriveds* of the *transform* of the *decomposition*,

$$D_{U,D,z}^{-1}(D_{U,D,z}(A)) \quad \subseteq \quad D_{U,D^{\mathrm{T}},z}^{-1}(A * D^{\mathrm{T}})$$

so the set of *iso-decompositions* is *law-like* with an *iso-derivedence* of

$$\frac{|D_{U,\mathrm{i},D,z}^{-1}(D_{U,D,z}(A))|}{|D_{U,\mathrm{i},D^{\mathrm{T}},z}^{-1}(A * D^{\mathrm{T}})|}$$

In the case where the *decomposition* consists of a root node only, $D = \{((\emptyset, T), \emptyset)\}$, the *iso-decompositions* equals the *iso-deriveds*,

$$D_{U,D,z}^{-1}(D_{U,D,z}(A)) \quad = \quad D_{U,T,z}^{-1}(A * T)$$

In this case the set of *iso-decompositions* is maximally *law-like*.

Given the *fud decomposition* of *one functional definition sets* $D \in \mathcal{D}_{\mathrm{F},U} = \mathcal{D}_{\mathrm{F}} \cap \mathrm{trees}(\mathcal{S}_U \times \mathcal{F}_{U,1})$, such that $\mathrm{und}(D) \subseteq V$, let the *component-derived-set* function valued function of the *substrate histograms* $D_{U,D,\mathrm{F},z} \in \mathcal{A}_{U,V,z} \to (\mathcal{A}_U \to \mathrm{P}(\mathcal{A}_U))$ be defined

$$D_{U,D,\mathrm{F},z} = \{(A, \{(C, \{A * C * T_F : T \in F\}) : (C,F) \in \mathrm{cont}(D)\}) : A \in \mathcal{A}_{U,V,z}\}$$

The set of *iso-fud-decompositions* can be related to the set of intersections of *iso-deriveds* for each *slice*

$$\forall B \in D^{-1}_{U,D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A))\ \forall (C,F) \in \mathrm{cont}(D)$$
$$(B * C \in \bigcap_{T \in F} D^{-1}_{U,T_F,z_C}(A * C * T_F))$$

or

$$\forall (C,F) \in \mathrm{cont}(D)\ (D^{-1}_{U,D,\mathrm{F},z_C}(D_{U,D,\mathrm{F},z_C}(A * C)) \subseteq \bigcap_{T \in F} D^{-1}_{U,T_F,z_C}(A * C * T_F))$$

If the *top transform* exists the set of *slice iso-fud-decompositions* is a subset of the *slice iso-deriveds*,

$$\forall (C,F) \in \mathrm{cont}(D)\ (\exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F)) \implies$$
$$D^{-1}_{U,D,\mathrm{F},z_C}(D_{U,D,\mathrm{F},z_C}(A * C)) \subseteq D^{-1}_{U,F^{\mathrm{T}},z_C}(A * C * F^{\mathrm{T}}))$$

so in some cases the set of *slice iso-fud-decompositions* is *law-like* with an *iso-derivedence* of

$$\frac{|D^{-1}_{U,\mathrm{i},D,\mathrm{F},z_C}(D_{U,D,\mathrm{F},z_C}(A * C))|}{|D^{-1}_{U,\mathrm{i},F^{\mathrm{T}},z_C}(A * C * F^{\mathrm{T}})|}$$

If the *top transform* exists for all of the *fuds*, then the set of *iso-fud decompositions* is a subset of the *iso-deriveds*,

$$\forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F)) \implies$$
$$D^{-1}_{U,D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A)) \subseteq D^{-1}_{U,D^{\mathrm{T}},z}(A * D^{\mathrm{T}})$$

In this case the set of *iso-fud-decompositions* is *law-like* with an *iso-derivedence* of

$$\frac{|D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A))|}{|D^{-1}_{U,\mathrm{i},D^{\mathrm{T}},z}(A * D^{\mathrm{T}})|}$$

The set of *slice iso-fud-decompositions* is a subset of the *slice iso-abstracts*

$$\forall (C,F) \in \mathrm{cont}(D)\ (D^{-1}_{U,D,\mathrm{F},z_C}(D_{U,D,\mathrm{F},z_C}(A * C)) \subseteq Y^{-1}_{U,F^{\mathrm{T}},\mathrm{W},z_C}((A * C * F^{\mathrm{T}})^{\mathrm{X}}))$$

so the set of *slice iso-fud-decompositions* is *entity-like* with an *iso-abstractence* of

$$\frac{|D^{-1}_{U,\mathrm{i},D,\mathrm{F},z_C}(D_{U,D,\mathrm{F},z_C}(A * C))|}{|Y^{-1}_{U,\mathrm{i},F^{\mathrm{T}},\mathrm{W},z_C}((A * C * F^{\mathrm{T}})^{\mathrm{X}})|}$$

Thus when the *model* is extended to *fuds* or *fud decompositions*, the *iso-set* or *slice iso-set* corresponding to the *iso-deriveds* is only sometimes *law-like*, although it is always at least *entity-like*.

Similarly, in the *formal-abstract*-pair-valued function of the *substrate histograms*, $Y_{U,T,z}$, the *model* is a *transform*. Again consider extending the *model* to (i) *fuds*, (ii) *decompositions*, and (iii) *fud decompositions*.

Let *substrate histogram* $A \in \mathcal{A}_{U,V,z}$ be in *system* U and have *variables* V and *size* z. Given the *one functional definition set* $F \in \mathcal{F}_{U,1}$, such that $\mathrm{und}(F) \subseteq V$, let the *abstract* set valued function of the *substrate histograms* $Y_{U,F,\mathrm{W},z} \in \mathcal{A}_{U,V,z} \to \mathrm{P}(\mathcal{A}_U)$ be defined

$$Y_{U,F,\mathrm{W},z} = \{(A, \{(A * T_F)^{\mathrm{X}} : T \in F\}) : A \in \mathcal{A}_{U,V,z}\}$$

where $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$.

The set of *iso-fud-abstracts* is the intersection of the *iso-abstracts* of each *transform*

$$Y^{-1}_{U,F,\mathrm{W},z}(\{(A * T_F)^{\mathrm{X}} : T \in F\}) \quad = \quad \bigcap_{T \in F} Y^{-1}_{U,T_F,\mathrm{W},z}((A * T_F)^{\mathrm{X}})$$

The set of *iso-fud-abstracts* is a subset of the *iso-abstracts*,

$$Y^{-1}_{U,F,\mathrm{W},z}(\{(A * T_F)^{\mathrm{X}} : T \in F\}) \quad \subseteq \quad Y^{-1}_{U,F^{\mathrm{T}},\mathrm{W},z}((A * F^{\mathrm{T}})^{\mathrm{X}})$$

So the set of *iso-fud-abstracts* is *entity-like* with an *iso-abstractence* of

$$\frac{|Y^{-1}_{U,F,\mathrm{W},z}(\{(A * T_F)^{\mathrm{X}} : T \in F\})|}{|Y^{-1}_{U,\mathrm{i},F^{\mathrm{T}},\mathrm{W},z}((A * F^{\mathrm{T}})^{\mathrm{X}})|}$$

In the case where the *fud* is a singleton, $F = \{T\}$, the *iso-fud-abstracts* equals the *iso-abstracts*,

$$Y^{-1}_{U,\{T\},\mathrm{W},z}(\{(A * T)^{\mathrm{X}}\}) \quad = \quad Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

and the *iso-fud-abstracts* is maximally *entity-like*.

In the case where the *fud* consists of a single *layer* of *partition transforms*, $F \in \mathcal{F}_{U,\mathrm{P}}$ such that $|\mathrm{der}(F)| = |F|$, the *iso-fud-abstracts* equals the *iso-abstracts*, which is the intersection of the *iso-deriveds* of the *partition transforms*,

$$\begin{aligned} Y^{-1}_{U,F,\mathrm{W},z}(\{(A * T_F)^{\mathrm{X}} : T \in F\}) \quad &= \quad Y^{-1}_{U,F^{\mathrm{T}},\mathrm{W},z}((A * F^{\mathrm{T}})^{\mathrm{X}}) \\ &= \quad \bigcap_{T \in F} D^{-1}_{U,T,z}(A * T) \end{aligned}$$

In this case the *iso-fud-abstracts* is maximally *entity-like*.

Let the *formal* set valued function of the *substrate histograms* $Y_{U,F,\mathrm{V},z} \in \mathcal{A}_{U,V,z} \to \mathrm{P}(\mathcal{A}_U)$ be defined

$$Y_{U,F,\mathrm{V},z} = \{(A, \{A^{\mathrm{X}} * T_F : T \in F\}) : A \in \mathcal{A}_{U,V,z}\}$$

where $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$.

The set of *iso-fud-formals* is the intersection of the *iso-formals* of each *transform*

$$Y^{-1}_{U,F,\mathrm{V},z}(\{A^{\mathrm{X}} * T_F : T \in F\}) \;=\; \bigcap_{T \in F} Y^{-1}_{U,T_F,\mathrm{V},z}(A^{\mathrm{X}} * T_F)$$

If the *top transform* exists the set of *iso-fud-formals* is a subset of the *iso-formals*,

$$\exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F)) \implies Y^{-1}_{U,F,\mathrm{V},z}(\{A^{\mathrm{X}} * T_F : T \in F\}) \subseteq Y^{-1}_{U,F^{\mathrm{T}},\mathrm{V},z}(A^{\mathrm{X}} * F^{\mathrm{T}})$$

Let the *formal*-*abstract*-pair set valued function of the *substrate histograms* $Y_{U,F,z} \in \mathcal{A}_{U,V,z} \to \mathrm{P}(\mathcal{A}_U \times \mathcal{A}_U)$ be defined

$$Y_{U,F,z} = \{(A, \{(A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}) : T \in F\}) : A \in \mathcal{A}_{U,V,z}\}$$

where $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$.

The set of *iso-fud-independents* is the intersection of the *iso-transform independents* of each *transform*

$$Y^{-1}_{U,F,z}(\{(A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}) : T \in F\}) \;=\; \bigcap_{T \in F} Y^{-1}_{U,T_F,z}((A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}))$$

The set of *iso-fud-independents* is the intersection of the *iso-fud-formals* and the *iso-fud-abstracts*

$$Y^{-1}_{U,F,z}(\{(A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}) : T \in F\}) \;=\; Y^{-1}_{U,F,\mathrm{V},z}(\{A^{\mathrm{X}} * T_F : T \in F\}) \;\cap\; Y^{-1}_{U,F,\mathrm{W},z}(\{(A * T_F)^{\mathrm{X}} : T \in F\})$$

If the *top transform* exists the set of *iso-fud-independents* is a subset of the *iso-transform independents*,

$$\exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F)) \implies Y^{-1}_{U,F,z}(\{(A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}) : T \in F\}) \subseteq Y^{-1}_{U,F^{\mathrm{T}},z}((A^{\mathrm{X}} * F^{\mathrm{T}}, (A * F^{\mathrm{T}})^{\mathrm{X}}))$$

The set of *iso-fud-independents* is a subset of the *iso-abstracts*,

$$Y^{-1}_{U,F,z}(\{(A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}) : T \in F\}) \quad \subseteq \quad Y^{-1}_{U,F^{\mathrm{T}},\mathrm{W},z}((A * F^{\mathrm{T}})^{\mathrm{X}})$$

So the set of *iso-fud-independents* is *entity-like* with an *iso-abstractence* of

$$\frac{|Y^{-1}_{U,F,z}(\{(A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}) : T \in F\})|}{|Y^{-1}_{U,\mathrm{i},F^{\mathrm{T}},\mathrm{W},z}((A * F^{\mathrm{T}})^{\mathrm{X}})|}$$

In the case where the *fud* is a singleton, $F = \{T\}$, the *iso-fud-independents* equals the *iso-transform-independents*,

$$Y^{-1}_{U,\{T\},z}(\{(A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})\}) \quad = \quad Y^{-1}_{U,T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))$$

Given the *decomposition* of *one functional transforms* $D \in \mathcal{D}_U = \mathcal{D} \cap \mathrm{trees}(\mathcal{S}_U \times \mathcal{T}_{U,\mathrm{f},1})$, such that $\mathrm{und}(D) \subseteq V$, let the *component-formal-abstract*-pair function valued function of the *substrate histograms* $Y_{U,D,z} \in \mathcal{A}_{U,V,z} \to (\mathcal{A}_U \to (\mathcal{A}_U \times \mathcal{A}_U))$ be defined

$$Y_{U,D,z} = \{(A, \{(C, ((A*C)^{\mathrm{X}}*T, (A*C*T)^{\mathrm{X}})) : (C,T) \in \mathrm{cont}(D)\}) : A \in \mathcal{A}_{U,V,z}\}$$

where $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$.

The set of *iso-decomposition-independents* can be related to the set *iso-transform-independents* for each *slice*,

$$\begin{aligned}&\forall B \in Y^{-1}_{U,D,z}(Y_{U,D,z}(A))\ \forall (C,T) \in \mathrm{cont}(D)\\ &\qquad\qquad (B * C \in Y^{-1}_{U,T,z_C}(((A * C)^{\mathrm{X}} * T, (A * C * T)^{\mathrm{X}})))\end{aligned}$$

where $z_C = \mathrm{size}(A * C)$.

In the case where the *decomposition* consists of a root node only, $D = \{((\emptyset, T), \emptyset)\}$, the *iso-decomposition-independents* equals the *iso-transform-independents*,

$$Y^{-1}_{U,D,z}(Y_{U,D,z}(A)) \quad = \quad Y^{-1}_{U,T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))$$

Given the *fud decomposition* of *one functional definition sets* $D \in \mathcal{D}_{\mathrm{F},U} = \mathcal{D}_{\mathrm{F}} \cap \mathrm{trees}(\mathcal{S}_U \times \mathcal{F}_{U,1})$, such that $\mathrm{und}(D) \subseteq V$, let the *component-formal-abstract*-pair set function valued function of the *substrate histograms* $Y_{U,D,\mathrm{F},z} \in \mathcal{A}_{U,V,z} \to (\mathcal{A}_U \to \mathrm{P}(\mathcal{A}_U \times \mathcal{A}_U))$ be defined

$$\begin{aligned}&Y_{U,D,\mathrm{F},z} \ = \\ &\quad \{(A, \{(C, \{((A * C)^{\mathrm{X}} * T_F, (A * C * T_F)^{\mathrm{X}}) : T \in F\}) :\\ &\qquad\qquad (C,F) \in \mathrm{cont}(D)\}) : A \in \mathcal{A}_{U,V,z}\}\end{aligned}$$

The set of *iso-fud-decomposition-independents* can be related to the set of intersections of *iso-transform-independents* for each *slice*

$$\forall B \in Y^{-1}_{U,D,\mathrm{F},z}(Y_{U,D,\mathrm{F},z}(A))\ \forall (C,F) \in \mathrm{cont}(D)$$
$$(B * C \in \bigcap_{T \in F} Y^{-1}_{U,T_F,z_C}(((A * C)^{\mathrm{X}} * T_F, (A * C * T_F)^{\mathrm{X}})))$$

or

$$\forall (C,F) \in \mathrm{cont}(D)$$
$$(Y^{-1}_{U,D,\mathrm{F},z_C}(Y_{U,D,\mathrm{F},z_C}(A * C)) \subseteq \bigcap_{T \in F} Y^{-1}_{U,T_F,z_C}(((A * C)^{\mathrm{X}} * T_F, (A * C * T_F)^{\mathrm{X}})))$$

The set of *slice iso-fud-decomposition-independents* is a subset of the *slice iso-abstracts*

$$\forall (C,F) \in \mathrm{cont}(D)\ (Y^{-1}_{U,D,\mathrm{F},z_C}(Y_{U,D,\mathrm{F},z_C}(A * C)) \subseteq Y^{-1}_{U,F^{\mathrm{T}},\mathrm{W},z_C}((A * C * F^{\mathrm{T}})^{\mathrm{X}}))$$

so the set of *slice iso-fud-decomposition-independents* is *entity-like* with an *iso-abstractence* of

$$\frac{|Y^{-1}_{U,\mathrm{i},D,\mathrm{F},z_C}(Y_{U,D,\mathrm{F},z_C}(A * C))|}{|Y^{-1}_{U,\mathrm{i},F^{\mathrm{T}},\mathrm{W},z_C}((A * C * F^{\mathrm{T}})^{\mathrm{X}})|}$$

3.12.6 Integral iso-sets and entropy

The set of *integral substrate histograms* in *system* U, of *variables* V and *size* z is defined in section 'Iso-sets', above, as

$$\mathcal{A}_{U,\mathrm{i},V,z} = \{A : A \in \mathcal{A}_{U,\mathrm{i}},\ A^{\mathrm{U}} = V^{\mathrm{C}},\ \mathrm{size}(A) = z\}$$

Its cardinality is the cardinality of weak compositions $|\mathrm{C}'(V^{\mathrm{C}}, z)|$

$$|\mathcal{A}_{U,\mathrm{i},V,z}| = \frac{(z + v - 1)!}{z!\ (v-1)!}$$

where $v = |V^{\mathrm{C}}|$.

In the case where the *size* is less than or equal to the *volume*, $z \leq v$, the logarithm of the cardinality of weak compositions may be approximated

$$\begin{aligned} \ln \frac{(z + v - 1)!}{z!\ (v-1)!} &= \overline{z} \ln v - \underline{z} \ln z \\ &\approx z \ln \frac{v}{z} \end{aligned}$$

by abuse of notation. If the *size*, z, is fixed, log-cardinality varies with the logarithm of the *volume*, $\ln v$.

In the case where the *size* is greater than the *volume*, $z > v$, the log-cardinality approximates

$$\begin{aligned}\ln \frac{(z+v-1)!}{z!\ (v-1)!} &\approx \overline{v} \ln z - \underline{v} \ln v \\ &\approx v \ln \frac{z}{v}\end{aligned}$$

If the *volume*, v, is fixed, the log-cardinality varies with the logarithm of the *size*, $\ln z$.

The logarithm of the cardinality of weak compositions may also be analysed by means of Stirling's approximation in the case where $z \gg \ln z$ and $v \gg \ln v$,

$$\begin{aligned}\ln \frac{(z+v-1)!}{z!\ (v-1)!} &\approx (z+v)\ln(z+v) \ - \ z \ln z \ - \ v \ln v \\ &= z \ln \frac{z+v}{z} \ + \ v \ln \frac{z+v}{v} \\ &\approx (z \ln \frac{v}{z} \ : \ z < v) \ + \\ &\quad\ (2z \ln 2 \ : \ z = v) \ + \\ &\quad\ (v \ln \frac{z}{v} \ : \ z > v)\end{aligned}$$

The cardinality of the set of *integral iso-deriveds* is the product of the weak compositions of the *components*,

$$|D^{-1}_{U,\mathrm{i},T,z}(A * T)| \ = \ \prod_{(R,C) \in T^{-1}} \frac{((A*T)_R + |C| - 1)!}{(A*T)_R!\ (|C|-1)!}$$

The *integral iso-deriveds* is a subset of the *integral substrate histograms*, $D^{-1}_{U,\mathrm{i},T,z}(A * T) \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$, so the cardinality of the set of *integral iso-deriveds* is bounded

$$1 \ \leq \ |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \ \leq \ \frac{(z+v-1)!}{z!\ (v-1)!}$$

The logarithm of the cardinality of the set of *integral iso-deriveds* is

$$\begin{aligned}\ln |D^{-1}_{U,\mathrm{i},T,z}(A*T)| &= \sum_{(R,C) \in T^{-1}} \ln \frac{((A*T)_R + |C|-1)!}{(A*T)_R!\ (|C|-1)!} \\ &= \sum_{(R,\cdot) \in T^{-1}} \ln \frac{((A*T)_R + (V^{\mathrm{C}}*T)_R-1)!}{(A*T)_R!\ ((V^{\mathrm{C}}*T)_R-1)!} \\ &= \sum_{(\cdot,C) \in T^{-1}} \ln \frac{(\mathrm{size}(A*C) + |C|-1)!}{\mathrm{size}(A*C)!\ (|C|-1)!}\end{aligned}$$

The *integral iso-deriveds log-cardinality* is approximately bounded

$$0 \ \leq\ \ln |D^{-1}_{U,\mathrm{i},T,z}(A*T)| \ \leq\ (z+v)\ln(z+v) \ -\ z \ln z \ -\ v \ln v$$

In the case where the *volume* is much greater than one, $v \gg 1$, the *integral iso-deriveds log-cardinality* is approximately proportional to the negative *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}&\ln |D^{-1}_{U,\mathrm{i},T,z}(A*T)| \\ &\quad \approx \sum_{(R,\cdot) \in T^{-1}} (A*T + V^{\mathrm{C}}*T)_R \ln (A*T + V^{\mathrm{C}}*T)_R \\ &\qquad\quad - \sum_{(R,\cdot) \in T^{-1}} (A*T)_R \ln (A*T)_R \ -\ \sum_{(R,\cdot) \in T^{-1}} (V^{\mathrm{C}}*T)_R \ln (V^{\mathrm{C}}*T)_R \\ &\quad = (z+v)\ln(z+v) \ -\ z \ln z \ -\ v \ln v \\ &\qquad\quad - \big((z+v) \times \mathrm{entropy}(A*T + V^{\mathrm{C}}*T) \\ &\qquad\qquad\qquad -z \times \mathrm{entropy}(A*T) \ - v \times \mathrm{entropy}(V^{\mathrm{C}}*T)\big)\end{aligned}$$

In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, where the *derived counts* or *component sizes* are generally less than their *component cardinalities*, $A*T < V^{\mathrm{C}}*T$ or $\forall (\cdot,C) \in T^{-1}\ (\mathrm{size}(A*C) < |C|)$, then the *integral iso-deriveds log-cardinality* varies against the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned}\ln |D^{-1}_{U,\mathrm{i},T,z}(A*T)| &\sim \sum_{(R,\cdot) \in T^{-1}} (A*T)_R \ln \frac{(V^{\mathrm{C}}*T)_R}{(A*T)_R} \\ &\sim -z \times \mathrm{entropyRelative}(A*T, V^{\mathrm{C}}*T)\end{aligned}$$

Similarly, in the domain where the *size* is greater than the *volume*, $z > v$, where the *derived counts* or *component sizes* are generally greater than their *component cardinalities*, $A*T > V^{\mathrm{C}}*T$ or $\forall (\cdot,C) \in T^{-1}\ (\mathrm{size}(A*C) > |C|)$, then the *integral iso-deriveds log-cardinality* varies against the *volume* scaled *component cardinality size relative entropy*,

$$\begin{aligned}\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| &\sim \sum_{(R,\cdot) \in T^{-1}} (V^{\mathrm{C}} * T)_R \ln \frac{(A * T)_R}{(V^{\mathrm{C}} * T)_R} \\ &\sim -v \times \mathrm{entropyRelative}(V^{\mathrm{C}} * T, A * T)\end{aligned}$$

In both domains the *integral iso-deriveds log-cardinality* varies against the *relative entropy*. That is, *integral iso-deriveds log-cardinality* is minimised when (a) the *cross entropy* is maximised and (b) the *component entropy* is minimised. The *cross entropy* is maximised when high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

In the case where the *derived* is *independent*, $A * T = (A * T)^{\mathrm{X}}$, the cardinality of the set of *integral iso-abstracts* equals the cardinality of the set of *integral iso-deriveds*,

$$\begin{aligned}|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| &= |D^{-1}_{U,\mathrm{i},T,z}((A * T)^{\mathrm{X}})| \\ &= \prod_{(R,C) \in T^{-1}} \frac{((A * T)^{\mathrm{X}}_R + |C| - 1)!}{(A * T)^{\mathrm{X}}_R!\ (|C| - 1)!}\end{aligned}$$

and so in this case the *integral iso-abstracts log-cardinality* is approximately proportional to the negative *abstract size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}&\ln |Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| \\ &\quad\approx\ (z + v) \ln(z + v) \ -\ z \ln z\ -\ v \ln v \\ &\qquad\quad -\ \big((z + v) \times \mathrm{entropy}((A * T)^{\mathrm{X}} + V^{\mathrm{C}} * T) \\ &\qquad\qquad\qquad -z \times \mathrm{entropy}((A * T)^{\mathrm{X}})\ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)\big)\end{aligned}$$

Conjecture that the logarithm of the cardinality of the *integral iso independents* corresponding to A^{X} varies with the *size* scaled *independent entropy*,

$$\ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| \ \sim\ z \times \mathrm{entropy}(A^{\mathrm{X}})$$

The conjecture is suggested by considering the *state* $S \in V^{\mathrm{CS}}$ having minimum *count* in the *independent*, $S \in \mathrm{mind}(A^{\mathrm{X}})$, which therefore is also the minimum of the *perimeter*, $\forall w \in V\ (S \cap \mathrm{mind}(Q_A(w)) \neq \emptyset)$. The minimum *count* of the minimum *state*, $\mathrm{minr}(\{(w, Q_A(w)(S\%\{w\}) : w \in V\}) \in \mathbf{N}$, limits the number of ways the *iso-independents* can be cumulatively constructed. See 'Deltas and Perturbations', below, for a discussion of the construction of the *iso-independents* from a sequence of *circuit deltas*.

The positive correlation between the *integral iso-independents log-cardinality* and the *independent entropy* is consistent with the negative correlation between the *integral iso-abstracts log-cardinality* and the *relative entropy*,

$$\begin{aligned}
&\ln |Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| \\
&\quad\sim\quad - \big((z+v) \times \mathrm{entropy}((A * T)^{\mathrm{X}} + V^{\mathrm{C}} * T) \\
&\qquad\qquad -z \times \mathrm{entropy}((A * T)^{\mathrm{X}}) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)\big) \\
&\quad\sim\quad z \times \mathrm{entropy}((A * T)^{\mathrm{X}})
\end{aligned}$$

As the *transform* tends to *value full functional*, $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}}$, the *abstract* tends to the *independent*, $(A * T_{\mathrm{s}})^{\mathrm{X}} = A^{\mathrm{X}}$, and the *abstract entropy*, $\mathrm{entropy}((A * T)^{\mathrm{X}})$, tends to the *independent entropy*, $\mathrm{entropy}(A^{\mathrm{X}})$.

The *integral iso-transform-independents* is a subset of the *integral iso-abstracts*, $Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}})) \subseteq Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$, so the *integral iso-transform-independents log-cardinality* varies with the *integral iso-abstracts log-cardinality*,

$$\ln |Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))| \ \sim\ \ln |Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|$$

Conjecture that, in the case where *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, the *integral iso-transform-independents log-cardinality* varies against the *abstract size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}
&\ln |Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))| \ \sim \\
&\quad - \big((z+v) \times \mathrm{entropy}((A * T)^{\mathrm{X}} + V^{\mathrm{C}} * T) \\
&\qquad\qquad -z \times \mathrm{entropy}((A * T)^{\mathrm{X}}) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)\big)
\end{aligned}$$

and with the *size* scaled *abstract entropy*,

$$\ln |Y^{-1}_{U,\mathrm{i},T,z}((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}))| \ \sim\ z \times \mathrm{entropy}((A * T)^{\mathrm{X}})$$

3.12.7 Shuffled history

Closely related to the *independent histogram* are the *shuffles* of the non-empty *history* H of *histogram* $A = \mathrm{histogram}(H)$ of *size* $z = |H| > 0$, having at least one *variable* $n \geq 1$ where $V = \text{vars}(H)$ and $n = |V|$. The result is a set of *histories* of the same *size* z where the *variable values* are *shuffled* between the *states*.

Define shuffles $\in \mathcal{H} \to \text{P}(\mathcal{H})$ as the monoidal concatenation of the set of *reduced histories* having *shuffle* prefixed *event identifiers*

$$\text{shuffles}(H) :=$$
$$\prod\{\{\{((X,x),T) : (x,(\cdot,T)) \in X\} : X \in \text{ids}(H) \cdot (H\%\{v\})\} : v \in V\}$$

where $(\%) \in \mathcal{H} \times \text{P}(\mathcal{V}) \to \mathcal{H}$ is defined $H\%W := \{(x, S\%W)) : (x,S) \in H\}$, the monoidal product is defined $\prod X = \text{fold1}((*), \text{flip}(Q))$ for some $Q \in \text{enums}(X)$, and the monoidal product operator $(*) \in \mathcal{H} \times \mathcal{H} \to \mathcal{H}$ is

$$G * I := \{((X \cup Y, x), S \cup T) : ((X,x),S) \in G,\ ((Y,y),T) \in I,\ x = y\}$$

The shuffles function is defined where $n \geq 1$.

The *shuffles* all have the same *size*, $\forall G \in \text{shuffles}(H)\ (|G| = z)$. The *histograms* of the *shuffles* include the *histogram* of the given *history*, $A \in \{\text{his}(G) : G \in \text{shuffles}(H)\}$, where his = histogram. If *histogram* A is *singleton*, $|A^{\text{F}}| = 1$, then there is only one *shuffle histogram*, $\{\text{his}(G) : G \in \text{shuffles}(H)\} = \{A\}$. The cardinality of the *shuffles* is $|\text{shuffles}(H)| = z!^n$. The cardinality of the *shuffle histograms* is less than or equal to the cardinality of the *shuffles*, $|\{\text{his}(G) : G \in \text{shuffles}(H)\}| \leq |\text{shuffles}(H)|$. If *history* H is *mono-variate*, $n = 1$, then there is only one *shuffle histogram*, $\{\text{his}(G) : G \in \text{shuffles}(H)\} = \{A\}$. All *histograms* of the *shuffles* of H are *congruent*,

$$\forall G \in \text{shuffles}(H)\ \lozenge B = \text{his}(H)\ (\text{congruent}(B, A))$$

All the *histograms* of the *shuffles* of H have the same *independent histogram*

$$\forall G \in \text{shuffles}(H)\ \lozenge B = \text{his}(G)\ (B^{\text{X}} \equiv A^{\text{X}})$$

The *scaled sum* of the *histograms* of any subset of the *shuffled histories* also has the same *independent*

$$\forall P \subseteq \text{shuffles}(H)\ \lozenge B = \text{scalar}(1/|P|) * \sum_{G \in P} \text{his}(G)\ (B^{\text{X}} \equiv A^{\text{X}})$$

The *scaled sum* of the *histograms* of the *shuffled histories* equals the *independent*

$$\text{scalar}(1/z!^n) * \sum(\text{his}(G) : G \in \text{shuffles}(H)) \equiv A^{\text{X}}$$

3.13 Rolls

A *roll*, $R \in \mathcal{S} \to \mathcal{S}$, is a function on the *cartesian* set of *states* of some set of *variables* V in *system* U, $R \in V^{\mathrm{CS}} \to V^{\mathrm{CS}}$.

Define the set of *rolls* as $\mathrm{rolls} \subset \mathcal{S} \to \mathcal{S}$. Define $\mathrm{vars} \in (\mathcal{S} \to \mathcal{S}) \to \mathrm{P}(\mathcal{V})$ as $\mathrm{vars}(R) := \{v : S \in \mathrm{dom}(R) \cup \mathrm{ran}(R),\ v \in \mathrm{vars}(S)\}$. *Rolls* are constrained such that $\forall R \in \mathrm{rolls}\ \forall S \in \mathrm{dom}(R) \cup \mathrm{ran}(R)\ (\mathrm{vars}(S) = \mathrm{vars}(R))$.

Define the application of a *roll* R in *variables* V to a *histogram* A having the same *variables* $\mathrm{vars}(R) = \mathrm{vars}(A) = V$ as $\mathrm{roll} \in \mathrm{rolls} \times \mathcal{A} \to \mathcal{A}$

$$\begin{aligned}
\mathrm{roll}(R, A) &:= \{(S, c) : (S, c) \in A,\ S \notin \mathrm{dom}(R)\} + \sum_{S \in A^{\mathrm{S}} \cap \mathrm{dom}(R)} \{(R_S, A_S)\} \\
&= \sum_{S \in A^{\mathrm{S}} \setminus \mathrm{dom}(R)} \{(S, A_S)\} + \sum_{S \in A^{\mathrm{S}} \cap \mathrm{dom}(R)} \{(R_S, A_S)\}
\end{aligned}$$

Define $(*) \in \mathcal{A} \times \mathrm{rolls} \to \mathcal{A}$ as $A * R := \mathrm{roll}(R, A)$. The application of the empty *roll* is defined $A * \emptyset = A$. (Note that the operator has a type ambiguity between the empty *roll*, $\emptyset \in \mathrm{rolls}$, and the *empty histogram*, $\emptyset \in \mathcal{A}$.) $A * R$ is undefined if $\mathrm{vars}(R) \neq \mathrm{vars}(A)$. If A is *scalar* then the only applicable *roll*, apart from the empty *roll*, is $\{(\emptyset, \emptyset)\}$ and so a *scalar rolls* to itself, $\mathrm{scalar}(z) * \{(\emptyset, \emptyset)\} = \mathrm{scalar}(z)$. The *rolled histogram* $A * R$ is *congruent* to the *underlying histogram* A, $\mathrm{congruent}(A, A * R)$. That is, the application is *size-conserving*, $\mathrm{size}(A * R) = \mathrm{size}(A)$ and in the same *variables*, $\mathrm{vars}(A * R) = \mathrm{vars}(A)$. The part of A for which the *states* $Q \subset \mathrm{states}(A)$ are neither in the domain nor range of R, $Q = \mathrm{states}(A) \setminus \mathrm{dom}(R) \setminus \mathrm{ran}(R)$, is unchanged under application, $A * Q^{\mathrm{U}} * R = A * R * Q^{\mathrm{U}} = A * Q^{\mathrm{U}}$. If $A \in \mathcal{A}_U$ and $R \in \mathrm{cartesian}(U)(\mathrm{vars}(A)) \to \mathrm{cartesian}(U)(\mathrm{vars}(A)) \subset \mathrm{rolls}$ then $A * R \in (A^{\mathrm{CS}} \to \mathbf{Q}_{\geq 0}) \subset \mathcal{A}_U$.

Define the *identity roll* $\mathrm{id}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{rolls}$ as $\mathrm{id}(U)(V) := \{(S, S) : S \in V^{\mathrm{CS}}\}$. Application of the *identity roll* leaves a *histogram* A unchanged, $A * \mathrm{id}(U)(\mathrm{vars}(A)) = A$.

A *roll* $R \in \mathrm{rolls}$ is *circular* if there exists a source *state* which is also a target *state*, $\mathrm{dom}(R) \cap \mathrm{ran}(R) \neq \emptyset$. The *identity roll* on *variables* V, $\mathrm{id}(U)(V)$, is *circular*.

The set of *rolls* in *substrate variables* V of *system* U, $V^{\mathrm{CS}} \to V^{\mathrm{CS}} \subset \mathrm{rolls}$, can be constructed

$$V^{\mathrm{CS}} \to V^{\mathrm{CS}} = \{R : R \subseteq V^{\mathrm{CS}} \times V^{\mathrm{CS}},\ |\mathrm{dom}(R)| = |R|\} = \prod_{S \in V^{\mathrm{CS}}} \{S\} \times V^{\mathrm{CS}}$$

The *substrate rolls* includes the *empty roll*, $\emptyset \in V^{\mathrm{CS}} \to V^{\mathrm{CS}}$. The cardinality of the set of *substrate rolls* is bounded $|V^{\mathrm{CS}} \to V^{\mathrm{CS}}| \leq 2^y y^y$ where $y = |V^{\mathrm{C}}|$. The subset of the *substrate rolls* which have cardinality equal to the *volume* are the *substrate complete rolls*, $\{R \in V^{\mathrm{CS}} \to V^{\mathrm{CS}},\ |R| = |V^{\mathrm{CS}}|\}$. The cardinality of the *substrate complete rolls* is $|V^{\mathrm{CS}} :\to V^{\mathrm{CS}}| = y^y$.

A list of *rolls*, for example $L \in \mathcal{L}(V^{\mathrm{CS}} \to V^{\mathrm{CS}})$, in *variables* V, can be applied to a *histogram* A in sequence, because the application of each *roll* results in a *congruent histogram* to which a successive *roll* may be applied. Define $\mathrm{roll} \in \mathcal{L}(\mathrm{rolls}) \times \mathcal{A} \to \mathcal{A}$ as $\mathrm{roll}(L, A) := \mathrm{roll}(\mathrm{sequence}(L), A)$ and $\mathrm{roll} \in \mathcal{K}(\mathrm{rolls}) \times \mathcal{A} \to \mathcal{A}$ as

$$\begin{aligned} \mathrm{roll}((R, X), A) &:= \mathrm{roll}(X, A * R) \\ \mathrm{roll}(\emptyset, A) &:= A \end{aligned}$$

$\mathrm{roll}(L, A)$ is undefined unless all of the *rolls* are in the same *variables* as the *histogram*, $\forall R \in \mathrm{set}(L)\ (\mathrm{vars}(R) = \mathrm{vars}(A))$. Define $(*) \in \mathcal{A} \times \mathcal{L}(\mathrm{rolls}) \to \mathcal{A}$ as $A * L := \mathrm{roll}(L, A)$. The application of the empty list of *rolls* is defined $A * \emptyset = A$. (Again note the operator type ambiguity between $\emptyset \in \mathcal{L}(\mathrm{rolls})$ and $\emptyset \in \mathcal{A}$.) The application of a list of *rolls* to a *histogram* is left associative, $A * L = A * L_1 * L_2 \ldots * L_l = (A * L_1) * L_2 \ldots * L_l = ((A * L_1) * L_2) \ldots * L_l$, where $l = |L|$. The *list rolled histogram* $A * L$ is *congruent* to the *underlying histogram* A, $\mathrm{congruent}(A, A * L)$. If $A \in \mathcal{A}_U$ and $L \in \mathcal{L}(A^{\mathrm{CS}} \to A^{\mathrm{CS}}) \subset \mathcal{L}(\mathrm{rolls})$ then $A * L \in (A^{\mathrm{CS}} \to \mathbf{Q}_{\geq 0}) \subset \mathcal{A}_U$.

A pair of *rolls*, $R_1, R_2 \in \mathrm{rolls}$, in the same *variables*, $\mathrm{vars}(R_1) = \mathrm{vars}(R_2) = V$, is a pair of endomorphic functions and hence can be composed to form a single *roll* $R_2 \circ R_1 = \mathrm{compose}(R_1, R_2) \in \mathrm{rolls}$ (see Appendix 'Function composition'). The function composition is here defined as an outer join $R_2 \circ R_1$. Define $\mathrm{compose} \in (\mathcal{S} \to \mathcal{S}) \times (\mathcal{S} \to \mathcal{S}) \to (\mathcal{S} \to \mathcal{S})$ as

$$\begin{aligned} &\mathrm{compose}(R_1, R_2) := \\ &\quad \{(S_1, T_2) : (S_1, T_1) \in R_1,\ (S_2, T_2) \in R_2,\ S_2 = T_1\} \cup \\ &\quad \{(S_1, T_1) : (S_1, T_1) \in R_1,\ T_1 \notin \mathrm{dom}(R_2)\} \cup \\ &\quad \{(S_2, T_2) : (S_2, T_2) \in R_2,\ S_2 \notin \mathrm{dom}(R_1)\} \end{aligned}$$

Define $R_2 \circ R_1 :=$ compose(R_1, R_2). compose(R_1, R_2) is undefined if vars$(R_1) \neq$ vars(R_2). The domain of a composition is the union of the domains of the arguments, dom$(R_2 \circ R_1) =$ dom$(R_1) \cup$ dom(R_2). A sequence of compositions of *rolls* is right associative, $R_3 \circ R_2 \circ R_1 = R_3 \circ (R_2 \circ R_1)$.

If the composition of a *roll* R in *variables* V and the *identity roll* $\mathrm{id}(U)(V)$ is equal to the *identity roll*, $R \circ \mathrm{id}(U)(V) = \mathrm{id}(U)(V)$, then R is said to be an *identity equivalent roll*.

A list of *rolls*, $L \in \mathcal{L}(\mathrm{rolls})$, in *variables* V, $\forall R \in \mathrm{set}(L)\ (\mathrm{vars}(R) = V)$, can be composed recursively, compose $\in \mathcal{L}(\mathrm{rolls}) \to \mathrm{rolls}$. For example, $\mathrm{compose}(\{(1, R_1), (2, R_2), (3, R_3)\}) = R_3 \circ R_2 \circ R_1$. The application of a *roll list* to a *histogram* A, $\mathrm{roll}(L, A)$, is equal to the application of the joined list, $A * L = A * \mathrm{compose}(L)$. For example, let *roll list* $L = \{(1, R_1), (2, R_2), (3, R_3)\}$, then $A*L = A*R_1*R_2*R_3 = A*(R_3 \circ R_2 \circ R_1)$.

A *roll list* L is *unique* where the unioned list has the same cardinality as the sum of the cardinalities of the *rolls*, $|\bigcup \mathrm{set}(L)| = \sum_{i \in \{1 \ldots |L|\}} |R_i|$. That is, each map of the *rolls* appears only once, $\forall i \in \{1 \ldots |L| - 1\}\ \forall i \in \{i+1 \ldots |L|\}\ (L_i \cap L_j = \emptyset)$.

A *roll list* L is *functional* where the unioned list is functional, $\bigcup \mathrm{set}(L) \in \mathcal{S} \to \mathcal{S}$.

A *roll list* L is *non-circular* where no source *state* subsequently appears as a target *state*,

$$\forall i \in \{1 \ldots |L|\}\ \forall S \in \mathrm{dom}(L_i)\ (S \notin \{S_2 : R \in \mathrm{set}(L_{\{i \ldots |L|\}}),\ (S_1, S_2) \in R\})$$

The composition of a *non-circular functional roll list*, $L \in \mathcal{L}(V^{\mathrm{CS}} \to V^{\mathrm{CS}})$, has disjoint domain and range, $\mathrm{dom}(R) \cap \mathrm{ran}(R) = \emptyset$ where $R = \mathrm{compose}(L)$. Thus $R \in \mathrm{dom}(R) \to (V^{\mathrm{CS}} \setminus \mathrm{dom}(R))$.

A *unique non-circular functional roll list* L is such that no source *state* subsequently appears as either a source or target *state*,

$$\begin{aligned} &\forall i \in \{1 \ldots |L|\}\ \Diamond X = \bigcup \mathrm{set}(L_{\{i+1 \ldots |L|\}}) \\ &\quad (\mathrm{dom}(L_i) \cap (\mathrm{ran}(L_i) \cup \mathrm{dom}(X) \cup \mathrm{ran}(X)) = \emptyset) \end{aligned}$$

A *roll* $R \in$ rolls having *variables* V can be converted to a *partition* $P \in \mathcal{R}_U$, by taking the functional inverse of the *roll* stuffed with the *identity roll*, $P = \mathrm{ran}(\mathrm{inverse}(R'))$ where $R' = R \circ \mathrm{id}(U)(V)$. A *roll list*, $L \in$

$\mathcal{L}(V^{\text{CS}} \to V^{\text{CS}})$, can be mapped to a reverse *partition* sequence. That is, a list of *partitions* such that each is succeeded by parent *partitions*. Let $K = \{(i, \text{ran}(\text{inverse}(\text{compose}(L_{\{1 \ldots i\}}) \circ \text{id}(U)(V)))) : i \in \{1 \ldots |L|\}\} \in \mathcal{L}(\mathcal{R}_U)$. Then $|K| \geq 2 \implies \forall i \in \{1 \ldots |K|\}\ \forall j \in \{i+1 \ldots |K|\}\ (\text{parent}(K_j, K_i))$.

A *roll* $R \in \text{rolls}$ having *variables* V can converted to a *transform* in $\mathcal{T}_{U,V}$. One method is to create a *partition transform* $P^{\text{T}} \in \mathcal{T}_{U,V}$ on the *partition* $P \in \mathcal{R}_U$ of the *cartesian states* of the *variables*, $P \in \text{B}(V^{\text{CS}})$, implied by the functional inverse, $P = \text{ran}(\text{inverse}(R \circ \text{id}(U)(V)))$. This *transform* has a single *derived variable*, $|\text{der}(P^{\text{T}})| = 1$, and therefore the *derived histogram* is *independent*, $A * P^{\text{T}} = (A * P^{\text{T}})^{\text{X}}$, when applied to some *underlying histogram* A in *variables* V.

Another method is create a *partition-set* and thence a *fud* of *partition transforms*. Each *partition transform* corresponds to an *underlying variable*, $v \in V$, by *reducing* the target *state* to that *variable*. Each *derived variable partitions* the entire *cartesian* set of *states* of all the *variables*, V^{CS}, not just the *cartesian* of the *underlying variable*, $\{v\}^{\text{CS}}$. Define $\text{transform}(U, V) \in \text{rolls} \to \mathcal{T}_{U,\text{f},1}$ as

$$\begin{aligned}&\text{transform}(U, V)(R) := \\ &\quad \{P^{\text{T}} : v \in V,\ P = \text{ran}(\text{inverse}(\{(S_1, S_2\%\{v\}) : (S_1, S_2) \in R'\}))\}^{\text{T}}\end{aligned}$$

where $\text{vars}(R) = V$ or $R = \emptyset$, and $R' = R \circ \text{id}(U)(V)$ is the given *roll* stuffed with the *identity roll*. Define $R^{\text{T}} := \text{transform}(U)(R)$ where the *system* U and the *substrate variables* V are implicit. The *transforms* of *rolls* in *variables* V form a subset of the *substrate transforms set*, $\{R^{\text{T}} : R \in V^{\text{CS}} \to V^{\text{CS}}\} \subset \mathcal{T}_{U,V}$. The cardinality of the *derived variables* of the *transform* of a *roll* is less than or equal to the cardinality of the *underlying variables*, $|W| \leq |V|$, where $W = \text{der}(R^{\text{T}})$. The *valency* of each of the *derived variables* is less than or equal to that of its corresponding *underlying variable*, $\forall v \in V\ (|\text{ran}(\text{inverse}(\{(S_1, S_2\%\{v\}) : (S_1, S_2) \in R'\}))| \leq |U_v|)$. Thus the *derived volume* is less than or equal to the *underlying volume* $|W^{\text{C}}| \leq |V^{\text{C}}|$. In the case of non-empty *substrate variables*, $V \neq \emptyset$, if the cardinality of *derived* and *underlying variables* is the same, $|W| = |V|$, then the *volume* of the *derived histogram* is equal to the *size* of the *effective cartesian subvolume* formed by the *independent* of the *roll* of the *cartesian histogram* $|W^{\text{C}}| = |(V^{\text{C}} * R)^{\text{XF}}|$. If the *roll* of the *cartesian histogram* is equal to the *cartesian histogram*, $(V^{\text{C}} * R)^{\text{XF}} = V^{\text{C}}$, then the *volume* of the *derived histogram* is equal to the *volume* of the *underlying*, $|W^{\text{C}}| = |V^{\text{C}}|$. In this case, the *roll transform* is *left total*, $X\%V = V^{\text{C}}$ where $X = \text{his}(R^{\text{T}})$, and the *underlying volume* equals the *derived volume*, $|W^{\mathrm{C}}| = |V^{\mathrm{C}}|$, so the *roll transform* is a *frame transform*. The *transform* is also *right total*, $X\%W = W^{\mathrm{C}}$, and hence *frame full functional*, if the *roll* is an *identity equivalent roll*, $R \circ \mathrm{id}(U)(V) = \mathrm{id}(U)(V)$, or if the *roll* is otherwise *circular*. If the *roll* is an *identity equivalent roll*, $R \circ \mathrm{id}(U)(V) = \mathrm{id}(U)(V)$, then the *roll transform* is a *value full functional transform* or *reframe transform*. In this case the *roll transform* is the singleton element of the *strong self non-overlapping substrate self-cartesian transforms set*, $\{R^{\mathrm{T}}\} = \{\{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}}\} = \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,s}}\} \subseteq \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n,s}}$.

The cardinality of the set of *transforms* of *rolls* is bounded by the cardinality of the subset of the *substrate rolls* which have cardinality equal to the *volume*, $|V^{\mathrm{CS}} :\to V^{\mathrm{CS}}| = y^y$. That is, $|\{R^{\mathrm{T}} : R \in V^{\mathrm{CS}} \to V^{\mathrm{CS}}\}| \leq y^y$. This may be compared to the cardinality of its superset, the *substrate transforms set*, $|\mathcal{T}_{U,V}| = 2^{\mathrm{bell}(y)}$.

3.13.1 Value rolls

A *value roll* is equivalent to a special case of a *roll* $R \in V^{\mathrm{CS}} \to V^{\mathrm{CS}}$ on *variables* V. Let $v \in V$ be one of the *variables*, and $s, t \in U_v$ be source and target *values* of v. The *value roll* (V, v, s, t) has a corresponding *roll* R such that all *states* incident on the source *value* s are mapped to the target *value* t, $R = \{(S,\ S \setminus \{(v, s)\} \cup \{(v, t)\}) : S \in V^{\mathrm{CS}},\ S_v = s\}$. Define the set of *value rolls* $\mathrm{rollValues}(U) \subset \mathrm{P}(\mathcal{V}_U) \times \mathcal{V}_U \times \mathcal{W}_U \times \mathcal{W}_U$ such that $\forall (V, v, s, t) \in \mathrm{rollValues}(U)\ ((v \in V) \wedge (\{s, t\} \subseteq U_v))$.

In order to construct a *non-circular roll* from a *value roll*, define $\mathrm{roll}(U) \in \mathrm{rollValues}(U) \to (\mathrm{rolls} \cap (\mathcal{S}_U \to \mathcal{S}_U))$ as

$$
\begin{aligned}
&\mathrm{roll}(U)((V, v, s, t)) := \\
&\quad \{(S,\ S \setminus \{(v, s)\} \cup \{(v, t)\}) : S \in \mathrm{cartesian}(U)(V),\ (v, s) \in S\} \setminus \mathrm{id}(U)(V)
\end{aligned}
$$

If $t = s$ then the *roll* is empty, $\mathrm{roll}(U)((V, v, s, s)) = \emptyset$. Define $(V, v, s, t)^{\mathrm{R}} := \mathrm{roll}(U)((V, v, s, t))$.

The *substrate value rolls set* in *variables* V of *system* U is

$$
\{(V, v, s, t) : v \in V,\ s, t \in U_v\} \subseteq \mathrm{rollValues}(U)
$$

The *substrate value rolls set* has cardinality $|\{(V, v, s, t) : v \in V,\ s, t \in U_v\}| = \sum_{v \in V} |U_v|^2$. In the case of *regular variables* of *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_v| : v \in V\}$, the cardinality is nd^2.

The *independent* of the application of a *value roll* (V, v, s, t) to a *histogram* A is equal to the application of a *value roll* to the *independent histogram*

$$(A * (V, v, s, t)^{\mathrm{R}})^{\mathrm{X}} = A^{\mathrm{X}} * (V, v, s, t)^{\mathrm{R}}$$

Thus the application of a *value roll* to an *independent histogram* is also *independent*,

$$A^{\mathrm{X}} * (V, v, s, t)^{\mathrm{R}} = (A^{\mathrm{X}} * (V, v, s, t)^{\mathrm{R}})^{\mathrm{X}}$$

Let $\mathcal{J}_{U,V}$ be the infinite *substrate value roll lists set* in *variables* V and *system* U, $\mathcal{J}_{U,V} = \{L : L \in \mathcal{L}(\mathrm{rollValues}(U)),\ (\forall (W, v, s, t) \in \mathrm{set}(L)\ (W = V))\}$. The list of *value rolls* $J \in \mathcal{J}_{U,V}$ can be converted into a list of *rolls*, $\mathrm{map}(\mathrm{roll}(U), J) = \{(i, (V, v, s, t)^{\mathrm{R}}) : (i, (V, v, s, t)) \in J\} \subset \mathcal{L}(\mathrm{rolls})$. Define $J^{\mathrm{R}} := \mathrm{map}(\mathrm{roll}(U), J)$. A list of *value rolls* J may be composed indirectly, $\mathrm{compose}(J^{\mathrm{R}})$. The application of a *value roll list* to a *histogram* A, $\mathrm{roll}(J^{\mathrm{R}}, A)$, is equal to the application of the joined list, $A * J^{\mathrm{R}} = A * \mathrm{compose}(J^{\mathrm{R}})$. *Rolled independent histograms* remain *independent* $A^{\mathrm{X}} * J^{\mathrm{R}} = (A^{\mathrm{X}} * J^{\mathrm{R}})^{\mathrm{X}}$.

A list of *value rolls* $J \in \mathcal{J}_{U,V}$ in *variables* V can be rearranged without altering its composition $\mathrm{compose}(J^{\mathrm{R}})$ so long as the order of the *value rolls* remains the same for each *variable*. That is, the set of lists filtered by *variable* can be concatenated in any order. Let $\mathrm{filtv}(U) \in \mathcal{V}_U \times \mathcal{L}(\mathrm{rollValues}(U)) \to \mathcal{L}(\mathrm{rollValues}(U))$ be defined as $\mathrm{filtv}(U)(v, J) := \mathrm{filter}(\{(N,\ w = v) : N \in \mathrm{rollValues}(U),\ (\cdot, w, \cdot, \cdot) = N\}, J)$. Then $\forall K \in \mathcal{L}(V)\ (|K| = |V| \wedge \mathrm{set}(K) = V \implies \mathrm{compose}((\mathrm{concat}(\{(i, \mathrm{filtv}(U)(v, J)) : (i, v) \in K\}))^{\mathrm{R}}) = \mathrm{compose}(J^{\mathrm{R}}))$.

Consider a *non-circular functional* list of *value rolls* $J \in \mathcal{J}_{U,V}$ in the same *variable* $v \in V$, $J = \mathrm{filtv}(U)(v, J)$. The cumulative initial sub-lists of J can be mapped to a reverse partition sequence. Let $l = |J|$. Let $K = \{(i, \{(s, t)\}) : (i, (\cdot, \cdot, s, t)) \in J\} \in \mathcal{L}(U_v \to U_v)$. Let $\mathrm{id}(U)(v) = \{(u, u) : u \in U_v\} \in U_v \to U_v$. Let $L = \{(i, \mathrm{ran}(\mathrm{inverse}(\mathrm{compose}(K_{\{1 \ldots i\}}) \circ \mathrm{id}(U)(v)))) : i \in \{1 \ldots l\}\} \in \mathcal{L}(\mathrm{B}(U_v))$. Then $l \geq 1 \implies \mathrm{parent}(L_1, \mathrm{id}(U)(v))$ and $l \geq 2 \implies \forall i \in \{1 \ldots l-1\}\ (\mathrm{parent}(L_{i+1}, L_i))$. If, in addition, J is constrained to be a list of *non-identity value rolls*, $\forall (\cdot, \cdot, s, t) \in \mathrm{set}(J)\ (s \neq t)$, then $l \geq 1 \implies |L_1| = |U_v| - 1$ and $l \geq 2 \implies \forall i \in \{1 \ldots l-1\}\ (|L_{i+1}| = |L_i| - 1)$. The maximum length of such a list is $d - 1$ where $d = |U_v|$. The cardinality of the set of all such *value* partition lists of maximum length is $d!(d-1)!/2^{d-1}$. The maximum length of a *unique non-circular functional* list of *non-identity value rolls* in *regular variables* V, having *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$, is $n(d-1)$. The cardinality of the set of possible *value rolls* at the head of such a list is $nd(d-1)$. The cardinality of the set of possible *value* partitions at the head is half of this, $nd(d-1)/2$, because of the degeneracy of $\{(s,t)\}$ and $\{(t,s)\}$. The cardinality of the set of *value* partitions which correspond to the composed initial sublists of the lists in $\mathcal{J}_{U,V}$ is $\sum((\text{bell}(d)-1)^k : k \in \{0\ldots n\})$.

A list of *value rolls* $J \in \mathcal{J}_{U,V}$ in *variables* V can be converted into a *transform*, $J^{\text{RT}} = \text{transform}(U)(\text{map}(\text{roll}(U), J)))$, by first converting each *value roll* to a *roll*, then composing and finally converting to a *transform* as defined above. An alternative method is to construct a *fud* $F \in \mathcal{F}_{U,\text{P}}$ of *partition transforms* such that each *partition variable* corresponds to one of the *underlying variables*. That is, such that there is a surjective map between the *underlying variables* and the *partition variables*, $\text{und}(F) \rightarrow: \text{der}(F)$. This method highlights the fact that the resultant *fud* F is a *non-overlapping* partition of the *underlying variables*. First compose by *variable* by filtering the *value roll* list. Define $\text{compv}(U) \in \mathcal{V}_U \times \mathcal{L}(\text{rollValues}(U)) \to (\mathcal{W}_U \to \mathcal{W}_U)$ as $\text{compv}(U)(v,J) := \text{compose}(\{(i,\{(s,t)\}) : (i,(\cdot,\cdot,s,t)) \in \text{filtv}(U)(v,J)\})$, which is such that $\text{compv}(U)(v,J) \in U_v \to U_v$. Then partition the *values*, define $\text{partv}(U) \in \mathcal{V}_U \times \mathcal{L}(\text{rollValues}(U)) \to \text{P}(\text{P}(\mathcal{W}_U))$ as $\text{partv}(U)(v,J) := \text{ran}(\text{inverse}(\text{compv}(U)(v,J) \circ \text{id}(U)(v)))$, which is such that $\text{partv}(U)(v,J) \in \text{B}(U_v)$. Then $F = \{\{\{v\} \times C : C \in \text{partv}(U)(v,J)\}^{\text{T}} : v \in V\} \in \mathcal{F}_{U,\text{P}}$. Define $\text{transform}(U) \in \mathcal{L}(\text{rollValues}(U)) \to \mathcal{T}_{U,\text{f},1}$ as

$$\text{transform}(U)(J) := \{\{\{\{(v,u)\} : u \in C\} : C \in \text{partv}(U)(v,J)\}^{V\text{T}} : v \in V\}^{\text{T}}$$

which is defined when $J \in \mathcal{J}_{U,V}$ and $J \neq \emptyset$. Define $J^{\text{T}} := \text{transform}(U)(J)$. Thus $J^{\text{T}} \in \mathcal{T}_{U,V}$. Define $(V,v,s,t)^{\text{T}} := \{(1,(V,v,s,t))\}^{\text{T}}$.

The *value roll list transform* equals the *roll list transform*, $J^{\text{T}} = J^{\text{RT}}$. Therefore the constraints on J^{RT} also apply to J^{T}. That is, the cardinality of the *derived variables* is less than or equal to the cardinality of the *underlying variables*, $|W| \leq |V|$, where $W = \text{der}(J^{\text{T}})$. The *valency* of each of the *derived variables* is less than or equal to that of its corresponding *underlying variable*, $\forall v \in V\ (|\text{partv}(U)(v,J)| \leq |U_v|)$. The *value roll list transform* is *non-overlapping*, $\neg\text{overlap}(J^{\text{T}})$. The *transform* is constructed from a *non-overlapping fud* of *partition transforms* and hence the *transform* must be *right total*, $(X\%W)^{\text{F}} = W^{\text{C}}$ where $(X,W) = J^{\text{T}}$. Thus the *derived volume* is less than or equal to the *underlying volume*, $|W^{\text{C}}| \leq |V^{\text{C}}|$. In the case of non-empty *substrate variables*, $V \neq \emptyset$, if the cardinalities of the *derived* and *underlying variables* is the same, $|W| = |V|$, then the *volume* of the *derived histogram* is equal to the *size* of the *effective cartesian sub-volume* formed by the *roll* of the *cartesian histogram* $|W^{\mathrm{C}}| = |(V^{\mathrm{C}} * J^{\mathrm{R}})^{\mathrm{F}}|$. The *value roll list transform* can only be *full functional* if the stuffed *roll* is the *identity roll* $J^{\mathrm{R}} \circ \mathrm{id}(U)(V) = \mathrm{id}(U)(V)$. In this case the *value roll list transform* is a *value full functional transform* or *reframe transform*. It is the singleton of the *strong self non-overlapping substrate self-cartesian transforms set*, $\{J^{\mathrm{T}}\} = \{\{\{v\}^{\mathrm{CS\{\}}V\mathrm{T}} : v \in V\}^{\mathrm{T}}\} = \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n,s}}\} \subseteq \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n,s}}$.

The set of *transforms* of *composed value rolls*, $\{J^{\mathrm{T}} : J \in \mathcal{J}_{U,V}\} \subset \mathcal{T}_{U,V}$ is the *self non-overlapping substrate transforms set* $\mathcal{T}_{U,V,\mathrm{n,s}}$ which is a subset of the *non-overlapping substrate transforms set*,

$$\{J^{\mathrm{T}} : J \in \mathcal{J}_{U,V}\} = \mathcal{T}_{U,V,\mathrm{n,s}} \subseteq \mathcal{T}_{U,V,\mathrm{n}}$$

If V is *regular* having *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$, the cardinality of $\mathcal{T}_{U,V,\mathrm{n,s}}$ is bounded $|\mathcal{T}_{U,V,\mathrm{n,s}}| \leq \mathrm{bell}(d)^n$. This may be compared to the cardinality of its superset, the *non-overlapped substrate transforms set*, $|\mathcal{T}_{U,V,\mathrm{n}}| < 2 \times \mathrm{bell}(n) \times \mathrm{bell}(d^n) + 1$, and thence to the cardinality of its superset, the *substrate transforms set*, $|\mathcal{T}_{U,V}| = 2^{\mathrm{bell}(d^n)}$

$$\mathrm{bell}(d)^n < 2 \times \mathrm{bell}(n) \times \mathrm{bell}(d^n) \leq 2^{\mathrm{bell}(d^n)}$$

It may also be compared to the cardinality of its superset, the set of *transforms* of *rolls*, $|\{R^{\mathrm{T}} : R \in V^{\mathrm{CS}} \to V^{\mathrm{CS}}\}| \leq d^{nd^n}$

$$\mathrm{bell}(d)^n < d^{nd^n}$$

Finally compare it to the cardinality of its superset, the *fixed dimension substrate transforms set*, $|\{T : T \in \mathcal{T}_{U,V},\ |\mathrm{der}(T)| = n\}| = \mathrm{bell}(d^n)^{\underline{n}}/n!$

$$\mathrm{bell}(d)^n < \frac{1}{n!}\mathrm{bell}(d^n)^{\underline{n}}$$

The subset of the *transforms* of *composed value rolls*, $\{J^{\mathrm{T}} : J \in \mathcal{J}_{U,V}\} = \mathcal{T}_{U,V,\mathrm{n,s}}$, where the *value roll list* has only one *value roll*, $J = \{(1, (V, v, s, t))\}$ and $t \neq s$, is the intersection of the *substrate decremented transforms set* and the *self non-overlapping substrate transforms set*

$$\begin{aligned}
&\{J^{\mathrm{T}} : J \in \mathcal{J}_{U,V},\ |J| = 1,\ (\cdot, \cdot, s, t) = J_1,\ t \neq s\} \\
&\quad = \mathcal{T}_{U,V,-} \cap \mathcal{T}_{U,V,\mathrm{n,s}} \\
&\quad = \{(\{Q^{V\mathrm{T}}\} \cup \{\{u\}^{\mathrm{CS\{\}}V\mathrm{T}} : u \in V \setminus \{x\}\})^{\mathrm{T}} : x \in V,\ Q \in \mathrm{decs}(\{x\}^{\mathrm{CS\{\}}})\}
\end{aligned}$$

where decs = decrements $\in \mathcal{R}_U \to \mathrm{P}(\mathcal{R}_U)$. If the *substrate* V is *regular* having *valency* d, then $|\mathcal{T}_{U,V,-} \cap \mathcal{T}_{U,V,\mathrm{n,s}}| = nd(d-1)/2$, which is the cardinality of the set of possible *value* partitions at the head of a *non-circular functional* list of *non-identity value rolls.*

As mentioned above, the set of *transforms* of *composed value rolls* is the *self non-overlapping substrate transforms set*, $\{J^{\mathrm{T}} : J \in \mathcal{J}_{U,V}\} = \mathcal{T}_{U,V,\mathrm{n,s}}$. In section 'Substrate structures', above, the *self non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n,s}}$, can be constructed from *linear fuds* of *multi-partition transforms*, $L \in \mathcal{L}(\mathcal{T}_{U,\mathrm{P}^*})$, where the first *transform* is the *value full functional transform* and subsequent *transforms* are *strong self non-overlapping substrate decremented transforms.* In the case of non-empty *substrate*, $V \neq \emptyset$,

$$\begin{aligned}\mathcal{T}_{U,V,\mathrm{n,s}} = \{(\bigcup \mathrm{set}(L))^{\mathrm{TNT}} : M = \{\{v\}^{\mathrm{CS}\{\}} : v \in V\},\\ L \in \mathrm{subpaths}(\{(M, \mathrm{tdec}(U)(M))\})\}\end{aligned}$$

where the tree of *self non-overlapping substrate decremented partition-sets* is defined $\mathrm{tdec}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathrm{P}(\mathcal{R}_U))$ as

$$\begin{aligned}&\mathrm{tdec}(U)(M) := \{(N, \mathrm{tdec}(U)(N)) :\\ &\quad w \in M,\ Q \in \mathrm{decs}(\{w\}^{\mathrm{CS}\{\}}),\ N = \{Q\} \cup \{\{u\}^{\mathrm{CS}\{\}} : u \in M,\ u \neq w\}\}\end{aligned}$$

A subset of the non-empty *unique non-circular functional substrate value roll lists* can be defined such that the source *value* is greater than the target *value* where there exists some *order* on the *values* of each of the *variables* $D_v \in \mathrm{enums}(U_v)$. Define the *decrementing value roll lists*

$$\begin{aligned}\mathcal{J}_{U,V,-} = \{J : J \in \mathcal{J}_{U,V},\ J \neq \emptyset,\ \forall (\cdot,\cdot,s,t) \in \mathrm{set}(J)\ (s > t),\\ \forall (i,(\cdot,v_i,s_i,t_i)),(j,(\cdot,v_j,s_j,t_j)) \in J\\ (v_i = v_j \implies (i \leq j \implies t_j \neq s_i) \wedge (i \neq j \implies s_j \neq s_i))\}\end{aligned}$$

The *decrementing value roll lists*, $\mathcal{J}_{U,V,-}$, can be constructed recursively by means of a tree of *value rolls* $\mathrm{tdecrv}(U,V) \in \mathcal{J}_{U,V} \to \mathrm{trees}(\mathrm{rollValues}(U))$, defined as

$$\begin{aligned}&\mathrm{tdecrv}(U,V)(J) :=\\ &\quad \{((V,v,s,t), \mathrm{tdecrv}(U,V)(J \cup \{(|J|+1,(V,v,s,t))\})) :\\ &\qquad v \in V,\ s,t \in U_v,\ s > t,\\ &\qquad X = \{x : (\cdot,w,x,\cdot) \in \mathrm{set}(J),\ w = v\},\ s \notin X,\ t \notin X\}\end{aligned}$$

Each source *value* may be *rolled* no more than once for each *variable*, and only to lesser target *values.* The *decrementing value roll lists* is $\mathcal{J}_{U,V,-} =$

subpaths(tdecrv$(U,V)(\emptyset)$). The *decrementing value roll lists*, $\mathcal{J}_{U,V,-}$, is a finite set, corresponding bijectively to *linear fuds* of *strong self non-overlapping substrate decremented transforms*,

$$\mathcal{J}_{U,V,-} :\leftrightarrow: \text{subpaths}(\text{tdec}(U)(V))$$

because the construction trees map bijectively,

$$\text{places}(\text{tdecrv}(U,V)(\emptyset)) :\leftrightarrow: \text{places}(\text{tdec}(U)(V))$$

To construct *value rolls* in a *decrementing value roll list*, $J \in \mathcal{J}_{U,V,-}$, given the application of the *decrementing value roll list* so far, $V^{\mathrm{C}} * J_{\{1 \ldots i\}}$, it is only necessary to test that the *source* and *target values* on the *perimeter* are non-zero. That is, putative *value roll* $(\cdot, v, s, t)$ may be added, $J_{i+1} = (V, v, s, t)$, if $(V^{\mathrm{C}} * J^{\mathrm{R}}_{\{1 \ldots i\}} \ \% \ \{v\})(\{(v, s)\}) \neq 0$ and $(V^{\mathrm{C}} * J^{\mathrm{R}}_{\{1 \ldots i\}} \ \% \ \{v\})(\{(v, t)\}) \neq 0$.

3.14 Deltas and Perturbations

A *delta* is a pair of *histograms* $(D, I) \in \mathcal{A} \times \mathcal{A}$ in the same *variables* V, $\text{vars}(D) = \text{vars}(I) = V$. The *application* of a *delta* to a *histogram* A in *variables* V is the *subtraction* of D followed by the *addition* of I. The resultant *histogram*, $A - D + I$, is a *perturbation* of A. An *effective delta* is such that the *perturbation* is no more *effective* than the given *histogram*, $(A - D + I)^{\mathrm{F}} \leq A^{\mathrm{F}}$. A *zero delta* is such that $A - D + I \equiv A$.

If $D \leq A$ then $\text{size}(A - D + I) = \text{size}(A) - \text{size}(D) + \text{size}(I)$. If $D \leq A$ and the *delta* is *congruent*, $\text{size}(D) = \text{size}(I)$, then the *histogram* and its *perturbation* are *congruent*, $\text{congruent}(A,\ A - D + I)$.

If $D \leq A$ and the *congruent delta histograms* are *effective singletons*, $|D^{\mathrm{F}}| = |I^{\mathrm{F}}| = 1$, of unit *size*, $\text{size}(D) = \text{size}(I) = 1$, then the *congruent perturbation*, $A - D + I$, is an *event perturbation*. The set of *event perturbations* of *histogram* A in *variables* V and *system* U is

$$\{A - \{R\}^{\mathrm{U}} + \{S\}^{\mathrm{U}} : (R, d) \in A,\ d \geq 1,\ S \in A^{\mathrm{CS}}\}$$

The set of *effective event perturbations* is

$$\{A - \{R\}^{\mathrm{U}} + \{S\}^{\mathrm{U}} : (R, d) \in A,\ d \geq 1,\ (S, c) \in A,\ c > 0\}$$

which is independent of *system*.

Let $(D, I) \in \mathcal{A} \times \mathcal{A}$ be a *delta* of *histogram* A such that $D \leq A$. A *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ is a functor (or monoid homomorphism) of the *delta application* operator, $(A * T) - (D * T) + (I * T) = (A - D + I) * T$ where $\mathrm{vars}(D) = \mathrm{vars}(I) = \mathrm{und}(T) = \mathrm{vars}(A)$.

Let $N_{(D,I)} \in \mathcal{E}_U$ be a *histogram expression* of *delta* $(D, I) \in \mathcal{A} \times \mathcal{A}$ having *variables* $\mathrm{vars}(D) = \mathrm{vars}(I) = V$. The *expression application* to *histogram* A in *variables* V is such that $N_{(D,I)}(A) = A - D + I \in \mathcal{A}$. See appendix 'Histogram expressions' for a definition of $N_{(D,I)}$.

The *application* of a *roll* $R \in V^{\mathrm{CS}} \to V^{\mathrm{CS}} \subset \mathrm{rolls}$ to a *histogram* A is equivalent to the *application* of a *congruent delta* $(D, I) \in \mathcal{A} \times \mathcal{A}$ in *variables* V. Define $\mathrm{delta} \in \mathrm{rolls} \times \mathcal{A} \to (\mathcal{A} \times \mathcal{A})$ as

$$\mathrm{delta}(R, A) := \left(\sum_{S \in A^{\mathrm{S}} \cap \mathrm{dom}(R)} \{(S, A_S)\}, \sum_{S \in A^{\mathrm{S}} \cap \mathrm{dom}(R)} \{(R_S, A_S)\} \right)$$

which is defined if $\mathrm{vars}(R) = \mathrm{vars}(A)$. Thus the *congruent perturbation* equals the *rolled histogram*, $A - D + I = A * R$, where $(D, I) = \mathrm{delta}(R, A)$.

A *value roll* $(V, w, s, t) \in \mathrm{rollValues}(U)$ is a special case of a *roll*, $(V, w, s, t)^{\mathrm{R}} \in V^{\mathrm{CS}} \to V^{\mathrm{CS}} \subset \mathrm{rolls}$, and hence implies an equivalent *congruent perturbation* to the *histogram*, $A - D + I = A * (V, w, s, t)^{\mathrm{R}}$, where $(D, I) = \mathrm{delta}((V, w, s, t)^{\mathrm{R}}, A)$. Similarly a *value roll list* $J \in \mathcal{J}_{U,V} \subset \mathcal{L}(\mathrm{rollValues}(U))$ on *variables* V in *system* U implies an equivalent *congruent perturbation*, $A - D + I = A * J^{\mathrm{R}}$, where $(D, I) = \mathrm{delta}(J^{\mathrm{R}}, A)$.

Consider the subset of the *substrate transforms set*, $\mathcal{T}_{U,V}$, on *variables* V in *system* U, which are *transforms* of *value roll lists*, $\{J^{\mathrm{T}} : J \in \mathcal{J}_{U,V}\} \subset \mathcal{T}_{U,V}$. The *application* of the *delta* (D, I) corresponding the *value roll list* $J \in \mathcal{J}_{U,V}$, $(D, I) = \mathrm{delta}(J^{\mathrm{R}}, A)$, to the *histogram* A is isomorphic to the *application* of the *transform* J^{T}, $A * J = A - D + I \cong A * J^{\mathrm{T}}$, but not equal because the *derived variables* are not equal to the *underlying variables*, $\mathrm{der}(J^{\mathrm{T}}) \neq \mathrm{und}(J^{\mathrm{T}})$, and hence $\mathrm{vars}(A * J) \neq \mathrm{vars}(A * J^{\mathrm{T}})$.

As shown above, the *application* of a *value roll* $(V, w, s, t) \in \mathrm{rollValues}(U)$ to an *independent histogram* A^{X} in *variables* V, conserves *independence*, $A^{\mathrm{X}} * (V, w, s, t)^{\mathrm{R}} = (A^{\mathrm{X}} * (V, w, s, t)^{\mathrm{R}})^{\mathrm{X}}$. Thus the corresponding *delta* $(D, I) = \mathrm{delta}((V, w, s, t)^{\mathrm{R}}, A)$ also conserves *independence*, $A^{\mathrm{X}} - D + I = (A^{\mathrm{X}} - D + I)^{\mathrm{X}}$. Similarly for *value roll lists*, $A^{\mathrm{X}} * J^{\mathrm{R}} = (A^{\mathrm{X}} * J^{\mathrm{R}})^{\mathrm{X}} = A^{\mathrm{X}} - D + I =$

$(A^{\mathrm{X}} - D + I)^{\mathrm{X}}$ where $(D, I) = \mathrm{delta}(J^{\mathrm{R}}, A)$.

Let $N_R \in \mathcal{E}_U$ be a *histogram expression* of *roll* $R \in \mathrm{rolls}$ having *variables* V. The *expression application* to *histogram* A in *variables* V is such that $N_R(A) = A * R \in \mathcal{A}$. See appendix 'Histogram expressions' for a definition of N_R.

Let $Y_{U,\mathrm{i},V,z} \in \mathcal{A}_{U,\mathrm{i},V,z} \to \mathcal{A}_{U,V,z}$ be the subset of the *independent* function, $Y_{U,\mathrm{i},V,z} = \{(A, A^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\} \subset \mathrm{independent}$. The set of *integral iso-independents* of *integral histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ is $Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$. Let $Q_A \subset \mathcal{A}_{\mathrm{i}} \times \mathcal{A}_{\mathrm{i}}$ be the subset of the *integral congruent deltas* which conserve *iso-independence*, $\forall (D, I) \in Q_A\ (A - D + I \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))$. The *perimeters* of each of the *iso-independent perturbations* are equal, $\forall (D, I) \in Q_A\ \forall w \in V\ ((A - D + I)\%\{w\} = A\%\{w\})$. None of the *iso-independent deltas* result in *event perturbations* except for the *zero delta*, $\forall (D, I) \in Q_A\ (\mathrm{size}(I) \neq 1)$.

Define the *circuit deltas* as the subset of *iso-independent deltas* having *size* less than or equal to two, $C_A = \{(D, I) : (D, I) \in Q_A,\ \mathrm{size}(I) \leq 2\}$. The *circuit deltas* may be defined explicitly,

$$C_A = \{(\{S, T\}^{\mathrm{U}}, \{S\%(V \setminus W) \cup T\%W,\ T\%(V \setminus W) \cup S\%W\}^{\mathrm{U}}) : S, T \in A^{\mathrm{FS}},\ W \subseteq V\}$$

Conjecture that all of the *iso-independent deltas* are linear *sums* of the *circuit deltas*.

$$\forall (D, I) \in Q_A\ \exists L \in \mathcal{L}(C_A)\ ((D = \sum X : i \in \{1 \ldots |L|\},\ (X, \cdot) = L_i) \wedge (I = \sum Y : i \in \{1 \ldots |L|\},\ (\cdot, Y) = L_i))$$

Value roll deltas $A - D + I = A * (V, w, s, t)^{\mathrm{R}}$ cannot be *iso-independent deltas*, $A - D + I \notin Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, because the *iso-independents* are equivalence classes of the *independent*. That is, $A^{\mathrm{X}} * (V, w, s, t)^{\mathrm{R}} \notin Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$.

Similarly, given some *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ where $\mathrm{und}(T) = V$ and $W = \mathrm{der}(T)$, define $Y_{U,\mathrm{i},T,z} \in \mathcal{A}_{U,\mathrm{i},V,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})$ as $Y_{U,\mathrm{i},T,z} = \{(A, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$. The finite set of *integral iso-transform-independents* of $((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})$ is $Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$. A *delta* (D, I) is *iso-transform-independence conserving* with respect to T if $A - D + I \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$. In the stronger case of (i) the *delta* is *iso-independence conserving*, $A - D + I \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, so that $(A-D+I)^{\mathrm{X}} = A^{\mathrm{X}}$ and $(A-D+I)^{\mathrm{X}}*T = A^{\mathrm{X}}*T$, and (ii) the *transformed applied delta* is *iso-independence conserving*, $(A-D+I)*T \in Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})$, so that the *delta* is *iso-abstract*, $((A-D+I)*T)^{\mathrm{X}} = (A*T)^{\mathrm{X}}$, then the *delta* is *iso-transform-independence conserving*.

If the *transform* is a *self partition*, $T^{\mathrm{P}} = V^{\mathrm{CS}\{\}}$, then the set of *integral iso-transform-independents* equals the set of *integral iso-independents* in the *underlying variables*, $Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}})) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and hence only *iso-independent deltas*, $A - D + I \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, are *iso-transform-independence conserving* in this case.

If the *transform* is a *unary partition*, $T^{\mathrm{P}} = \{V^{\mathrm{CS}}\}$, then the set of *integral iso-transform-independents* equals the *integral congruent support* in the *underlying variables*, $Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}})) = \mathcal{A}_{U,\mathrm{i},V,z}$, and any *congruent delta* is *iso-transform-independence conserving* in this case.

If the *transform* is a *value roll transform*, $T = J^{\mathrm{T}}$ where $J = \{(1,(V,w,s,t))\} \in \mathcal{J}_{U,V}$, then the *value roll delta*, $A-D+I = A*(V,w,s,t)^{\mathrm{R}}$, is *iso-transform-independence conserving*, $A*(V,w,s,t)^{\mathrm{R}} \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}}))$, because $(A-D+I)*T = A*T$ and $(A-D+I)^{\mathrm{X}}*T = A^{\mathrm{X}}*T$.

As discussed in section 'Transforms and Independent', above, given a *substrate histogram* $A \in \mathcal{A}_{U,V,z}$ and a *substrate transform* $T \in \mathcal{T}_{U,V}$, having *derived variables* $W = \mathrm{der}(T)$, the case where the *formal* equals the *abstract*, $A^{\mathrm{X}}*T = (A*T)^{\mathrm{X}}$, is equivalent to requiring each *partition variable derived histogram* to equal the *partition variable derived histogram* of the *independent*,

$$
\begin{aligned}
&A^{\mathrm{X}}*T = (A*T)^{\mathrm{X}} \iff \\
&\quad\quad \forall P \in W\ (A*P^{\mathrm{T}} = A^{\mathrm{X}}*P^{\mathrm{T}}) \\
&\quad = \ \forall P \in W\ \forall C \in P\ (\mathrm{size}(A*C^{\mathrm{U}}) = \mathrm{size}(A^{\mathrm{X}}*C^{\mathrm{U}}))
\end{aligned}
$$

If the *transform* is also *non-overlapping*, then the constraint can be expressed in terms of the *contraction* of the *partition variable*,

$$
\begin{aligned}
&\neg\mathrm{overlap}(T)\ \wedge\ A^{\mathrm{X}}*T = (A*T)^{\mathrm{X}} \iff \\
&\quad\quad \forall P \in W\ (A\%V_P*P^{\%\mathrm{T}} = A^{\mathrm{X}}\%V_P*P^{\%\mathrm{T}})
\end{aligned}
$$

So the subset of *substrate histograms* which are such that the *formal* equals the *abstract* may be partly generated from the *integral congruent deltas* which preserve *iso-independence* but which are constrained to the *partition component*. In the case where the *transform* is *non-overlapping*,

$$\begin{aligned}
&\neg\text{overlap}(T) \ \wedge\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies \\
&\quad \{A : A \in \mathcal{A}_{U,V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\} \\
&\supseteq\ \{(A^{\mathrm{X}}\%V_P - D + I) * (\hat{A}^{\mathrm{X}}\%(V \setminus V_P)) : A \in \mathcal{A}_{U,V,z}, \\
&\qquad\qquad P \in W,\ C \in P,\ (D,I) \in Q_{A\%V_P},\ D^{\mathrm{F}} \subseteq C^{\mathrm{U}},\ I^{\mathrm{F}} \subseteq C^{\mathrm{U}}\}
\end{aligned}$$

3.15 Histogram expression lists

A *histogram expression list* $M \in \mathcal{L}(\mathcal{E}_U)$ may be applied to an argument *histogram* $A \in \mathcal{A}_U$ in sequence to produce a list of *histograms*. Define *histogram expression list application* $M(A) \in \mathcal{L}(\mathcal{A}_U)$ such that $M(A)_1 = A$, $|M(A)| = |M| + 1$ and $\forall i \in \{1 \ldots |M|\}\ (M(A)_{i+1} = M_i(M(A)_i))$. The *application* of a *histogram expression list*, $M(A)$, is (i) *size conserving* if the *histograms* have the same *size* as the argument, $\forall B \in \text{set}(M(A))\ (\text{size}(B) = \text{size}(A))$, (ii) *variables conserving* if the *histograms* have the same *variables* as the argument, $\forall B \in \text{set}(M(A))\ (\text{vars}(B) = \text{vars}(A))$, (iii) *congruent* if the *histograms* are *congruent* to the argument, $\forall B \in \text{set}(M(A))\ (\text{congruent}(B, A))$, (iv) *independence conserving* if the *histograms* are all *independent*, $\forall B \in \text{set}(M(A))\ (B = B^{\mathrm{X}})$, (v) *iso-independence conserving* if the *histograms* are all in the same *iso-independent* component, $\text{set}(M(A)) \subseteq Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, (vi) *circular* if the last *histogram* is equal to the first, $M(A)_l = A$ where $l = |M|+1$.

Given the initial *histogram* $A \in \mathcal{A}_U$ and two *histogram expression lists* $R, S \in \mathcal{L}(\mathcal{E}_U)$ which are such that the final *histograms* are equal, $R(A)_l = S(A)_m$, where $l = |R(A)|$ and $m = |S(A)|$, then the longer list may be said to have *smaller histogram expressions*. That is, if $m > l$ then S has *smaller histogram expressions* than R, and vice-versa.

Given a list of *histograms* $L \in \mathcal{L}(\mathcal{A})$ in the same set of *variables* V, $\forall B \in \text{set}(L)\ (\text{vars}(B) = V)$, a list of *deltas* can be implied, $P = \{(L_i - L_{i+1}, L_{i+1} - L_i) : i \in \{1 \ldots |L| - 1\}\} \in \mathcal{L}(\mathcal{A} \times \mathcal{A})$. Let $N_{(D,I)} \in \mathcal{E}_U$ be a *histogram expression* of *delta* $(D, I) \in \mathcal{A} \times \mathcal{A}$ having *variables* $\text{vars}(D) = \text{vars}(I) = V$. If the *histogram expression list* of *deltas*, $N_P = \{(i, N_{(D,I)}) : (i, (D, I)) \in P\} \in \mathcal{L}(\mathcal{E}_U)$, is *applied* to the first *histogram* in list L then the list is recovered, $N_P(L_1) = L$. The *delta histogram expression list*, N_P, is *congruent* with respect to L_1.

Given two *histogram expression lists* of *deltas* $N_P, N_Q \in \mathcal{L}(\mathcal{E}_U)$, where $P, Q \in$

$\mathcal{L}(\mathcal{A} \times \mathcal{A})$, which are such that the final *histograms* are equal, $N_P(A)_p = N_Q(A)_q$, where $p = |N_P(A)|$ and $q = |N_Q(A)|$, then the longer list may be said to have *smaller delta expressions*. That is, if $q > p$ then N_Q has *smaller delta expressions* than N_P, and vice-versa.

If a *histogram expression list* $M \in \mathcal{L}(\mathcal{E}_U)$ is *size conserving* but not *congruent* with respect to initial *histogram* $A \in \mathcal{A}_U$ then in some cases there may exist a corresponding list of *one functional transforms* $X \in \mathcal{L}(\mathcal{T}_{U,\mathrm{f},1})$ such that $\forall i \in \{1 \ldots |M|\}\ (A * X_i = M(A)_{i+1})$. In addition, if the *histogram expression list*, M, consists of *transform histogram expressions*, $N_T \in \mathcal{E}_U$ which are such that $N_T(B) = B * T$, constructed from a list of *one functional transforms* $Y \in \mathcal{L}(\mathcal{T}_{U,\mathrm{f},1})$ such that $M = \{(i, N_T) : (i, T) \in Y\}$, then the *transforms* of Y may be viewed as the changes between the *transforms* of X, $(X_i, X_{i-1}) \cong Y_i$. Again, the longer the list, $M(A)$, the *smaller* the *transform histogram expressions*, $\mathrm{set}(M)$.

Given a *histogram expression list* $M \in \mathcal{L}(\mathcal{E}_U)$ and an initial *histogram* $A \in \mathcal{A}_U$, an *orbit* is the pair of *histogram expression list applications* initialised by the *histogram* itself, A, and the *independent histogram*, A^{X}. That is, $(M(A), M(A^{\mathrm{X}}))$. The lengths of the *orbit* lists are equal, $|M(A)| = |M(A^{\mathrm{X}})|$. If it is the case that the *orbit* map is such that $\forall i \in \{1 \ldots l\}\ (M(A)_i^{\mathrm{X}} = M(A^{\mathrm{X}})_i)$, where $l = |M(A)|$, then the *histogram expression list*, M, is *independence conserving*, $\forall B \in \mathrm{set}(M(A^{\mathrm{X}}))\ (B = B^{\mathrm{X}})$, and the *orbit* is said to be an *independent function*, $\mathrm{map}(\mathrm{independent}, M(A)) = M(A^{\mathrm{X}})$. If $M(A)$ is *iso-independence conserving*, $\mathrm{set}(M(A)) \subseteq Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, and the *orbit* is an *independent function* then the *histogram expressions* are identity functions when *applied* to the *independent* argument, $\mathrm{set}(M(A^{\mathrm{X}})) = \{A^{\mathrm{X}}\}$.

Let $N_R \in \mathcal{E}_U$ be a *histogram expression* of *roll* $R \in \mathrm{rolls}$ having *variables* V. The *orbit* of the non-empty *value roll list* $J \in \mathcal{J}_{U,V}$ is *functionally independent*. Let $M_J = \{(i, N_{(V,w,s,t)^{\mathrm{R}}}) : (i, (V, w, s, t)) \in J\}$, then *orbit* $(M_J(A), M_J(A^{\mathrm{X}}))$ is such that $\{(i, M_J(A)_i^{\mathrm{X}}) : i \in \{1 \ldots l\}\} = M_J(A^{\mathrm{X}})$ and hence is an *independent function*. The *value roll histogram expression list*, M_J, is not *iso-independence conserving* with respect to A and hence $M_J(A^{\mathrm{X}})$ is not singleton, $|M_J(A^{\mathrm{X}})| > 1$.

The *transforms* of a *non-circular fud* $F \in \mathcal{F}$ can be arranged in a list of *layer fuds* $L = \mathrm{inverse}(\{(T, \mathrm{layer}(F, \mathrm{der}(T))) : T \in F\}) \in \mathcal{L}(\mathrm{P}(F))$. A *linear fud* is a *non-circular fud* such that the *underlying variables* of the *transforms* in each *layer fud* are the *derived variables* of the *layer fud* immediately below, $\forall i \in \{2 \ldots |L|\}\ (\mathrm{und}(L_i) \subseteq \mathrm{der}(L_{i-1}))$. Each of the *layer fuds*,

$\mathrm{set}(L) \in \mathrm{B}(F)$, can be combined into a single *transform*. Thus a *linear fud* may be represented as a *histogram expression list* of *transform expressions*, $M_F = \{(i, N_{G^{\mathrm{T}}}) : (i, G) \in L\} \in \mathcal{L}(\mathcal{E}_U)$, which is such that $M_F(A)_l = A * F^{\mathrm{T}}$ where $\mathrm{und}(F) = \mathrm{vars}(A)$ and $l = |M_F(A)|$. The *transform expressions* of M_F may be viewed as changes between the cumulative *fuds* at each *layer*. Let $T_{\{1 \ldots i\}} = \mathrm{transform}(\bigcup \mathrm{set}(\mathrm{take}(i, L))) = \{T : T \in F,\ i \leq \mathrm{layer}(F, \mathrm{der}(T))\}^{\mathrm{T}}$. Then $M_F(A)_i = A * T_{\{1 \ldots i\}}$ and $M_F(i)(A * T_{\{1 \ldots i-1\}}) = A * T_{\{1 \ldots i\}}$. If the *layer transforms*, $\{(i, G^{\mathrm{T}}) : (i, G) \in L\} \in \mathcal{L}(\mathcal{T}_{U,\mathrm{f},1})$, are *non-overlapping*, $\forall i \in \{1 \ldots |L|\}\ (\neg\mathrm{overlap}(L_i^{\mathrm{T}}))$, then the *application* of the *linear fud histogram expression list*, $M_F(A^{\mathrm{X}})$, is *independence conserving*, $M_F(A^{\mathrm{X}})_i * L_i^{\mathrm{T}} = (M_F(A^{\mathrm{X}})_i * L_i^{\mathrm{T}})^{\mathrm{X}}$. However, the *orbit*, $(M_F(A), M_F(A^{\mathrm{X}}))$, is not necessarily an *independent function*. This is because it is not always the case that $(M_F(A)_i * L_i^{\mathrm{T}})^{\mathrm{X}} = M_F(A^{\mathrm{X}})_i * L_i^{\mathrm{T}}$ even if L_i^{T} is *non-overlapping*.

A *non-circular fud* $H \in \mathcal{F}$ that is not necessarily a *linear fud* can be viewed as a *histogram expression list* of *histogram expressions* that *multiply* the argument by the *product* of the *histograms* of the *layer fud transforms*. Let $N_G \in \mathcal{E}_U$, where $G \in \mathcal{F}$, be such that $N_G(B) = B * \prod_{(X,\cdot) \in G} X$. Then $M_H = \{(i, N_G) : (i, G) \in L\} \in \mathcal{L}(\mathcal{E}_U)$, where $L = \mathrm{inverse}(\{(T, \mathrm{layer}(H, \mathrm{der}(T))) : T \in H\}) \in \mathcal{L}(\mathrm{P}(H))$. Then M_H is such that $M_H(A)_l \ \%\ \mathrm{der}(H) = A * H^{\mathrm{T}}$ where $l = |M_H(A)|$. The cardinality of the *variables* of the *histograms* of the *applied* list, $M_H(A)$, increase as *derived variables* are added in each *layer*, $\forall i \in \{1 \ldots l\}\ (|\mathrm{vars}(M_H(A)_{i+1})| \geq |\mathrm{vars}(M_H(A)_i)|)$. The *histogram expression list*, M_H, is *size conserving*. Also the *underlying histogram* is conserved, $\{B\%V : B \in M_H(A)\} = \{A\}$ where $V = \mathrm{vars}(A)$.

A *distinct decomposition* $X \in \mathcal{D}_{\mathrm{d},U}$ having *underlying variables* V has a *components* tree, $\mathrm{components}(U)(X) \in \mathrm{trees}(\mathrm{P}(V^{\mathrm{CS}}))$. Let $L \in \mathcal{L}(\mathrm{P}(V^{\mathrm{CS}}))$ be one of the paths, $L \in \mathrm{paths}(\mathrm{components}(U)(X))$. Each successive *component* on the path is a subset of the previous *component*, $\forall i \in \{1 \ldots |L| - 1\}\ (L_{i+1} \subseteq L_i)$. Thus there exists a *delta histogram expression list* $N_L = \{(i, N_{(D,I)}) : i \in \{1 \ldots |L| - 1\},\ (D, I) = (L_i^{\mathrm{U}} - L_{i+1}^{\mathrm{U}}, \emptyset)\} \in \mathcal{L}(\mathcal{E}_U)$ where $N_{(D,I)} \in \mathcal{E}_U$ is a *histogram expression* of *delta* $(D, I) \in \mathcal{A} \times \mathcal{A}$ having *variables* $\mathrm{vars}(D) = \mathrm{vars}(I) = V$. The *application* to the first *component* L_1 recovers the *components* path, $N_L(L_1^{\mathrm{U}}) = \{(i, C^{\mathrm{U}}) : (i, C) \in L\}$. If the path's *components* are *cartesian sub-volumes*, $\forall C \in \mathrm{set}(L)\ (C^{\mathrm{UF}} = C^{\mathrm{UXF}})$, then the *expression list application* is *independence conserving*.

3.16 Distinct geometry sized cardinal substrate histograms

Let the set of *sized cardinal substrate histograms* $\mathcal{A}_z$ be the set of *complete integral cardinal substrate histograms* of *size* z and *dimension* less than or equal to the *size* such that the *independent* is *completely effective*

$$\mathcal{A}_z = \{A : A \in \mathcal{A}_{\mathrm{c}} \cap \mathcal{A}_{\mathrm{i}},\ \mathrm{size}(A) = z,\ |V_A| \leq z,\ A^{\mathrm{U}} = A^{\mathrm{XF}} = A^{\mathrm{C}}\}$$

where $A^{\mathrm{CS}} = \mathrm{cartesian}(U_A)(V_A)$ and $U_A = \mathrm{implied}(\mathrm{implied}(A))$ and $V_A = \mathrm{vars}(A)$. There is no single *system* that contains all the *substrate histograms.* The infinite *implied system*, U_A where $A \in \mathcal{A}_z$, contains the *substrate variables*, $V_A \subset \mathrm{vars}(U_A)$, and all the *partition variables* in the *power functional definition set* on V_A, $\forall F \in \mathcal{F}_{U_A,V_A}\ (\mathrm{vars}(F) \subset \mathrm{vars}(U_A))$.

The set of *substrate histograms* of zero *size* is empty, $\mathcal{A}_0 = \emptyset$. The set of *substrate histograms* of *size* one is a singleton of the *mono-variate, mono-valent histogram*, $\mathcal{A}_1 = \{\{\{(1,1)\}\}^{\mathrm{U}}\}$. The set of *substrate histograms* of a given *size* is finite, $|\mathcal{A}_z| < \infty$. The cardinality of *substrate histograms* of a given non-zero *size* has a lower bound implied by the strong compositions of the *reductions*

$$|\mathcal{A}_z| \geq \left(\sum_{d \in \{1 \ldots z\}} |\mathrm{C}(\{1 \ldots d\}, z)|\right)^z = \left(\sum_{d \in \{1 \ldots z\}} \frac{(z-1)!}{(d-1)!(z-d)!}\right)^z$$

The finite set of *sized cardinal substrate histograms* may be constructed explicitly by constructing *cardinal systems* and *cardinal histories* in the *systems*,

$$\begin{aligned}\mathcal{A}_z = \{A :\ & x \in \{1 \ldots z\},\ V = \{1 \ldots x\},\\ & U \in \prod_{v \in V} \{v\} \times \{\{1 \ldots u\} : u \in \{1 \ldots z\}\},\\ & H \in \{1 \ldots z\} :\to V^{\mathrm{CS}},\\ & A = \mathrm{histogram}(H) + V^{\mathrm{CZ}},\ A^{\mathrm{XF}} = A^{\mathrm{C}}\}\end{aligned}$$

Each *substrate histogram* $A \in \mathcal{A}_z$ has $|V_A|! \prod_{w \in V_A} |U_A(w)|!$ *cardinal substrate permutations.* These *frame* mappings partition the *substrate histograms* into equivalence classes having the same *geometry.* Let P_z be the partition, $P_z \in \mathrm{B}(\mathcal{A}_z)$, such that the components of P_z are the equivalence classes by *cardinal substrate permutation*, $\forall C \in P_z\ \forall A \in C\ (|C| = |V_A|! \prod_{w \in V_A} |U_A(w)|)$.

Each of the *substrate histograms* in a component of P_z, that are equivalent by *cardinal substrate permutation*, have the same *entropy*, $\forall C \in P_z\ \forall A, B \in C\ (\text{entropy}(A) = \text{entropy}(B))$.

A subset $X_z \subset \mathcal{A}_z$ of the *substrate histograms* can be defined such that each element of X_z is uniquely chosen from a component of P_z, so that $|X_z| = |P_z|$ and $\exists M \in X_z \leftrightarrow P_z\ \forall (A, C) \in M\ (A \in C)$. The set X_z of *substrate histograms*, which are distinct by *geometry*, then forms a support of a uniform *probability function*, $\{(A, 1/|P_z|) : A \in X_z\} \in \mathcal{P}$, upon which the calculation of expectation and variance of derived *substrate structures* could be made.

Instead of choosing a *distinct geometry* subset, $X_z \subset \mathcal{A}_z$, as the support, consider weighting a support of $\mathcal{A}_z$. Let the *geometry-weighted function* $Q_z \in \mathcal{A}_z \to \mathbf{Q}_{>0}$ be defined

$$Q_z = \{(A, \frac{1}{|V_A|! \prod_{w \in V_A} |U_A(w)|!}) : A \in \mathcal{A}_z\}$$

which is such that $\forall C \in P\ \forall A \in C\ (Q_z(A) = 1/|C|)$. The *geometry-weighted probability function* $\hat{Q}_z \in \mathcal{P}$ is the normalised *geometry-weighted function*, $\hat{Q}_z = \text{normalise}(Q_z)$. The *geometry-weighted probability function*, $\hat{Q}_z$, can be calculated without explicitly calculating P_z itself. Using the weightings of the support of *substrate histograms*, $\mathcal{A}_z$, avoids the need to specify the selection of distinct *substrate histograms*, X_z, from the *permutation* equivalence classes, P_z.

If the *substrate histograms* are partitioned, for example to analyse correlations grouped by low or high *entropy*, then the partition should be a parent partition of P_z. That is, the *substrate histograms* partition should be independent of *cardinal substrate permutation*.

Define the central moment functions of the *geometry-weighted probability function*, $\hat{Q}_z$, that operate on real-valued functions of the *sized cardinal substrate histograms*, $\mathcal{A}_z \to \mathbf{R}$. In the cases where the real-valued functions are not left total, the *geometry-weighted probability function* $\hat{Q}_z \in \mathcal{P}$ is renormalised for the subset of the *substrate histograms*. Define the function $\text{ex}(z) \in (\mathcal{A}_z \to \mathbf{R}) \to \mathbf{R}$ as

$$\text{ex}(z)(F) := \text{expected}(\hat{R}_z)(F)$$

where

$$\hat{R}_z = \text{normalise}(\{(A, \frac{1}{|V_A|! \prod_{w \in V_A} |U_A(w)|!}) : A \in \text{dom}(F)\}) \in \mathcal{P}$$

and $F \neq \emptyset$. Define the function $\text{var}(z) \in (\mathcal{A}_z \to \mathbf{R}) \to \mathbf{R}$ as

$$\text{var}(z)(F) := \text{variance}(\hat{R}_z)(F)$$

Define the function $\text{cov}(z) \in (\mathcal{A}_z \to \mathbf{R}) \times (\mathcal{A}_z \to \mathbf{R}) \to \mathbf{R}$ as

$$\text{cov}(z)(F, G) := \text{covariance}(\hat{R}_z)(F, G)$$

where $\text{dom}(F) \cap \text{dom}(G) \neq \emptyset$ and

$$\hat{R}_z = \text{normalise}(\{(A, \frac{1}{|V_A|! \prod_{w \in V_A} |U_A(w)|!}) : A \in \text{dom}(F) \cap \text{dom}(G)\}) \in \mathcal{P}$$

Define the function $\text{corr}(z) \in (\mathcal{A}_z \to \mathbf{R}) \times (\mathcal{A}_z \to \mathbf{R}) \to \mathbf{R}$ as

$$\text{corr}(z)(F, G) := \text{correlation}(\hat{R}_z)(F, G)$$

The correlation is defined only if it is the case that both variances are non-zero, $\text{var}(z)(\text{filter}(\text{dom}(G), F)) > 0$ and $\text{var}(z)(\text{filter}(\text{dom}(F), G)) > 0$.

Also, let the arithmetic binary operators on reals, $(+), (-), (\times), (/) \in \mathbf{R}^2 \to \mathbf{R}$, be lifted to operators on real-valued functions. That is, the addition operator is lifted $F + G := \{(x, F_x + G_x) : x \in \text{dom}(F) \cap \text{dom}(G)\}$. The subtraction operator is lifted, $F - G := \{(x, F_x - G_x) : x \in \text{dom}(F) \cap \text{dom}(G)\}$. The multiplication operator is lifted, $F * G := \{(x, F_x \times G_x) : x \in \text{dom}(F) \cap \text{dom}(G)\}$. The divison operator is lifted $F/G := \{(x, F_x/G_x) : x \in \text{dom}(F) \cap \text{dom}(G),\ G_x \neq 0\}$.

3.17 Distribution over histograms

The set of *distributions* $\mathcal{Q}$ is a set of positive rational valued finite functions of *integral histograms* that have common *variables*. That is, $\mathcal{Q} \subset \mathcal{A}_\text{i} \to \mathbf{Q}_{\geq 0}$ such that $\forall Q \in \mathcal{Q}\ (|Q| < \infty)$ and $\forall Q \in \mathcal{Q}\ \forall A \in \text{dom}(Q)\ (\text{vars}(A) = \text{vars}(Q))$ where $\text{vars} \in \mathcal{Q} \to \text{P}(\mathcal{V})$ is defined as $\text{vars}(Q) := \bigcup\{\text{vars}(A) : A \in \text{dom}(Q)\}$.

The elements of the range of a *distribution* $Q \in \mathcal{Q}$, $\text{ran}(Q)$, are called the *frequencies*. The domain of Q, $\text{dom}(Q)$, is called the *support*. The elements of the *support* are called *sample histograms* if there is associated with Q a *distribution histogram* E having the same *variables*, $\text{vars}(E) = \text{vars}(Q)$, and some non-zero integral *size* $z_E \in \mathbf{N}_{>0}$. A *draw* is a pair of (i) the *distribution histogram*, E, and (ii) some non-zero integral *size* $z \in \mathbf{N}_{>0}$, $(E, z) \in \mathcal{A}_\text{i} \times \mathbf{N}_{>0}$.

Define the set of *complete distributions* $\mathcal{Q}_U \subset \mathcal{Q} \cap (\mathcal{A}_U \to \mathbf{Q}_{\geq 0})$ such that all the *sample histograms* of the domain are *complete* in *system* U

$$\forall Q \in \mathcal{Q}_U\ \forall A \in \mathrm{dom}(Q)\ (A^{\mathrm{U}} = A^{\mathrm{C}})$$

Define the set of *congruent distributions* of integral *size* $z \in \mathbf{N}$, $\mathcal{Q}_z \subset \mathcal{Q}$, such that all the *sample histograms* of the domain are *congruent*

$$\forall Q \in \mathcal{Q}_z\ \forall A \in \mathrm{dom}(Q)\ (\mathrm{size}(A) = z)$$

Define the set of *constructible distributions* which have *integral frequencies*, $\mathcal{Q}_{\mathrm{i}} \subset \mathcal{Q}$,

$$\mathcal{Q}_{\mathrm{i}} = \{Q : Q \in \mathcal{Q},\ \mathrm{ran}(Q) \subset \mathbf{N}\}$$

The *histograms* of *shuffles* of a *history* H that is *congruent* to the *support* of a *congruent distribution* $Q \in \mathcal{Q}_z$, $\mathrm{vars}(H) = \mathrm{vars}(Q)$ and $|H| = z$, may also be in the *support* because they are *congruent* and *integral*, $\{\mathrm{histogram}(G) : G \in \mathrm{shuffles}(H)\} \subset \mathcal{A}_{\mathrm{i}}$. The *independent histograms* of *histograms* in the *support* may be in the *support* where they are *integral*, $\{A^{\mathrm{X}} : A \in \mathrm{dom}(Q)\} \cap \mathcal{A}_{\mathrm{i}}$.

The *integral congruent support* $\mathcal{A}_{U,\mathrm{i},V,z}$ of *size* z and *variables* V in *system* U is the finite set of all *complete congruent integral histograms*

$$\mathcal{A}_{U,\mathrm{i},V,z} = \{A : A \in \mathcal{A}_{U,\mathrm{i}},\ A^{\mathrm{U}} = V^{\mathrm{C}},\ \mathrm{size}(A) = z\}$$

The *integral congruent support* can be constructed recursively, $\mathcal{A}_{U,\mathrm{i},V,z} = \{A + \{S\}^{\mathrm{U}} : A \in \mathcal{A}_{U,\mathrm{i},V,z-1},\ S \in V^{\mathrm{CS}}\}$ where $\mathcal{A}_{U,\mathrm{i},V,0} = \{V^{\mathrm{CZ}}\}$. The cardinality of the *integral congruent support* is the cardinality of weak compositions $|\mathrm{C}'(V^{\mathrm{C}}, z)|$

$$|\mathcal{A}_{U,\mathrm{i},V,z}| = \frac{(z+v-1)!}{z!\ (v-1)!}$$

where $v = |V^{\mathrm{C}}|$.

A *stuffed congruent distribution* Q has a domain of the *integral congruent support*, $Q \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{Q}_U \cap \mathcal{Q}_z$.

The *multiple support* $\mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}$ of maximum *count* z and *variables* V in *system* U is the finite set of all *complete integral histograms*

$$\mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}} = \{A : A \in \mathcal{A}_{U,\mathrm{i}},\ A^{\mathrm{U}} = V^{\mathrm{C}},\ \mathrm{ran}(A) \subseteq \{0\ldots z\}\}$$

The minimum *size* of the *multiple support histograms* is zero and the maximum *size* is vz where $v = |V^{\mathrm{C}}|$. That is, $\forall A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}\ (0 \leq \mathrm{size}(A) \leq vz)$.

The maximum *count* is $\forall A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}$ $(\mathrm{maxr}(A) \leq z)$. The *histograms* of a *multiple support* are not all *congruent*, but the *integral congruent supports* of *sizes* less than or equal to z are subsets, $\forall i \in \{0\ldots z\}$ $(\mathcal{A}_{U,\mathrm{i},V,i} \subset \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}})$. The *multiple supports* of maximum *counts* less than z are subsets, $\forall i \in \{0\ldots z-1\}$ $(\mathcal{A}_{U,\mathrm{i},V,\{0\ldots i\}} \subset \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}})$. The cardinality of the *multiple support* is $|\mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}| = (z+1)^v$.

A *distribution* $Q \in \mathcal{Q}$ is a *probability distribution* if it is in the set of *probability functions*, $Q \in \mathcal{P}$. That is, the sum of the range is 1, $\forall Q \in \mathcal{Q} \cap \mathcal{P}$ $(\mathrm{sum}(Q) = 1)$.

The positive rational *sum* of a *distribution* Q is $\mathrm{sum}(Q) \in \mathbf{Q}_{\geq 0}$.

Define the *modal set* of a *distribution* as $\mathrm{modes} \in \mathcal{Q} \to \mathrm{P}(\mathcal{A})$ as

$$\mathrm{modes}(Q) := \mathrm{maxd}(Q)$$

Define the *mean histogram* of a *distribution* as $\mathrm{mean} \in \mathcal{Q} \to \mathcal{A}$ as

$$\mathrm{mean}(Q) := \sum(\mathrm{scalar}(f) * A : (A,f) \in Q)/\mathrm{scalar}(\mathrm{sum}(Q))$$

The *mean* is undefined if the *frequencies* sum to zero, $\mathrm{sum}(Q) = 0$. The *mean* of a *complete distribution* is a *complete histogram*, $M^{\mathrm{U}} = M^{\mathrm{C}}$ where $M = \mathrm{mean}(Q)$ and $Q \in \mathcal{Q}_U$. The *mean* of a *congruent distribution* $Q \in \mathcal{Q}_z$ has *size* equal to z, $\mathrm{size}(M) = z$, and so is *congruent* to the *sample histograms* in the *support*, $\forall Q \in \mathcal{Q}_z\ \forall A \in \mathrm{dom}(Q)$ $(\mathrm{congruent}(A,M))$. The *mean histogram* is not necessarily *integral* and so is not necessarily in the *support*, $M \notin \mathcal{A}_{\mathrm{i}} \implies M \notin \mathrm{dom}(Q)$. If Q is a *complete probability distribution*, $Q \in \mathcal{Q}_U \cap \mathcal{P}$, then the *mean histogram* is the *histogram* of expected *counts*

$$\mathrm{mean}(Q) = \{(S, \mathrm{expected}(Q)(\{(A, A_S) : A \in \mathrm{dom}(Q)\})) : S \in V^{\mathrm{CS}}\}$$

where $V = \mathrm{vars}(Q)$.

Define the *variance* of the *counts* of a *state* in a *complete distribution* as $\mathrm{var}(U) \in \mathcal{Q}_U \to (\mathcal{S}_U \to \mathbf{Q}_{\geq 0})$ as

$$\mathrm{var}(U)(Q) := \{(S, \sum(f/\mathrm{sum}(Q) \times (A_S - M_S)^2 : (A,f) \in Q)) : S \in V^{\mathrm{CS}}\}$$

where $V = \mathrm{vars}(Q)$ and $M = \mathrm{mean}(Q)$. The *variance* is undefined if the *frequencies* sum to zero, $\mathrm{sum}(Q) = 0$. Note that in the case of uniform *distribution*, $|\mathrm{ran}(Q)| = 1$, the variance of each *state*, $\mathrm{var}(U)(Q)(S)$, is the population variance, not the sample variance. Although the range of the *variance* is a subset of the positive rationals, it is not treated here as a *histogram*, but grouped along with the higher central moments of the *distribution* which are not necessarily positive. If Q is a *probability distribution*, $Q \in \mathcal{P}$, then the *variance* is the variance of the *counts*

$$\mathrm{var}(U)(Q) = \{(S, \mathrm{variance}(Q)(\{(A, A_S) : A \in \mathrm{dom}(Q)\})) : S \in V^{\mathrm{CS}}\}$$

Similarly the *covariance* of the *counts* of a pair of *states* in a *complete probability distribution* is $\mathrm{cov}(U) \in (\mathcal{Q}_U \cap \mathcal{P}) \to ((\mathcal{S}_U \times \mathcal{S}_U) \to \mathbf{Q})$

$$\begin{aligned} \mathrm{cov}(U)(Q) := \{&((S, R), \\ &\quad \mathrm{covariance}(Q)(\{(A, A_S) : A \in \mathrm{dom}(Q)\},\ \{(A, A_R) : A \in \mathrm{dom}(Q)\})) : \\ &\quad S, R \in V^{\mathrm{CS}}\} \end{aligned}$$

The *moment generating function* of the *counts* of *states* having *moment parameters* $T \in \mathcal{S} \to \mathbf{R}$ in a *probability distribution* is $\mathrm{mgf} \in (\mathcal{Q} \cap \mathcal{P}) \to ((\mathcal{S} \to \mathbf{R}) \to \mathbf{R})$

$$\mathrm{mgf}(Q)(T) := \mathrm{expected}(Q)(\{(A, \exp(\sum_{S \in A^{\mathrm{S}}} T_S A_S)) : A \in \mathrm{dom}(Q)\})$$

where $\mathrm{dom}(T) = \mathrm{dom}(\mathrm{mean}(Q))$ and exp is the exponential function.

3.17.1 Historical distributions

Consider the subsets of non-empty *history* $H_E \in \mathcal{H} \setminus \{\emptyset\}$ of cardinality z, $\{G : G \subseteq H_E,\ |G| = z\} \subset \mathcal{H}$. The *historical distribution drawn without replacement* from H_E is the *distribution* of these *sample histories* over the *histograms*,

$$\mathrm{count}(\{(\mathrm{histogram}(G), G) : G \subseteq H_E,\ |G| = z\}) \in \mathcal{Q}_z$$

where $\mathrm{histogram}(G) := \{(S, |C|) : (S, C) \in G^{-1}\}$, and $\mathrm{count}(X) := \{(a, |\{c : (b, c) \in X,\ b = a\}|) : a \in \mathrm{dom}(X)\}$.

The *historical distribution* can equally well be defined in terms of a subset of the *histogram* function. Let $I \subset$ histogram be the *histogram* valued function of all possible subsets of the *history* H_E of cardinality z,

$$I \ = \ \{(G, \mathrm{histogram}(G)) : G \subseteq H_E,\ |G| = z\}$$

Then the *historical distribution* is $\{(A, |D|) : (A, D) \in I^{-1}\} \in \mathcal{Q}_z$.

The *event identifiers*, $\mathrm{dom}(H_E)$, serve merely to make the subsets unique so the *distribution* may be defined in terms of an arbitrarily constructed *history* of the *distribution histogram* $E = \mathrm{histogram}(H_E)$, $\mathrm{count}(\{(\mathrm{histogram}(G), G) : G \subseteq \mathrm{history}(E),\ |G| = z\}) = \mathrm{count}(\{(\mathrm{histogram}(G), G) : G \subseteq H_E,\ |G| = z\})$. The set of *historical distributions* $Q_{\mathrm{h}} \in \mathcal{A}_{\mathrm{i}} \times \mathbf{N} \to \mathcal{Q}_{\mathrm{i}}$ is the set of *constructible distributions* parameterised by the *without replacement draw* $(E, z) \in \mathcal{A}_{\mathrm{i}} \times \mathbf{N}$

$$Q_{\mathrm{h}}(E, z) = \mathrm{count}(\{(\mathrm{histogram}(G), G) : G \subseteq \mathrm{history}(E),\ |G| = z\}) \in \mathcal{Q}_{\mathrm{i}} \cap \mathcal{Q}_z$$

The *size* of the *distribution histogram* is $z_E = \mathrm{size}(E) = |H_E| > 0$. The *draw size* must be less than or equal to the *distribution size*, $z \leq z_E$. All of the *sample histograms* are less than or equal to the *distribution histogram*, $\forall A \in \mathrm{dom}(Q_{\mathrm{h}}(E, z))\ (A \leq E)$. The maximum *count* in the *sample histograms* is less than or equal to the *draw size* $\forall A \in \mathrm{dom}(Q_{\mathrm{h}}(E, z))\ (\mathrm{maxr}(A) \leq z)$.

The *without replacement* character of the *draw* can be shown by a recursive definition that *draws* one *event* from the implied *history* at each step. Define $\mathrm{drawnr} \in \mathcal{H} \times \mathbf{N} \times \mathcal{H} \to \mathrm{P}(\mathcal{H})$ as $\mathrm{drawnr}(H, z, G) := \bigcup\{\mathrm{drawnr}(H \setminus \{e\}, z-1, G \cup \{e\}) : e \in H\}$ where $\mathrm{drawnr}(H, 0, G) := \{G\}$. Then $Q_{\mathrm{h}}(E, z) = \mathrm{count}(\{(\mathrm{histogram}(X), X) : X \in \mathrm{drawnr}(\mathrm{history}(E), z, \emptyset)\})$.

The *sum* of a *historical distribution* $Q_{\mathrm{h}}(E, z)$ is the combination of z *drawn* from z_E

$$\mathrm{sum}(Q_{\mathrm{h}}(E, z)) = \binom{z_E}{z} = \frac{z_E!}{z!\ (z_E - z)!} \in \mathbf{N}_{>0}$$

Each *frequency* of a *sample histogram* $A \in \mathrm{dom}(Q_{\mathrm{h}}(E, z))$ in a *historical distribution* is the product of the combinations in which the subset A_S is *drawn* from E_S for all of the *states*

$$Q_{\mathrm{h}}(E, z)(A) = \prod_{S \in A^{\mathrm{S}}} \binom{E_S}{A_S} = \prod_{S \in A^{\mathrm{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!} \in \mathbf{N}_{>0}$$

If the *historical distribution histogram* E is *incomplete*, that is, it does not contain all of its *cartesian states*, $E^{\mathrm{U}} \neq E^{\mathrm{C}}$, for some *system* U, then all of the element *histograms* in the *support* must necessarily be *incomplete* in that *system*, $E^{\mathrm{U}} \neq E^{\mathrm{C}} \implies \forall z \in \{1 \ldots z_E\}\ \forall A \in \mathrm{dom}(Q_{\mathrm{h}}(E, z))\ (A^{\mathrm{U}} \neq A^{\mathrm{C}})$. Even if the *historical distribution histogram* is *complete*, $E^{\mathrm{U}} = E^{\mathrm{C}}$, many of its element *histograms* would be *incomplete* because each must have non-zero *counts*, $\mathrm{minr}(A) > 0$. In fact, only when the *size* z is greater than or equal to the *distribution histogram's volume*, $z \geq |E^{\mathrm{C}}|$, can any of the element *histograms* be *complete*, $z < |E^{\mathrm{C}}| \implies \forall A \in \mathrm{dom}(Q_{\mathrm{h}}(E,z))\ (A^{\mathrm{U}} \neq A^{\mathrm{C}})$.

Furthermore, the *support* of a *historical distribution* cannot equal the *integral congruent support*, $\mathrm{dom}(Q_{\mathrm{h}}(E,z)) \neq \mathcal{A}_{U,\mathrm{i},V,z}$ where $V = \mathrm{vars}(E)$, because the *integral congruent support* contains *histograms* with zero *counts*, $\exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ (0 \in \mathrm{ran}(A))$. The *stuffed historical distribution* $Q_{\mathrm{h},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{Q}_U \cap \mathcal{Q}_z$ can be constructed from a *historical distribution*, $Q_{\mathrm{h}}(E,z)$, by *completing* the *support histograms* and *stuffing* with the disjoint subset of the *integral congruent support* with zero *frequencies*

$$\begin{aligned} Q_{\mathrm{h},U}(E,z) &= \{(A + A^{\mathrm{CZ}}, f) : (A,f) \in Q_{\mathrm{h}}(E,z)\} \cup \\ &\quad (\mathcal{A}_{U,\mathrm{i},V,z} \setminus \{A + A^{\mathrm{CZ}} : A \in \mathrm{dom}(Q_{\mathrm{h}}(E,z))\}) \times \{0\} \end{aligned}$$

where $V = \mathrm{vars}(E)$. The *stuffed historical distribution support* equals the *integral congruent support*, $\mathrm{dom}(Q_{\mathrm{h},U}(E,z)) = \mathcal{A}_{U,\mathrm{i},V,z}$, so each of the *histograms* in the *support* is unique by *histogram equivalence*

$$\forall A, B \in \mathrm{dom}(Q_{\mathrm{h},U}(E,z))\ (A \equiv B \implies A = B)$$

The *sum* of a *stuffed historical distribution* equals the *sum* of its corresponding *historical distribution*

$$\mathrm{sum}(Q_{\mathrm{h},U}(E,z)) = \mathrm{sum}(Q_{\mathrm{h}}(E,z)) = \frac{z_E!}{z!\ (z_E - z)!}$$

The *stuffed historical distribution* can be defined explicitly

$$Q_{\mathrm{h},U}(E,z) = \{(A, \mathrm{if}(A \leq E, \prod_{S \in A^{\mathrm{FS}}} \frac{E_S!}{A_S!\ (E_S - A_S)!}, 0)) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The *stuffed historical probability distribution* $\hat{Q}_{\mathrm{h},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{Q}_U \cap \mathcal{Q}_z \cap \mathcal{P}$ is defined

$$\begin{aligned} \hat{Q}_{\mathrm{h},U}(E,z) &= \mathrm{normalise}(Q_{\mathrm{h},U}(E,z)) \\ &= \{(A, f/\mathrm{sum}(Q_{\mathrm{h},U}(E,z))) : (A,f) \in Q_{\mathrm{h},U}(E,z)\} \end{aligned}$$

where $z_E > 0$ and $z > 0$. The *stuffed historical probability distribution* is a multivariate hypergeometric distribution. The *sum* of the *stuffed historical probability distribution* is one

$$\mathrm{sum}(\hat{Q}_{\mathrm{h},U}(E,z)) = 1$$

The *mean* of the *stuffed historical probability distribution* is

$$\text{mean}(\hat{Q}_{\text{h},U}(E,z)) = \text{scalar}(z) * P$$

where $P = E/\text{scalar}(z_E) + E^{\text{CZ}}$. The *mean* is *congruent* to the *histograms* of the *integral congruent support*, $\forall A \in \mathcal{A}_{U,\text{i},V,z}$ (congruent(A,M)), where $M = \text{mean}(\hat{Q}_{\text{h},U}(E,z))$. If the *mean* is *integral* then it is in the *support*, $M \in \mathcal{A}_\text{i} \implies M \in \mathcal{A}_{U,\text{i},V,z}$.

The *variance* of *state* S in the *stuffed historical probability distribution* is

$$\text{var}(U)(\hat{Q}_{\text{h},U}(E,z))(S) = z\frac{z_E - z}{z_E - 1}P_S(1 - P_S)$$

The *covariance* of a pair of *states* (S,R), where $R \neq S$, in the *stuffed historical probability distribution* is

$$\text{cov}(U)(\hat{Q}_{\text{h},U}(E,z))((S,R)) = -z\frac{z_E - z}{z_E - 1}P_S P_R$$

3.17.2 Multinomial distributions

Let the power of a set, X^n, be defined as the set of all n-tuples of the elements of the set, $X^n = \prod(\{1 \ldots n\} \times \{X\}) = \{L : L \in \mathcal{L}(X),\ |L| = n\}$. The cardinality of the power is $|X^n| = |X|^n$.

Consider the set of lists of the *events drawn with replacement* from non-empty *history* $H_E \in \mathcal{H} \setminus \{\emptyset\}$ of cardinality z, $H_E^z = \{L : L \in \mathcal{L}(H_E),\ |L| = z\}$. Construct from this set, H_E^z, a set of *sample histories* of cardinality z with new *event identifiers* modified to include the position, $X = \{G : L \in H_E^z,\ G = \{((i,x),S) : (i,(x,S)) \in L\}\} \subset \mathcal{H}$. The set of *sample histories*, X, is bijective with the set of *event* lists, $|X| = |H_E^z|$. The *multinomial distribution drawn with replacement* from H_E is the *distribution* of these *sample histories* over the *histograms*,

$$\text{count}(\{(\text{histogram}(G), G) : G \in X\}) \in \mathcal{Q}_z$$

The *distribution* is the same for all *histories* for which $E = \text{histogram}(H_E)$ is the *distribution histogram*, $\text{count}(\{(\text{histogram}(G), G) : L \in \text{history}(E)^z,\ G = \{((i,x),S) : (i,(x,S)) \in L\}\}) = \text{count}(\{(\text{histogram}(G), G) : G \in X\})$. The set of *multinomial distributions* $Q_\text{m} \in \mathcal{A}_\text{i} \times \mathbf{N} \to \mathcal{Q}_\text{i}$ is the set of *constructible distributions* parameterised by the *with replacement draw* $(E,z) \in \mathcal{A}_\text{i} \times \mathbf{N}$

$$\begin{aligned} Q_\text{m}(E,z) = \text{count}(\{&(\text{histogram}(G), G) : \\ & L \in \text{history}(E)^z,\ G = \{((i,x),S) : (i,(x,S)) \in L\}\}) \in \mathcal{Q}_\text{i} \cap \mathcal{Q}_z \end{aligned}$$

The *distribution* can also be written, $Q_{\mathrm{m}}(E, z) = \mathrm{count}(\{(\mathrm{count}(\{(S, i) : (i, (\cdot, S)) \in L\}), L) : L \in \mathrm{history}(E)^z\})$. The *size* of the *distribution histogram* is $z_E = \mathrm{size}(E) = |H_E|$. The *draw size*, z, is not constrained by the *distribution size*, z_E, $z \in \mathbf{N}$. The maximum *count* in the *sample histograms* is less than or equal to the *draw size* $\forall A \in \mathrm{dom}(Q_{\mathrm{m}}(E, z))\ (\mathrm{maxr}(A) \leq z)$.

The *multinomial distribution* can be constructed in steps of *with replacement draws* of one *event*. This contrasts with the *non replacement draw* of the *historical distribution*. Define $\mathrm{drawwr} \in \mathcal{H} \times \mathbf{N} \times \mathcal{H} \to \mathrm{P}(\mathcal{H})$ as $\mathrm{drawwr}(H, z, G) := \bigcup\{\mathrm{drawwr}(H, z{-}1, G \cup \{((z, x), S)\}) : (x, S) \in H\}$ where $\mathrm{drawwr}(H, 0, G) := \{G\}$. Then the *multinomial distribution* is such that $Q_{\mathrm{m}}(E, z) = \mathrm{count}(\{(\mathrm{histogram}(X), X) : X \in \mathrm{drawwr}(\mathrm{history}(E), z, \emptyset)\})$. Here the *event identifier* is prefixed with a sequence position, so that

$$|\mathrm{drawwr}(\mathrm{history}(E), z, \emptyset)| = |H_E^z|$$

The *sum* of a *multinomial distribution* $Q_{\mathrm{m}}(E, z)$ is equal to $|H_E^z| = |H_E|^z$

$$\mathrm{sum}(Q_{\mathrm{m}}(E, z)) = z_E^z \in \mathbf{N}$$

Each *frequency* of a *sample histogram* $A \in \mathrm{dom}(Q_{\mathrm{m}}(E, z))$ in a *multinomial distribution* is the product of (i) the *multinomial coefficient* which is the combination in which the subsets of cardinality A_S are chosen from a set of cardinality z for all *states*, and (ii) the cardinality of the lists *drawn with replacement* from H_E equivalent to a permutation defined by some order on the *states* $D \in \mathrm{enums}(A^{\mathrm{S}})$, $\mathrm{concat}(\{(i, \{1 \ldots A_S\} \times \{S\}) : (S, i) \in D\}) \in \{G : L \in H_E^z,\ G = \{(i, S) : (i, (\cdot, S)) \in L\}\}$

$$Q_{\mathrm{m}}(E, z)(A) = \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} E_S^{A_S} \in \mathbf{N}_{>0}$$

The *generalised multinomial distribution* $Q_{\mathrm{m},U}(E, z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \subset \mathcal{Q}_U \cap \mathcal{Q}_z$ is a *stuffed congruent distribution* that can be constructed from a *multinomial distribution*, $Q_{\mathrm{m}}(E, z)$, by *completing* the *support histograms* and *stuffing* with the disjoint subset of the *integral congruent support* with zero *frequencies*

$$\begin{aligned} Q_{\mathrm{m},U}(E, z) \ &= \ \{(A + A^{\mathrm{CZ}}, f) : (A, f) \in Q_{\mathrm{m}}(E, z)\} \cup \\ &\quad\ (\mathcal{A}_{U,\mathrm{i},V,z} \setminus \{A + A^{\mathrm{CZ}} : A \in \mathrm{dom}(Q_{\mathrm{m}}(E, z))\}) \times \{0\} \end{aligned}$$

where $V = \mathrm{vars}(E)$. The *generalised multinomial distribution support* equals the *integral congruent support*, $\mathrm{dom}(Q_{\mathrm{m},U}(E, z)) = \mathcal{A}_{U,\mathrm{i},V,z}$, so each of the *histograms* in the *support* is unique by *histogram equivalence*

$$\forall A, B \in \mathrm{dom}(Q_{\mathrm{m},U}(E, z))\ (A \equiv B \implies A = B)$$

The *sum* of a *generalised multinomial distribution* equals the *sum* of its corresponding *multinomial distribution*

$$\mathrm{sum}(Q_{\mathrm{m},U}(E,z)) = \mathrm{sum}(Q_{\mathrm{m}}(E,z)) = z_E^z$$

The definition of the set of *generalised multinomial distributions* is generalised to allow parameterisation by *non-integral distribution histograms*, $Q_{\mathrm{m},U} \in \mathcal{A} \times \mathbf{N} \to \mathcal{Q}$. Contrast this to parameterisation by *draw*, $\mathcal{A}_{\mathrm{i}} \times \mathbf{N}$, which is defined only for *integral distribution histograms*. Define the *generalised multinomial distribution*, $Q_{\mathrm{m},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \subset \mathcal{Q}_U \cap \mathcal{Q}_z$, explicitly as

$$Q_{\mathrm{m},U}(E,z) := \{(A, \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} E_S^{A_S}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

where $V = \mathrm{vars}(E)$ and E is *complete* $E^{\mathrm{U}} = E^{\mathrm{C}}$. Define $0^0 = 0! = 1! = 1$ so that the *multinomial coefficient* is defined for zero A_S. Define $0^x = 0$ where $x \neq 0$. The *multinomial coefficient* is integral and is greater than or equal to one

$$\forall A \in \mathcal{A}_{U,\mathrm{i},V,z} \ \left(\frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \in \mathbf{N}_{>0}\right)$$

The *generalised multinomial probability distribution* $\hat{Q}_{\mathrm{m},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{Q}_U \cap \mathcal{Q}_z \cap \mathcal{P}$ is defined

$$\begin{aligned}
\hat{Q}_{\mathrm{m},U}(E,z) &= \mathrm{normalise}(Q_{\mathrm{m},U}(E,z)) \\
&= \{(A, f/\mathrm{sum}(Q_{\mathrm{m},U}(E,z))) : (A,f) \in Q_{\mathrm{m},U}(E,z)\} \\
&= \{(A, \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}
\end{aligned}$$

where $\mathrm{sum}(Q_{\mathrm{m},U}(E,z)) > 0$ which implies that $z_E > 0$ and $z > 0$. The *sum* of the *generalised multinomial probability distribution* is one

$$\mathrm{sum}(\hat{Q}_{\mathrm{m},U}(E,z)) = \sum_{A \in \mathcal{A}_{U,\mathrm{i},V,z}} \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S} = 1$$

The *generalised multinomial probability distribution* may be expressed in terms of a *probability distribution histogram* $\hat{E} = E/\mathrm{scalar}(z_E) \in \mathcal{A} \cap \mathcal{P}$,

$$\hat{Q}_{\mathrm{m},U}(E,z) = \{(A, z! \prod_{S \in A^{\mathrm{S}}} \frac{\hat{E}_S^{A_S}}{A_S!}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The *frequencies* of the *generalised multinomial probability distribution* can be approximated by means of the Stirling approximation

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = z! \prod_{S \in A^{\mathrm{S}}} \frac{\hat{E}_S^{A_S}}{A_S!} \approx \prod_{S \in A^{\mathrm{FS}}} \left(\frac{\hat{E}_S}{\hat{A}_S}\right)^{A_S}$$

where $\hat{A} = \mathrm{resize}(1, A)$. Compare this approximation to the same term for a *scaled draw size* kz and *scaled sample histogram* $\mathrm{scalar}(k) * A$ where $k \in \mathbf{N}_{>0}$

$$\begin{aligned}
\hat{Q}_{\mathrm{m},U}(E,kz)(\mathrm{scalar}(k) * A) &= (kz)! \prod_{S \in A^{\mathrm{S}}} \frac{\hat{E}_S^{kA_S}}{(kA_S)!} \\
&\approx \left(\prod_{S \in A^{\mathrm{FS}}} \left(\frac{\hat{E}_S}{\hat{A}_S}\right)^{A_S}\right)^k \\
&= (\hat{Q}_{\mathrm{m},U}(E,z)(A))^k
\end{aligned}$$

The probability of *drawing* A in each of k *draws* (E,z) of sets of lists of the *events drawn with replacement* from *history* $H_E \in \mathcal{H}$ of cardinality z, having total cardinality $|H_E^z|^k$, is equal to the probability of *drawing* A once from (E,z) raised to the power k, $(\hat{Q}_{\mathrm{m},U}(E,z)(A))^k$. This is approximately equal to the probability of *drawing* $\mathrm{scalar}(k) * A$ from (E,kz) in the set of lists of the *events drawn with replacement* from *history* H_E of cardinality kz, having the same total cardinality $|H_E^{kz}| = |H_E^z|^k$. The approximation is best when the *multinomial coefficient* is minimised. This is the case when the *entropy* of A, $\mathrm{entropy}(A)$, is low, for example when A is *diagonal*.

Noting that the cardinality of the *integral congruent support* is less than or equal to the cardinality of the *scaled integral congruent support*, $|\mathcal{A}_{U,\mathrm{i},V,z}| \leq |\mathcal{A}_{U,\mathrm{i},V,kz}|$, the sum of the *generalised multinomial probability distribution* can be approximated in terms of the *scaled generalised multinomial probability distribution*

$$\mathrm{sum}(\hat{Q}_{\mathrm{m},U}(E,z)) \approx \sum_{A \in \mathcal{A}_{U,\mathrm{i},V,z}} (\hat{Q}_{\mathrm{m},U}(E,kz)(\mathrm{scalar}(k) * A))^{1/k}$$

The *mean* of the *generalised multinomial probability distribution* is

$$\mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E,z)) = \mathrm{scalar}(z) * \hat{E}$$

The *integral mean multinomial probability distribution conjecture* states that if the *mean* of the *multinomial probability distribution* is *integral* then it is also *modal*

$$\mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E,z)) \in \mathcal{A}_i \implies \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E,z)) \in \mathrm{modes}(\hat{Q}_{\mathrm{m},U}(E,z))$$

See the discussion in 'Minimum Alignment', below, which generalises the *multinomial probability distribution* to be a probability density function (by using the gamma function), and then shows that *non-integral means* can have probability density less than the *modes*, in the case of negative *alignment*.

Consider the subset of the *integral congruent support* which consists of the *histograms* bracketing the *mean* $M = \text{mean}(\hat{Q}_{\text{m},U}(E,z))$ by floor and ceiling *counts*, $\{A : A \in \mathcal{A}_{U,\text{i},V,z},\ \forall S \in A^{\text{S}}\ (A_S \in \{\lfloor M_S \rfloor, \lceil M_S \rceil\})\}$. It is the case that there exist *multinomial probability distributions* such that this bracketing subset is not a superset of the *modes*

$$\exists (E,z) \in \mathcal{A}_{\text{i}} \times \mathbf{N}\ \Diamond M = \text{mean}(\hat{Q}_{\text{m},U}(E,z))$$
$$\exists A \in \text{modes}(\hat{Q}_{\text{m},U}(E,z))\ \exists S \in A^{\text{S}}\ (A_S \notin \{\lfloor M_S \rfloor, \lceil M_S \rceil\})$$

The *variance* of *state* S in the *generalised multinomial probability distribution* is

$$\text{var}(U)(\hat{Q}_{\text{m},U}(E,z))(S) = z\hat{E}_S(1 - \hat{E}_S)$$

The *covariance* of a pair of *states* (S,R), where $R \neq S$, in the *generalised multinomial probability distribution* is

$$\text{cov}(U)(\hat{Q}_{\text{m},U}(E,z))((S,R)) = -z\hat{E}_S\hat{E}_R$$

The *moment generating function* of the *generalised multinomial probability distribution* is

$$\text{mgf}(U)(\hat{Q}_{\text{m},U}(E,z))(T) = \left(\sum_{S \in V^{\text{CS}}} \hat{E}_S e^{T_S}\right)^z$$

where $T \in V^{\text{CS}} \to \mathbf{R}$.

Compare the *historical* and *multinomial distributions*. The *multinomial coefficient* can be separated from the *permutorial* part in both *distributions* showing that the *historical distribution frequency*, $Q_{\text{h}}(E,z)(A)$, is less than or equal to the *multinomial distribution frequency*, $Q_{\text{m}}(E,z)(A)$, of the *histogram*, A,

$$\begin{aligned}
Q_{\text{h}}(E,z)(A) &= \prod_{S \in A^{\text{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!} \\
&= \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!} \frac{1}{z!} \prod_{S \in A^{\text{S}}} E_S^{\underline{A_S}} \\
&\leq \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!} \prod_{S \in A^{\text{S}}} E_S^{A_S} \\
&= Q_{\text{m}}(E,z)(A)
\end{aligned}$$

where $A \leq E$ and $x^{\underline{n}}$ is the falling factorial.

In the *multinomial distribution*, *drawn with replacement*, the *permutorial* is the product of the *counts* of the *states* of the *distribution histogram* raised to the power of the *count* of the corresponding *state* of the *sample histogram*, $E_S^{A_S}$. The *permutorial* of the *historical distribution*, *drawn without replacement*, is the same except that the power is falling factorial, $E_S^{\underline{A_S}}$.

The *multinomial distribution frequency* is also larger than the *historical distribution frequency* because of the factor of $z!$. This arises because in order to have integral *frequencies*, $Q_{\mathrm{m}}(E,z)(A) \in \mathbf{N}$, the *multinomial distribution* must modify the *event identifiers* with the position in the lists in $H_E^z \in \mathcal{L}(\mathcal{L}(H_E))$. If the *historical distribution* was defined to be constructed from a list of modified *histories* $J = \{G : L \in \mathcal{L}(H_E),\ |\mathrm{set}(L)| = |L| = z,\ G = \{((i,x),S) : (i,(x,S)) \in L\}\}$ rather than from subsets $K = \{G : G \subseteq H_E,\ |G| = z\}$ then the same factor, $z!$, would also appear in the *historical distribution*, $|J| = z_E^{\underline{z}}$ and $|K| = z_E^{\underline{z}}/z!$, hence $|J| = z!|K|$.

Compare the *stuffed historical* and *multinomial probability distributions*

$$\begin{aligned}
\hat{Q}_{\mathrm{m},U}(E,z)(A) &= \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S} \\
&= \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \frac{1}{z_E^z} \prod_{S \in A^{\mathrm{S}}} E_S^{A_S} \\
\hat{Q}_{\mathrm{h},U}(E,z)(A) &= \frac{z!\ (z_E - z)!}{z_E!} \prod_{S \in A^{\mathrm{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!} \\
&= \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \frac{1}{z_E^{\underline{z}}} \prod_{S \in A^{\mathrm{S}}} E_S^{\underline{A_S}}
\end{aligned}$$

Here the scaling factor of $z!$ disappears.

As the ratio z_E/z increases the *generalised multinomial probability distribution* increasingly approximates to the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{m},U}(E,z) \approx \hat{Q}_{\mathrm{h},U}(E,z)$. The ratio of *multinomial frequency* to the *historical frequency* for *distribution histogram* $E \in \mathcal{A}_{U,\mathrm{i},V,z_E}$ and *sample his-*

togram $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ such that $A \leq E$ and $A \in \mathrm{dom}(Q_{\mathrm{h},U}(E,z))$ is

$$\begin{aligned}
\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\hat{Q}_{\mathrm{h},U}(E,z)(A)} &= \frac{z_E!}{(z_E - z)!z_E^z} \prod_{S \in A^{\mathrm{S}}} \frac{(E_S - A_S)!}{E_S!} E_S^{A_S} \\
&= \prod_{i \in \{1 \ldots z\}} \left(\frac{z_E - z + i}{z_E}\right) \prod_{S \in A^{\mathrm{S}},\ j \in \{1 \ldots A_S\}} \left(\frac{E_S}{E_S - A_S + j}\right)
\end{aligned}$$

Calculating the special case of the *scaled cartesian distribution histogram* $E = \mathrm{scalar}(z_E/v) * V^{\mathrm{C}}$ and *sample histogram* $A = \mathrm{scalar}(z/v) * V^{\mathrm{C}}$, where $v = |V^{\mathrm{C}}|$, $z/v \in \mathbf{N}_{>0}$ and $z_E/z \in \mathbf{N}_{>0}$, then

$$\begin{aligned}
\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\hat{Q}_{\mathrm{h},U}(E,z)(A)} &= \prod_{i \in \{1 \ldots z\}} \left(\frac{z_E - z + i}{z_E}\right) \prod_{j \in \{1 \ldots z/v\}} \left(\frac{z_E}{z_E - z + vj}\right)^v \\
&\approx \prod_{i \in \{1 \ldots z\}} \left(\frac{z_E - z + i}{z_E}\right) \prod_{i \in \{1 \ldots z\}} \left(\frac{z_E}{z_E - z + i}\right) \\
&= 1
\end{aligned}$$

Taking the root is less of an approximation where z is small, $z \ll z_E/v$.

A constraint exists on the *induction* of the *multinomial distribution histogram* $E \in \mathcal{A}_{\mathrm{i},U}$, of *size* z_E, given a *sample histogram* $A \in \mathrm{dom}(\hat{Q}_{\mathrm{m},U}(E,z))$ and *draw size* $z = \mathrm{size}(A)$. If A is assumed to be *modal*, $A \in \mathrm{modes}(\hat{Q}_{\mathrm{m},U}(E,z))$, and A is not *completely uniform*, $A \neq \mathrm{resize}(z, A^{\mathrm{C}})$, then E cannot be *completely uniform*, $E \neq \mathrm{resize}(z_E, E^{\mathrm{C}})$. That is

$$A \in \mathrm{modes}(\hat{Q}_{\mathrm{m},U}(E,z)) \wedge (A \neq Z_{z/v} * V^{\mathrm{C}}) \implies (E \neq Z_{z_E/v} * V^{\mathrm{C}})$$

where $V = \mathrm{vars}(E)$, $v = |V^{\mathrm{CS}}|$ and $Z_x = \mathrm{scalar}(x)$. The *mean* M of the *completely uniform distribution histogram* $\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}}, z)$ is also *completely uniform*, $M = \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}}, z)) = Z_{z/v} * V^{\mathrm{C}}$. The *probability* of the *mean histogram* is the *frequency*

$$\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}}, z)(M) = \frac{z!}{(\frac{z}{v}!)^v} \left(\frac{1}{v}\right)^z$$

Let perturb $\in \mathcal{A} \to \mathrm{P}(\mathcal{A})$ be the set of *effective event perturbations* of a *histogram*, excluding the given *histogram*,

$$\begin{aligned}
&\mathrm{perturb}(A) := \\
&\quad \{A + \{S\}^{\mathrm{U}} - \{R\}^{\mathrm{U}} : (S,c) \in A,\ c > 0,\ (R,d) \in A,\ S \neq R,\ d \geq 1\}
\end{aligned}$$

All *perturbations* of the *mean* are less *probable* than the *mean*

$$\forall A \in \text{perturb}(M)\ (\hat{Q}_{\text{m},U}(V^{\text{C}}, z)(A) < \hat{Q}_{\text{m},U}(V^{\text{C}}, z)(M))$$

because

$$\frac{1}{\left(\frac{z}{v} - 1\right)!\left(\frac{z}{v} + 1\right)!} < \frac{1}{\left(\frac{z}{v}!\right)^2}$$

All *sample histograms* can be placed in a list $L \in \mathcal{L}(\text{dom}(\hat{Q}_{\text{m},U}(E, z)))$ beginning with the *mean* and followed by *perturbations* of the previous item. That is, $L_1 = M$ and $\forall i \in \{1 \ldots |L| - 1\}\ (L_{i+1} \in \text{perturb}(L_i))$. Each step on paths constructed for a *completely uniform distribution* is subject to the inequality above, $\forall i \in \{1 \ldots |L|-1\}\ (\hat{Q}_{\text{m},U}(V^{\text{C}}, z)(L_{i+1}) < \hat{Q}_{\text{m},U}(V^{\text{C}}, z)(L_i))$. Thus all *sample histograms* except for the *mean* are less *probable* than the *mean* in a *completely uniform distribution*. Furthermore, the *mean* is *modal* according to the conjecture above for *integral means* in the *multinomial distribution*, $M \in \text{modes}(\hat{Q}_{\text{m},U}(V^{\text{C}}, z))$. Therefore the set of *modes* of a *completely uniform distribution histogram* is a singleton of the *mean histogram*, $\text{modes}(\hat{Q}_{\text{m},U}(V^{\text{C}}, z)) = \{M\}$. Hence the constraint on *induction*.

The *entropy* (defined in appendix 'Entropy and Gibbs' inequality') of the *distribution histogram* E scaled by the *draw size* z approximates to the *sum variance* of the *generalised multinomial probability distribution*

$$\begin{aligned}
\text{sum}(\text{var}(U)(\hat{Q}_{\text{m},U}(E, z))) &= \sum_{S \in V^{\text{CS}}} z\hat{E}_S(1 - \hat{E}_S) \\
&\sim -z \sum_{S \in E^{\text{FS}}} \hat{E}_S \ln \hat{E}_S \\
&= z \times \text{entropy}(\hat{E}) \\
&= z \times \text{entropy}(E)
\end{aligned}$$

where $\hat{E} = \text{resize}(1, E)$ and $V = \text{vars}(E)$. The approximate proportionality depends on the first term of the Taylor series, $\ln x \approx (x - 1)$, and so is a very coarse approximation. Increasing *entropy* tends to increase the *sum variance*. Conversely, increasing *entropy* tends to decrease the absolute *sum covariance*, $\text{sum}(\text{cov}(U)(\hat{Q}_{\text{m},U}(E, z))) - \text{sum}(\text{var}(U)(\hat{Q}_{\text{m},U}(E, z))) = \sum(-z\hat{E}_S\hat{E}_R : S, R \in V^{\text{CS}},\ R \neq S)$, becoming less negative.

Conjecture that the cardinality of the *modal set* of the *multinomial distribution* tends to increase with increasing *entropy* for constant *draw size*

$$\begin{aligned}
|\text{modes}(\hat{Q}_{\text{m},U}(E, z))| &\sim \text{sum}(\text{var}(U)(\hat{Q}_{\text{m},U}(E, z))) \\
&\sim z \times \text{entropy}(E)
\end{aligned}$$

Note that the *entropy* is of the *distribution histogram*, $\text{entropy}(E)$, not the *distribution* itself, $\text{entropy}(\hat{Q}_{\text{m},U}(E,z))$.

Consider the special case of a *draw* of *size* z from two *uniform distribution histograms* E and D having *volumes* v and kv respectively where $k \in \mathbf{N}_{>1}$. The ratio of the *sum variance* of the *generalised multinomial probability distribution* of these *draws* is

$$\begin{aligned}
\frac{\text{sum}(\text{var}(U)(\hat{Q}_{\text{m},U}(D,z)))}{\text{sum}(\text{var}(U)(\hat{Q}_{\text{m},U}(E,z)))} &= \frac{\sum z\hat{D}_R(1-\hat{D}_R) : R \in D^{\text{S}}}{\sum z\hat{E}_S(1-\hat{E}_S) : S \in E^{\text{S}}} \\
&= \frac{kvz\frac{1}{kv}\left(1-\frac{1}{kv}\right)}{vz\frac{1}{v}\left(1-\frac{1}{v}\right)} \\
&= \frac{v-\frac{1}{k}}{v-1} \\
&> 1
\end{aligned}$$

where $\hat{D} = \text{resize}(1,D)$ and $\hat{E} = \text{resize}(1,E)$. Thus the *sum variance* increases with increasing *volume*, but the effect is not very large. Similarly, the ratios of the *entropies* is

$$\begin{aligned}
\frac{\text{entropy}(D)}{\text{entropy}(E)} &= \frac{-\sum \hat{D}_R \ln \hat{D}_R : R \in D^{\text{FS}}}{-\sum \hat{E}_S \ln \hat{E}_S : S \in E^{\text{FS}}} \\
&= \frac{kv\frac{1}{kv}\ln\frac{1}{kv}}{v\frac{1}{v}\ln\frac{1}{v}} \\
&= 1 + \frac{\ln k}{\ln v} \\
&> 1
\end{aligned}$$

Consider a *one functional transform* $T \in \mathcal{T}_{U,\text{f},1}$ having *underlying variables* equal to the *variables* V of *uniform distribution histogram* D, $\text{und}(T) = V = \text{vars}(D)$. The *entropy* of the *application* of T to D must be less than or equal to the *entropy* of D, $\text{entropy}(V^{\text{C}} * T) \leq \text{entropy}(V^{\text{C}})$, because the *derived volume* is less than or equal to the *underlying volume*, $|W^{\text{C}}| \leq |V^{\text{C}}|$ where $W = \text{der}(T)$. The *entropies* are equal when W is a *frame* of V and thus T is *full functional*.

The log of the *generalised multinomial probability distribution*, $\ln \circ \hat{Q}_{\text{m},U}(E,z) \in \mathcal{A}_{U,\text{i},V,z} \to \mathbf{R}$, can be approximated by the *sized negative relative entropy* between the *sample histogram* and the *distribution histogram* by means of the

Stirling approximation

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = z! \prod_{S \in A^{\mathrm{S}}} \frac{\hat{E}_S^{A_S}}{A_S!} \approx \prod_{S \in A^{\mathrm{FS}}} \left(\frac{\hat{E}_S}{\hat{A}_S}\right)^{A_S}$$

where $\hat{E} = \mathrm{resize}(1,E)$ and $\hat{A} = \mathrm{resize}(1,A)$. So

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{m},U}(E,z)(A) &\approx \sum_{S \in A^{\mathrm{FS}}} A_S \ln \frac{\hat{E}_S}{\hat{A}_S} \\
&= -z \sum_{S \in A^{\mathrm{FS}}} \hat{A}_S \ln \frac{\hat{A}_S}{\hat{E}_S} \\
&= -z \times \mathrm{entropyRelative}(\hat{A},\hat{E})
\end{aligned}$$

where $A^{\mathrm{F}} \leq E^{\mathrm{F}}$. By Gibbs' inequality, the logarithm is maximised when the *sample histogram* is the *mean*, $A = \mathrm{scalar}(z/z_E) * E$. Thus the *mean* is *modal* within the approximation. The *sized negative relative entropy* can be thought of as the similarity of the *sample histogram* to the *distribution histogram*. This can be seen by comparing the *sized negative relative entropy*,

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{m},U}(E,z)(A) &\approx -z \sum_{S \in A^{\mathrm{FS}}} \hat{A}_S \ln \frac{\hat{A}_S}{\hat{E}_S} \\
&\approx z \sum_{S \in A^{\mathrm{FS}}} \hat{A}_S (1 - \frac{\hat{A}_S}{\hat{E}_S})
\end{aligned}$$

where $\hat{A}_S \approx \hat{E}_S$, to the *sum variance*,

$$\mathrm{sum}(\mathrm{var}(U)(\hat{Q}_{\mathrm{m},U}(E,z))) = z \sum_{S \in V^{\mathrm{CS}}} \hat{E}_S(1 - \hat{E}_S)$$

In any case, it can be seen that the log of the *generalised multinomial probability distribution* varies with the *sized entropy* of the *sample histogram* as well as the *sized negative relative entropy*,

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{m},U}(E,z)(A) &\approx -z \sum_{S \in A^{\mathrm{FS}}} \hat{A}_S \ln \frac{\hat{A}_S}{\hat{E}_S} \\
&= z \times \mathrm{entropy}(A) + z \sum_{S \in A^{\mathrm{FS}}} \hat{A}_S \ln \hat{E}_S
\end{aligned}$$

In the case where the *states* are uniformly probable, the *distribution histogram* is the *cartesian*, V^{C}. In this case the *generalised multinomial probability distribution* is proportional to the *multinomial coefficient*,

$$\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}}, z)(A) \quad = \quad \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \frac{1}{v^z}$$

and the logarithm of the *multinomial probability distribution* varies with the *sized entropy*,

$$\ln \hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}}, z)(A) \quad \sim \quad z \times \mathrm{entropy}(A)$$

This case is equivalent to the case discussed in 'Histogram entropy', above. Let $I \subset \mathrm{histogram}$ be the *histogram* valued function of all possible *histories* of *size* z in *variables* V,

$$I \quad = \quad \{(H, \mathrm{histogram}(H)) : H \in \{1 \ldots z\} :\to V^{\mathrm{CS}}\}$$

Let W be the cardinality of *histories* for each *histogram*,

$$W \quad = \quad \{(A, |D|) : (A, D) \in I^{-1}\}$$

The *histogram probability function*, $\hat{W} \in \mathcal{P}$, equals the *cartesian-distributed multinomial probability distribution*,

$$\hat{W}(A) \;=\; \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \frac{1}{v^z} \;=\; \hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}}, z)(A)$$

That is, the case where the *histogram* is *drawn* from the *uniform cartesian distribution histogram* is equivalent to uniformly probable *state*.

Consider functions of the *sample histograms*. Conjecture that the expected *independent histogram* of a *generalised multinomial probability distribution* equals the *scaled independent distribution histogram*

$$\{(S, \mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E, z))(\{(A, A^{\mathrm{X}}_S) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\})) : S \in V^{\mathrm{CS}}\} = \mathrm{scalar}(z) * \hat{E}^{\mathrm{X}} = \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z))$$

where $\hat{E} = E/\mathrm{scalar}(z_E)$. Conjecture that the *counts* of the *states* of the *sample histograms* are positively correlated with the *counts* of the *states* of the *independent* of the *sample histograms*, but that the correlation is less than one

$$\forall S \in V^{\mathrm{CS}}\ (0 \leq X(E, z, S) < 1)$$

where *independent* correlation function $X(E, z, S)$ is defined for *effective volumes* greater than one, $|E^{\mathrm{F}}| > 1$

$$X(E, z, S) := \\ \quad \mathrm{correlation}(\hat{Q}_{\mathrm{m},U}(E, z))(\{(A, A_S) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}, \{(A, A^{\mathrm{X}}_S) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\})$$

Compare the positive correlation between a *state* of the *sample histograms* and the corresponding *state* of the *independent sample histograms* to the negative correlation between different *states* of the *sample histograms*,

$$\mathrm{cov}(U)(\hat{Q}_{\mathrm{m},U}(E, z))((S, R)) = -z\hat{E}_S\hat{E}_R$$

Conjecture that the covariance between different *states* of the *independent* of *sample histograms* is less negative

$$\forall S, R \in V^{\mathrm{CS}}\ (R \neq S \implies |Y(E, z, S, R)| \leq |\mathrm{cov}(U)(\hat{Q}_{\mathrm{m},U}(E, z))((S, R))|)$$

where *independent* covariance function $Y(E, z, S, R)$ is defined for $R \neq S$

$$Y(E, z, S, R) := \\ \quad \mathrm{covariance}(\hat{Q}_{\mathrm{m},U}(E, z))(\{(A, A^{\mathrm{X}}_R) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}, \{(A, A^{\mathrm{X}}_S) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\})$$

Consider the logarithm of the factorial function when interpolated by means of the unit-translated gamma function. The unit-translated gamma function is defined $(\Gamma_!) \in \mathbf{R} \to \mathbf{R}$ as $\Gamma_! x = \Gamma(x + 1)$ which is such that $\forall x \in \mathbf{N}\ (\ln \Gamma_! x = \ln \Gamma(x + 1) = \ln x!)$. The gamma function is log convex and hence the expected logarithm of the factorial of the *counts* of the *states* of the *sample histograms* is greater than or equal to the logarithm of the factorial of the *counts* of the *states* of the *mean histogram* by Jensen's inequality

$$\forall S \in V^{\mathrm{CS}}\ (\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E, z))(\{(A, \ln A_S!) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \geq \ln \Gamma_! M_S)$$

where the *mean histogram* is $M = \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E, z))$. Conjecture that the expected logarithm of the factorial of the *counts* of the *states* of the *sample histograms* is greater than or equal to the expected logarithm of the factorial of the *counts* of the *states* of the *independent sample histograms*

$$\forall S \in V^{\mathrm{CS}}\ (\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E, z))(\{(A, \ln A_S!) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \geq \\ \mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E, z))(\{(A, \ln \Gamma_! A^{\mathrm{X}}_S) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}))$$

3.17.3 Multiple binomial distributions

Consider the set of lists of the *events drawn with replacement* from *history* $H_E \in \mathcal{H}$ of cardinality z, $H_E^z = \{L : L \in \mathcal{L}(H_E),\ |L| = z\} \subset \mathcal{L}(H_E)$. Whereas above the lists where modified to construct *multinomial sample histories drawn with replacement*, here they are modified to form *binomial sample histories drawn with replacement* of some given *state* $S \in V^{\mathrm{CS}}$ where $V = \mathrm{vars}(H_E)$. Replace all *states* not equal to S with a dummy *empty state* $\emptyset \in \mathcal{S}$ to define a set of *sample histories*, $X = \{G : L \in H_E^z,\ G = \{((i,x),R) : (i,(x,R)) \in L,\ R = S\} \cup \{((i,x),\emptyset) : (i,(x,R)) \in L,\ R \neq S\}\} \subset \mathcal{X} \to \mathcal{S}$. Strictly speaking the set X is not a set of *histories* because $\mathrm{vars}(\emptyset) \neq V$ (except in the case where the *history* H_E is *scalar*, $V = \emptyset$), but the construction serves to show how the *draw* is binomial.

Now consider H_E^z raised to the power of the *volume* $v = |V^{\mathrm{C}}|$, $(H_E^z)^v = \{L : L \in \mathcal{L}(H_E^z),\ |L| = v\} \subset \mathcal{L}(\mathcal{L}(H_E))$. Each of the lists of this set represents an independent *binomial draw* for each of the *states* in the *volume*. Let $D \in \mathrm{enums}(V^{\mathrm{CS}})$ be some map between the *states* of V and the elements of this set, $\forall L \in (H_E^z)^v\ (D \in V^{\mathrm{CS}} \leftrightarrow \mathrm{dom}(L))$. Construct from this set, $(H_E^z)^v$, a set of *sample histories* of cardinality less than or equal to z^v, $Y = \{G : L \in (H_E^z)^v,\ G = \{((i,j,x),S) : (S,i) \in D,\ (j,(x,R)) \in L_i,\ R = S\}\} \subset \mathcal{H}$. Here Y is a set of *multiple binomial sample histories drawn with replacement*. The *events* of the *states* corresponding to the dummy *empty state* are not included in this construction.

The set of *multiple binomial distributions* $Q_{\mathrm{b},U} \in \mathcal{A}_{\mathrm{i},U} \times \mathbf{N} \to \mathcal{Q}_U \cap \mathcal{Q}_{\mathrm{i}}$ is the set of *constructible distributions* parameterised by the *with replacement draw* $(E,z) \in \mathcal{A}_{\mathrm{i},U} \times \mathbf{N}$

$$\begin{aligned}Q_{\mathrm{b},U}(E,z) = \mathrm{count}(\{(\mathrm{histogram}(G) + V^{\mathrm{CZ}}, G) : L \in (H_E^z)^v,\\ G = \{((i,j,x),S) : (S,i) \in D,\ (j,(x,R)) \in L_i,\ R = S\}\}) \in \mathcal{Q}_U \cap \mathcal{Q}_{\mathrm{i}}\end{aligned}$$

where $V = \mathrm{vars}(A)$ and $D \in \mathrm{enums}(V^{\mathrm{CS}})$. The *support* of the *multiple binomial distribution* is the *multiple support*, $\mathrm{dom}(Q_{\mathrm{b},U}(E,z)) = \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}$. Hence the *sample size* is not constrained to be equal to the *draw size*. The minimum *size* of the *sample histograms* is zero and the maximum *size* is vz, $\forall A \in \mathrm{dom}(Q_{\mathrm{b},U}(E,z))\ (0 \leq \mathrm{size}(A) \leq vz)$. The maximum *count* in the *sample histograms* is less than or equal to the *draw size*

$$\forall A \in \mathrm{dom}(Q_{\mathrm{m}}(E,z))\ (\mathrm{maxr}(A) \leq z)$$

The *sum* of a *multiple binomial distribution* is

$$\mathrm{sum}(Q_{\mathrm{b},U}(E,z)) = z_E^{zv}$$

The set of *generalised multiple binomial distributions* $Q_{\mathrm{b},U} \in \mathcal{A}_U \times \mathbf{N} \to \mathcal{Q}_U$ is the set of *distributions* parameterised by *non-integral distribution histograms* as well as *integral draw distribution histograms*. They are defined explicitly, $Q_{\mathrm{b},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}} \to \mathbf{Q}_{\geq 0}) \subset \mathcal{Q}_U$, as

$$Q_{\mathrm{b},U}(E,z) := \{(A, \prod_{S \in V^{\mathrm{CS}}} \frac{z!}{A_S!(z-A_S)!} E_S^{A_S} (z_E - E_S)^{z-A_S}) : A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}\}$$

where $V = \mathrm{vars}(E)$.

The *generalised multiple binomial probability distribution* is $\hat{Q}_{\mathrm{b},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{Q}_U \cap \mathcal{P}$

$$\begin{aligned}
\hat{Q}_{\mathrm{b},U}(E,z) &= \mathrm{normalise}(Q_{\mathrm{b},U}(E,z)) \\
&= \{(A, \prod_{S \in V^{\mathrm{CS}}} \frac{z!}{A_S!(z-A_S)!} \hat{E}_S^{A_S} (1 - \hat{E}_S)^{z-A_S}) : A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}\}
\end{aligned}$$

where $\hat{E} = E/\mathrm{scalar}(z_E)$.

The *mean* of the *generalised multiple binomial probability distribution* is

$$\mathrm{mean}(\hat{Q}_{\mathrm{b},U}(E,z)) = \mathrm{scalar}(z) * \hat{E}$$

If the *mean* of the *generalised multiple binomial probability distribution* is *integral* then it is the element of the singleton *modal* set

$$M \in \mathcal{A}_i \implies \mathrm{modes}(\hat{Q}_{\mathrm{b},U}(E,z)) = \{M\}$$

where $M = \mathrm{mean}(\hat{Q}_{\mathrm{b},U}(E,z))$. If the *mean* is not *integral* then the *modal* set is a singleton consisting of the *floor* of the *mean*

$$\mathrm{modes}(\hat{Q}_{\mathrm{b},U}(E,z)) = \{\mathrm{floor}(A)\}$$

The *variance* of *state* S in the *generalised multiple binomial probability distribution* is

$$\mathrm{var}(U)(\hat{Q}_{\mathrm{b},U}(E,z))(S) = z\hat{E}_S(1 - \hat{E}_S)$$

The *covariance* of a pair of *states* (S,R), where $R \neq S$, is zero because the *states* are independently *drawn*,

$$\mathrm{cov}(U)(\hat{Q}_{\mathrm{b},U}(E,z))((S,R)) = 0$$

The *moment generating function* of the *generalised multiple binomial probability distribution* is

$$\mathrm{mgf}(U)(\hat{Q}_{\mathrm{b},U}(E,z))(T) = \prod_{S \in V^{\mathrm{CS}}} (1 - \hat{E}_S + \hat{E}_S e^{T_S})$$

where $\hat{E} = E/\mathrm{scalar}(z_E)$ and $T \in V^{\mathrm{CS}} \to \mathbf{R}$.

The *mean* and *variance* of the *generalised multiple binomial probability distribution* equals the *mean* and *variance* of the *generalised multinomial probability distribution*

$$\mathrm{mean}(\hat{Q}_{\mathrm{b},U}(E,z)) = \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E,z))$$

and

$$\mathrm{var}(U)(\hat{Q}_{\mathrm{b},U}(E,z)) = \mathrm{var}(U)(\hat{Q}_{\mathrm{m},U}(E,z))$$

However the *covariance* is not equal, $\mathrm{cov}(U)(\hat{Q}_{\mathrm{m},U}(E,z))((S,R)) \neq 0$ where $E_S, E_R > 0$. Increasing *entropy* of the *distribution histogram*, $\mathrm{entropy}(E)$, tends to decrease the absolute *sum covariance* of joint *states*, $R \neq S$, of the *generalised multinomial probability distribution*

$$|\mathrm{sum}(\mathrm{cov}(U)(\hat{Q}_{\mathrm{m},U}(E,z))) - \mathrm{sum}(\mathrm{var}(U)(\hat{Q}_{\mathrm{m},U}(E,z)))|$$

So as *entropy* increases the moments converge and the *generalised multiple binomial probability distribution* increasingly approximates to the *generalised multinomial probability distribution* where the *supports* intersect

$$\{(A,f) : (A,f) \in \hat{Q}_{\mathrm{b},U}(E,z),\ \mathrm{size}(A) = z\} \approx \hat{Q}_{\mathrm{m},U}(E,z)$$

Similarly, the conjecture above that the covariance between different *states* of the *independent* of *sample histograms* in the *generalised multinomial probability distribution* is less negative than between different *states* of the *sample histograms* suggests that the *generalised multiple binomial probability distribution* approximates to the *generalised multinomial probability distribution* better for the *independent sample histograms*

$$\{(A,f) : (A,f) \in \hat{Q}_{\mathrm{b},U}(E,z),\ \mathrm{size}(A) = z,\ A = A^{\mathrm{X}}\} \approx \{(A,f) : (A,f) \in \hat{Q}_{\mathrm{m},U}(E,z),\ A = A^{\mathrm{X}}\}$$

The *multiple Poisson probability function* is $\hat{Q}_{\mathrm{p},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}} \to \mathbf{R}_{\geq 0}) \cap \mathcal{P}$ is defined

$$\hat{Q}_{\mathrm{p},U}(E,z) = \{(A, \prod_{S \in V^{\mathrm{CS}}} \frac{e^{-z\hat{E}_S}(z\hat{E}_S)^{A_S}}{A_S!}) : A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}\}$$

where $\hat{E} = E/\mathrm{scalar}(z_E)$. Here the Poisson distribution parameter is $z\hat{E}_S$. This function is not a *distribution* because it is real-valued. Hence it cannot be constructed from finite H_E. However at large z and v it approximates to the *generalised multiple binomial probability distribution*. Using Stirling's approximation,

$$\begin{aligned}
\hat{Q}_{\mathrm{p},U}(E,z) &= \{(A, \prod_{S \in V^{\mathrm{CS}}} \frac{e^{-z\hat{E}_S}(z\hat{E}_S)^{A_S}}{A_S!}) : A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}\} \\
&= \{(A,\ e^{-z}z^z \prod_{S \in V^{\mathrm{CS}}} \frac{\hat{E}_S^{A_S}}{A_S!}) : A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}\} \\
&\approx \{(A,\ z! \prod_{S \in V^{\mathrm{CS}}} \frac{\hat{E}_S^{A_S}}{A_S!}) : A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}\} \\
&\approx \{(A, \prod_{S \in V^{\mathrm{CS}}} \frac{z!}{A_S!(z-A_S)!}\hat{E}_S^{A_S}(1-\hat{E}_S)^{z-A_S}) : A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}\} \\
&= \hat{Q}_{\mathrm{b},U}(E,z)
\end{aligned}$$

The *multiple Poisson probability function* also approximates to the *generalised multinomial probability distribution* at large z and v

$$\begin{aligned}
\{(A,\ z! \prod_{S \in V^{\mathrm{CS}}} \frac{\hat{E}_S^{A_S}}{A_S!}) : A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}\} &\supset \{(A,\ z! \prod_{S \in V^{\mathrm{CS}}} \frac{\hat{E}_S^{A_S}}{A_S!}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\} \\
&= \hat{Q}_{\mathrm{m},U}(E,z)
\end{aligned}$$

3.17.4 Uniform distributions

The *uniform distributions* in *system* U are parameterised by a pair of (i) a set of *draw variables* $V \subset \mathcal{V}_U$ and (ii) a non-zero integral *draw size* $z \in \mathbf{N}_{>0}$. In terms of sets of *histories*, the *uniform distribution* is a *constructible distribution* $Q_{\mathrm{u},U} \in \mathcal{V}_U \times \mathbf{N}_{>0} \to \mathcal{Q}_{\mathrm{i}}$ is defined

$$\begin{aligned}
Q_{\mathrm{u},U}(V,z) &= \mathrm{count}(\{(\mathrm{histogram}(G), G) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ G = \mathrm{history}(A)\}) \\
&= \mathcal{A}_{U,\mathrm{i},V,z} \times \{1\}
\end{aligned}$$

The *support* of *uniform distributions* is the *integral congruent support*. That is, $\mathrm{dom}(Q_{\mathrm{u},U}(V,z)) = \mathcal{A}_{U,\mathrm{i},V,z}$.

The *uniform probability distribution* $\hat{Q}_{\mathrm{u},U}(V,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{Q}_U \cap \mathcal{Q}_z \cap \mathcal{P}$ is defined

$$\begin{aligned}\hat{Q}_{\mathrm{u},U}(V,z) &= \mathcal{A}_{U,\mathrm{i},V,z} \times \{1/|\mathcal{A}_{U,\mathrm{i},V,z}|\} \\ &= \mathcal{A}_{U,\mathrm{i},V,z} \times \left\{\frac{z!\ (v-1)!}{(z+v-1)!}\right\}\end{aligned}$$

where $v = |V^{\mathrm{CS}}|$. The *modal set* of a *uniform probability distribution* is the entire *support*, $\mathrm{modes}(\hat{Q}_{\mathrm{u},U}(V,z)) = \mathcal{A}_{U,\mathrm{i},V,z}$. The *mean histogram* is the *uniform histogram*, $\mathrm{mean}(\hat{Q}_{\mathrm{u},U}(V,z)) = \mathrm{scalar}(z/v) * V^{\mathrm{C}}$ where $v = |V^{\mathrm{C}}|$. Compare this to the *mean histogram* of the *multiple support*, $\mathrm{mean}(\mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}} \times \{1\}) = \mathrm{scalar}(z/2) * V^{\mathrm{C}}$.

The *supports* of the *uniform probability distribution* $\hat{Q}_{\mathrm{u},U}(V,z)$ and the *generalised multinomial probability distribution* of the *uniform distribution histogram* $\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}},z)$ are equal

$$\mathrm{dom}(\hat{Q}_{\mathrm{u},U}(V,z)) = \mathrm{dom}(\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}},z)) = \mathcal{A}_{U,\mathrm{i},V,z}$$

but the *distributions* are not equal, $\hat{Q}_{\mathrm{u},U}(V,z) \neq \hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}},z)$, except in the trivial case of *mono-variate, mono-valent* V.

3.17.5 Iso-independent conditional multinomial distributions

The discussion 'Historical distributions', above, shows how the *historical distribution*, $Q_{\mathrm{h}}(E,z)$, is derived from subsets of the *events drawn without replacement* from *history* $H_E \in \mathcal{H}$ of cardinality z, $\{G : G \subseteq H_E,\ |G| = z\}$. The set of *historical distributions* $Q_{\mathrm{h}} \in \mathcal{A}_{\mathrm{i}} \times \mathbf{N} \to \mathcal{Q}_{\mathrm{i}}$ is the set of *constructible distributions* parameterised by a *without replacement draw* $(E,z) \in \mathcal{A}_{\mathrm{i}} \times \mathbf{N}$

$$Q_{\mathrm{h}}(E,z) = \mathrm{count}(\{(\mathrm{histogram}(G),G) : G \subseteq \mathrm{history}(E),\ |G| = z\}) \in \mathcal{Q}_{\mathrm{i}} \cap \mathcal{Q}_z$$

Similarly, the discussion 'Multinomial distributions', above, shows how the *multinomial distribution*, $Q_{\mathrm{m}}(E,z)$, is derived from the set of lists of the *events drawn with replacement* from *history* $H_E \in \mathcal{H}$ of cardinality z, $H_E^z = \{L : L \in \mathcal{L}(H_E),\ |L| = z\}$. The set of *multinomial distributions* $Q_{\mathrm{m}} \in \mathcal{A}_{\mathrm{i}} \times \mathbf{N} \to \mathcal{Q}_{\mathrm{i}}$ is the set of *constructible distributions* parameterised by a *with replacement draw* $(E,z) \in \mathcal{A}_{\mathrm{i}} \times \mathbf{N}$

$$\begin{aligned}Q_{\mathrm{m}}(E,z) = \mathrm{count}(\{&(\mathrm{histogram}(G),G) : \\ &L \in \mathrm{history}(E)^z,\ G = \{((i,x),S) : (i,(x,S)) \in L\}\}) \in \mathcal{Q}_{\mathrm{i}} \cap \mathcal{Q}_z\end{aligned}$$

Both cases construct an intermediate set of *histories* $I \subset \mathcal{H}$. In the *historical* case the intermediate *histories* are simply subsets of the given *history*, $I = \{G : G \subseteq H_E,\ |G| = z\} \in \mathrm{P}(\mathcal{H})$. In the *multinomial* case the intermediate *histories* are constructed by prefixing by list position the *event identifiers* of lists of *events* of the given *history*, $I = \{G : L \in H_E^z,\ G = \{((i,x),S) : (i,(x,S)) \in L\}\} \in \mathrm{P}(\mathcal{H})$. The cardinalities of the components of the partition of I implied by the inverse of the *histogram* function, histogram $\in \mathcal{H} \to \mathcal{A}$, form the *frequencies* of the *constructible distribution*, $\{(A,|C|) : (A,C) \in \mathrm{inverse}(\mathrm{filter}(I,\mathrm{his}))\} = \mathrm{count}(\{(\mathrm{his}(G),G) : G \in I\}) \in \mathrm{Q_i}$, where his = histogram. So the count partition is $\mathrm{ran}(\mathrm{inverse}(\mathrm{filter}(I,\mathrm{his}))) = \mathrm{ran}(\mathrm{inverse}(\{(G,\mathrm{his}(G)) : G \in I\})) \in \mathrm{B}(I)$.

Given some partition $P \in \mathrm{B}(I)$ of the intermediate set of *drawn histories* $I \in \mathrm{P}(\mathcal{H})$, a second intermediate set $J \in \mathrm{P}(\mathcal{H})$ having a corresponding partition $R \in \mathrm{B}(J)$ may be constructed such that (i) there exists a bijection $M \in P \cdot R$ between the partitions, and (ii) the cardinalities of the components of R are uniform, $|\{|C'| : C' \in R\}| = 1$. Each component $C' \in R$ corresponding to a component $C \in P$, that is, $(C,C') \in M$, may be constructed by prefixing each of the *event identifiers* of each of the *histories* in C with the sets of *histories* in the product of the remaining components, $\prod_{D \in P \setminus \{C\}} D \subset \mathrm{P}(\mathcal{H})$. Define hiso $\in \mathrm{P}(\mathrm{P}(\mathcal{H})) \to \mathrm{P}(\mathrm{P}(\mathcal{H}))$ as

$$\mathrm{hiso}(P) := \{C' : C \in P,\ C' = \{\{((N,x),S) : (x,S) \in G\} : G \in C,\ N \in \prod_{D \in P \setminus \{C\}} D\}\}$$

Then $R = \mathrm{hiso}(P)$ and $J = \bigcup R$. If the argument P to the *history iso* function is a partition, $P \in \mathrm{B}(\bigcup P)$, then the cardinality of each of the components of the resultant partition is $\prod_{C \in P} |C|$. The sum of the cardinalities is $|J| = \sum_{C' \in R} |C'| = |P| \prod_{C \in P} |C|$. The resultant second intermediate set of *histories* J forms a new *distribution*, $\mathrm{count}(\{(\mathrm{his}(G),G) : G \in J\} \in \mathcal{Q}_\mathrm{i}$.

Now consider a *with replacement draw* parameterised by (i) a *distribution histogram* E in *variables* V and *system* U, and (ii) an *independent histogram* A^X in *variables* V having integral *size* $z = \mathrm{size}(A^\mathrm{X}) \in \mathbf{N}$. That is, *draw* parameters $(E, A^\mathrm{X}) \in \mathcal{A}_{U,\mathrm{i}} \times \mathcal{A}_U$. The *independent histogram size*, z, defines the *draw size*. The *independent histogram* is also constrained such that the *integral iso-independent* set defined by it is non-empty, $|Y_{U,\mathrm{i},V,z}^{-1}(A^\mathrm{X})| > 0$, where $Y_{U,\mathrm{i},V,z} = \{(B, B^\mathrm{X}) : B \in \mathcal{A}_{U,\mathrm{i},V,z}\} \subset$ independent is the *integral congruent independent function*.

Define the *iso-independent conditional multinomial distribution* $Q_{\mathrm{m,y}}(E, A^{\mathrm{X}})$ as the *distribution* derived from the subset of the lists of the *events drawn with replacement*, H_E^z, that are constrained to the set of *integral iso-independents* defined by the *draw* parameter A^{X},

$$\{L : L \in \mathcal{L}(H_E),$$
$$G = \{((i,x),S) : (i,(x,S)) \in L\},\ B = \mathrm{histogram}(G),\ B^{\mathrm{X}} \equiv A^{\mathrm{X}}\} \subseteq H_E^z$$

That is

$$\begin{aligned} &Q_{\mathrm{m,y}}(E, A^{\mathrm{X}}) = \mathrm{count}(\{(B,G) : \\ &\quad L \in \mathrm{history}(E)^z,\ G = \{((i,x),S) : (i,(x,S)) \in L\}, \\ &\quad B = \mathrm{histogram}(G),\ B^{\mathrm{X}} \equiv A^{\mathrm{X}}\}) \in \mathcal{Q}_z \end{aligned}$$

The *iso-independent conditional multinomial distribution*, $Q_{\mathrm{m,y}}(E, A^{\mathrm{X}})$, is a subset of the corresponding *multinomial distribution* of the same *size*, $Q_{\mathrm{m}}(E, z)$. That is, $Q_{\mathrm{m,y}}(E, A^{\mathrm{X}}) = \{(B,f) : (B,f) \in Q_{\mathrm{m}}(E,z),\ B^{\mathrm{X}} \equiv A^{\mathrm{X}}\}$. Thus $\mathrm{sum}(Q_{\mathrm{m,y}}(E, A^{\mathrm{X}})) \leq \mathrm{sum}(Q_{\mathrm{m}}(E,z))$. The *iso-independent conditional multinomial distribution* can be defined explicitly for $B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$ as

$$Q_{\mathrm{m,y}}(E, A^{\mathrm{X}})(B) = Q_{\mathrm{m}}(E,z)(B) = \frac{z!}{\prod_{S \in B^{\mathrm{S}}} B_S!} \prod_{S \in B^{\mathrm{S}}} E_S^{B_S} \in \mathbf{N}_{>0}$$

The *stuffed iso-independent conditional multinomial distribution*

$$Q_{\mathrm{m,y},U}(E, A^{\mathrm{X}}) \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \to \mathbf{Q}_{\geq 0} \subset \mathcal{Q}_U \cap \mathcal{Q}_z$$

can be constructed from an *iso-independent conditional multinomial distribution*, $Q_{\mathrm{m,y}}(E, A^{\mathrm{X}})$, by *completing* the *support histograms* and *stuffing* with the disjoint subset of the *integral congruent support* that are *iso-independent histograms* with zero *frequencies*

$$\begin{aligned} Q_{\mathrm{m,y},U}(E, A^{\mathrm{X}}) &= \{(B + B^{\mathrm{CZ}}, f) : (B,f) \in Q_{\mathrm{m,y}}(E, A^{\mathrm{X}})\} \cup \\ &\quad (Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \setminus \{B + B^{\mathrm{CZ}} : B \in \mathrm{dom}(Q_{\mathrm{m,y}}(E, A^{\mathrm{X}}))\}) \times \{0\} \end{aligned}$$

The *stuffed iso-independent conditional multinomial probability distribution* $\hat{Q}_{\mathrm{m,y},U}(E, A^{\mathrm{X}}) \in (Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \to \mathbf{Q}_{\geq 0}) \cap \mathcal{Q}_U \cap \mathcal{Q}_z \cap \mathcal{P}$ is defined

$$\begin{aligned} \hat{Q}_{\mathrm{m,y},U}(E, A^{\mathrm{X}}) &= \mathrm{normalise}(Q_{\mathrm{m,y},U}(E, A^{\mathrm{X}})) \\ &= \{(B, f/\mathrm{sum}(Q_{\mathrm{m,y},U}(E, A^{\mathrm{X}}))) : (B,f) \in Q_{\mathrm{m,y},U}(E, A^{\mathrm{X}})\} \end{aligned}$$

where $z_E > 0$ and $z > 0$.

Finally the *generalised iso-independent conditional multinomial probability distribution* over the entire *integral congruent support*, $\mathcal{A}_{U,\text{i},V,z}$, can be constructed by treating each of the *iso-independent* components of the partition implied by the *integral congruent independent* function, $\text{ran}(Y^{-1}_{U,\text{i},V,z}) \in \text{B}(\mathcal{A}_{U,\text{i},V,z})$, as equally probable. Define the *generalised iso-independent conditional multinomial probability distributions* parameterised by both *integral* and *non-integral distribution histograms*, $\hat{Q}_{\text{m,y},U} \in \mathcal{A}_U \times \mathbf{N} \to \mathcal{Q}_U$, explicitly $\hat{Q}_{\text{m,y},U}(E,z) \in (\mathcal{A}_{U,\text{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P} \subset \mathcal{Q}_U \cap \mathcal{Q}_z$ as

$$\hat{Q}_{\text{m,y},U}(E,z) := \text{normalise}(\{(A, \frac{Q_{\text{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})} Q_{\text{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\text{i},V,z}\})$$

which is such that

$$\hat{Q}_{\text{m,y},U}(E,z) = \text{normalise}(\bigcup(\hat{Q}_{\text{m,y},U}(E,A^{\text{X}}) : A^{\text{X}} \in \text{ran}(Y_{U,\text{i},V,z})))$$

In the case of *integral distribution histogram*, $E \in \mathcal{A}_\text{i}$, this definition of the *generalised iso-independent conditional multinomial probability distribution*, $\hat{Q}_{\text{m,y},U}(E,z)$, implies a corresponding *constructible distribution* $Q_{\text{m,y},U}(E,z) \in (\mathcal{A}_{U,\text{i},V,z} \to \mathbf{N}) \subset \mathcal{Q}_\text{i}$, by scaling the *frequencies* of $\hat{Q}_{\text{m,y},U}(E,z)$ by a factor,

$$Q_{\text{m,y},U}(E,z)(A) := \left(|\text{ran}(Y_{U,\text{i},V,z})| \prod_{A^{\text{X}} \in \text{ran}(Y_{U,\text{i},V,z})} \sum_{B \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})} Q_{\text{m},U}(E,z)(B) \right) \times \hat{Q}_{\text{m,y},U}(E,z)(A)$$

The *iso-independent* function, $Y_{U,\text{i},V,z} \in \mathcal{A}_{U,\text{i},V,z} \to \mathcal{A}_{U,V,z} \subset \text{independent}$, implies a function of *histories*, $\{(G, Y_{U,\text{i},V,z}(A + A^{\text{CZ}})) : G \in I,\ A = \text{his}(G)\} = \{(G, A^{\text{X}} + A^{\text{CZ}}) : G \in I,\ A = \text{his}(G)\} \in \mathcal{H} \to \mathcal{A}_{U,V,z}$, where the *multinomial* intermediate set of *histories* is $I = \{G : L \in H^z_E,\ G = \{((i,x),S) : (i,(x,S)) \in L\}\} \in \text{P}(\mathcal{H})$ and his = histogram. This in turn implies a partition of *histories* $P = \text{ran}(\text{inverse}(\{(G, A^{\text{X}} + A^{\text{CZ}}) : G \in I,\ A = \text{his}(G)\})) \in \text{B}(I)$. This partition of *histories* is a parent partition of the partition of *histories* implied by the *histogram* function, $\text{parent}(P, \text{ran}(\text{inverse}(\text{filter}(I, \text{his}))))$. The scaling factor is equal to the sum of the cardinalities of the resultant partition of the *history iso* function

$$\sum_{C' \in \text{hiso}(P)} |C'| = |P| \prod_{C \in P} |C| = |\text{ran}(Y_{U,\text{i},V,z})| \prod_{A^{\text{X}} \in \text{ran}(Y_{U,\text{i},V,z})} \sum_{B \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})} Q_{\text{m},U}(E,z)(B)$$

That is, the *frequencies* of the *multinomial distribution*, $Q_{\text{m},U}(E,z)$, are partitioned by the *iso-independent* function, $Y_{U,\text{i},V,z}$. The sum of the *frequencies* of each component of this partition is scaled. The scaling of the *frequencies* of *histograms* corresponds to the repetition of *histories* in the components of partition $R = \mathrm{hiso}(P) \in \mathrm{B}(J)$, of the second intermediate set of *histories* $J = \bigcup R \subset \mathcal{H}$, having uniform component cardinalities, $|\{|C'| : C' \in R\}| = 1$. The resultant second intermediate set of *histories*, J, forms the *generalised iso-independent conditional multinomial distribution*, $Q_{\mathrm{m,y},U}(E,z) = \mathrm{count}(\{(\mathrm{his}(G), G) : G \in J\} \in \mathcal{Q}_{\mathrm{i}}$.

The upper bound to the cardinality of the second intermediate *history* set, $|J|$, which is the sum of the *generalised iso-independent conditional multinomial distribution*, is

$$\mathrm{sum}(Q_{\mathrm{m,y},U}(E,z)) \leq r \left(\frac{z_E^z}{r}\right)^r$$

where $z_E^z \geq r$ and

$$r = |\mathrm{ran}(Y_{U,\mathrm{i},V,z})| = \prod_{w \in V} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!}$$

The probability of an *integral congruent support histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ may be compared between the *generalised iso-independent conditional multinomial probability distribution*, $\hat{Q}_{\mathrm{m,y},U}(E,z)$, and the *generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m},U}(E,z)$. The cardinality of the range of the *integral congruent independent function* is

$$|\mathrm{ran}(Y_{U,\mathrm{i},V,z})| = \prod_{w \in V} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!}$$

If the sum of the *iso-independent* probabilities of A is greater than the fraction implied by this cardinality

$$\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B) > \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|}$$

then the *generalised iso-independent conditional multinomial probability* is less than the *generalised multinomial probability*, $\hat{Q}_{\mathrm{m,y},U}(E,z)(A) < \hat{Q}_{\mathrm{m},U}(E,z)(A)$, and vice-versa. In the case where the *independent histogram* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$ and therefore an *iso-independent*, $A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and such that

$$\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}}) > \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|}$$

then $\hat{Q}_{\mathrm{m,y},U}(E,z)(A) < \hat{Q}_{\mathrm{m},U}(E,z)(A)$. Furthermore, it is conjectured above that the logarithm of the cardinality of the *integral iso-independents* corresponding to A^{X} varies with the *size* scaled *independent entropy*,

$$\ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| \quad \sim \quad z \times \mathrm{entropy}(A^{\mathrm{X}})$$

Therefore conjecture that the *generalised iso-independent conditional multinomial probability* tends to be less than the *generalised multinomial probability*, $\hat{Q}_{\mathrm{m,y},U}(E,z)(A) < \hat{Q}_{\mathrm{m},U}(E,z)(A)$, when the *entropy* of the *independent histogram*, $\mathrm{entropy}(A^{\mathrm{X}})$, is high, and vice-versa

$$\ln \hat{Q}_{\mathrm{m,y},U}(E,z)(A) - \ln \hat{Q}_{\mathrm{m},U}(E,z)(A) \quad \sim \quad -\ \mathrm{entropy}(A^{\mathrm{X}})$$

The *generalised iso-independent conditional multinomial probability distribution* is constructed by normalising each of the components of the *iso-independent* partition of the *integral congruent support*, $\mathrm{ran}(Y^{-1}_{U,\mathrm{i},V,z}) \in \mathrm{B}(\mathcal{A}_{U,\mathrm{i},V,z})$. The same method can be applied to construct a *conditional multinomial probability distribution* given any partition. Consider the *integral iso-transform-independent* partition, $\mathrm{ran}(Y^{-1}_{U,\mathrm{i},T,z}) \in \mathrm{B}(\mathcal{A}_{U,\mathrm{i},V,z})$, given *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ where $\mathrm{und}(T) = V$ and $W = \mathrm{der}(T)$. The *integral iso-transform-independent* function is defined $Y_{U,\mathrm{i},T,z} \in \mathcal{A}_{U,\mathrm{i},V,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})$ as $Y_{U,\mathrm{i},T,z} = \{(A,((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$. Now the subset of the lists of the *events drawn with replacement*, H_E^z, that are constrained to the set of *integral iso-transform-independents* defined by the *draw* parameter $((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})$ is

$$\begin{aligned}
&\{L : L \in \mathcal{L}(H_E),\\
&\quad G = \{((i,x),S) : (i,(x,S)) \in L\},\ B = \mathrm{histogram}(G),\\
&\quad B^{\mathrm{X}} * T \equiv A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} \equiv (A * T)^{\mathrm{X}}\} \subseteq H_E^z
\end{aligned}$$

The *generalised iso-transform-independent conditional multinomial probability distribution* over the *integral congruent support*, $\mathcal{A}_{U,\mathrm{i},V,z}$, is defined $\hat{Q}_{\mathrm{m,y},T,U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P} \subset \mathcal{Q}_U \cap \mathcal{Q}_z$ as

$$\begin{aligned}
&\hat{Q}_{\mathrm{m,y},T,U}(E,z) =\\
&\quad \mathrm{normalise}(\{(A, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}}))} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\})
\end{aligned}$$

In the case of *integral distribution histogram*, $E \in \mathcal{A}_{\mathrm{i}}$, this definition of the *generalised iso-transform-independent conditional multinomial probability distribution*, $\hat{Q}_{\mathrm{m,y},T,U}(E,z)$, implies a corresponding *constructible distribution* $Q_{\mathrm{m,y},T,U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{N}) \subset \mathcal{Q}_{\mathrm{i}}$, by scaling the *frequencies* of

$\hat{Q}_{\mathrm{m,y},U}(E,z)$ by a factor,

$$Q_{\mathrm{m,y},T,U}(E,z)(A) :=$$
$$\left(|\mathrm{ran}(Y_{U,\mathrm{i},T,z})| \prod_{X \in \mathrm{ran}(Y_{U,\mathrm{i},T,z})} \sum_{B \in Y^{-1}_{U,\mathrm{i},T,z}(X)} Q_{\mathrm{m},U}(E,z)(B)\right) \times \hat{Q}_{\mathrm{m,y},T,U}(E,z)(A)$$

3.17.6 Likely histograms

Let $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ be an *integral substrate histogram* in *system* U having non-empty *variables* $V = \mathrm{vars}(A) \neq \emptyset$ and non-zero *size* $z = \mathrm{size}(A) > 0$. The *maximum likelihood estimate* for the *distribution histogram* of the *generalised multinomial probability* of the *histogram*, A, is the *mean* or *histogram* itself,

$$\{A\} = \mathrm{maxd}(\{(D, Q_{\mathrm{m},U}(D,z)(A)) : D \in \mathcal{A}_{U,V,z}\})$$

The *maximum likelihood estimate* $A_{\mathrm{x}} \in \mathcal{A}_{U,V,z}$ for the *distribution histogram* of the sum of the *generalised multinomial probabilities* of the *integral iso-independents* of the *histogram*, A, is defined

$$\{A_{\mathrm{x}}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))) : D \in \mathcal{A}_{U,V,z}\})$$

where the *integral iso-independents* is

$$Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \quad = \quad \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} = A^{\mathrm{X}}\}$$

Conjecture that the *maximum likelihood estimate* for the *multinomial probability* of the membership of a *histogram* in the *iso-independents* is simply the *independent*, $A_{\mathrm{x}} = A^{\mathrm{X}}$.

The *independent* is in the *iso-independents*, $A^{\mathrm{X}} \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})$. If the *independent* is *integral*, it is in the *integral iso-independents*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$.

The *histogram independent*, A^{X}, is the *maximum likelihood estimate* of the *distribution histogram* of the total *multinomial probability* of the subset of the *integral substrate histograms* which are such that the *independent* equals the *histogram independent*,

$$\{A^{\mathrm{X}}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))) : D \in \mathcal{A}_{U,V,z}\})$$

Now consider the case where the membership of the *iso-independents* is a given. Define the *dependent histogram* $A^{\mathrm{Y}} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-independent*,

$$\{A^{\mathrm{Y}}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}) : D \in \mathcal{A}_{U,V,z}\})$$

The *dependent histogram*, A^{Y}, is only defined if there is a unique maximum,

$$|\mathrm{max}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}) : D \in \mathcal{A}_{U,V,z}\})| = 1$$

The *dependent histogram* equals the *histogram* if the *histogram* is *independent*, $A = A^{\mathrm{X}} \implies A^{\mathrm{Y}} = A = A^{\mathrm{X}}$.

In the case where the *histogram* is not *independent*, $A \neq A^{\mathrm{X}}$, and the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, then the *independent* term appears in the denominator,

$$\begin{aligned} &A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \implies \\ &\qquad 0 < Q_{\mathrm{m},U}(D,z)(A^{\mathrm{X}}) < \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})) \end{aligned}$$

The *histogram* term, $Q_{\mathrm{m},U}(D,z)(A)$, appears in both the numerator and the denominator, so while (a) the *maximum likelihood estimate* for the numerator alone is just the *mean*, A, and (b) the *maximum likelihood estimate* for the denominator alone is the *independent*, A^{X}, optimisation overall in the *iso-independent* case tends to minimise the *independent* term, $Q_{\mathrm{m},U}(D,z)(A^{\mathrm{X}})$, in the denominator, while maximising the *histogram* term, $Q_{\mathrm{m},U}(D,z)(A)$, in the numerator. That is, in the denominator,

$$\begin{aligned} 0 &< \hat{Q}_{\mathrm{m},U}(A^{\mathrm{Y}},z)(A^{\mathrm{X}}) \\ &\leq \hat{Q}_{\mathrm{m},U}(A^{\mathrm{Y}},z)(A) \\ &< \sum(\hat{Q}_{\mathrm{m},U}(A^{\mathrm{Y}},z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})) \\ &\leq 1 \end{aligned}$$

So the overall *maximum likelihood estimate*, which is the *dependent*, is near the *histogram*, $A^{\mathrm{Y}} \sim A$, only in as much as it is far from the *independent*, $A^{\mathrm{Y}} \nsim A^{\mathrm{X}}$.

The *dependent*, A^{Y}, is sometimes not computable. Although the *substrate* *histograms* are countably infinite, $\mathcal{A}_{U,V,z} \leftrightarrow \mathbf{N}$, the maximisation never terminates. An approximation to the continuous case may be made by using a scaling factor. The *scaled complete integral congruent histograms* equals the *complete congruent histograms* in the limit

$$\lim_{k \to \infty} \{A/Z_k : A \in \mathcal{A}_{U,\mathrm{i},V,kz}\} = \mathcal{A}_{U,V,z}$$

where $k \in \mathbf{N}_{>0}$ and $Z_k = \mathrm{scalar}(k)$. The finite approximation to the *dependent* is

$$\{A^{\mathrm{Y}}_k\} = \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})$$

which is defined if the maximisation is a singleton. The approximation, $A^{\mathrm{Y}}_k \approx A^{\mathrm{Y}}$, improves as k tends to infinity.

Given a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ having *underlying variables* equal to the *variables* V of the *substrate histogram*, the *transform-independent* $A^{\mathrm{X}(T)} \in \mathcal{A}_{U,V,z}$ is defined as the *maximum likelihood estimate* for the *distribution histogram* of the sum of the *generalised multinomial probabilities* of the *integral iso-transform-independents* of the *histogram*, A,

$$\{A^{\mathrm{X}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A))) : D \in \mathcal{A}_{U,V,z}\})$$

where the *integral iso-transform-independents* is abbreviated

$$\begin{aligned}\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) &= Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \\ &= \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}\end{aligned}$$

In the case where the *integral iso-transform-independents* equals the *integral substrate histograms*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) = \mathcal{A}_{U,\mathrm{i},V,z}$, there is no unique maximum, and the *transform-independent* is defined as the *scaled normalised cartesian*,

$$\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) = \mathcal{A}_{U,\mathrm{i},V,z} \implies A^{\mathrm{X}(T)} := Z_A * \hat{V}^{\mathrm{C}}$$

where $Z_A = \mathrm{scalar}(\mathrm{size}(A))$ and $\hat{X} := \mathrm{normalise}(X)$. Otherwise, this definition assumes that there is always a unique maximum,

$$\forall A \in \mathcal{A}_{U,\mathrm{i},V,z}\ \forall T \in \mathcal{T}_{U,V}\ (\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) \neq \mathcal{A}_{U,\mathrm{i},V,z} \implies (|\max(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A))) : D \in \mathcal{A}_{U,V,z}\})| = 1))$$

The *transform-independent*, $A^{\mathrm{X}(T)}$, is called the *independent analogue* of the *iso-transform-independents*.

The *transform-independent* is sometimes not computable. The *likelihood function* of the sum of the *multinomial probabilities* is a polynomial, so the roots of the derivative are algebraic rather than rational. The finite approximation to the algebraic case for the *transform-independent* for some $k \in \mathbf{N}_{>0}$ is

$$\{A_k^{\mathrm{X}(T)}\} = \\ \mathrm{maxd}(\{(D/Z_k, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A))) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, where $T_{\mathrm{f}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$, the *iso-transform-independents* equals the *iso-independents*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{f}},z}(((A^{\mathrm{X}} * T_{\mathrm{f}}),(A * T_{\mathrm{f}})^{\mathrm{X}})) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and the *transform independent* equals the *independent*, $A^{\mathrm{X}(T_{\mathrm{f}})} = A^{\mathrm{X}}$.

At the other extreme where the *transform* is *unary*, $T = T_{\mathrm{u}}$, where $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$, the *iso-transform-independents* equals the *substrate histograms*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{u}},z}(((A^{\mathrm{X}} * T_{\mathrm{u}}),(A * T_{\mathrm{u}})^{\mathrm{X}})) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *transform independent* equals the *scaled normalised cartesian*, $A^{\mathrm{X}(T_{\mathrm{u}})} = Z_A * \hat{V}^{\mathrm{C}}$.

It is only in the case where the *formal* of the *transform-independent* equals the *formal*, $A^{\mathrm{X}(T)\mathrm{X}} * T = A^{\mathrm{X}} * T$ and the *abstract* of the *transform-independent* equals the *abstract*, $(A^{\mathrm{X}(T)} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$ that the *transform-independent* is in the *iso-transform-independents*,

$$(A^{\mathrm{X}(T)\mathrm{X}} * T = A^{\mathrm{X}} * T) \wedge ((A^{\mathrm{X}(T)} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}) \\ \iff A^{\mathrm{X}(T)} \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))$$

This is the case if the *transform* is *full functional*, $A^{\mathrm{X}(T_{\mathrm{f}})} = A^{\mathrm{X}}$, or *unary*, $A^{\mathrm{X}(T_{\mathrm{u}})} = Z_A * \hat{V}^{\mathrm{C}}$.

The *integral iso-transform-independents* have the same *transform-independent*,

$$\forall B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))\ (B^{\mathrm{X}(T)} = A^{\mathrm{X}(T)})$$

so the relation is functional

$$\{((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ T \in \mathcal{T}_{U,V}\} :\to \{A^{\mathrm{X}(T)} : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ T \in \mathcal{T}_{U,V}\}$$

Conjecture that, in the case where the *independent* is not *cartesian*, the relation is strictly iso-morphic,

$$\begin{aligned}\{((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ \hat{A}^{\mathrm{X}} \neq \hat{V}^{\mathrm{C}},\ T \in \mathcal{T}_{U,V}\} :\leftrightarrow: \\ \{A^{\mathrm{X}(T)} : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ \hat{A}^{\mathrm{X}} \neq \hat{V}^{\mathrm{C}},\ T \in \mathcal{T}_{U,V}\}\end{aligned}$$

While the *independent* is in the *iso-independents*, $A^{\mathrm{X}} \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, it is only in the *iso-transform-independents* if the *formal independent* equals the *abstract*,

$$(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$$

The *transform-independent* is only in the *iso-independents* if its *independent* equals the *independent*,

$$A^{\mathrm{X}(T)\mathrm{X}} = A^{\mathrm{X}} \implies A^{\mathrm{X}(T)} \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})$$

The degree to which the *iso-transform-independents* is said to be *aligned-like* is the *iso-independence*,

$$\frac{|\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} = \frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

As the *iso-independence* increases, the *transform-independent*, $A^{\mathrm{X}(T)}$, depends less on the *transform*, T, and tends to the *independent*, A^{X}.

The *lifted transform-independent* $A^{\mathrm{X}(T)'} \in \mathcal{A}_{U,W,z}$ is defined

$$\begin{aligned}&\{A^{\mathrm{X}(T)'}\} = \\ &\quad \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B') : B \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A))) : D \in \mathcal{A}_{U,W,z}\})\end{aligned}$$

where the *lifted integral iso-transform-independents* is abbreviated

$$\begin{aligned}&\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A) \\ &\quad = \{B * T : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\} \\ &\quad = \{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\} \\ &\quad = \{B * T : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}\end{aligned}$$

The *derived iso-independence* of the *integral lifted iso-transform-independents* is

$$\frac{|\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

As the *derived iso-independence* increases, the *lifted transform-independent*, $A^{\mathrm{X}(T)'}$, tends to the *abstract*, $(A * T)^{\mathrm{X}}$.

Conjecture that the *maximum likelihood estimate* for the *integral iso-formals* is the *naturalised formal*, $A^{\mathrm{X}} * T * T^{\dagger}$,

$$\begin{aligned}&\{A^{\mathrm{X}} * T * T^{\dagger}\} = \\ &\quad \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T))) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

where the *integral iso-set* does not equal the *integral substrate histograms*, $Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \neq \mathcal{A}_{U,\mathrm{i},V,z}$.

The *naturalised formal*, $A^{\mathrm{X}} * T * T^{\dagger}$, is the *independent analogue* of the *iso-formals*.

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *iso-formals* equals the *iso-independents*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{f}},\mathrm{V},z}(A^{\mathrm{X}} * T_{\mathrm{f}}) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and the *naturalised formal* equals the *independent*, $A^{\mathrm{X}} * T_{\mathrm{f}} * T_{\mathrm{f}}^{\dagger} = A^{\mathrm{X}}$. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *iso-formals* equals the *substrate histograms*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{u}},\mathrm{V},z}(A^{\mathrm{X}} * T_{\mathrm{u}}) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *naturalised formal* equals the *scaled normalised cartesian*, $A^{\mathrm{X}} * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger} = Z_A * \hat{V}^{\mathrm{C}}$.

The *naturalised formal* is not necessarily in the *iso-formals*,

$$(A^{\mathrm{X}} * T * T^{\dagger})^{\mathrm{X}} * T = A^{\mathrm{X}} * T \iff A^{\mathrm{X}} * T * T^{\dagger} \in Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T)$$

The *naturalised formal* is an *iso-formal* if the *transform* is *full functional*, $A^{\mathrm{X}} * T_{\mathrm{f}} * T_{\mathrm{f}}^{\dagger} = A^{\mathrm{X}}$, or *unary*, $A^{\mathrm{X}} * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger} = Z_A * \hat{V}^{\mathrm{C}}$.

Similarly, conjecture that the *maximum likelihood estimate* for the *integral iso-abstracts* is the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$,

$$\begin{aligned}&\{(A * T)^{\mathrm{X}} * T^{\dagger}\} = \\ &\quad \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}))) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

where the *integral iso-set* does not equal the *integral substrate histograms*, $Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \neq \mathcal{A}_{U,\mathrm{i},V,z}$.

The *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, is the *independent analogue* of the *iso-abstracts*.

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *iso-abstracts* equals the *iso-independents*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{f}},\mathrm{W},z}((A*T_{\mathrm{f}})^{\mathrm{X}}) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and the *naturalised abstract* equals the *independent*, $(A*T_{\mathrm{f}})^{\mathrm{X}}*T_{\mathrm{f}}^{\dagger} = A^{\mathrm{X}}$. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *iso-abstracts* equals the *substrate histograms*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{u}},\mathrm{W},z}((A * T_{\mathrm{u}})^{\mathrm{X}}) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *naturalised abstract* equals the *scaled normalised cartesian*, $(A * T_{\mathrm{u}})^{\mathrm{X}} * T_{\mathrm{u}}^{\dagger} = Z_A * \hat{V}^{\mathrm{C}}$.

The *naturalised abstract* is in the *iso-abstracts*, $(A * T)^{\mathrm{X}} * T^{\dagger} \in Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$, because $(((A * T)^{\mathrm{X}} * T^{\dagger}) * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$.

In the case where the *formal* equals the *abstract*, *independent analogue* of the *iso-formals* equals the *independent analogue* of the *iso-abstracts*,

$$A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} * T * T^{\dagger} = (A * T)^{\mathrm{X}} * T^{\dagger}$$

The *iso-transform-independents* is the intersection of the *iso-formals* and the *iso-abstracts*, so conjecture that, in this case, the *independent analogue* of the *iso-transform-independents* equals that of the *iso-formals* and the *iso-abstracts*,

$$A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}(T)} = A^{\mathrm{X}} * T * T^{\dagger} = (A * T)^{\mathrm{X}} * T^{\dagger}$$

In this case the *transform-independent* is *formal*, $\mathrm{formal}(A^{\mathrm{X}(T)}, T)$,

$$A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}(T)} * T = A^{\mathrm{X}} * T$$

and *abstract*, $\mathrm{abstract}(A^{\mathrm{X}(T)}, T)$,

$$A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}(T)} * T = (A * T)^{\mathrm{X}} = (A^{\mathrm{X}(T)} * T)^{\mathrm{X}}$$

The *transform-independent*, $A^{\mathrm{X}(T)}$, is the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the subset of the *substrate histograms* which are such that the *formal* equals the *histogram formal* and the *abstract* equals the *histogram abstract*,

$$\{A^{\mathrm{X}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A))) : D \in \mathcal{A}_{U,V,z}\})$$

Now consider the case where the membership of the *iso-transform-independents* is a given. Define the *transform-dependent* $A^{\mathrm{Y}(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-transform-independent*,

$$\{A^{\mathrm{Y}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *transform-dependent*, $A^{\mathrm{Y}(T)}$, is the *dependent analogue* of the *iso-transform-independents*.

The *transform-dependent histogram*, $A^{\mathrm{Y}(T)}$, is only defined if there is a unique maximum,

$$|\mathrm{max}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}) : D \in \mathcal{A}_{U,V,z}\})| = 1$$

The *transform-dependent histogram* equals the *histogram* if the *histogram* equals the *transform-independent histogram*, $A = A^{\mathrm{X}(T)} \implies A^{\mathrm{Y}(T)} = A = A^{\mathrm{X}(T)}$.

The *transform-dependent*, $A^{\mathrm{Y}(T)}$, is sometimes not computable. The finite approximation to the *transform-dependent* is

$$\{A_k^{\mathrm{Y}(T)}\} =$$
$$\mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *iso-transform-independents* equals the *iso-independents*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{f}},z}(((A^{\mathrm{X}} * T_{\mathrm{f}}), (A * T_{\mathrm{f}})^{\mathrm{X}})) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and the *transform-dependent* equals the *dependent*, $A^{\mathrm{Y}(T_{\mathrm{f}})} = A^{\mathrm{Y}}$. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *iso-transform-independents* equals the *substrate histograms*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{u}},z}(((A^{\mathrm{X}} * T_{\mathrm{u}}), (A * T_{\mathrm{u}})^{\mathrm{X}})) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *transform-dependent* equals the *histogram*, $A^{\mathrm{Y}(T_{\mathrm{u}})} = A$.

The *maximum likelihood estimate* for the numerator alone is the *histogram*, A, and the *maximum likelihood estimate* for the denominator alone is the *transform-independent*, $A^{\mathrm{X}(T)}$, so the overall *maximum likelihood estimate*, which is the *transform-dependent*, is near the *histogram*, $A^{\mathrm{Y}(T)} \sim A$, only in as much as it is far from the *transform-independent*, $A^{\mathrm{Y}(T)} \nsim A^{\mathrm{X}(T)}$.

As the *iso-independence*,

$$\frac{|\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) \ \cap \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) \ \cup \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} = \frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \ \cap \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \ \cup \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

increases, the *transform-dependent*, $A^{\mathrm{Y}(T)}$, depends less on the *transform*, T, and tends to the *dependent*, A^{Y}.

The *maximum likelihood estimate* for the *integral iso-abstracts* is conjectured above to be the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$,

$$\{(A * T)^{\mathrm{X}} * T^{\dagger}\} =$$
$$\mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}))) : D \in \mathcal{A}_{U,V,z}\})$$

The *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, is the *independent analogue* of the *iso-abstracts*.

The *lifted abstract-independent* $A^{\mathrm{U}(T)'} \in \mathcal{A}_{U,W,z}$ is defined

$$\{A^{\mathrm{U}(T)'}\} = \\ \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B') : B' \in \mathrm{isowl}(U)(T,A))) : D \in \mathcal{A}_{U,W,z}\})$$

where the *lifted integral iso-abstracts* is abbreviated

$$\mathrm{isowl}(U)(T,A) := \{B * T : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\}$$

The *lifted iso-abstracts* is a subset of the *derived iso-independents*, $\{B * T : B \in Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})\} \subseteq Y^{-1}_{U,W,z}((A * T)^{\mathrm{X}})$, so the *derived iso-independence* of the *integral lifted iso-abstracts* is

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

As the *derived iso-independence* increases, the *lifted abstract-independent*, $A^{\mathrm{U}(T)'}$, tends to the *abstract*, $(A * T)^{\mathrm{X}}$.

Now consider the case where the membership of the *iso-abstracts* is a given. Define the *abstract-dependent* $A^{\mathrm{W}(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-abstract*,

$$\{A^{\mathrm{W}(T)}\} = \\ \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})}) : D \in \mathcal{A}_{U,V,z}\})$$

The *abstract-dependent*, $A^{\mathrm{W}(T)}$, is the *dependent analogue* of the *iso-abstracts*.

The *abstract-dependent histogram*, $A^{\mathrm{W}(T)}$, is only defined if there is a unique maximum,

$$|\max(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})}) : D \in \mathcal{A}_{U,V,z}\})| = 1$$

The *abstract-dependent histogram* equals the *histogram* if the *histogram* equals the *naturalised abstract*, $A = (A * T)^{\mathrm{X}} * T^{\dagger} \implies A^{\mathrm{W}(T)} = A = (A * T)^{\mathrm{X}} * T^{\dagger}$.

The *abstract-dependent*, $A^{\mathrm{W}(T)}$, is sometimes not computable. The finite approximation to the *abstract-dependent* is

$$\{A_k^{\mathrm{W}(T)}\} = \\ \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *iso-abstracts* equals the *iso-independents*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{f}},\mathrm{W},z}((A * T_{\mathrm{f}})^{\mathrm{X}}) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and the *abstract-dependent* equals the *dependent*, $A^{\mathrm{W}(T_{\mathrm{f}})} = A^{\mathrm{Y}}$. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *iso-abstracts* equals the *substrate histograms*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{u}},\mathrm{W},z}((A * T_{\mathrm{u}})^{\mathrm{X}}) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *abstract-dependent* equals the *histogram*, $A^{\mathrm{W}(T_{\mathrm{u}})} = A$.

The *maximum likelihood estimate* for the numerator alone is the *histogram*, A, and the *maximum likelihood estimate* for the denominator alone is the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, so the overall *maximum likelihood estimate*, which is the *abstract-dependent*, is near the *histogram*, $A^{\mathrm{W}(T)} \sim A$, only in as much as it is far from the *naturalised abstract*, $A^{\mathrm{W}(T)} \nsim (A * T)^{\mathrm{X}} * T^{\dagger}$.

Similar to the case of the *iso-transform-independents*, above, consider the case of the *iso-partition-independents*. Given a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ having *underlying variables* equal to the *variables* V of the *substrate histogram*, the *partition-independent* $A^{\mathrm{P}(T)} \in \mathcal{A}_{U,V,z}$ is defined as the *maximum likelihood estimate* for the *distribution histogram* of the sum of the *generalised multinomial probabilities* of the *integral iso-partition-independents* of the *histogram*, A,

$$\{A^{\mathrm{P}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A))) : D \in \mathcal{A}_{U,V,z}\})$$

where the *integral iso-partition-independents* is abbreviated

$$\mathrm{isop}(U)(T,A) \ := \ Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

and the *iso-partition-independents* is such that

$$Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \\ = \ \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ (B^{\mathrm{X}} * T)^{\mathrm{X}} = (A^{\mathrm{X}} * T)^{\mathrm{X}},\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}$$

In the case where the *transform* is a *substrate transform*, $T \in \mathcal{T}_{U,V}$,

$$Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \\ = \ \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ \forall P \in W\ (B^{\mathrm{X}} * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}} \ \wedge \ B * P^{\mathrm{T}} = A * P^{\mathrm{T}})\}$$

In the case where the *integral iso-partition-independents* equals the *integral substrate histograms*, $\mathrm{isop}(U)(T,A) = \mathcal{A}_{U,\mathrm{i},V,z}$, there is no unique maximum, and the *partition-independent* is defined as the *scaled normalised cartesian*,

$$\mathrm{isop}(U)(T,A) = \mathcal{A}_{U,\mathrm{i},V,z} \implies A^{\mathrm{P}(T)} := Z_A * \hat{V}^{\mathrm{C}}$$

where $Z_A = \mathrm{scalar}(\mathrm{size}(A))$ and $\hat{X} := \mathrm{normalise}(X)$. Otherwise, this definition assumes that there is always a unique maximum,

$$\begin{aligned}&\forall A \in \mathcal{A}_{U,\mathrm{i},V,z}\ \forall T \in \mathcal{T}_{U,V}\ (\mathrm{isop}(U)(T,A) \neq \mathcal{A}_{U,\mathrm{i},V,z} \implies \\ &\quad (|\max(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A))) : D \in \mathcal{A}_{U,V,z}\})| = 1))\end{aligned}$$

The *partition-independent*, $A^{\mathrm{P}(T)}$, is called the *independent analogue* of the *iso-partition-independents*.

The *partition-independent* is sometimes not computable. The finite approximation for the *partition-independent* for some $k \in \mathbf{N}_{>0}$ is

$$\begin{aligned}&\{A^{\mathrm{P}(T)}_k\} = \\ &\quad \mathrm{maxd}(\{(D/Z_k, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A))) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})\end{aligned}$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, where $T_{\mathrm{f}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$, the *iso-partition-independents* equals the *iso-independents*, $\mathrm{isop}(U)(T_{\mathrm{f}},A) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and the *partition-independent* equals the *independent*, $A^{\mathrm{P}(T_{\mathrm{f}})} = A^{\mathrm{X}}$.

At the other extreme where the *transform* is *unary*, $T = T_{\mathrm{u}}$, where $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$, the *iso-partition-independents* equals the *substrate histograms*, $\mathrm{isop}(U)(T_{\mathrm{u}},A) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *partition-independent* equals the *scaled normalised cartesian*, $A^{\mathrm{P}(T_{\mathrm{u}})} = Z_A * \hat{V}^{\mathrm{C}}$.

It is only in the case where the *formal independent* of the *partition-independent* equals the *formal independent*, $(A^{\mathrm{P}(T)\mathrm{X}} * T)^{\mathrm{X}} = (A^{\mathrm{X}} * T)^{\mathrm{X}}$ and the *abstract* of the *partition-independent* equals the *abstract*, $(A^{\mathrm{P}(T)} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$ that the *partition-independent* is in the *iso-partition-independents*,

$$\begin{aligned}&((A^{\mathrm{P}(T)\mathrm{X}} * T)^{\mathrm{X}} = (A^{\mathrm{X}} * T)^{\mathrm{X}}) \wedge ((A^{\mathrm{P}(T)} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}) \\ &\quad \iff A^{\mathrm{X}(T)} \in Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})\end{aligned}$$

This is the case if the *transform* is *full functional*, $A^{\mathrm{P}(T_{\mathrm{f}})} = A^{\mathrm{X}}$, or *unary*, $A^{\mathrm{P}(T_{\mathrm{u}})} = Z_A * \hat{V}^{\mathrm{C}}$.

The *integral iso-partition-independents* have the same *partition-independent*,

$$\forall B \in \mathrm{isop}(U)(T,A)\ (B^{\mathrm{P}(T)} = A^{\mathrm{P}(T)})$$

so the relation is functional

$$\{((A^{\mathrm{X}}*T)^{\mathrm{X}}, (A*T)^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ T \in \mathcal{T}_{U,V}\} :\to \{A^{\mathrm{P}(T)} : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ T \in \mathcal{T}_{U,V}\}$$

Conjecture that, in the case where the *independent* is not *cartesian*, the relation is strictly iso-morphic,

$$\begin{aligned}\{((A^{\mathrm{X}} * T)^{\mathrm{X}}, (A * T)^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ \hat{A}^{\mathrm{X}} \neq \hat{V}^{\mathrm{C}},\ T \in \mathcal{T}_{U,V}\} :\leftrightarrow: \\ \{A^{\mathrm{P}(T)} : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ \hat{A}^{\mathrm{X}} \neq \hat{V}^{\mathrm{C}},\ T \in \mathcal{T}_{U,V}\}\end{aligned}$$

While the *independent* is in the *iso-independents*, $A^{\mathrm{X}} \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, it is only in the *iso-partition-independents* if the *formal independent* equals the *abstract*,

$$(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

The *partition-independent* is only in the *iso-independents* if its *independent* equals the *independent*,

$$A^{\mathrm{P}(T)\mathrm{X}} = A^{\mathrm{X}} \implies A^{\mathrm{P}(T)} \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})$$

Conjecture that the *maximum likelihood estimate* for the *integral iso-formal-independents* is the *naturalised formal independent*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\dagger}$,

$$\begin{aligned}&\{(A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\dagger}\} = \\ &\quad \mathrm{maxd}(\{(D, \textstyle\sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}))) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

where the *integral iso-set* does not equal the *integral substrate histograms*, $Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \neq \mathcal{A}_{U,\mathrm{i},V,z}$.

The *naturalised formal independent*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\dagger}$, is the *independent analogue* of the *iso-formal-independents*.

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *iso-formal-independents* equals the *iso-independents*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{f}},\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}}*T_{\mathrm{f}})^{\mathrm{X}}) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and the *naturalised formal independent* equals the *independent*, $(A^{\mathrm{X}} * T_{\mathrm{f}})^{\mathrm{X}} * T_{\mathrm{f}}^{\dagger} = A^{\mathrm{X}}$. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *iso-formal-independents* equals the *substrate histograms*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{u}},\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}}*T_{\mathrm{u}})^{\mathrm{X}}) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *naturalised formal independent* equals the *scaled normalised cartesian*, $(A^{\mathrm{X}} * T_{\mathrm{u}})^{\mathrm{X}} * T_{\mathrm{u}}^{\dagger} = Z_A * \hat{V}^{\mathrm{C}}$.

The *naturalised formal independent* is not necessarily in the *iso-formals*,

$$(((A^{\mathrm{X}}*T)^{\mathrm{X}}*T^{\dagger})^{\mathrm{X}}*T)^{\mathrm{X}} = (A^{\mathrm{X}}*T)^{\mathrm{X}} \iff (A^{\mathrm{X}}*T)^{\mathrm{X}}*T^{\dagger} \in Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}}*T)^{\mathrm{X}})$$

The *naturalised formal independent* is an *iso-formal-independent* if the *transform* is *full functional*, $(A^{\mathrm{X}} * T_{\mathrm{f}})^{\mathrm{X}} * T_{\mathrm{f}}^{\dagger} = A^{\mathrm{X}}$, or *unary*, $(A^{\mathrm{X}} * T_{\mathrm{u}})^{\mathrm{X}} * T_{\mathrm{u}}^{\dagger} = Z_A * \hat{V}^{\mathrm{C}}$.

In the case where the *formal independent* equals the *abstract*, *independent analogue* of the *iso-formal-independents* equals the *independent analogue* of the *iso-abstracts*,

$$(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies (A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\dagger} = (A * T)^{\mathrm{X}} * T^{\dagger}$$

The *iso-partition-independents* is the intersection of the *iso-formal independents* and the *iso-abstracts*, so conjecture that, in this case, the *independent analogue* of the *iso-partition-independents* equals that of the *iso-formal independents* and the *iso-abstracts*,

$$(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies A^{\mathrm{P}(T)} = (A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\dagger} = (A * T)^{\mathrm{X}} * T^{\dagger}$$

The *partition-independent*, $A^{\mathrm{P}(T)}$, is the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the subset of the *substrate histograms* which are such that the *formal independent* equals the *histogram formal independent* and the *abstract* equals the *histogram abstract*,

$$\{A^{\mathrm{P}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A))) : D \in \mathcal{A}_{U,V,z}\})$$

Now consider the case where the membership of the *iso-partition-independents* is a given. Define the *partition-dependent* $A^{\mathrm{R}(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-partition-independent*,

$$\{A^{\mathrm{R}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *partition-dependent*, $A^{\mathrm{R}(T)}$, is the *dependent analogue* of the *iso-partition-independents*.

The *partition-dependent histogram*, $A^{\mathrm{R}(T)}$, is only defined if there is a unique maximum,

$$|\mathrm{max}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A)}) : D \in \mathcal{A}_{U,V,z}\})| = 1$$

The *partition-dependent histogram* equals the *histogram* if the *histogram* equals the *partition-independent histogram*, $A = A^{\mathrm{P}(T)} \implies A^{\mathrm{R}(T)} = A = A^{\mathrm{P}(T)}$.

The *partition-dependent*, $A^{\mathrm{R}(T)}$, is sometimes not computable. The finite approximation to the *partition-dependent* is

$$\begin{aligned}&\{A_k^{\mathrm{R}(T)}\} = \\ &\quad \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A)}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})\end{aligned}$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *iso-partition-independents* equals the *iso-independents*, $\mathrm{isop}(U)(T_{\mathrm{f}}, A) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and the *partition-dependent* equals the *dependent*, $A^{\mathrm{R}(T_{\mathrm{f}})} = A^{\mathrm{Y}}$. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *iso-partition-independents* equals the *substrate histograms*, $\mathrm{isop}(U)(T_{\mathrm{u}}, A) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *partition-dependent* equals the *histogram*, $A^{\mathrm{R}(T_{\mathrm{u}})} = A$.

The *maximum likelihood estimate* for the numerator alone is the *histogram*, A, and the *maximum likelihood estimate* for the denominator alone is the *partition-independent*, $A^{\mathrm{P}(T)}$, so the overall *maximum likelihood estimate*, which is the *partition-dependent*, is near the *histogram*, $A^{\mathrm{R}(T)} \sim A$, only in as much as it is far from the *partition-independent*, $A^{\mathrm{R}(T)} \nsim A^{\mathrm{P}(T)}$.

The *maximum likelihood estimate* for the *integral iso-independents* of *histogram A* is conjectured above to be the *independent*, A^{X},

$$\{A^{\mathrm{X}}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))) : D \in \mathcal{A}_{U,V,z}\})$$

Similarly, conjecture that, given *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$, the *maximum likelihood estimate* for the *integral iso-deriveds* is the *naturalisation*, $A * T * T^{\dagger}$,

$$\begin{aligned}&\{A * T * T^{\dagger}\} = \\ &\quad \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T))) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

where the *integral iso-set* does not equal the *integral substrate histograms*, $D^{-1}_{U,\mathrm{i},T,z}(A * T) \neq \mathcal{A}_{U,\mathrm{i},V,z}$.

The *naturalisation*, $A*T*T^{\dagger}$, is the *independent analogue* of the *iso-deriveds*.

The *derived* of the *naturalisation* equals the *derived*, $(A*T*T^{\dagger})*T = A*T$, and the *components* of the *naturalisation* are *uniform*, $\forall C \in T^{\mathrm{P}}(((A * T * T^{\dagger})*C^{\mathrm{U}})^{\wedge} = (V^{\mathrm{C}}*C^{\mathrm{U}})^{\wedge})$. The *naturalisation* is a member of the *iso-deriveds*, $A * T * T^{\dagger} \in D^{-1}_{U,T,z}(A * T)$. If the *naturalisation* is *integral* it is a member of the *integral iso-deriveds*,

$$A * T * T^{\dagger} \in \mathcal{A}_{\mathrm{i}} \implies A * T * T^{\dagger} \in D^{-1}_{U,\mathrm{i},T,z}(A * T)$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *integral iso-deriveds* is a singleton, $D^{-1}_{U,\mathrm{i},T_{\mathrm{f}},z}(A * T_{\mathrm{f}}) = \{A\}$, and the *naturalisation* equals the *histogram*, $A * T_{\mathrm{f}} * T_{\mathrm{f}}^{\dagger} = A$. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *integral iso-deriveds* equals the *substrate histograms*, $D^{-1}_{U,\mathrm{i},T_{\mathrm{u}},z}(A* T_{\mathrm{u}}) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *naturalisation* equals the *scaled normalised cartesian*, $A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger} = Z_A * \hat{V}^{\mathrm{C}}$.

Now consider the case where the membership of the *iso-deriveds* is a given. Define the *derived-dependent* $A^{\mathrm{D}(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-derived*,

$$\{A^{\mathrm{D}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D, z)(A)}{\sum Q_{\mathrm{m},U}(D, z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *derived-dependent*, $A^{\mathrm{D}(T)}$, is the *dependent analogue* of the *iso-deriveds*.

The *derived-dependent histogram*, $A^{\mathrm{D}(T)}$, is only defined if there is a unique maximum,

$$|\mathrm{max}(\{(D, \frac{Q_{\mathrm{m},U}(D, z)(A)}{\sum Q_{\mathrm{m},U}(D, z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}) : D \in \mathcal{A}_{U,V,z}\})| = 1$$

The *derived-dependent histogram* equals the *histogram* if the *histogram* equals the *naturalisation histogram*, $A = A * T * T^{\dagger} \implies A^{\mathrm{D}(T)} = A = A * T * T^{\dagger}$.

The *derived-dependent*, $A^{\mathrm{D}(T)}$, is sometimes not computable. The finite approximation to the *derived-dependent* is

$$\{A_k^{\mathrm{D}(T)}\} = \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *integral iso-deriveds* is a singleton, $D^{-1}_{U,\mathrm{i},T_{\mathrm{f}},z}(A * T_{\mathrm{f}}) = \{A\}$, and the *derived-dependent* is undefined. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *integral iso-deriveds* equals the *substrate histograms*, $D^{-1}_{U,\mathrm{i},T_{\mathrm{u}},z}(A * T_{\mathrm{u}}) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the *derived-dependent* equals the *histogram*, $A^{\mathrm{D}(T_{\mathrm{u}})} = A$.

The *maximum likelihood estimate* for the numerator alone is the *histogram*, A, and the *maximum likelihood estimate* for the denominator alone is the *naturalisation*, $A * T * T^{\dagger}$, so the overall *maximum likelihood estimate*, which is the *derived-dependent*, is near the *histogram*, $A^{\mathrm{D}(T)} \sim A$, only in as much as it is far from the *naturalisation*, $A^{\mathrm{D}(T)} \nsim A * T * T^{\dagger}$.

Now consider the case where the *model* is extended from *transforms* to (i) *fuds*, (ii) *decompositions* and (iii) *fud decompositions*.

Let F be a *one functional definition set*, $F \in \mathcal{F}_{U,1}$, such that the *underlying* are a subset of the *histogram variables*, $\mathrm{und}(F) \subseteq V$, and there exists a *top transform*, $\exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$. The *derived* set valued function of the *substrate histograms* is

$$D_{U,F,z} = \{(A, \{A * T_F : T \in F\}) : A \in \mathcal{A}_{U,V,z}\}$$

where $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$. In this case the *top transform* exists so the set of *iso-fuds* is a subset of the *iso-deriveds*,

$$D^{-1}_{U,F,z}(\{A * T_F : T \in F\}) \subseteq D^{-1}_{U,F^{\mathrm{T}},z}(A * F^{\mathrm{T}})$$

Define the *fud-independent* $A^{\mathrm{E_F}(F)} \in \mathcal{A}_{U,V,z}$, as the *maximum likelihood estimate* for the *distribution histogram* of the sum of the *generalised multinomial probabilities* of the *integral iso-fuds* of the *histogram*, A,

$$\{A^{\mathrm{E_F}(F)}\} = \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},F,z}(D_{U,F,z}(A)))) : E \in \mathcal{A}_{U,V,z}\})$$

The *fud-independent*, $A^{\mathrm{E_F}(F)}$, is the *independent analogue* of the *iso-fuds*.

Like the *transform-independent*, $A^{\mathrm{X}(T)}$, the *fud-independent*, $A^{\mathrm{E_F}(F)}$, is sometimes not computable. It approximates, however, to the arithmetic *average* of the *naturalisations*,

$$A^{\mathrm{E_F}(F)} \quad \approx \quad Z_{1/|F|} * \sum_{T \in F} A * T_F * T_F^{\dagger}$$

Define the *fud-dependent* $A^{\mathrm{D_F}(F)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-fud*,

$$\{A^{\mathrm{D_F}(F)}\} = \\ \mathrm{maxd}(\{(E, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum Q_{\mathrm{m},U}(E,z)(B) : B \in D_{U,\mathrm{i},F,z}^{-1}(D_{U,F,z}(A))}) : E \in \mathcal{A}_{U,V,z}\})$$

The *fud-dependent*, $A^{\mathrm{D_F}(F)}$, is the *dependent analogue* of the set of the *iso-fuds*.

The *fud-dependent* equals the *histogram* only if the *histogram* equals the *fud-independent*,

$$A = A^{\mathrm{E_F}(F)} \implies A^{\mathrm{D_F}(F)} = A$$

The *maximum likelihood estimate* is near the *histogram*, $A^{\mathrm{D_F}(F)} \sim A$, only in as much as it is far from the *fud-independent*, $A^{\mathrm{D_F}(F)} \nsim A^{\mathrm{E_F}(F)}$.

Let D be a *decomposition* of *one functional transforms*, $D \in \mathcal{D}_U = \mathcal{D} \cap \mathrm{trees}(\mathcal{S}_U \times \mathcal{T}_{U,\mathrm{f},1})$, such that the *underlying* are a subset of the *histogram variables*, $\mathrm{und}(D) \subseteq V$. The *component-derived* relation valued function of the *substrate histograms* is

$$D_{U,D,z} = \{(A, \{(C, A * C * T) : (C,T) \in \mathrm{cont}(D)\}) : A \in \mathcal{A}_{U,V,z}\}$$

where $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$.

The set of *iso-decompositions* is a subset of the *iso-deriveds* of the *transform* of the *decomposition*,

$$D_{U,D,z}^{-1}(D_{U,D,z}(A)) \quad \subseteq \quad D_{U,D^{\mathrm{T}},z}^{-1}(A * D^{\mathrm{T}})$$

Define the *decomposition-independent* $A^{\mathrm{E_D}(D)} \in \mathcal{A}_{U,V,z}$, as the *maximum likelihood estimate* for the *distribution histogram* of the sum of the *generalised* *multinomial probabilities* of the *integral iso-decompositions* of the *histogram*, A,

$$\{A^{\mathrm{E_D}(D)}\} = \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},D,z}(D_{U,D,z}(A)))) : E \in \mathcal{A}_{U,V,z}\})$$

The *decomposition-independent*, $A^{\mathrm{E_D}(D)}$, is the *independent analogue* of the *iso-decompositions*.

The *decomposition-independent*, $A^{\mathrm{E_D}(D)}$, is sometimes not computable. It approximates, however, to the *scaled sum* of the *slice naturalisations*,

$$A^{\mathrm{E_D}(D)} \approx Z_z * \left(\sum_{(C,T)\in\mathrm{cont}(D)} A * C * T * T^{\dagger}\right)^{\wedge}$$

where $()^{\wedge}$ = normalise.

Define the *decomposition-dependent* $A^{\mathrm{D_D}(D)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-decomposition*,

$$\{A^{\mathrm{D_D}(D)}\} = \mathrm{maxd}(\{(E, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},D,z}(D_{U,D,z}(A))}) : E \in \mathcal{A}_{U,V,z}\})$$

The *decomposition-dependent*, $A^{\mathrm{D_D}(D)}$, is the *dependent analogue* of the set of the *iso-decompositions*.

The *decomposition-dependent* equals the *histogram* only if the *histogram* equals the *decomposition-independent*,

$$A = A^{\mathrm{E_D}(D)} \implies A^{\mathrm{D_D}(D)} = A$$

The *maximum likelihood estimate* is near the *histogram*, $A^{\mathrm{D_D}(D)} \sim A$, only in as much as it is far from the *decomposition-independent*, $A^{\mathrm{D_D}(D)} \nsim A^{\mathrm{E_D}(D)}$.

Let D be a *fud decomposition* of *one functional definition sets*, $D \in \mathcal{D}_{\mathrm{F},U} = \mathcal{D}_{\mathrm{F}} \cap \mathrm{trees}(\mathcal{S}_U \times \mathcal{F}_{U,1})$, such that the *underlying* are a subset of the *histogram variables*, $\mathrm{und}(D) \subseteq V$, and the *top transform* exists for all of the *fuds*, $\forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$. The *component-derived-set* relation valued function of the *substrate histograms* is

$$D_{U,D,\mathrm{F},z} = \{(A, \{(C, \{A*C*T_F : T \in F\}) : (C,F) \in \mathrm{cont}(D)\}) : A \in \mathcal{A}_{U,V,z}\}$$

In this case the *top transforms* exist, so the set of *iso-fud-decompositions* is a subset of the *iso-deriveds* of the *transform* of the *decomposition*,

$$D^{-1}_{U,D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A)) \quad\subseteq\quad D^{-1}_{U,D^{\mathrm{T}},z}(A * D^{\mathrm{T}})$$

Define the *fud-decomposition-independent* $A^{\mathrm{E}_{\mathrm{D,F}}(D)} \in \mathcal{A}_{U,V,z}$, as the *maximum likelihood estimate* for the *distribution histogram* of the sum of the *generalised multinomial probabilities* of the *integral iso-fuds* of the *histogram*, A,

$$\begin{aligned}&\{A^{\mathrm{E}_{\mathrm{D,F}}(D)}\} \;=\\ &\quad \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A)))) : E \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *fud-decomposition-independent*, $A^{\mathrm{E}_{\mathrm{D,F}}(F)}$, is the *independent analogue* of the *iso-fuds*.

The *fud-decomposition-independent*, $A^{\mathrm{E}_{\mathrm{D,F}}(D)}$, is sometimes not computable. It approximates, however, to the *scaled sum* of the *slice* arithmetic *average* of the *naturalisations*,

$$A^{\mathrm{E}_{\mathrm{D,F}}(D)} \quad\approx\quad Z_z * \left(\sum_{(C,F)\in\mathrm{cont}(D)} \left(Z_{1/|F|} * \sum_{T\in F} A * C * T_F * T_F^{\dagger}\right)\right)^{\wedge}$$

Define the *fud-decomposition-dependent* $A^{\mathrm{D}_{\mathrm{D,F}}(D)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-fud-decomposition*,

$$\begin{aligned}&\{A^{\mathrm{D}_{\mathrm{D,F}}(D)}\} =\\ &\quad \mathrm{maxd}(\{(E, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A))}) : E \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *fud-decomposition-dependent*, $A^{\mathrm{D}_{\mathrm{D,F}}(D)}$, is the *dependent analogue* of the set of the *iso-fud-decompositions*.

The *fud-decomposition-dependent* equals the *histogram* only if the *histogram* equals the *fud-decomposition-independent*,

$$A = A^{\mathrm{E}_{\mathrm{D,F}}(D)} \implies A^{\mathrm{D}_{\mathrm{D,F}}(D)} = A$$

The *maximum likelihood estimate* is near the *histogram*, $A^{\mathrm{D}_{\mathrm{D,F}}(D)} \sim A$, only in as much as it is far from the *decomposition-independent*, $A^{\mathrm{D}_{\mathrm{D,F}}(D)} \nsim A^{\mathrm{E}_{\mathrm{D,F}}(D)}$.

Conjecture that the *maximum likelihood estimate* for the *integral iso-components* is the *unnaturalisation*, $V_z^{\mathrm{C}} * T * T^{\odot A}$,

$$\begin{aligned}\{V_z^{\mathrm{C}} * T * T^{\odot A}\} = \\ \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in C^{-1}_{U,\mathrm{i},T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}))) \\ : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

where the *integral iso-set* does not equal the *integral substrate histograms*, $C^{-1}_{U,\mathrm{i},T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}) \neq \mathcal{A}_{U,\mathrm{i},V,z}$.

The *unnaturalisation*, $V_z^{\mathrm{C}} * T * T^{\odot A}$, is the *independent analogue* of the *iso-components*.

Now consider the case where the membership of the *iso-components* is a given. Define the *components-dependent* $A^{\mathrm{C}(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-components*,

$$\begin{aligned}\{A^{\mathrm{C}(T)}\} = \\ \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in C^{-1}_{U,\mathrm{i},T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\})}) \\ : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *components-dependent*, $A^{\mathrm{C}(T)}$, is the *dependent analogue* of the *iso-components*.

The *maximum likelihood estimate* is near the *histogram*, $A^{\mathrm{C}(T)} \sim A$, only in as much as it is far from the *unnaturalisation*, $A^{\mathrm{C}(T)} \nsim V_z^{\mathrm{C}} * T * T^{\odot A}$.

The set of *iso-liftisations* is defined (in section 'Iso-sets', above) as the intersection of the *iso-formals* and *iso-deriveds*

$$Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap D^{-1}_{U,T,z}(A * T)$$

Define the *liftisation*, $A^{\mathrm{K}(T)} \in \mathcal{A}_{U,V,z}$, as the *maximum likelihood estimate* for the *distribution histogram* of the sum of the *generalised multinomial probabilities* of the *integral iso-liftisations* of the *histogram*, A,

$$\{A^{\mathrm{K}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isol}(U)(T,A))) : D \in \mathcal{A}_{U,V,z}\})$$

where $\mathrm{isol}(U)(T,A) := Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap D^{-1}_{U,\mathrm{i},T,z}(A * T)$, and the *integral iso-set* does not equal the *integral substrate histograms*, $\mathrm{isol}(U)(T,A) \neq \mathcal{A}_{U,\mathrm{i},V,z}$.

The *liftisation*, $A^{\mathrm{K}(T)}$, is the *independent analogue* of the *iso-liftisations*.

Like the *transform-independent*, $A^{\mathrm{X}(T)}$, the *liftisation*, $A^{\mathrm{K}(T)}$, is sometimes not computable.

As for the set of *iso-deriveds* and the corresponding *independent analogue*, the *naturalisation*, $A * T * T^{\dagger}$, in the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *liftisation* equals the *histogram*, $A^{\mathrm{K}(T_{\mathrm{f}})} = A$, and in the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *liftisation* equals the *scaled normalised cartesian*, $A^{\mathrm{K}(T_{\mathrm{u}})} = Z_A * \hat{V}^{\mathrm{C}}$. In the case where the *formal* of the *naturalisation* equals the *formal*, the *liftisation* equals the *naturalisation*, $(A * T * T^{\dagger})^{\mathrm{X}} * T = A^{\mathrm{X}} * T \implies A^{\mathrm{K}(T)} = A * T * T^{\dagger}$, otherwise the *liftisation* varies between the *naturalised formal*, $A^{\mathrm{K}(T)} \sim A^{\mathrm{X}} * T * T^{\dagger}$, and the *naturalisation*, $A^{\mathrm{K}(T)} \sim A * T * T^{\dagger}$.

It is only in the case where the *formal* of the *liftisation* equals the *formal*, $A^{\mathrm{K}(T)\mathrm{X}} * T = A^{\mathrm{X}} * T$ and the *derived* of the *liftisation* equals the *derived*, $A^{\mathrm{K}(T)} * T = A * T$, that the *liftisation* is in the *iso-liftisations*,

$$\begin{aligned}&(A^{\mathrm{K}(T)\mathrm{X}} * T = A^{\mathrm{X}} * T) \land (A^{\mathrm{K}(T)} * T = A * T) \\ &\qquad\qquad \iff A^{\mathrm{K}(T)} \in Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap D^{-1}_{U,T,z}(A * T)\end{aligned}$$

Consider the case where the membership of the *iso-liftisations* is a given. Define the *liftisation-dependent* $A^{\mathrm{L}(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-liftisation*,

$$\begin{aligned}&\{A^{\mathrm{L}(T)}\} = \\ &\quad \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isol}(U)(T,A)}) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *liftisation-dependent*, $A^{\mathrm{L}(T)}$, is the *dependent analogue* of the set of the *iso-liftisations*.

The *maximum likelihood estimate* is near the *histogram*, $A^{\mathrm{L}(T)} \sim A$, only in as much as it is far from the *liftisation*, $A^{\mathrm{L}(T)} \nsim A^{\mathrm{K}(T)}$.

The set of *iso-idealisations* is (i) the intersection of the *iso-component-independents* and the *iso-derived* which is (ii) a subset of the intersection of the *iso-independents* and *iso-deriveds* which is (iii) a subset of the *iso-liftisations* which is (iv) a subset of the *iso-transform-independents* which is (v) a subset of the *iso-partition-independents* which is (vi) a subset of the *iso-abstracts*,

$$\begin{aligned} Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) &= C^{-1}_{U,T,\mathrm{x},z}(\{(A * C^{\mathrm{U}})^{\mathrm{X}\wedge} : C \in T^{\mathrm{P}}\}) \cap D^{-1}_{U,T,z}(A * T) \\ &\subseteq Y^{-1}_{U,V,z}(A^{\mathrm{X}}) \cap D^{-1}_{U,T,z}(A * T) \\ &\subseteq Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap D^{-1}_{U,T,z}(A * T) \\ &\subseteq Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \\ &\subseteq Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \cap Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \\ &\subseteq Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \end{aligned}$$

Conjecture that, given *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$, the *maximum likelihood estimate* for the *integral iso-idealisations* is the *idealisation*, $A*T*T^{\dagger A}$,

$$\begin{aligned} &\{A * T * T^{\dagger A}\} = \\ &\quad \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isoi}(U)(T,A))) : D \in \mathcal{A}_{U,V,z}\}) \end{aligned}$$

where $\mathrm{isoi}(U)(T,A) := Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A*T*T^{\dagger A})$ and the *integral iso-set* does not equal the *integral substrate histograms*, $\mathrm{isoi}(U)(T,A) \neq \mathcal{A}_{U,\mathrm{i},V,z}$.

The *idealisation*, $A*T*T^{\dagger A}$, is the *independent analogue* of the *iso-idealisations*.

The *derived* of the *idealisation* equals the *derived*, $(A*T*T^{\dagger A})*T = A*T$, the *independent* of the *idealisation* equals the *independent*, $(A*T*T^{\dagger A})^{\mathrm{X}} = A^{\mathrm{X}}$, and the *components* of the *idealisation* are *independent*, $\forall C \in T^{\mathrm{P}}((A*T*T^{\dagger A}) * C^{\mathrm{U}} = ((A*T*T^{\dagger A}) * C^{\mathrm{U}})^{\mathrm{X}})$. If the *idealisation* is *integral* it is a member of the *integral iso-idealisations*,

$$A * T * T^{\dagger A} \in \mathcal{A}_{\mathrm{i}} \implies A * T * T^{\dagger A} \in Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})$$

In addition, it is a member of the *integral iso-independents*, *integral iso-deriveds*, and the *integral iso-transform-independents*,

$$\begin{aligned} &A * T * T^{\dagger A} \in \mathcal{A}_{\mathrm{i}} \implies \\ &\quad A * T * T^{\dagger A} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \cap D^{-1}_{U,\mathrm{i},T,z}(A * T) \cap \\ &\qquad\qquad Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \end{aligned}$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *integral iso-idealisations* is a singleton, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{f}},\dagger,z}(A*T_{\mathrm{f}}*T_{\mathrm{f}}^{\dagger A}) = \{A\}$, and the *idealisation* equals the *histogram*, $A * T_{\rm f} * T_{\rm f}^{\dagger A} = A$. In the case where the *transform* is *unary*, $T = T_{\rm u}$, the *integral iso-idealisations* equals the *iso-independents*, $Y^{-1}_{U,{\rm i},T_{\rm u},\dagger,z}(A * T_{\rm u} * T_{\rm u}^{\dagger A}) = Y^{-1}_{U,{\rm i},V,z}(A^{\rm X})$, and the *idealisation* equals the *independent*, $A * T_{\rm u} * T_{\rm u}^{\dagger A} = A^{\rm X}$.

Consider the case where the membership of the *iso-idealisations* is a given. Define the *idealisation-dependent* $A^{\dagger(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-idealisation*,

$$\{A^{\dagger(T)}\} = \\ \mathrm{maxd}(\{(D, \frac{Q_{{\rm m},U}(D,z)(A)}{\sum Q_{{\rm m},U}(D,z)(B) : B \in \mathrm{isoi}(U)(T,A)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *idealisation-dependent*, $A^{\dagger(T)}$, is the *dependent analogue* of the set of the *iso-idealisations*.

The *idealisation-dependent* equals the *histogram* if the *histogram* is *ideal*, $\mathrm{ideal}(A,T) \implies A^{\dagger(T)} = A = A * T * T^{\dagger A}$.

The *maximum likelihood estimate* is near the *histogram*, $A^{\dagger(T)} \sim A$, only in as much as it is far from the *idealisation*, $A^{\dagger(T)} \nsim A * T * T^{\dagger A}$.

The *derived-dependent* is intermediate between the *idealisation-dependent* and the *transform-dependent*, $A^{\dagger(T)} \sim A^{{\rm D}(T)} \sim A^{{\rm Y}(T)}$.

The set of *iso-surrealisations* is the intersection of the *iso-abstracts* and *iso-components*

$$Y^{-1}_{U,T,{\rm W},z}((A * T)^{\rm X}) \ \cap \ C^{-1}_{U,T,z}(\{(A * C^{\rm U})^{\wedge} : C \in T^{\rm P}\})$$

Conjecture that the *maximum likelihood estimate* for the *integral iso surrealisations* is the *surrealisation*, $(A * T)^{\rm X} * T^{\odot A}$,

$$\{(A * T)^{\rm X} * T^{\odot A}\} = \\ \mathrm{maxd}(\{(D, \sum (Q_{{\rm m},U}(D,z)(B) : B \in \mathrm{isor}(U)(T,A))) : D \in \mathcal{A}_{U,V,z}\})$$

where $\mathrm{isor}(U)(T,A) := Y^{-1}_{U,{\rm i},T,{\rm W},z}((A * T)^{\rm X}) \ \cap \ C^{-1}_{U,{\rm i},T,z}(\{(A * C^{\rm U})^{\wedge} : C \in T^{\rm P}\})$ and the *integral iso-set* does not equal the *integral substrate histograms*, $\mathrm{isor}(U)(T,A) \neq \mathcal{A}_{U,{\rm i},V,z}$.

The *surrealisation*, $(A * T)^{\mathrm{X}} * T^{\odot A}$, is the *independent analogue* of the *iso-surrealisations*.

The *surrealisation* is an *iso-surrealisation*, $(A * T)^{\mathrm{X}} * T^{\odot A} \in Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \cap C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\})$.

The *surrealisation* varies as the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\odot A} \sim (A * T)^{\mathrm{X}} * T^{\dagger}$, and the *unnaturalisation*, $(A * T)^{\mathrm{X}} * T^{\odot A} \sim V_z^{\mathrm{C}} * T * T^{\odot A}$.

Consider the case where the membership of the *iso-surrealisations* is a given. Define the *surrealisation-dependent* $A^{\odot(T)} \in \mathcal{A}_{U,V,z}$ as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-surrealisation*,

$$\begin{aligned}&\{A^{\odot(T)}\} = \\ &\quad \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isor}(U)(T,A)}) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *surrealisation-dependent*, $A^{\odot(T)}$, is the *dependent analogue* of the set of the *iso-surrealisations*.

The *surrealisation-dependent* equals the *histogram* if the *histogram* is *surreal*,

$$\mathrm{abstract}(A,T) \implies A^{\odot(T)} = A = (A * T)^{\mathrm{X}} * T^{\odot A}$$

The *maximum likelihood estimate* is near the *histogram*, $A^{\odot(T)} \sim A$, only in as much as it is far from the *surrealisation*, $A^{\odot(T)} \nsim (A * T)^{\mathrm{X}} * T^{\odot A}$.

The *surrealisation-dependent* varies with the *components-dependent*, $A^{\odot(T)} \sim A^{\mathrm{C}(T)}$.

The set of *iso-extremes* is defined in section 'Iso-sets', above, as the union of the *iso-liftisations* and the *iso-surrealisations*,

$$\begin{aligned}&(Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \;\cap\; D^{-1}_{U,T,z}(A * T)) \;\cup \\ &(Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \;\cap\; C^{-1}_{U,T,z}(\{(A * C^{\mathrm{U}})^{\wedge} : C \in T^{\mathrm{P}}\}))\end{aligned}$$

The set of *iso-extremes* is not, strictly speaking, an *iso-set*, because there is no *independent analogue* valued function for which it is the component of the implied partition. A *dependent analogue* can be defined, however. Define the *midisation*, $A^{\mathrm{M}(T)} \in \mathcal{A}_{U,V,z}$, as the *maximum likelihood estimate* for the *distribution histogram* of the (i) the *multinomial probability* of the *histogram* relative to (ii) the sum of the *multinomial probabilities* of the union of (a) the *integral iso-liftisations* and (b) the *integral iso-surrealisations*,

$$\{A^{\mathrm{M}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isolr}(U)(T,A)}) : D \in \mathcal{A}_{U,V,z}\})$$

where $\mathrm{isolr}(U)(T,A) := \mathrm{isol}(U)(T,A) \cup \mathrm{isor}(U)(T,A)$.

The *midisation* is sometimes not computable. The finite approximation to the continuous case for the *midisation* for some $k \in \mathbf{N}_{>0}$ is

$$\{A_k^{\mathrm{M}(T)}\} = \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isolr}(U)(T,A)}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the union of the *iso-liftisations* and the *iso-surrealisations* equals the *iso-independents*, and so the *midisation* equals the *dependent*, $A^{\mathrm{M}(T_{\mathrm{f}})} = A^{\mathrm{Y}}$. In the case where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the union of the *iso-liftisations* and the *iso-surrealisations* also equals the *iso-independents*, and so the *midisation* equals the *dependent*, $A^{\mathrm{M}(T_{\mathrm{u}})} = A^{\mathrm{Y}}$.

The *maximum likelihood estimate* is near the *histogram*, $A^{\mathrm{M}(T)} \sim A$, only in as much as it is far from the *liftisation*, $A^{\mathrm{M}(T)} \nsim A^{\mathrm{K}(T)}$ and the *surrealisation*, $A^{\mathrm{M}(T)} \nsim (A * T)^{\mathrm{X}} * T^{\odot A}$. The *midisation* varies as the *liftisation-dependent*, $A^{\mathrm{M}(T)} \sim A^{\mathrm{L}(T)}$ and the *surrealisation-dependent*, $A^{\mathrm{M}(T)} \sim A^{\odot(T)}$. In as much as the *idealisation-dependent* varies with the *liftisation-dependent*, $A^{\dagger(T)} \sim A^{\mathrm{L}(T)}$, the *midisation* varies as the *idealisation-dependent*, $A^{\mathrm{M}(T)} \sim A^{\dagger(T)}$.

Conjecture that the *entropies* of the *likely histograms* are subject to the following inequalities. First via the *independent*,

$$\begin{aligned} & \text{entropy}(V_z^{\mathrm{C}} * T * T^{\dagger}) = \text{entropy}(V_z^{\mathrm{C}}) \\ \geq\ & \text{entropy}((A * T)^{\mathrm{X}} * T^{\dagger}) \\ \geq\ & \text{entropy}(A^{\mathrm{P}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{X}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{X}}) \\ \geq\ & \text{entropy}(A) \\ \geq\ & \text{entropy}(A^{\mathrm{Y}}) \\ \geq\ & \text{entropy}(A^{\mathrm{Y}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{R}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{W}(T)}) \\ \geq\ & \text{entropy}(Z_A) = 0 \end{aligned}$$

Now via the *idealisation*,

$$\begin{aligned} & \text{entropy}((A * T)^{\mathrm{X}} * T^{\dagger}) \\ \geq\ & \text{entropy}(A * T * T^{\dagger}) \\ \geq\ & \text{entropy}(A^{\mathrm{K}(T)}) \\ \geq\ & \text{entropy}(A * T * T^{\odot A^{\mathrm{X}}}) \\ \geq\ & \text{entropy}(A * T * T^{\dagger A}) \\ \geq\ & \text{entropy}(A) \\ \geq\ & \text{entropy}(A^{\dagger(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{L}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{D}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{Y}(T)}) \end{aligned}$$

Now via the *surrealisation*,

$$\begin{aligned} & \text{entropy}(V_z^{\mathrm{C}} * T * T^{\dagger}) = \text{entropy}(V_z^{\mathrm{C}}) \\ \geq\ & \text{entropy}(V_z^{\mathrm{C}} * T * T^{\odot A}) \\ \geq\ & \text{entropy}(A^{\mathrm{X}} * T * T^{\odot A}) \\ \geq\ & \text{entropy}(A) \\ \geq\ & \text{entropy}(A^{\odot(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{C}(T)}) \end{aligned}$$

Finally via the *midisation*,

$$\begin{aligned} & \text{entropy}(A^{\mathrm{K}(T)}) \\ \geq\ & \text{entropy}(A) \\ \geq\ & \text{entropy}(A^{\mathrm{M}(T)}) \end{aligned}$$

and

$$\begin{aligned} & \text{entropy}((A * T)^{\mathrm{X}} * T^{\odot A}) \\ \geq\ & \text{entropy}(A) \\ \geq\ & \text{entropy}(A^{\mathrm{M}(T)}) \end{aligned}$$

The *entropy* of the *dependent analogue* is conjectured to be less than or equal to *entropy* of the *histogram*. For example,

$$\begin{aligned} & \text{entropy}(A) \\ \geq\ & \text{entropy}(A^{\dagger(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{L}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{D}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{Y}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{R}(T)}) \\ \geq\ & \text{entropy}(A^{\mathrm{W}(T)}) \end{aligned}$$

The *entropy* of *entity-like dependents* varies against the cardinality of the *iso-set*,

$$\text{entropy}(A^{\mathrm{I}(T)}) \sim -|I|$$

where $A \in I$ and $I \subseteq Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$.

In the case where the *dependent analogue* is in the *iso-set* and the *iso-set* is *law-like*, the difference in *entropy* must be in the *entropy* of the *components*. In the case of *derived-dependent*,

$$\begin{aligned} & A^{\mathrm{D}(T)} \in D^{-1}_{U,T,z}(A * T) \implies \\ & \quad A^{\mathrm{D}(T)} * T = A * T \implies \text{entropy}(A^{\mathrm{D}(T)} * T) = \text{entropy}(A * T) \end{aligned}$$

So

$$\begin{aligned} & A^{\mathrm{D}(T)} \in D^{-1}_{U,T,z}(A * T) \implies \\ & \quad \sum(\text{entropy}(A^{\mathrm{D}(T)} * C) : (R, C) \in T^{-1},\ (A^{\mathrm{D}(T)} * T)_R > 0) \\ & \leq\ \sum(\text{entropy}(A * C) : (R, C) \in T^{-1},\ (A * T)_R > 0) \end{aligned}$$

Similarly for the *idealisation-dependent*,

$$\begin{aligned}
&A^{\dagger(T)} \in D_{U,T,z}^{-1}(A * T) \implies \\
&\qquad \sum(\text{entropy}(A^{\dagger(T)} * C) : (R,C) \in T^{-1},\ (A^{\dagger(T)} * T)_R > 0) \\
&\quad \leq \quad \sum(\text{entropy}(A * C) : (R,C) \in T^{-1},\ (A * T)_R > 0)
\end{aligned}$$

The *midisation* varies with the *histogram*, $A^{\mathrm{M}(T)} \sim A$, in as much as it varies against the *liftisation*, $A^{\mathrm{M}(T)} \nsim A^{\mathrm{K}(T)}$ and the *surrealisation*, $A^{\mathrm{M}(T)} \nsim (A*T)^{\mathrm{X}}*T^{\odot A}$. The *multinomial probability* with respect to the *midisation* is maximised at the *mean*, so conjecture that, in the case where the *midisation* is *integral*, $A^{\mathrm{M}(T)} \in \mathcal{A}_{\mathrm{i}}$, the *multinomial probability* of the *midisation* with respect to the *midisation* varies as the *multinomial probability* of the *histogram* divided by the sum of the *multinomial probabilities* of the *liftisation* and the *surrealisation*,

$$\begin{aligned}
&Q_{\mathrm{m},U}(A^{\mathrm{M}(T)}, z)(A^{\mathrm{M}(T)}) \\
&\quad \sim \quad \frac{Q_{\mathrm{m},U}(A^{\mathrm{M}(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{M}(T)}, z)(B) : B \in \mathrm{isolr}(U)(T,A)} \\
&\quad \sim \quad \frac{Q_{\mathrm{m},U}(A^{\mathrm{M}(T)}, z)(A)}{Q_{\mathrm{m},U}(A^{\mathrm{M}(T)}, z)(A^{\mathrm{K}(T)}) + Q_{\mathrm{m},U}(A^{\mathrm{M}(T)}, z)((A * T)^{\mathrm{X}} * T^{\odot A})}
\end{aligned}$$

Dividing out the *permutorial* leaves the *multinomial coefficient*, so, after taking the logarithm, conjecture that the *entropy* of the *midisation* varies very approximately as the *entropy* of the *histogram* less the *entropies* of the *liftisation* and the *surrealisation*,

$$\begin{aligned}
\text{entropy}(A^{\mathrm{M}(T)}) \quad \sim \quad \text{entropy}(A) \ &- \ \text{entropy}(A^{\mathrm{K}(T)}) \\
&- \ \text{entropy}((A * T)^{\mathrm{X}} * T^{\odot A})
\end{aligned}$$

The *idealisation* is in the *iso-liftisations*, $A * T * T^{\dagger A} \in Y_{U,T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \cap D_{U,T,z}^{-1}(A * T)$, so, insofar as the *entropy* of the *idealisation* approximates to the *entropy* of the *liftisation*, $\text{entropy}(A * T * T^{\dagger A}) \approx \text{entropy}(A^{\mathrm{K}(T)})$, the *entropy* of the *midisation* varies computably as the *entropy* of the *histogram* less the *entropies* of the *idealisation* and the *surrealisation*,

$$\begin{aligned}
\text{entropy}(A^{\mathrm{M}(T)}) \quad \sim \quad \text{entropy}(A) \ &- \ \text{entropy}(A * T * T^{\dagger A}) \\
&- \ \text{entropy}((A * T)^{\mathrm{X}} * T^{\odot A})
\end{aligned}$$

In section 'Converse action entropy', above, it is conjectured that the sum of the *entropies* of the *contentisation* and the *neutralisation* varies as the sum of the *entropies* of the *histogram* and the *independent*

$$\text{entropy}(A^{\mathrm{X}} * T * T^{\odot A}) + \text{entropy}(A * T * T^{\odot A^{\mathrm{X}}}) \sim \text{entropy}(A) + \text{entropy}(A^{\mathrm{X}})$$

So, insofar as the *neutralisation entropy* approximates to the *idealisation entropy*, $\text{entropy}(A * T * T^{\odot A^{\mathrm{X}}}) \approx \text{entropy}(A * T * T^{\dagger A})$, the *entropy* of the *midisation* varies as the computable difference between the *entropies* of the *contentisation* and the *surrealisation* less the *entropy* of the *independent*,

$$\begin{aligned} \text{entropy}(A^{\mathrm{M}(T)}) \;\; \sim \;\; \text{entropy}(A^{\mathrm{X}} * T * T^{\odot A}) \; - \; \text{entropy}((A * T)^{\mathrm{X}} * T^{\odot A}) \\ - \; \text{entropy}(A^{\mathrm{X}}) \end{aligned}$$

If the *histogram*, A, is a given, then the *entropy* of the *independent*, $\text{entropy}(A^{\mathrm{X}})$, is a constant with respect to varying *transform*. If, in addition, the *formal* is constrained to be *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, then the *entropy* of the *midisation* varies as the difference between the *entropies* of the *formal independent* and the *abstract*,

$$\text{entropy}(A^{\mathrm{M}(T)}) \;\; \sim \;\; \text{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \; - \; \text{entropy}((A * T)^{\mathrm{X}})$$

Since the *entropy* of the doubly-*independent formal independent* is greater than or equal to that of the singly-*independent abstract*, $\text{entropy}((A^{\mathrm{X}}*T)^{\mathrm{X}}) \geq \text{entropy}((A*T)^{\mathrm{X}})$, the minimisation of the *midisation entropy*, $\text{entropy}(A^{\mathrm{M}(T)})$, tends to minimise the positive difference between the *entropies* of the *formal independent* and the *abstract*, $\text{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}}) - \text{entropy}((A * T)^{\mathrm{X}}) \geq 0$, and so the *abstract* tends to equal the *formal independent*, which equals the *formal*, $(A * T)^{\mathrm{X}} = (A^{\mathrm{X}} * T)^{\mathrm{X}} = A^{\mathrm{X}} * T$. That is, in the case where the *formal* is constrained to be *independent*, for example when *non-overlapping*, $\neg\text{overlap}(T) \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, then the minimisation of the *midisation entropy*, $\text{entropy}(A^{\mathrm{M}(T)})$, tends to *formal-abstract equivalence*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$.

This is consistent with the difference between the *partition contentisation entropy* and the *partition surrealisation entropy* in the case where the *transform* is a *substrate transform* $T \in \mathcal{T}_{U,V}$, having *derived variables* $W = \text{der}(T)$. Given a *partition variable* $P \in W$, the difference is

$$\begin{aligned} & \text{entropy}(A^{\mathrm{X}} * P^{\mathrm{T}} * P^{\mathrm{T}\odot A}) \; - \; \text{entropy}((A * P^{\mathrm{T}})^{\mathrm{X}} * P^{\mathrm{T}\odot A}) \\ = \;\; & \text{entropy}(A^{\mathrm{X}} * P^{\mathrm{T}} * P^{\mathrm{T}\odot A}) \; - \; \text{entropy}(A) \end{aligned}$$

where the *partition contentisation entropy* is between the *independent entropy* and the *histogram entropy*, $\text{entropy}(A^{\mathrm{X}}) \geq \text{entropy}(A^{\mathrm{X}} * P^{\mathrm{T}} * P^{\mathrm{T}\odot A}) \geq$

entropy(A). When the *formal independent* equals the *abstract*, the *partition transform* is *formal* and the *partition contentisation* equals the *histogram*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies A * P^{\mathrm{T}} = A^{\mathrm{X}} * P^{\mathrm{T}} \implies A^{\mathrm{X}} * P^{\mathrm{T}} * P^{\mathrm{T}\odot A} = A$, and so the difference between the *partition contentisation entropy* and the *partition surrealisation entropy* reduces to zero.

For a given *histogram*, A, the *midisation* varies against the computable sum of the *entropies* of the *idealisation* and the *surrealisation*,

$$\mathrm{entropy}(A^{\mathrm{M}(T)}) \quad \sim \quad -\ (\mathrm{entropy}(A * T * T^{\dagger A}) + \mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\odot A}))$$

Conjecture that the sum of the *entropies* of the *idealisation* and the *surrealisation* is greater than or equal to the sum of the *entropies* of the *histogram* and the *independent*

$$\mathrm{entropy}(A * T * T^{\dagger A}) + \mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\odot A}) \ \geq \ \mathrm{entropy}(A) + \mathrm{entropy}(A^{\mathrm{X}})$$

In the case where the *transform* is *self*, for example a *value full functional transform* $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}}$, the *idealisation* equals the *histogram* and the *surrealisation* equals the *independent*, so the sums are equal,

$$\mathrm{entropy}(A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger A}) + \mathrm{entropy}((A * T_{\mathrm{s}})^{\mathrm{X}} * T_{\mathrm{s}}^{\odot A}) \ = \ \mathrm{entropy}(A) + \mathrm{entropy}(A^{\mathrm{X}})$$

In the case where the *transform* is *unary*, for example $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, the *idealisation* equals the *independent* and the *surrealisation* equals the *histogram*, so the sums are equal,

$$\mathrm{entropy}(A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A}) + \mathrm{entropy}((A * T_{\mathrm{u}})^{\mathrm{X}} * T_{\mathrm{u}}^{\odot A}) \ = \ \mathrm{entropy}(A^{\mathrm{X}}) + \mathrm{entropy}(A)$$

Conjecture that there exists some intermediate *substrate transform* $T_{\mathrm{m}} \in \mathcal{T}_{U,V}$ which is neither *self* nor *unary*, $T_{\mathrm{m}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$, such that the sum of the *entropies* of the *idealisation* and the *surrealisation* is maximised,

$$T_{\mathrm{m}} \in \mathrm{maxd}(\{(T, \mathrm{entropy}(A * T * T^{\dagger A}) + \mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\odot A})) : T \in \mathcal{T}_{U,V}\})$$

So in some cases the *mid transform*, T_{m}, minimises the *midisation entropy*,

$$T_{\mathrm{m}} \in \mathrm{mind}(\{(T, \mathrm{entropy}(A^{\mathrm{M}(T)})) : T \in \mathcal{T}_{U,V}\})$$

The *mid idealisation entropy* approximates to the *mid surrealisation entropy*,

$$\mathrm{entropy}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A}) \quad \approx \quad \mathrm{entropy}((A * T_{\mathrm{m}})^{\mathrm{X}} * T_{\mathrm{m}}^{\odot A})$$

but the *mid idealisation derived entropy* is less than or equal to the *mid surrealisation derived entropy*,

$$\begin{aligned} & \text{entropy}(A * T_{\text{m}} * T_{\text{m}}^{\dagger A} * T_{\text{m}}) \;=\; \text{entropy}(A * T_{\text{m}}) \\ \leq \;& \text{entropy}((A * T_{\text{m}})^{\text{X}} * T_{\text{m}}^{\odot A} * T_{\text{m}}) \;=\; \text{entropy}((A * T_{\text{m}})^{\text{X}}) \end{aligned}$$

In the case where the minimisation constrains the *formal* to be *independent*,

$$T_{\text{m}} \in \text{mind}(\{(T, \text{entropy}(A^{\text{M}(T)})) : T \in \mathcal{T}_{U,V},\ A^{\text{X}} * T = (A^{\text{X}} * T)^{\text{X}}\})$$

the *mid transform* tends to be such that the *formal* equals the *abstract*, $A^{\text{X}} * T_{\text{m}} = (A * T_{\text{m}})^{\text{X}}$. In these cases, the *mid derived entropy* is less than or equal to the *independent mid derived entropy*, $\text{entropy}(A * T_{\text{m}}) \leq \text{entropy}((A * T_{\text{m}})^{\text{X}}) = \text{entropy}(A^{\text{X}} * T_{\text{m}})$.

While the minimisation of the *midisation entropy*, where the *formal* is *independent*, does not preclude a decrease in the *mid derived entropy*, $\text{entropy}(A * T_{\text{m}})$, conjecture that, overall, the *mid component size cardinality relative entropy* is nonetheless small, $\text{entropyRelative}(A * T_{\text{m}}, V^{\text{C}} * T_{\text{m}}) \approx 0$. That is, the minimisation of the *midisation entropy*, where the *formal* is *independent*, tends to a *partition* of the *states* such that the *sizes* of the *components* tend to be synchronised with their *cardinalities*.

This may be seen by considering the case where the *histogram* is *fully diagonalised*, $\text{diagonalFull}(U)(A)$, and *uniform*, $|\text{ran}(A * A^{\text{F}})| = 1$. In this case the *independent* is *cartesian*, $A^{\text{X}} = V_z^{\text{C}}$, where $V_z^{\text{C}} = \text{scalar}(z/v) * V^{\text{C}}$. In the case where the *transform* is a *substrate transform*, $T \in \mathcal{T}_{U,V}$, and the *formal independent* equals the *abstract*, each *partition derived* must equal the *partition formal*, $(A^{\text{X}} * T)^{\text{X}} = (A * T)^{\text{X}} \iff \forall P \in \text{der}(T)\ (A * P^{\text{T}} = A^{\text{X}} * P^{\text{T}})$. This is satisfied by any *partition* that (i) *partitions* the *effective states* along the *diagonal* into *components*, $P :\leftrightarrow: A^{\text{F}}$, and (ii) is such that the *component cardinalities* are uniform, $|\{|C| : C \in P\}| = 1$. Let $F \subset \mathcal{T}_{U,V}$ be the set of *transforms*,

$$F \;=\; \{T : T \in \mathcal{T}_{U,V},\ \forall P \in \text{der}(T)\ (P :\leftrightarrow: A^{\text{F}} \wedge |\{|C| : C \in P\}| = 1)\}$$

All of these *transforms* are such that the *formal independent* equals the *abstract*, $\forall T \in F\ ((A^{\text{X}} * T)^{\text{X}} = (A * T)^{\text{X}})$, and all are such that the *cross entropy* equals the *derived entropy*, $\forall T \in F\ (\text{entropyCross}(A * T, V^{\text{C}} * T) = \text{entropy}(A * T))$, and so have zero *relative entropy*, $\forall T \in F\ (\text{entropyRelative}(A * T_{\text{m}}, V^{\text{C}} * T_{\text{m}}) = 0)$. Each of the *derived variables* of the *value full functional transform partitions* the *diagonal* and has uniform *component cardinality*,

$\{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}} \in F$. As the constraints on the *histogram* are relaxed such that (i) the *diagonal* is not *uniform*, so the *independent* is no longer *cartesian*, $A^{\mathrm{X}} \neq V_z^{\mathrm{C}}$, and (ii) *effective off-diagonal states* are allowed, so bijections, $P :\leftrightarrow: A^{\mathrm{F}}$ are fewer, the cardinality of *transforms* that are such that the *formal independent* equals the *abstract* decreases, but the *sizes* of the *components* still tend to be synchronised with their *cardinalities* and so *relative entropy* remains small.

The *idealisation independent* equals the *independent*, so the *idealisation entropy* is less than or equal to the *independent entropy*, $\mathrm{entropy}(A^{\mathrm{X}}) = \mathrm{entropy}((A * T * T^{\dagger A})^{\mathrm{X}}) \geq \mathrm{entropy}(A * T * T^{\dagger A})$. Equality occurs in the case where the *transform* is *unary*, $\mathrm{entropy}(A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A}) = \mathrm{entropy}(A^{\mathrm{X}})$. The *idealisation entropy* is greater than or equal to the *histogram entropy*, $\mathrm{entropy}(A * T * T^{\dagger A}) \geq \mathrm{entropy}(A)$. Equality occurs in the case where the *transform* is *self*, $\mathrm{entropy}(A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger A}) = \mathrm{entropy}(A)$. Therefore the *mid idealisation entropy* is between the *independent entropy* and the *histogram entropy*,

$$\mathrm{entropy}(A^{\mathrm{X}}) \;\geq\; \mathrm{entropy}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A}) \;\geq\; \mathrm{entropy}(A)$$

The *idealisation* equals the *sum* of its *independent commponents*, $A*T*T^{\dagger A} = \sum_{(\cdot,C) \in T^{-1}} (A * C)^{\mathrm{X}}$. Consider the case where the *idealisation* is *integral*, for example if it were the *histogram* of some *history*. In this case the *components* are *independent* and *integral*,

$$A * T * T^{\dagger A} \in \mathcal{A}_{\mathrm{i}} \implies \forall (\cdot, C) \in T^{-1}\ ((A * C)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}})$$

As shown in section 'Independent histograms', above, the logarithm of the cardinality of *integral independent histograms* of a given set of *variables* V and *size* z varies against the *volume* $v = |V^{\mathrm{C}}|$,

$$\ln |\{A : A \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{XF}} = V^{\mathrm{C}},\ \mathrm{size}(A) = z,\ A = A^{\mathrm{X}}\}| \;\sim\; -v$$

Given that the total *component cardinality* is the *volume*, $\sum_{(R,\cdot) \in T^{-1}} (V^{\mathrm{C}} * T)_R = \sum_{(\cdot,C) \in T^{-1}} |C| = v$, and the total *component size* is the *size*, $\sum_{(R,\cdot) \in T^{-1}} (A * T)_R = \sum_{(\cdot,C) \in T^{-1}} \mathrm{size}(A * C) = z$, a minimisation of the *integral idealisation entropy* from the *mid transform* tends to increase the *mid component size cardinality cross entropy*,

$$\mathrm{entropyCross}(A * T_{\mathrm{m}}, V^{\mathrm{C}} * T_{\mathrm{m}}) \;\sim\; -\ \mathrm{entropy}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A})$$

The *mid idealisation entropy* approximates to the *mid surrealisation entropy*, $\mathrm{entropy}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A}) \approx \mathrm{entropy}((A * T_{\mathrm{m}})^{\mathrm{X}} * T_{\mathrm{m}}^{\odot A})$, but the *mid idealisation* *derived entropy* is less than or equal to the *mid surrealisation derived entropy*, $\mathrm{entropy}(A*T_{\mathrm{m}}) \leq \mathrm{entropy}((A*T_{\mathrm{m}})^{\mathrm{X}})$. So in the case where the minimisation of the *integral idealisation entropy* from the *mid transform* does not increase the *derived entropy*, $\mathrm{entropy}(A*T_{\mathrm{m}})$, the *mid component size cardinality relative entropy* varies against the *mid idealisation entropy*,

$$\mathrm{entropyRelative}(A*T_{\mathrm{m}}, V^{\mathrm{C}}*T_{\mathrm{m}}) \quad \sim \quad -\ \mathrm{entropy}(A*T_{\mathrm{m}}*T_{\mathrm{m}}^{\dagger A})$$

That is, a minimisation of the *integral idealisation entropy* starting from the *mid transform* is a maximisation of the *relative entropy*. So high *size components* tend to be low *cardinality components* and low *size components* tend to be high *cardinality components*.

This may be contrasted with the minimisation of the *midisation entropy*, $\mathrm{entropy}(A^{\mathrm{M}(T)})$, where the *formal* is *independent*, $A^{\mathrm{X}}*T = (A^{\mathrm{X}}*T)^{\mathrm{X}}$, which has low *relative entropy*, $\mathrm{entropyRelative}(A*T_{\mathrm{m}}, V^{\mathrm{C}}*T_{\mathrm{m}}) \approx 0$, because the *formal independent* approximates to the *abstract*, $(A^{\mathrm{X}}*T)^{\mathrm{X}} \approx (A*T)^{\mathrm{X}}$. That is, a minimisation of the *midisation entropy*, where the *formal* is *independent*, tends to decrease the *relative entropy*, but a subsequent minimisation of the *integral idealisation entropy* tends to increase the *relative entropy*.

The *transform-independent*, $A^{\mathrm{X}(T)}$, is the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of membership of the *iso-transform-independents*,

$$\{A^{\mathrm{X}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A))) : D \in \mathcal{A}_{U,V,z}\})$$

The corresponding the *transform-dependent*, $A^{\mathrm{Y}(T)}$, is the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-transform-independent*,

$$\begin{aligned}\{(A^{\mathrm{Y}(T)}, \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)})\} = \\ \max(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

Conjecture that the logarithm of the *maximum conditional multinomial probability* varies against the corresponding logarithm of the *conditional probability* where the *distribution histogram* is not the *dependent analogue*, $A^{\mathrm{Y}(T)}$, but the *independent analogue*, $A^{\mathrm{X}(T)}$,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \sim - \ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}$$

This may be seen by considering first the *conditional probability* with respect to the *dependent-analogue* on the left hand side of this relation. If the *independent-analogue* is in the *iso-set*, $A^{\mathrm{X}(T)} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, and the *histogram* is not equal to the *independent-analogue*, $A \neq A^{\mathrm{X}(T)}$, then the terms of the denominator with respect to the *dependent-analogue* are such that

$$\begin{aligned} 0 &< \hat{Q}_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A^{\mathrm{X}(T)}) \\ &< \hat{Q}_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A) \\ &< \sum(\hat{Q}_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)) \\ &\leq 1 \end{aligned}$$

That is, the *independent-analogue* term is less than the numerator,

$$Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A^{\mathrm{X}(T)}) < Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A)$$

Conjecture that *independent-analogue* term is the least of the denominator,

$$\forall B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\ (Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A^{\mathrm{X}(T)}) \leq Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(B))$$

Now consider the *conditional probability* with respect to the *independent-analogue* on the right hand side of this relation. Here the terms of the denominator with respect to the *independent-analogue* are such that

$$\begin{aligned} 0 &< \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(A) \\ &< \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(A^{\mathrm{X}(T)}) \\ &< \sum(\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)) \\ &\leq 1 \end{aligned}$$

Conjecture that *independent-analogue* term is the greatest of this denominator,

$$\forall B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\ (Q_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(A^{\mathrm{X}(T)}) \geq Q_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(B))$$

Conjecture that the probability of *drawing* the *histogram* from the *dependent-analogue* is greater than that from the *independent-analogue*, $\hat{Q}_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A) >$

$\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(A)$, and so the numerator of the right hand side is less than that of the left hand side. Conjecture further, however, that the probability of *drawing* the *independent-analogue* from the *independent-analogue* is greater than the probability of *drawing* the *histogram* from the *dependent-analogue*, $\hat{Q}_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(A) < \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(A^{\mathrm{X}(T)})$, and so the denominator of the right hand side is conjectured to be greater than that of the left hand side. Hence the *conditional probability* with respect to the *dependent-analogue* is conjectured to vary against the *conditional probability* with respect to the *independent-analogue*.

The *independent-analogue* term is the largest of the denominator with respect to the *independent-analogue*, $\forall B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\ (Q_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(A^{\mathrm{X}(T)}) \geq Q_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(B))$. So the logarithm of the *maximum conditional probability* is conjectured to vary against the logarithm of the *relative probability* with respect to the *independent-analogue*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \quad \sim \quad -\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(A)}{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(A^{\mathrm{X}(T)})}$$

Define the *distribution-relative multinomial space* of a *histogram* $A \in \mathcal{A}_{U,V,z}$ with respect to a *distribution histogram* $E \in \mathcal{A}_{U,V,z}$ as $\mathrm{spaceRelative} \in \mathcal{A} \to (\mathcal{A} \to \mathbf{R})$, the negative logarithm *relative multinomial probability density*,

$$\mathrm{spaceRelative}(E)(A) \quad := \quad -\ln \frac{\mathrm{mpdf}(U)(E, z)(A)}{\mathrm{mpdf}(U)(E, z)(E)}$$

In the case where the *histogram* and *distribution histogram* are *integral*, $A, E \in \mathcal{A}_{\mathrm{i}}$, the *distribution-relative multinomial space* is the negative logarithm *relative multinomial probability*,

$$\mathrm{spaceRelative}(E)(A) \quad := \quad -\ln \frac{Q_{\mathrm{m},U}(E, z)(A)}{Q_{\mathrm{m},U}(E, z)(E)}$$

So, in this case, the *relative space* of the *histogram* with respect to the *transform-independent* is

$$\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) \quad := \quad -\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(A)}{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(A^{\mathrm{X}(T)})}$$

and the logarithm of the *maximum conditional probability* is conjectured to vary with the *relative space*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \quad \sim \quad \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A)$$

Conjecture that in the case where an *independent analogue* is in the *iso-set*, and so is in the denominator, the *relative space* of a *histogram* with respect to the *independent analogue* is always positive and less than or equal to the *relative space* of the corresponding *dependent analogue* with respect to the *independent analogue*. So,

$$\begin{aligned}
A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies 0 &\leq \mathrm{spaceRelative}(A^{\mathrm{X}})(A) \\
&\leq \mathrm{spaceRelative}(A^{\mathrm{X}})(A^{\mathrm{Y}}) \\
A^{\mathrm{X}(T)} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) \implies 0 &\leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) \\
&\leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{Y}(T)}) \\
A * T * T^{\dagger A} \in \mathcal{A}_{\mathrm{i}} \implies 0 &\leq \mathrm{spaceRelative}(A * T * T^{\dagger A})(A) \\
&\leq \mathrm{spaceRelative}(A * T * T^{\dagger A})(A^{\dagger(T)}) \\
(A * T)^{\mathrm{X}} * T^{\odot A} \in \mathcal{A}_{\mathrm{i}} \implies 0 &\leq \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\odot A})(A) \\
&\leq \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\odot A})(A^{\odot(T)}) \\
A * T * T^{\dagger} \in \mathcal{A}_{\mathrm{i}} \implies 0 &\leq \mathrm{spaceRelative}(A * T * T^{\dagger})(A) \\
&\leq \mathrm{spaceRelative}(A * T * T^{\dagger})(A^{\mathrm{D}(T)}) \\
(A * T)^{\mathrm{X}} * T^{\dagger} \in \mathcal{A}_{\mathrm{i}} \implies 0 &\leq \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A) \\
&\leq \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A^{\mathrm{W}(T)})
\end{aligned}$$

Conjecture that the *relative space* of the *dependent* is less than or equal to that of the *transform-dependent* which, in turn, is less than or equal to that of the *abstract-dependent*,

$$\begin{aligned}
&\mathrm{spaceRelative}(A^{\mathrm{X}})(A^{\mathrm{Y}}) \\
\leq\ &\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{Y}(T)}) \\
\leq\ &\mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A^{\mathrm{W}(T)})
\end{aligned}$$

where $A^{\mathrm{X}(T)} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$.

Conjecture that, in the case where the *formal* equals the *abstract*, $A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}}$, the *relative space* of the *idealisation-dependent* is less than or equal to that of the *derived-dependent*, which, in turn, is less than or equal to that of the *transform-dependent*

$$\begin{aligned}
&\mathrm{spaceRelative}(A * T * T^{\dagger A})(A^{\dagger(T)}) \\
\leq\ &\mathrm{spaceRelative}(A * T * T^{\dagger})(A^{\mathrm{D}(T)}) \\
\leq\ &\mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A^{\mathrm{Y}(T)})
\end{aligned}$$

where $(A * T)^{\mathrm{X}} * T^{\dagger} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$.

Conjecture that the *relative space* of the *surrealisation-dependent* is less than or equal to that of the *components-dependent*,

$$\begin{aligned} & \text{spaceRelative}((A * T)^{\mathrm{X}} * T^{\odot A})(A^{\odot(T)}) \\ \leq\ & \text{spaceRelative}(V_z^{\mathrm{C}} * T * T^{\odot A})(A^{\mathrm{C}(T)}) \end{aligned}$$

3.18 Encoding space

Note that, in the following, the use of the terminology *space* is in the sense of computational space, i.e. the logarithm of the cardinality of some discrete set of states, rather than in the sense of the structured sets of mathematical spaces, such as topological space. See the appendices 'Coders' and 'Computers' for more formal definitions.

3.18.1 History space

The set of *histories* $\mathcal{H}_U \subset \mathcal{X} \to \mathcal{S}_U$ in finite *system* U is a superset of the *histories* $\mathcal{H}_{U,X} \subset \mathcal{H}_U$ where the domains of the *histories* are restricted to a finite subset of the *event identifiers* $X \subset \mathcal{X}$, so that $\mathcal{H}_{U,X} \subset X \to \mathcal{S}_U$. Now $\mathcal{H}_{U,X}$ is finite and can be constructed

$$\mathcal{H}_{U,X} = \bigcup \{X \to \text{cartesian}(U)(V) : V \subseteq \text{vars}(U)\}$$

$\mathcal{H}_{U,X}$ includes the *empty history* $\emptyset$. It also includes *non-variate histories* if there are any *events*, $|\{H : H \in \mathcal{H}_{U,X},\ \text{vars}(H) = \emptyset\}| > 0$ where $|X| > 0$.

The set of *states* in a *system* U is $\mathcal{S}_U = \{S : V \subseteq \text{vars}(U),\ S \in V^{\mathrm{CS}}\}$. In the case of a *regular system* U, having *dimension* $n = |U|$ and such that all the *variables* have the same *valency* $\{d\} = \{|W| : W \in \text{ran}(U)\}$, the cardinality of the set of *states* is

$$|\mathcal{S}_U| \ = \ \sum_{k \in \{0 \ldots n\}} \binom{n}{k} d^k \ < \ 2^n d^n \ \leq \ d^{2n}$$

where $d \geq 2$. The cardinality of the *histories* is

$$|\mathcal{H}_{U,X}| \ = \ 1 + \sum_{k \in \{0 \ldots n\}} \binom{n}{k} \sum_{z \in \{1 \ldots y\}} \binom{y}{z} d^{kz} \ \leq \ 2^{n+y} d^{ny}$$

where $y = |X|$.

Coders encapsulate the logic of *codes* of lists of a given *coder domain*, such that the *space* of the elements of the *coder domain* is computable. A *coder* $C \in \text{coders}(Y)$ on a *coder domain* Y defines the *code*, $Y :\leftrightarrow \mathbf{N}$, and the *space* function, $\text{space}(C) \in Y :\to \ln \mathbf{N}_{>0}$. The *encode* function, $\text{encode}(C) \in \mathcal{L}(Y) :\to \mathbf{N}$, and the *decode* function, $\text{decode}(C) \in \mathbf{N} \times \mathbf{N} \to: \mathcal{L}(Y)$, are functions of the *code* and the *space* such that $\forall L \in \mathcal{L}(Y)\ (\text{decode}(C)(|L|, \text{encode}(C)(L)) = L)$.

The *space* required to encode a list $L \in \mathcal{L}(Y)$ is the sum of the *spaces* of the list elements, $\sum_{i \in \{1 \ldots |L|\}} C^{\text{s}}(L_i)$, where $C^{\text{s}}(x) := \text{space}(C)(x)$. The *space* of a list is always sufficient to encode the list,

$$\forall L \in \mathcal{L}(Y)\ \big(\prod_{i \in \{1 \ldots |L|\}} e^{C^{\text{s}}(L_i)}\ >\ \text{encode}(C)(L)\big)$$

A *probability function* on the *coder domain* $P \in (Y :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ implies an *expected space*, $\text{expected}(P)(\text{space}(C))$. The *expected space* is greater than or equal to the *entropy* of the *probability function*, $\text{expected}(P)(\text{space}(C)) \geq \text{entropy}(P)$.

A *minimal coder* C_{m} is one where the *expected space* of a uniform *probability function* is the logarithm of the cardinality of the *coder domain*, $\text{expected}(Y \times \{1/|Y|\})(\text{space}(C_{\text{m}})) = \ln |Y|$. See appendix 'Coders' for a formal definition of *coders*.

History coders $\text{coders}(\mathcal{H}_{U,X})$ are *coders* where the *coder domain* is the set of all *histories* in *system* U and *identifier set* X. When $|\mathcal{H}_{U,X}|$ is small it is practicable to construct a *coder of histories* explicitly.

For example, if the *system* U contains a single, *mono-valent variable*, $U = \{(v, \{w\})\}$, so that there are only two *states* $\mathcal{S}_U = \{\emptyset, S\}$, where $S = \{(v, w)\}$, and such that there is only one *event identifier* $X = \{x\}$ then there are only three possible *histories*, $\mathcal{H}_{U,X} = \{\emptyset, \{(x, \emptyset)\}, \{(x, S)\}\}$. Choose an enumeration $N \in \text{enums}(\mathcal{H}_{U,X})$ and then we can construct a *history coder* $C \in \text{coders}(\mathcal{H}_{U,X})$ such that

$$\forall (H, i) \in N\ ((\text{encode}(C)(\{(1, H)\}) = i{-}1) \wedge (\text{decode}(C)(1, i{-}1) = \{(1, H)\}))$$

C is a *minimal coder* if we choose it such that $\forall H \in \mathcal{H}_{U,X}\ (\text{space}(C)(H) = \ln 3)$, because the *total space* of C is $|\mathcal{H}_{U,X}| \ln |\mathcal{H}_{U,X}| = 3 \ln 3$.

Even in the case when the *coder domain* $|\mathcal{H}_{U,X}|$ is small, the *decode* relation of a *coder* must be defined algorithmically (that is, recursively) because it is an infinite set. (In the case of *fixed-width coders* where the *space* is constant, this algorithm is straightforward, however.) The *decode* function implies constraints on the *space*, so the *code* and the *space* of *coders* are usually defined algorithmically too. We shall describe several algorithmic *coders* of *histories* and their *classifications*.

Let D_{X} be an order, $D_{\mathrm{X}} \in \mathrm{enums}(X)$, on the *event identifiers* X mapping them to the natural numbers, $D_{\mathrm{X}} \in X :\leftrightarrow \mathbf{N}$, by one of the enumerations $\mathrm{enums}(X) = X \cdot \{1 \ldots |X|\}$. Given order D_{X} we can choose an enumeration for any subset $Q \subset X$ such that $\mathrm{order}(D_{\mathrm{X}}, Q) \in \mathrm{enums}(Q)$.

Let D_{V} be an order $D_{\mathrm{V}} \in \mathrm{enums}(\mathrm{vars}(U))$ on the *variables* in the *system* U. Given order D_{V} and subset $V \subset \mathrm{vars}(U)$, we have $\mathrm{order}(D_{\mathrm{V}}, V) \in \mathrm{enums}(V)$. Let D_{S} be an order $D_{\mathrm{S}} \in \mathrm{enums}(\mathcal{S}_U)$ on the *states*. See appendix 'Constructing states order from variables and values orders' for an example of how D_{S} can be constructed. Given order D_{S} and *variables* V, we have the enumeration of the *cartesian* set of *states*

$$\mathrm{order}(D_{\mathrm{S}}, \mathrm{cartesian}(U)(V)) \in \mathrm{enums}(\mathrm{cartesian}(U)(V))$$

Note that the choice of orders on the *event identifiers* or *variables* or *states* can be entirely arbitrary, say, for example, lexical. There are no semantics imposed by any order, nor constraining the order. In particular, the *events* of a *history* are not necessarily chronological.

Consider the *index coder* of *histories*

$$C_{\mathrm{H}} = \mathrm{coderHistoryIndex}(U, X, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

A *history* $H \in \mathcal{H}_{U,X}$ can be encoded in a tuple $T_H \in \mathbf{N}^2 \times \mathbf{N}^2 \times \mathcal{L}(\mathbf{N})$. This tuple T_H can then be encoded into a natural number $E_H \in \{0 \ldots S_H - 1\}$. The *space* of the *history* is the *space* of this encoding, $\mathrm{space}(C_{\mathrm{H}})(H) = \mathrm{space}(S_H) = \ln S_H$.

The first pair of the tuple T_H encodes the set of *variables*. Each *history* $H \in \mathcal{H}_{U,X}$ has a subset $\mathrm{vars}(H) \subseteq \mathrm{vars}(U)$ of the *variables* of the *system*. Let $r = |\mathrm{vars}(U)|$. Let $V = \mathrm{vars}(H)$ and $n = |V|$. Given D_{V} choose an enumeration N in the enumerations of the set of subsets of the *variables* of cardinality n in *system* U

$$N \in \mathrm{enums}(\{W : W \in \mathrm{P}(\mathrm{vars}(U)),\ |W| = n\})$$

The pair encoding the *variables* of H is (n, N_V) where $n \in \{0 \ldots r\}$ and

$$N_V \in \{1 \ldots \frac{r!}{(r-n)!\ n!}\}$$

Define the *space* to encode *dimension* n in *system* U as spaceDimension $\in \mathcal{U} \to \ln \mathbf{N}_{>0}$

$$\text{spaceDimension}(U) := \ln(|\text{vars}(U)| + 1)$$

Define spaceSubset $\in \mathbf{N} \times \mathbf{N} \to \ln \mathbf{N}_{>0}$ as the *space* of a binomial combination

$$\text{spaceSubset}(a, b) := \ln \binom{a}{b} = \ln \frac{a!}{(a-b)!\ b!}$$

where $a \geq b$.

Define the *space* of the pair encoding a set of *variables* of *dimension* n in *system* U as spaceVariables$(U) \in \mathbf{N} \to \ln \mathbf{N}_{>0}$

$$\text{spaceVariables}(U)(n) := \text{spaceDimension}(U) + \text{spaceSubset}(|\text{vars}(U)|, n)$$

The *space* of (n, N_V) is spaceVariables$(U)(n)$.

Similarly, the second pair of the tuple T_H encodes the set of *event identifiers*. Each *history* $H \in \mathcal{H}_{U,X}$ has a subset $\text{ids}(H) \subseteq X$ of the *event identifier set*. Let $y = |X|$. Let $I = \text{ids}(H)$ and $z = |I| = |H|$. Given D_{X} choose an enumeration Z of the set of subsets of the *event identifiers* of cardinality z in *system* U

$$Z \in \text{enums}(\{Q : Q \in \mathrm{P}(X),\ |Q| = z\})$$

The pair encoding the *event identifiers* of H is (z, Z_I) where $z \in \{0 \ldots y\}$ and

$$Z_I \in \{1 \ldots \frac{y!}{(y-z)!z!}\}$$

Define the *space* of the *size*, z, of the set of *event identifiers* as spaceSize $\in \mathbf{N} \to \ln \mathbf{N}_{>0}$

$$\text{spaceSize}(y) := \ln(y + 1)$$

Define the *space* of the pair encoding a subset of the set of *event identifiers* as spaceIds $\in \mathbf{N} \times \mathbf{N} \to \ln \mathbf{N}_{>0}$

$$\text{spaceIds}(y, z) := \text{spaceSize}(y) + \text{spaceSubset}(y, z)$$

The *space* of (z, Z_I) is spaceIds(y, z).

Note that an alternative representation of membership of a subset is the use of the *fixed width* list of *bits*, $\mathcal{L}(\text{bits})$. In the case of *variables*, the length of the list is the cardinality of the *variables* in the *system*, $|\text{vars}(U)|$, and so the *space* required would be $r \ln 2$. In the case of the *event identifiers* the length of the list is the cardinality of the *event identifiers*, $|X|$, and so the *space* would be $y \ln 2$. The *fixed width* representation requires less *space* where entropy is high, that is where $n \approx r/2$ and $z \approx y/2$.

The last element $L \in \mathcal{L}(\mathbf{N})$ of the tuple T_H encodes the *states* of *events* in a list of *fixed width space*. Given order D_{S} and *variables* $V = \text{vars}(H)$, let M be an enumeration of the *cartesian* set of *states*, $M = \text{order}(D_{\text{S}}, \text{cartesian}(U)(V))$. Let $v = \text{volume}(U)(V)$. Then $|M| = v$ and hence $L \in \mathcal{L}(\{1 \ldots v\})$. Given order D_{X} and *event identifiers* $I = \text{ids}(H)$, let Q be an enumeration of the *event identifiers*, $Q = \text{order}(D_{\text{X}}, I)$. Then

$$L = \{(Q_x, M_S) : (x, S) \in H\}$$

Here M_S is the index number of the *state*, S, in the enumeration, M.

Define the *space* of the list encoding the *events* of *history* H in a *system* U as $\text{spaceEvents}(U) \in \mathcal{H}_U \to \ln \mathbf{N}_{>0}$

$$\text{spaceEvents}(U)(H) := z \ln v$$

where $z = |H|$, $V = \text{vars}(H)$ and $v = \text{volume}(U)(V)$.

Finally the tuple $T_H = ((n, N_V), (z, Z_I), L)$ can be encoded to $E_H \in \{0 \ldots S_H - 1\}$ where

$$S_H = (r+1) \times \frac{r!}{(r-n)!\ n!} \times (y+1) \times \frac{y!}{(y-z)!\ z!} \times v^z$$

The total *space* of the *index coder* C_{H} of a *history* H is the sum of the *variables space*, *ids space* and *events space*

$$\begin{aligned} \text{space}(C_{\text{H}})(H) \quad = \quad & \text{spaceVariables}(U)(|\text{vars}(H)|) + \\ & \text{spaceIds}(|X|, |H|) + \\ & \text{spaceEvents}(U)(H) \end{aligned}$$

A variation of the *index history coder*, C_{H}, is to encode only the *effective volume* in the *fixed width space*. Here the *history* $H \in \mathcal{H}_{U,X}$ is encoded in a tuple $T_H \in \mathbf{N}^2 \times \mathbf{N}^2 \times \mathbf{N}^2 \times \mathcal{L}(\mathbf{N})$ that has an extra pair of integers $(x, F_Q) \in$

$\mathbf{N}^2$ which encode (i) the *effective volume* $x = |A^{\mathrm{F}}|$, where $A = \mathrm{histogram}(H)$, and (ii) the index, F_Q, into an enumeration of the *effective states* subsets,

$$F_Q \in \{1 \ldots \frac{v!}{(v-x)!\ x!}\}$$

Now M is an enumeration of the *effective* set of *states*, $M = \mathrm{order}(D_{\mathrm{S}}, A^{\mathrm{FS}})$. Then $|M| = x$ and hence $L \in \mathcal{L}(\{1 \ldots x\})$.

The tuple $T_H = ((n, N_V), (z, Z_I), (x, F_Q), L)$ can be encoded to $E_H \in \{0 \ldots S_H - 1\}$ where

$$S_H = (r+1) \times \frac{r!}{(r-n)!\ n!} \times (y+1) \times \frac{y!}{(y-z)!\ z!} \times v \times \frac{v!}{(v-x)!\ x!} \times x^z$$

Given the set of *effective states*, $Q = A^{\mathrm{FS}}$, the *effective events* need only a *space* of $z \ln x$. Define the *space* of the *volume*, $v = |A^{\mathrm{C}}|$, as $\mathrm{spaceVolume} \in \mathbf{N}_{>0} \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceVolume}(v) := \ln v$$

Define the *effective space* of *histogram* A in a *system* U as $\mathrm{spaceEffective}(U) \in \mathcal{A}_U \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceEffective}(U)(A) := \mathrm{spaceVolume}(v) + \mathrm{spaceSubset}(v, x)$$

where $v = |A^{\mathrm{C}}|$ and $x = |A^{\mathrm{F}}|$. The *effective space* is undefined for zero *size*, $z = 0$.

Define the *effective events space* as $\mathrm{spaceEventsEffective}(U) \in \mathcal{H}_U \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceEventsEffective}(U)(H) := z \ln x$$

where $z = |H|$, $A = \mathrm{histogram}(H)$ and $x = |A^{\mathrm{F}}|$.

Define the *effective index history coder*

$$C_{\mathrm{H,F}} = \mathrm{coderHistoryIndexEffective}(U, X, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

The total *space* of the *effective index coder* $C_{\mathrm{H,F}}$ of a non-empty *history* $H \neq \emptyset$ is the sum of the *variables space*, *ids space*, *effective space* and *effective events space*

$$\begin{aligned} \mathrm{space}(C_{\mathrm{H,F}})(H) \quad = \quad & \mathrm{spaceVariables}(U)(|\mathrm{vars}(H)|) + \\ & \mathrm{spaceIds}(|X|, |H|) + \\ & \mathrm{spaceEffective}(U)(A) + \\ & \mathrm{spaceEventsEffective}(U)(H) \end{aligned}$$

where $A = \text{histogram}(H)$.

In the case where the *histogram* is *completely effective*, $A^{\text{F}} = A^{\text{C}}$, the *effective volume* equals the *volume*, $x = v$, so the *effective events space* equals the *events space*, $\text{spaceEventsEffective}(U)(H) = \text{spaceEvents}(U)(H)$. In this case the *space* of the *effective index coder* is greater than or equal to the *space* of the *index coder* $C^{\text{s}}_{\text{H,F}}(H) \geq C^{\text{s}}_{\text{H}}(H)$.

In the case where the *histogram* is a *singleton*, $x = |\text{ran}(H)| = |A^{\text{F}}| = 1$, the *effective events space* is zero, $\text{spaceEventsEffective}(U)(H) = 0$. The *effective space* is twice the *volume space*, $\text{spaceEffective}(U)(A) = 2 \ln v$, and so does not depend on the *size*, z. In the case where the *size* is greater than or equal to two, $z \geq 2$, the *space* of the *effective index coder* is less than or equal to the *space* of the *index coder* $C^{\text{s}}_{\text{H,F}}(H) \leq C^{\text{s}}_{\text{H}}(H)$.

3.18.2 Histogram space

Define the finite set of *trimmed integral histograms* in *system* U having *size* less than or equal to $y \in \mathbf{N}$ as

$$\mathcal{A}_{U,\text{i},\leq y} = \{\text{trim}(A) : A \in \mathcal{A}_{U,\text{i}},\ \text{size}(A) \leq y\}$$

In the case of a *regular system* U, having *dimension* $n = |U|$ and such that all the *variables* have the same *valency* $\{d\} = \{|W| : W \in \text{ran}(U)\}$, the cardinality of the set of *trimmed integral histograms* is such that $|\mathcal{A}_{U,\text{i},\leq y}| < y2^n |\mathcal{A}_{U,\text{i},\text{vars}(U),y}|$. Hence

$$|\mathcal{A}_{U,\text{i},\leq y}| < y2^n \frac{(y + d^n - 1)!}{y!\ (d^n - 1)!}$$

An explicit *minimal coder* $C_{\text{A,m}} \in \text{coders}(\mathcal{A}_{U,\text{i},\leq y})$ exists which, if given an order $D_{\text{A}} \in \mathcal{A}_{U,\text{i},\leq y} \leftrightarrow \mathbf{N}$ on the *histograms* in *system* U, simply enumerates the *coder domain* so that $\text{space}(C_{\text{A,m}})(A) = \ln |\mathcal{A}_{U,\text{i},\leq y}|$.

Consider a *coder* of *histograms*

$$C_{\text{A}} = \text{coderHistogram}(U, y, D_{\text{V}}, D_{\text{S}}) \in \text{coders}(\mathcal{A}_{U,\text{i},\leq y})$$

where $y = |X|$ is the cardinality of the *identifier set* X.

A *histogram* $A \in \mathcal{A}_{U,\text{i},\leq y}$ can be encoded in a tuple $T_A \in \mathbf{N}^2 \times \mathbf{N} \times \mathbf{N}$. This tuple T_A can then be encoded into a natural number $E_A \in \{0 \ldots S_A - 1\}$.

The *space* of the *histogram* is the *space* of this encoding, $\text{space}(C_A)(A) = \text{space}(S_A) = \ln S_A$.

The first pair of the tuple T_A encodes the set of *variables* in the same way as for the *history coder*, C_H. The *space* of the pair (n, N_V) is $\text{spaceVariables}(U)(n)$ where $V = \text{vars}(A)$ and $n = |V|$.

The second element of the tuple T_A encodes the *size* $z = \text{size}(A)$. Then $z \in \{0 \ldots y\}$ and the *space* of z is $\text{spaceSize}(y) = \ln(y + 1)$.

The last element of the tuple T_A encodes the set of *counts* of the *states*. The *coder* defined here has a *coder domain* of *trimmed histograms* rather than *complete histograms*. So, for example, the *histograms* of *histories*, $A = \text{histogram}(H)$, and the *histograms* of *unit transforms* $(X, \cdot) \in \mathcal{T}_U$, are members of the *coder domain*. Let D_A be an order on the *coder domain*

$$D_A \in \text{enums}(\mathcal{A}_{U,i,\leq y})$$

D_A can be chosen arbitrarily, or constructed from D_S. The *support* of a *multinomial distribution* is the set of *complete integral congruent histograms* in $\mathcal{A}_{U,i}$ having *variables* V and *size* z in *system* U, previously defined

$$\mathcal{A}_{U,i,V,z} := \{B : B \in \mathcal{A}_{U,i},\ B^U = V^C,\ \text{size}(B) = z\}$$

The *integral congruent support* consists of *complete histograms*. It forms a bijective map to the set of *trimmed integral congruent histograms*

$$\{\text{trim}(B) \in \mathcal{A}_{U,i,V,z}\} :\leftrightarrow: \mathcal{A}_{U,i,V,z}$$

The *integral congruent support* implies the equivalence classes in $\mathcal{A}_{U,i}$ such that $\text{trim}(A) \equiv A + A^{CZ}$. Thus for any *trimmed histogram* $A = \text{trim}(A)$ the *equivalent complete histogram* is in the *integral congruent support*, $A + A^{CZ} \in \mathcal{A}_{U,i,V,z}$. Given D_A, choose enumeration R of the enumerations of the *trimmed support* of the *multinomial distribution*

$$R \in \text{enums}(\{\text{trim}(B) : B \in \mathcal{A}_{U,i,V,z}\})$$

The *trimmed support* has the same cardinality as the *integral congruent support*, so the last element of the tuple T_A encoding the *counts* of the *states* of A is R_A which is such that

$$R_A \in \{1 \ldots \frac{(z + v - 1)!}{z!\ (v - 1)!}\}$$

R_A is the weak composition number of A. Define the *space* of a weak composition as $\mathrm{spaceCompositionWeak} \in \mathbf{N}_{>0} \times \mathbf{N} \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceCompositionWeak}(k,n) := \ln |\mathrm{C}'(\{1 \ldots k\},n)| = \ln \frac{(n+k-1)!}{n!\ (k-1)!}$$

Define the *space* of the encoding of the *counts* of the *states* of *histogram* A in a *system* U as $\mathrm{spaceCounts}(U) \in \mathcal{A}_{U,\mathrm{i}} \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceCounts}(U)(A) := \ln \frac{(z+v-1)!}{z!\ (v-1)!} = \mathrm{spaceCompositionWeak}(v,z)$$

where $z = \mathrm{size}(A)$, $V = \mathrm{vars}(A)$ and $v = \mathrm{volume}(U)(V)$. The *space* of R_A is $\mathrm{spaceCounts}(U)(A)$.

Finally the tuple $T_A = ((n, N_V), z, R_A)$ can be encoded to $E_A \in \{0 \ldots S_A - 1\}$ where

$$S_A = (r+1) \times \frac{r!}{(r-n)!\ n!} \times (y+1) \times \frac{(z+v-1)!}{z!\ (v-1)!}$$

The total *space* of the *coder* C_{A} of a *histogram* A is the sum of the *variables space*, *size space* and *counts space*

$$\begin{aligned}\mathrm{space}(C_{\mathrm{A}})(A) \quad = \quad & \mathrm{spaceVariables}(U)(|\mathrm{vars}(A)|) + \\ & \mathrm{spaceSize}(y) + \\ & \mathrm{spaceCounts}(U)(A)\end{aligned}$$

The *coder domain* $\mathcal{A}_{U,\mathrm{i},\leq y}$ of C_{A} excludes *integral histograms* having *zero states*, $A^{\mathrm{F}} \neq A^{\mathrm{U}}$, where $A \in \mathcal{A}_{U,\mathrm{i}}$ and $\mathrm{size}(A) \leq y$. However the *equivalent trimmed histogram*, $\mathrm{trim}(A) \equiv A$, will be in the *coder domain*, $\mathrm{trim}(A) \in \mathcal{A}_{U,\mathrm{i},\leq y}$, so there is always a means of *encoding equivalent histograms* although the *coder* cannot distinguish between them. An example is the *empty histogram* $\emptyset$ which is encoded, whereas its *completed equivalent* the *zero scalar histogram* $\{(\emptyset, 0)\}$ is not. However, $\mathrm{trim}(\{(\emptyset, 0)\}) = \emptyset$.

Also, note that the *histogram counts space* depends only on the *size* and *volume*, not the actual *counts* in a *histogram*. Nor does it depend on the cardinality of the *identifier set*.

Consider the *counts space* $\mathrm{spaceCounts}(U)(A)$ of *histogram* A in *system* U. Scale analysis would suggest that the *space* of a fixed width encoding $\mathcal{L}(\{1 \ldots z\})$ (similar to that of *history events space* $\mathcal{L}(\{1 \ldots v\})$) would be an upper bound, so that

$$\mathrm{spaceCounts}(U)(A) \leq v \ln z$$

where $z = \text{size}(A)$, $z > 0$, $v = \text{volume}(U)(V)$. This is in fact the case where $z > 1$ because

$$\frac{(z+v-1)!}{z!\ (v-1)!} = \frac{v}{z+v} \prod_{i \in \{1 \ldots v\}} \frac{z+i}{i} \leq z^v$$

By a symmetrical argument, if $z > 0$

$$\text{spaceCounts}(U)(A) \leq z \ln v$$

This is obvious in the case where A is *effectively complete* and so $z \geq v$, but also holds if $z < v$ because

$$\frac{(z+v-1)!}{z!\ (v-1)!} = \prod_{i \in \{1 \ldots z\}} \frac{v-1+i}{i} \leq v^z$$

Thus if A is the *histogram* of *history* H, $A = \text{histogram}(H)$, then the *history events space* is greater than the *histogram counts space* $\text{spaceCounts}(U)(A) \leq \text{spaceEvents}(U)(H)$.

In the case where $z = av^2$ where $a \in \mathbf{R}$ and $a \geq 1$

$$\frac{(z+v-1)!}{z!\ (v-1)!} = \frac{1}{av+1} \prod_{i \in \{1 \ldots v\}} \frac{av^2+i}{i} \geq (av+1)^{v-1}$$

Hence where $z \geq v^2$ and $z > 1$

$$(v-1) \ln v < \text{spaceCounts}(U)(A) \leq v \ln z < z \ln v$$

In the case where $v = az^2$ where $a \in \mathbf{R}$ and $a \geq 1$ and $z > 0$

$$\frac{(z+v-1)!}{z!\ (v-1)!} = \prod_{i \in \{1 \ldots z\}} \frac{az^2-1+i}{i} \geq (az)^z$$

Hence where $v \geq z^2$ and $z > 0$

$$z \ln z < \text{spaceCounts}(U)(A) \leq z \ln v$$

Similarly to the *effective index coder* $C_{\mathrm{H,F}}$, above, an *effective histogram coder* can be defined that encodes the *effective states* in a pair $(x, F_Q) \in \mathbf{N}^2$ added to the tuple, $T_A \in \mathbf{N}^2 \times \mathbf{N} \times \mathbf{N}^2 \times \mathbf{N}$. The pair encodes (i) the *effective volume* $x = |A^{\mathrm{F}}|$ and (ii) the index, F_Q, into an enumeration of the *effective states* subsets,

$$F_Q \in \{1 \ldots \frac{v!}{(v-x)!\ x!}\}$$

Given the set of *effective states*, $Q = A^{\mathrm{FS}}$, the *counts* can be encoded in a strong composition instead of a weak composition. The last element of the tuple, R_A, is now the strong composition number of A,

$$R_A \in \{1 \ldots \frac{(z-1)!}{(z-x)!\ (x-1)!}\}$$

The tuple $T_A = ((n, N_V), z, (x, F_Q), R_A)$ can be encoded to $E_A \in \{0 \ldots S_A - 1\}$ where

$$S_A = (r+1) \times \frac{r!}{(r-n)!\ n!} \times (y+1) \times v \times \frac{v!}{(v-x)!\ x!} \times \frac{(z-1)!}{(z-x)!\ (x-1)!}$$

Define the *space* of a strong composition as $\mathrm{spaceComposition} \in \mathbf{N}_{>0} \times \mathbf{N}_{>0} \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceComposition}(k, n) := \ln |\mathrm{C}(\{1 \ldots k\}, n)| = \ln \frac{(n-1)!}{(n-k)!\ (k-1)!}$$

Define the *effective counts space* as $\mathrm{spaceCountsEffective}(U) \in \mathcal{A}_{U,\mathrm{i}} \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceCountsEffective}(U)(A) := \ln \frac{(z-1)!}{(z-x)!\ (x-1)!} = \mathrm{spaceComposition}(x, z)$$

where $z = \mathrm{size}(A)$ and $x = |A^{\mathrm{F}}|$. The *effective counts space* is undefined for zero *size*, $z = 0$.

Define the *effective histogram coder*

$$C_{\mathrm{A,F}} = \mathrm{coderHistogramEffective}(U, y, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{A}_{U,\mathrm{i},\leq y})$$

The total *space* of the *effective histogram coder* $C_{\mathrm{A,F}}$ of a *non-zero histogram* A is the sum of the *variables space*, *size space*, *effective space* and *effective counts space*

$$\begin{aligned} \mathrm{space}(C_{\mathrm{A,F}})(A) \ &= \ \mathrm{spaceVariables}(U)(|\mathrm{vars}(A)|) + \\ &\quad\ \mathrm{spaceSize}(y) + \\ &\quad\ \mathrm{spaceEffective}(U)(A) + \\ &\quad\ \mathrm{spaceCountsEffective}(U)(A) \end{aligned}$$

The *effective histogram coder space*, $C^{\mathrm{s}}_{\mathrm{A,F}}(A)$, is less than the *histogram coder space*, $C^{\mathrm{s}}_{\mathrm{A}}(A)$, if $\mathrm{spaceEffective}(U)(A) + \mathrm{spaceCountsEffective}(U)(A) < \mathrm{spaceCounts}(U)(A)$,

$$\ln v + \ln \frac{v!}{(v-x)!\ x!} + \ln \frac{(z-1)!}{(z-x)!\ (x-1)!} < \ln \frac{(z+v-1)!}{z!\ (v-1)!}$$

In the case of an *effective singleton*, $x = |A^{\mathrm{F}}| = 1$, this simplifies to

$$2 \ln v < \ln \frac{(z+v-1)!}{z!\ (v-1)!}$$

In this case the *effective histogram coder space* does not depend on the *size*, z, so for given *volume* v there exists some *size* z such that the *effective histogram coder space* is less than the *histogram coder space*, $C^{\mathrm{s}}_{\mathrm{A,F}}(A) < C^{\mathrm{s}}_{\mathrm{A}}(A)$.

In the case of *unit uniform histogram*, $\mathrm{trim}(A) = A^{\mathrm{F}}$, the *effective volume* equals the *size*, $x = z$, and the inequality simplifies to

$$\frac{v\ v!}{(v-z)!} < \frac{(z+v-1)!}{(v-1)!}$$

In this case the *effective histogram coder space* is less than the *histogram coder space*, $C^{\mathrm{s}}_{\mathrm{A,F}}(A) < C^{\mathrm{s}}_{\mathrm{A}}(A)$, if the *size* is at least two, $z \geq 2$.

3.18.3 Classification space

Classifications are lossless transformations of *histories* and vice-versa

$$\begin{aligned}\forall H \in \mathcal{H}\ (\mathrm{history}(\mathrm{classification}(H)) = H)\\ \forall G \in \mathcal{G}\ (\mathrm{classification}(\mathrm{history}(G)) = G)\end{aligned}$$

Given a *system* $U \in \mathcal{U}$ and an *identifier set* $X \subset \mathcal{X}$, we can define $\mathcal{G}_{U,X} \subseteq \mathcal{S}_U \to (\mathrm{P}(X) \setminus \{\emptyset\})$. Now $\mathcal{G}_{U,X} :\leftrightarrow: \mathcal{H}_{U,X}$ and $\mathcal{H}_{U,X}$ is finite, so $\mathcal{G}_{U,X}$ is finite and can be constructed

$$\mathcal{G}_{U,X} = \bigcup \{\mathrm{cartesian}(U)(V) \leftrightarrow P : V \subset \mathrm{vars}(U),\ P \in \mathrm{B}(X)\} \cup \{\emptyset\}$$

Consider the *classification coder* of *histories*

$$C_{\mathrm{G}} = \mathrm{coderClassification}(U, X, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

The *coder domain* is the same as that of the *index coder*, C_{H}. A *classification* $G \in \mathcal{G}_{U,X}$, where $G = \mathrm{classification}(H)$ and $H \in \mathcal{H}_{U,X}$, can be encoded in a tuple $T_G \in \mathbf{N}^2 \times \mathbf{N}^2 \times \mathbf{N} \times \mathbf{N}$. This tuple T_G can then be encoded into a natural number $E_G \in \{0 \ldots S_G - 1\}$. The *space* of the *classification* is the *space* of this encoding, $\mathrm{space}(C_{\mathrm{G}})(H) = \mathrm{space}(S_G) = \ln S_G$.

The first pair of the tuple T_G encodes the set of *variables* in the same way as for the *history coder*, C_{H}, or the *histogram coder*, C_{A}. The *space* of the pair

(n, N_V) is spaceVariables$(U)(n)$ where $A = \text{histogram}(G)$, $V = \text{vars}(A)$ and $n = |V|$.

The second pair of the tuple T_G encodes the set of *event identifiers* in the same way as for the *history coder*, C_{H}. The *space* of the pair (z, Z_I) is spaceIds(y, z) where $y = |X|$ and $z = \text{size}(A)$.

The third element of the tuple T_G encodes the set of *counts* of the *states* in the same way as for the *histogram coder*, C_{A}. The *space* of the element R_A is spaceCounts$(U)(A)$.

The last element of the tuple T_G encodes the *classification* of the *events*. Let Q be the partition of *event identifiers* represented by non-empty $G \neq \emptyset$, $Q = \text{ran}(G) = \{C : (S, C) \in G\} \in \text{B}(\text{ids}(G))$. Given D_{X}, choose enumeration F of the enumerations of the partitions of the *event identifiers* corresponding to the *classification*

$F \in \text{enums}(\{P : P \in \text{B}(\text{ids}(G)),\ \exists X \in P :\leftrightarrow: Q\ \forall (Y, Z) \in X\ (|Y| = |Z|)\})$

The last element of the tuple T_G encoding the *classification* of the *event identifiers* of G is F_Q

$$F_Q \in \{1 \ldots \frac{z!}{\prod_{S \in \text{dom}(G)} |G_S|!}\}$$

Define the *space* of the encoding of the *classification* of the *event identifiers* of G having *histogram* $A = \text{histogram}(G)$ as spaceClassification $\in \mathcal{A}_{\text{i}} \to \ln \mathbf{N}_{>0}$

$$\text{spaceClassification}(A) := \ln z! - \sum_{S \in A^{\text{S}}} \ln A_S!$$

where $A \neq \emptyset$ and $z = \text{size}(A)$. Define spaceClassification$(\emptyset) := 0$. The *space* of F_Q is spaceClassification(histogram(G)).

This function can be defined more generically as spaceClassification $\in (\mathcal{X} \to \mathbf{N}) \to \ln \mathbf{N}_{>0}$

$$\text{spaceClassification}(Q) := \ln z! - \sum_{x \in \text{dom}(Q)} \ln Q_x!$$

where $z = \text{sum}(Q)$. With this definition spaceSubset is a special case of spaceClassification where $|Q| = 2$.

The *events classification space* is also called the *multinomial space* because the spaceClassification(histogram(G)) is the logarithm of the *multinomial coefficient* of its *histogram*

$$\text{spaceClassification}(A) = \ln z! - \sum_{S \in A^{\text{S}}} \ln A_S! = \ln \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!}$$

The *multinomial coefficient* forms part of the *multinomial distribution*,

$$Q_{\text{m}}(E, z)(A) = \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!} \prod_{S \in A^{\text{S}}} E_S^{A_S}$$

to count the permutations of the probabilities of the *distribution histogram's states*. The *multinomial coefficient* forms part of the *classification space* by counting the permutations or orders of the *events* within the partition by *state*.

Finally the tuple $T_G = ((n, N_V), (z, Z_I), R_A, F_Q)$ can be encoded to $E_G \in \{0 \ldots S_G - 1\}$ where

$$S_G = (r+1) \times \frac{r!}{(r-n)!\ n!} \times (y+1) \times \frac{y!}{(y-z)!\ z!} \times \frac{(z+v-1)!}{z!\ (v-1)!} \times \frac{z!}{\prod_{S \in \text{dom}(G)} |G_S|!}$$

where $z > 0$, otherwise

$$S_G = (r+1) \times \frac{r!}{(r-n)!\ n!} \times (y+1)$$

The total *space* of a *classification coder* of a *history* H is the sum of the *variables space*, *ids space*, *histogram counts space* and *events classification space*

$$\begin{aligned} \text{space}(C_{\text{G}})(H) \quad = \quad & \text{spaceVariables}(U)(|\text{vars}(H)|) + \\ & \text{spaceIds}(|X|, |H|) + \\ & \text{spaceCounts}(U)(A) + \\ & \text{spaceClassification}(A) \end{aligned}$$

where $A = \text{histogram}(H)$.

The *space* of a *non-empty non-scalar integral histogram* A in the *histogram coder*, C_{A}, is less than or equal to the *space* in the corresponding *classification* *coder*, $C_{\rm G}$, of its *history*, $A = \text{histogram}(H)$, because the *event identifiers* are aggregated into *counts*

$$S_G = (r+1) \times \frac{r!}{(r-n)!\ n!} \times (y+1) \times \frac{(z+v-1)!}{z!\ (v-1)!} \times \frac{y!}{(y-z)!\ \prod_{S \in \text{dom}(G)} |G_S|!}$$

and so

$$\text{space}(C_{\rm G})(H) = \text{space}(C_{\rm A})(A) + \ln \frac{y^{\underline{z}}}{\prod_{S \in A^{\rm S}} A_S!}$$

where $y^{\underline{z}}$ is the falling factorial, $y!/(y-z)!$. The second term is always positive because $y^{\underline{z}} \geq z!$. The part of the *space* of the *classification* that depends on the *volume*, and hence the *system*, is encapsulated in the *histogram space*.

Define a *generic classification coder* of *histories* $C_{\rm G,A}$ which is parameterised by an underlying *coder* of *histograms* $C \in \text{coders}(\mathcal{A}_{U,\rm i,\leq y})$

$$C_{\rm G,A} = \text{coderClassificationGeneric}(C, X, D_{\rm X}) \in \text{coders}(\mathcal{H}_{U,X})$$

The *generic classification coder* encodes the *variables space*, *size space* and *counts space* via the underlying *coder*. It adds the *ids space* (less *size space*) and *events classification space* such that

$$\begin{aligned} \text{space}(C_{\rm G,A})(H) \;\; = \;\; & \text{space}(C)(A) + \\ & \text{spaceIds}(y, |H|) - \text{spaceSize}(y) + \\ & \text{spaceClassification}(A) \end{aligned}$$

where $H \in \mathcal{H}_{U,X}$ and $A = \text{histogram}(H)$. In the case where the underlying *coder* is $C = C_{\rm A}$, which is constructed $C_{\rm A} = \text{coderHistogram}(U, y, D_{\rm V}, D_{\rm S})$, then the *generic classification coder space* equals the *classification coder space*, $\text{space}(C_{\rm G,A})(H) = \text{space}(C_{\rm G})(H)$.

Define the *effective classification coder* $C_{\rm G,F}$ with the *generic classification coder*, $C_{\rm G,A}$, parameterised by the *effective histogram coder*, $C_{\rm A,F}$,

$$C_{\rm G,F} = \text{coderClassificationGeneric}(C_{\rm A,F}, X, D_{\rm X}) \in \text{coders}(\mathcal{H}_{U,X})$$

The total *space* of the *effective classification coder* $C_{\rm G,F}$ of a non-empty *history* $H \neq \emptyset$ is the sum of the *variables space*, *ids space*, *effective space*, *effective counts space* and *events classification space*

$$\begin{aligned} \text{space}(C_{\rm G,F})(H) \;\; = \;\; & \text{spaceVariables}(U)(|\text{vars}(H)|) + \\ & \text{spaceIds}(|X|, |H|) + \\ & \text{spaceEffective}(U)(A) + \\ & \text{spaceCountsEffective}(U)(A) + \\ & \text{spaceClassification}(A) \end{aligned}$$

where $A = \text{histogram}(H)$.

The *effective classification coder space* is less than the *classification coder space* when the *effective histogram coder space* is less than the *histogram coder space*,

$$C^{\text{s}}_{\text{G,F}}(H) < C^{\text{s}}_{\text{G}}(H) \iff C^{\text{s}}_{\text{A,F}}(A) < C^{\text{s}}_{\text{A}}(A)$$

Compare the *classification coder*, C_{G}, and the *index coder*, C_{H}, using a couple of special cases.

In the case of *singleton classifications*, $A = \text{histogram}(G)$ such that $|A^{\text{F}}| = 1$, the *space* of the *classification* is less than or equal to the *space* of the *history* $H = \text{history}(G)$. That is, $\text{space}(C_{\text{G}})(H) \leq \text{space}(C_{\text{H}})(H)$

$$\begin{aligned}
&\text{spEv}(U)(H) - \text{spCt}(U)(A) - \text{spCl}(A) \\
&= \quad z \ln v - \ln \frac{(z+v-1)!}{z!\ (v-1)!} - \ln \frac{z!}{z!} \geq 0
\end{aligned}$$

because $\text{spCt}(U)(A) \leq z \ln v$, where $\text{spEv} = \text{spaceEvents}$, $\text{spCt} = \text{spaceCounts}$ and $\text{spCl} = \text{spaceClassification}$.

In the case of *completely effective uniform classifications*, $A = \text{histogram}(G)$ such that $A^{\text{F}} = A^{\text{C}}$ and $|\text{ran}(A)| = 1$, the *space* of the *classification* is greater than or equal to the *space* of the *history* $H = \text{history}(G)$. That is, $\text{space}(C_{\text{G}})(H) \geq \text{space}(C_{\text{H}})(H)$. Let $z = kv$ where $k \in \mathbf{N}$ and $k \geq 1$

$$\begin{aligned}
&\text{spEv}(U)(H) - \text{spCt}(U)(A) - \text{spCl}(A) \\
&= \quad z \ln v - \ln \frac{(z+v-1)!}{z!\ (v-1)!} - \ln \frac{z!}{((z/v)!)^v} \\
&= \quad kv \ln v - \ln(kv+v-1)! + \ln(v-1)! + v \ln k! \\
&= \quad \ln \prod_{c \in \{1 \ldots kv\}} \frac{v \sqrt[k]{k!}}{c+v-1} \leq 0
\end{aligned}$$

This is obviously true for $k = 1$. It is conjectured to be true for other values of k. Also the conjecture only applies to the case where there is an integral ratio $z/v \in \mathbf{N}$.

The special case of the *singleton classification*, $|A^{\text{F}}| = 1$, forms a minimum for the *events classification space*

$$\text{spaceClassification}(A) \geq 0$$

The second case of *completely effective uniform classifications*, $A = Z_{z/v} * V^{\mathrm{C}}$ where $Z_x = \mathrm{scalar}(x)$, forms a maximum

$$\mathrm{spaceClassification}(A) \leq \ln z! - v \ln(z/v)!$$

where $z/v \in \mathbf{N}$. This is a local maximum because $(z/v-1)!(z/v+1)! > ((z/v)!)^2$. Proof that it is everywhere a maximum can be shown by means of sequences of *perturbations*. See the similar case of the *mean* of the *multinomial probability distribution* of a *cartesian distribution* above. The *probability* of the *mean histogram* is the maximum *frequency*

$$\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}}, z)(M) = \frac{z!}{(\frac{z}{v}!)^v} \left(\frac{1}{v}\right)^z$$

where $M = \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}}, z)) = Z_{z/v} * V^{\mathrm{C}}$.

Apply Stirling's approximation to the maximum case

$$\begin{aligned} \ln z! - v \ln(z/v)! &< z \ln z - v \left(\frac{z}{v} \ln \frac{z}{v} - \frac{z}{v}\right) \\ &= z(\ln v + 1) \end{aligned}$$

Thus

$$\mathrm{spaceClassification}(A) < z(\ln v + 1)$$

The *classification coder* and *index coder* cannot differ by more than $z(\ln v+1)$ and so are always of the same order of complexity

$$\mathrm{space}(C_{\mathrm{G}}) \in \mathrm{O}(\mathrm{space}(C_{\mathrm{H}}), 2)$$

The appearance of the *multinomial coefficient* in the *classification coder* suggests a relationship with *entropy*. In fact, the *events classification space*, spaceClassification, can be approximated by the *sized entropy*, by the use of Stirling's approximation

$$\begin{aligned} \mathrm{spaceClassification}(A) &:= \ln z! - \sum_{S \in A^{\mathrm{S}}} \ln A_S! \\ &\approx z \ln z - z - \sum_{S \in A^{\mathrm{FS}}} (A_S \ln A_S - A_S) \\ &= z \ln z - z \sum_{S \in A^{\mathrm{FS}}} N_S \ln z N_S \\ &= -z \sum_{S \in A^{\mathrm{FS}}} N_S \ln N_S \\ &= z \times \mathrm{entropy}(A) \end{aligned}$$

where $z > 0$ and $N = \text{resize}(1, A)$. Thus *events classification space* increases with increasing *entropy*. This is consistent with the two special cases above. In the first case, the *singleton histogram* has low *entropy* and the *classification coder* requires less *space* than the *index coder*. In the second case, the *uniform histogram* has high *entropy* and hence the reverse is true.

Consider the *scaled entropy*, $z \times \text{entropy}(A)$, at the break-even *space*, where the *classification coder space* equals the *index coder space*, $C^{\text{s}}_{\text{G}}(H) = C^{\text{s}}_{\text{H}}(H)$,

$$\begin{aligned} z \times \text{entropy}(A) &\approx z \ln v \; - \; \ln \frac{(z+v-1)!}{z!\;(v-1)!} \\ &\approx z \ln v \; - \; ((z+v)\ln(z+v) \; - \; z \ln z \; - \; v \ln v) \end{aligned}$$

That is, if the *scaled entropy*, $z \times \text{entropy}(A)$, of a *history*, H, is greater than the break-even *scaled entropy*, $z \ln v - ((z+v)\ln(z+v) - z \ln z - v \ln v)$, then the *index coder* requires less *space* than the *classification coder*, $C^{\text{s}}_{\text{G}}(H) > C^{\text{s}}_{\text{H}}(H)$, and vice-versa.

Also, *entropy* can be the basis of a *coder*. See the Appendix 'Entropy encoding of states'.

The *events classification space* does not depend on the internal structure of the *states* in the *classification* G, only that the *states* form a functional domain. That is, *events classification space* does not depend on the *variables* in the *states* nor their *valencies*, except in respect of the total *volume* and the equivalence of *states*. To demonstrate, create a *unit functional transform* $T = (X, \{w\})$ with a single *derived variable* w such that if we choose an enumeration of the *cartesian states*, $Q \in \text{enums}(A^{\text{CS}})$, where $A = \text{histogram}(G)$ and such that $U_w = \text{ran}(Q)$ and $X = \{(S \cup \{(w,i)\}, 1) : (S,i) \in Q\}$, then we have the same *events classification space* $\text{spaceClassification}(A * T) = \text{spaceClassification}(A)$.

Conjecture that neither the *index coder* nor the *classification coder* of *histories* is a *minimal coder*,

$$\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ \forall C \in \{C_{\text{H}}, C_{\text{G}}\}\ (\sum_{H \in \mathcal{H}_{U,X}} \text{space}(C)(H) > |\mathcal{H}_{U,X}| \ln |\mathcal{H}_{U,X}|)$$

This is the case even for only *variate histories*. Let

$$\mathcal{H}_{U,X,\text{v}} = \{H : H \in \mathcal{H}_{U,X},\ |\text{vars}(H)| > 0\}$$

The *variate histories*, $\mathcal{H}_{U,X,\mathrm{v}}$, exclude *empty histories*, so $z > 0$. The *space* of the *coders* for the *variate coder domain* is calculated by subtracting the *space* required for *non-variate*

$$\mathrm{space}(C_{\mathrm{G,v}})(H) = \mathrm{space}(C_G)(H) - \ln \frac{r+1}{r} - \ln \frac{y+1}{y}$$

However, the *total space* of the *variate-classification coder* of *histories*, $C_{\mathrm{G,v}}$, is still greater than a *minimal coder* of $\mathcal{H}_{U,X,\mathrm{v}}$.

A *classification* $G \in \mathcal{G}_{U,X}$ is the inverse of its corresponding *history*, $G = \mathrm{classification}(H) = H^{-1}$ where $H \in \mathcal{H}_{U,X}$, and so it may be viewed as a relation between (i) the *effective states*, $\mathrm{dom}(G) = \mathrm{ran}(H) = A^{\mathrm{FS}}$, where $A = \mathrm{histogram}(H)$, and (ii) the components of a partition of the *event identifiers*, $\mathrm{ran}(G) \in \mathrm{B}(\mathrm{dom}(H))$. That is, $G \in A^{\mathrm{FS}} :\to: \mathrm{ran}(G)$. Thus a *classification* may be encoded by encoding (i) the *effective states*, A^{FS}, (ii) the permutation of the components of the partition, $x!$ where $x = |A^{\mathrm{FS}}|$, given some order on the *states*, D_{S}, and (iii) the partition index, $\{1 \ldots \mathrm{stir}(z,x)\}$ where $z = |H|$ and the Stirling number of the second kind is $\mathrm{stir}(n,k) := |\mathrm{S}(\{1 \ldots n\},k)|$.

Define the *partition classification coder*,

$$C_{\mathrm{G,B}} = \mathrm{coderClassificationPartition}(U,X,D_{\mathrm{V}},D_{\mathrm{S}},D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

Define the *space* of the encoding of the permutation and index of the *event identifiers* components as $\mathrm{spaceEventsComponents}(U) \in \mathcal{A}_{U,\mathrm{i}} \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceEventsComponents}(U)(A) := \ln x! + \ln \mathrm{stir}(z,x)$$

where $x = |A^{\mathrm{F}}|$ and $z = \mathrm{size}(A)$.

The total *space* of the *partition classification coder* $C_{\mathrm{G,B}}$ of a non-empty *history* $H \neq \emptyset$ is the sum of the *variables space*, *ids space*, *effective space*, and *events components space*

$$\begin{aligned}\mathrm{space}(C_{\mathrm{G,B}})(H) \;=\;& \mathrm{spaceVariables}(U)(|\mathrm{vars}(H)|) + \\ & \mathrm{spaceIds}(|X|,|H|) + \\ & \mathrm{spaceEffective}(U)(A) + \\ & \mathrm{spaceEventsComponents}(U)(A)\end{aligned}$$

where $A = \mathrm{histogram}(H)$.

The *partition classification coder*, $C_{\mathrm{G,B}}$, is less dependent on the *histogram*, A, than the *effective classification coder*, $C_{\mathrm{G,F}}$, because it depends on the *effectiveness* of the *states* and the total of the *counts*, rather than the individual *counts* of the *states*. That is, spaceEventsComponents$(U)(A)$, only depends on (i) the *size*, $z = \mathrm{size}(A)$, and (ii) the *effective volume*, $x = |A^{\mathrm{F}}|$, whereas spaceClassification(A) depends on the *histogram* as a relation, $A \in \mathcal{S} \to \mathbf{Q}_{\geq 0}$. In this, the *partition classification coder*, $C_{\mathrm{G,B}}$, resembles the *effective index coder*, $C_{\mathrm{H,F}}$, where the *effective events space*, spaceEventsEffective$(U)(H) := z \ln x$, only depends on the *size*, z and *effective volume*, x . Thus the *partition classification coder*, $C_{\mathrm{G,B}}$, is intermediate between the *effective index coder*, $C_{\mathrm{H,F}}$, and the *effective classification coder*, $C_{\mathrm{G,F}}$.

3.18.4 Derived history space

In section 'Histogram space', above, the *histogram coder* is constructed

$$C_{\mathrm{A}} = \mathrm{coderHistogram}(U, y, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{A}_{U,\mathrm{i},\leq y})$$

where $y = |X|$ is the cardinality of the *identifier set* $X \subset \mathcal{X}$, and $\mathcal{A}_{U,\mathrm{i},\leq y}$ is the set of *trimmed integral histograms* in *system* U having *size* less than or equal to y.

The *histogram coder*, C_{A}, is defined such that the *space* of *trimmed integral histogram* $A \in \mathcal{A}_{U,\mathrm{i},\leq y}$ is

$$\begin{aligned}\mathrm{space}(C_{\mathrm{A}})(A) &= \mathrm{spaceVariables}(U)(|V|) + \\ &\quad \mathrm{spaceSize}(y) + \\ &\quad \mathrm{spaceCounts}(U)(A)\end{aligned}$$

where $V = \mathrm{vars}(A)$.

The *histogram counts space* is the weak composition *space*,

$$\mathrm{spaceCounts}(U)(A) := \ln \frac{(z+v-1)!}{z!\ (v-1)!} = \mathrm{spaceCompositionWeak}(v, z)$$

where $z = \mathrm{size}(A)$, and $v = \mathrm{volume}(U)(V)$. In the case where the *size* is less than or equal to the *volume*, $z \leq v$, the *counts space* may be approximated

$$\begin{aligned}\ln \frac{(z+v-1)!}{z!\ (v-1)!} &= \overline{z} \ln v - \underline{z} \ln z \\ &\approx z \ln \frac{v}{z}\end{aligned}$$

by abuse of notation. If the *size*, z, is fixed, the *histogram counts space*, $\mathrm{spaceCounts}(U)(A)$, varies with the logarithm of the *volume*, $\ln v$.

In the case where the *size* is greater than the *volume*, $z > v$, the *counts space* approximates

$$\begin{aligned}
\ln \frac{(z+v-1)!}{z!\ (v-1)!} &\approx \overline{v} \ln z - \underline{v} \ln v \\
&\approx v \ln \frac{z}{v}
\end{aligned}$$

If the *volume*, v, is fixed, the *histogram counts space*, $\mathrm{spaceCounts}(U)(A)$, varies with the logarithm of the *size*, $\ln z$.

The weak composition *space* may also be analysed by means of Stirling's approximation,

$$\begin{aligned}
\ln \frac{(z+v-1)!}{z!\ (v-1)!} &\approx (z+v)\ln(z+v) \ - \ z \ln z \ - \ v \ln v \ - \ \ln \frac{z+v}{v} \\
&= z \ln \frac{z+v}{z} \ + \ (v-1) \ln \frac{z+v}{v} \\
&\approx (z \ln \frac{v}{z} \ : \ z < v) \ + \\
&\quad\ (2z \ln 2 \ : \ z = v) \ + \\
&\quad\ (v \ln \frac{z}{v} \ : \ z > v)
\end{aligned}$$

The following discussion considers how the *volume*, v, or *size*, z, may be broken into *components* by means of partitioning the *volume* with a *transform*, in order to reduce the overall *counts space*. That is, a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ has an *inverse*, $T^{-1} = \mathrm{inverse}(T)$, which implies a *partition* of the *volume* $T^{\mathrm{P}} \in \mathrm{B}(V^{\mathrm{CS}})$ where $T^{\mathrm{P}} = \{C^{\mathrm{S}} : C \in \mathrm{ran}(T^{-1})\}$.

Let the *substrate histogram coder* $C_{\mathrm{A},V} \in \mathrm{coders}(\mathcal{A}_{U,V,\mathrm{i},\leq y})$ be a *histogram coder*, C_{A}, but with given *variables* V. The *coder domain*, $\mathcal{A}_{U,V,\mathrm{i},\leq y}$, is the subset of the *trimmed integral histograms* of *size* less than or equal to y which are also in *variables* V. The *substrate histograms* is defined

$$\mathcal{A}_{U,V,\mathrm{i},\leq y} = \{\mathrm{trim}(A) : A \in \mathcal{A}_{U,\mathrm{i}},\ \mathrm{size}(A) \leq y,\ \mathrm{vars}(A) = V\}$$

which has cardinality

$$|\mathcal{A}_{U,V,\mathrm{i},\leq y}| = \sum_{z \in \{1 \ldots y\}} \frac{(z+v-1)!}{z!\ (v-1)!}$$

where $v = |V^{\mathrm{C}}|$.

The *substrate histogram coder* is such that

$$\mathrm{space}(C_{\mathrm{A},V})(A) \quad = \quad \mathrm{space}(C_{\mathrm{A}})(A) - \mathrm{spaceVariables}(U)(V)$$

The *substrate histogram coder space* is

$$\begin{aligned}\mathrm{space}(C_{\mathrm{A},V})(A) \quad = \quad & \mathrm{spaceSize}(y) + \\ & \mathrm{spaceCounts}(U)(A)\end{aligned}$$

Now consider the *derived substrate histogram coder* $C_{\mathrm{A},V,\mathrm{T}}$ given *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ in *variables* $V = \mathrm{und}(T)$. The *coder domain* is the subset of the *trimmed integral histograms* of *size* less than or equal to y which are also in *variables* V, $\mathcal{A}_{U,V,\mathrm{i},\leq y}$. The *derived substrate histogram coder* is instantiated

$$\begin{aligned}& C_{\mathrm{A},V,\mathrm{T}} = \\ & \quad \mathrm{coderHistogramSubstrateDerived}(U, y, T, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{A}_{U,V,\mathrm{i},\leq y})\end{aligned}$$

The *histogram*, $A \in \mathcal{A}_{U,V,\mathrm{i},\leq y}$, can be encoded in an intermediate tuple $T_A = (z, R'_{A*T}, L) \in \mathbf{N} \times \mathbf{N} \times \mathcal{L}(\mathbf{N})$.

There is no need to encode the *variables*, V, because these are defined by the *transform*, T, in the *derived substrate histogram coder* parameters. So the first element, $z \in \mathbf{N}$, of the tuple, T_A, encodes the *size* $z = \mathrm{size}(A) \in \{0 \ldots y\}$, which has *space* $\mathrm{spaceSize}(y) = \ln(y + 1)$.

The second element, $R'_{A*T} \in \mathbf{N}$, of the tuple, T_A, is the encoding of the *counts* of the *possible derived volume* of the *derived histogram*, $A * T$,

$$R'_{A*T} \in \{1 \ldots \frac{(z + w' - 1)!}{z!\ (w' - 1)!}\}$$

where $W = \mathrm{der}(T)$, *derived volume* $w = |W^{\mathrm{C}}|$ and *possible derived volume* $w' = |(V^{\mathrm{C}} * T)^{\mathrm{F}}| = |T^{-1}| \leq w$. This is similar to the encoding of the *counts* in the *histogram coder*, C_{A}, but the *derived coder* excludes necessarily *ineffective states*, $W^{\mathrm{CS}} \setminus (V^{\mathrm{C}} * T)^{\mathrm{FS}}$, that occur when the *transform* is *overlapped*, $\mathrm{overlap}(T)$. The *derived coder* can compute the *possible derived volume*, w', at instantiation because it depends only on the *transform*, T, which is a parameter of the constructor. The *space* is $\mathrm{spaceCountsDerived}(U)(A, T)$

where $\mathrm{spaceCountsDerived}(U) \in \mathcal{A}_{U,\mathrm{i}} \times \mathcal{T}_{U,\mathrm{f},1} \to \ln \mathbf{N}_{>0}$ is defined

$$\begin{aligned}\mathrm{spaceCountsDerived}(U)(A,T) &:= \ln \frac{(z+w'-1)!}{z!\ (w'-1)!} \\ &= \mathrm{spaceCompositionWeak}(w',z)\end{aligned}$$

The *possible derived volume* is less than or equal to the *derived volume*, $w' \leq w$, so the *derived counts space* is no greater than the *counts space* of the *derived histogram*, $\mathrm{spaceCountsDerived}(U)(A,T) \leq \mathrm{spaceCounts}(U)(A * T)$. The *possible derived volume* equals the *derived volume* if and only if the *transform* is *non-overlapped*, $\neg\mathrm{overlap}(T) \iff w' = w$, because it is only in this case that the *transform* is *right total*, $\mathrm{dom}(T^{-1}) = W^{\mathrm{CS}}$. In this case, the *derived counts space* equals the *counts space* of the *derived histogram*, $\neg\mathrm{overlap}(T) \iff \mathrm{spaceCountsDerived}(U)(A,T) = \mathrm{spaceCounts}(U)(A * T)$. The *possible derived volume* is less than or equal to the *underlying volume*, $w' \leq v$, so the *derived counts space* is no greater than the *counts space* of the *underlying histogram*, $\mathrm{spaceCountsDerived}(U)(A,T) \leq \mathrm{spaceCounts}(U)(A)$.

The last element, $L \in \mathcal{L}(\mathbf{N})$, of the tuple, T_A, is a list of the encodings of the *counts* of each of the *components* of the *transform inverse*, T^{-1},

$$L_{M'(R)} \in \{1 \ldots \frac{((A * T)_R + |C| - 1)!}{(A * T)_R!\ (|C| - 1)!}\}$$

where $(R,C) \in T^{-1}$, $M' = \mathrm{order}(D_{\mathrm{S}}, W^{\mathrm{CS}}) \in \mathrm{enums}(W^{\mathrm{CS}})$ and D_{S} is some *order* on the *states* in *system* U. Define the *partition components counts space* as $\mathrm{spaceCountsPartition} \in \mathcal{A} \times \mathcal{T}_{\mathrm{f}} \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceCountsPartition}(A,T) := \sum_{(R,C) \in T^{-1}} \ln \frac{((A * T)_R + |C| - 1)!}{(A * T)_R!\ (|C| - 1)!}$$

The *partition components counts space* required for an *ineffective component*, $(A * T)_R = 0$, is zero, regardless of the *component's cardinality*. That is, $\mathrm{spaceCompositionWeak}(|C|, 0) = 0$. The *partition components counts space* is also zero where the *component* is a singleton, $|C| = 1$, regardless of the *derived count*, $\mathrm{spaceCompositionWeak}(1, (A * T)_R) = 0$. If all of the *derived counts* are in singleton *components*, $\forall (R,C) \in T^{-1}\ ((A * T)_R > 0 \implies |C| = 1)$, then overall *partition components counts space* is zero, $\mathrm{spaceCountsPartition}(A,T) = 0$.

The *derived substrate histogram coder space* is

$$\begin{aligned}\mathrm{space}(C_{\mathrm{A},V,\mathrm{T}})(A) \quad = \quad & \mathrm{spaceSize}(y) + \\ & \mathrm{spaceCountsDerived}(U)(A,T) + \\ & \mathrm{spaceCountsPartition}(A,T)\end{aligned}$$

Compare the *derived substrate histogram coder*, $C_{\mathrm{A},V,\mathrm{T}}$, to the *substrate histogram coder*, $C_{\mathrm{A},V}$.

In the case of the *full functional transform*, $T = \{\{u\}^{\mathrm{CS\{\}T}} : u \in V\}^{\mathrm{T}}$, the *partition components counts space* is zero, $\mathrm{spaceCountsPartition}(A,T) = 0$, because the *derived volume* equals the *underlying volume*, $|W^{\mathrm{C}}| = |V^{\mathrm{C}}|$, and the *transform* is *right total*, $\forall(\cdot, C) \in T^{-1}\ (|C| = 1)$. In this case the *derived counts space* equals the *underlying counts space*,

$$\mathrm{spaceCountsDerived}(U)(A, \{\{u\}^{\mathrm{CS\{\}T}} : u \in V\}^{\mathrm{T}}) \quad = \quad \mathrm{spaceCounts}(U)(A)$$

and so the *derived substrate histogram coder space* equals the *substrate histogram coder space*, $C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) = C^{\mathrm{s}}_{\mathrm{A},V}(A)$.

Conversely, in the case of the *unary transform*, $T = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, the *derived counts space* is zero, $\mathrm{spaceCountsDerived}(U)(A,T) = 0$, because the *derived volume* is a singleton, $|W^{\mathrm{C}}| = 1$. The *partition components counts space* equals the *underlying counts space*,

$$\mathrm{spaceCountsPartition}(A, \{V^{\mathrm{CS}}\}^{\mathrm{T}}) \quad = \quad \mathrm{spaceCounts}(U)(A)$$

because there is only one *component*, $C = V^{\mathrm{C}}$. So in this case also, the *derived substrate histogram coder space* equals the *substrate histogram coder space*, $C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) = C^{\mathrm{s}}_{\mathrm{A},V}(A)$.

In the domain where the *size* is less than or equal to the *possible derived volume*, $z \leq w'$, the *derived counts space* varies with the *log possible derived volume*,

$$\mathrm{spaceCountsDerived}(U)(A,T) \quad \sim \quad \ln w'$$

In the domain where the *size* is greater than the *possible derived volume*, $z > w'$, the *derived counts space* varies with the *possible derived volume*,

$$\mathrm{spaceCountsDerived}(U)(A,T) \quad \sim \quad w'$$

The *partition components counts space* can be defined in terms of *derived state* or *component*,

$$\begin{aligned}
\text{spaceCountsPartition}(A,T) &= \sum_{(R,C) \in T^{-1}} \ln \frac{((A*T)_R + |C| - 1)!}{(A*T)_R!\ (|C|-1)!} \\
&= \sum_{(R,\cdot) \in T^{-1}} \ln \frac{((A*T)_R + (V^{\mathrm{C}} * T)_R - 1)!}{(A*T)_R!\ ((V^{\mathrm{C}} * T)_R - 1)!} \\
&= \sum_{(\cdot,C) \in T^{-1}} \ln \frac{(\text{size}(A*C) + |C| - 1)!}{\text{size}(A*C)!\ (|C|-1)!}
\end{aligned}$$

This is just the logarithm of the cardinality of the set of *integral iso-deriveds*,

$$\text{spaceCountsPartition}(A,T) = \ln |D^{-1}_{U,\mathrm{i},T,z}(A*T)|$$

The *integral iso-deriveds log-cardinality* is discussed in 'Integral iso-sets and entropy', above.

In the case where the *volume* is much greater than one, $v \gg 1$, the *partition components counts space* varies against the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}
&\text{spaceCountsPartition}(A,T) \sim \\
&\quad -\big((z+v) \times \text{entropy}(A*T + V^{\mathrm{C}} * T) \\
&\qquad\qquad -z \times \text{entropy}(A*T)\ - v \times \text{entropy}(V^{\mathrm{C}} * T)\big)
\end{aligned}$$

In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *partition components counts space* varies against the *size* scaled *component size cardinality relative entropy*,

$$\text{spaceCountsPartition}(A,T) \sim -z \times \text{entropyRelative}(A*T, V^{\mathrm{C}} * T)$$

Similarly, in the domain where the *size* is greater than the *volume*, $z > v$, the *partition components counts space* varies against the *volume* scaled *component cardinality size relative entropy*,

$$\text{spaceCountsPartition}(A,T) \sim -v \times \text{entropyRelative}(V^{\mathrm{C}} * T, A*T)$$

In both domains the *partition components counts space* varies against the *relative entropy*. That is, *partition components counts space* is minimised when (a) the *cross entropy* is maximised and (b) the *component entropy* is minimised. The *cross entropy* is maximised when high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

Consider the difference in *space* between the *derived substrate histogram coder*, $C_{\mathrm{A},V,\mathrm{T}}$, and the *substrate histogram coder*, $C_{\mathrm{A},V}$,

$$\begin{aligned} & C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A) \\ & \quad = \mathrm{spaceCountsDerived}(U)(A,T) + \mathrm{spaceCountsPartition}(A,T) - \\ & \qquad \mathrm{spaceCounts}(U)(A) \\ & \quad = \ln \frac{(z+w'-1)!}{z!\ (w'-1)!} + \sum_{(R,C) \in T^{-1}} \ln \frac{((A*T)_R + |C| - 1)!}{(A*T)_R!\ (|C|-1)!} - \\ & \qquad \ln \frac{(z+v-1)!}{z!\ (v-1)!} \end{aligned}$$

for each of three domains. First, in the domain where the *size* is less than or equal to the *possible derived volume*, $z \leq w' \leq v$, the difference varies with the *log possible derived volume* and against the *component size cardinality relative entropy*,

$$C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A) \sim \ln w' - \mathrm{entropyRelative}(A*T, V^{\mathrm{C}}*T)$$

In the domain where the *size* is between the *possible derived volume* and the *volume*, $w' \leq z \leq v$, the difference varies with the *possible derived volume* and against the *size* scaled *component size cardinality relative entropy*,

$$C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A) \sim w' - z \times \mathrm{entropyRelative}(A*T, V^{\mathrm{C}}*T)$$

Last, in the domain where the *size* is greater than the *volume*, $w' \leq v < z$, the difference varies with the *possible derived volume* and against the *volume* scaled *component cardinality size relative entropy*,

$$C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A) \sim w' - v \times \mathrm{entropyRelative}(V^{\mathrm{C}}*T, A*T)$$

So the *space* of the *derived substrate histogram coder*, $C_{\mathrm{A},V,\mathrm{T}}$, is minimised when (a) the *possible derived volume* is minimised, (b) the *component entropy* is minimised, and (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

For example, consider the *mono-variate tri-valent singleton cardinal substrate histogram* $A = \{(\{(u,1)\},z)\}$, with *binary partition transform* $T =$

$\{\{\{(u,1)\}\},\{\{(u,2)\},\{(u,3)\}\}\}^{\mathrm{T}}$ which *rolls* the *ineffective states* into a single *derived state*. The difference in *space* is exactly

$$\begin{aligned}
C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A) &= \ln \frac{(z+2-1)!}{z!\ (2-1)!} + \ln \frac{(z+1-1)!}{z!\ (1-1)!} + \ln \frac{(0+2-1)!}{0!\ (2-1)!} - \\
&\quad \ln \frac{(z+3-1)!}{z!\ (3-1)!} \\
&= -\ln(z+2) + \ln 2 \\
&< 0
\end{aligned}$$

where $z > 0$. This example has high *sizes* in low *cardinalities* and vice-versa, unconstrained by domain.

If the *derived histogram* is *independent*, $A * T = (A * T)^{\mathrm{X}}$, the *derived* equals the *abstract* and the *transform* is *surreal*, $A = (A * T)^{\mathrm{X}} * T^{\odot A}$. In this case the *derived histogram*, $A * T$, may be encoded by means of a *perimeter coder* of *histograms*, $C_{\mathrm{A,p}}$. The *space* of the encoding of the *perimeter* is

$$\mathrm{spacePerimeter}(U)(A * T) := \sum_{u \in W} \ln \frac{(z + |U_u| - 1)!}{z!\ (|U_u| - 1)!}$$

When compared to the *histogram coder*, C_{A}, the *counts space* of the *derived histogram* decreases, $C^{\mathrm{s}}_{\mathrm{A,p}}(A*T) - C^{\mathrm{s}}_{\mathrm{A}}(A*T) = \mathrm{spacePerimeter}(U)(A*T) - \mathrm{spaceCounts}(U)(A * T) \leq 0$. The *derived histogram* is *independent* and so tends to have higher *entropy*, $\mathrm{entropy}(A * T) = \mathrm{entropy}((A * T)^{\mathrm{X}})$. *Derived perimeter* encoding should be used in *derived coders* where the *derived entropy* is expected to be high.

In contrast, if the *partition components* are *independent*, $\forall C \in \mathrm{ran}(T^{-1})\ (A * C = (A * C)^{\mathrm{X}})$, the *transform* is *ideal*, $A = A * T * T^{\dagger A}$. In this case, each of the *components* may be encoded by means of a *perimeter histogram coder* having *space*

$$\sum_{(R,C) \in T^{-1}} \sum_{u \in V} \ln \frac{(z_R + |(C\%\{u\})^{\mathrm{F}}| - 1)!}{z_R!\ (|(C\%\{u\})^{\mathrm{F}}| - 1)!}$$

Note that the *component* may be larger than the *effective cartesian sub-volume*

$$(A * C)^{\mathrm{F}} = (A * C)^{\mathrm{XF}} \leq C$$

If the *component* is not a *cartesian sub-volume*, $C \neq C^{\mathrm{X}}$, the *perimeter histogram space* of the *component* may be greater than the *histogram space*.

Now consider *derived substrate history coders* given *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ in *variables* $V = \mathrm{und}(T)$. The *substrate coders' domain* $\mathcal{H}_{U,V,X} \subset \mathcal{H}_U$ is the subset of the *histories* in *system* U where the *event identifiers* are in the *event identifiers* set $X \subset \mathcal{X}$, and the *variables* are the given set $V \subseteq \mathrm{vars}(U)$,

$$\mathcal{H}_{U,V,X} = \{H : H \in \mathcal{H}_U,\ \mathrm{ids}(H) \subseteq X,\ \mathrm{vars}(H) = V\}$$

which has cardinality

$$|\mathcal{H}_{U,V,X}| = \sum_{z \in \{1 \ldots y\}} \binom{y}{z} v^z$$

where $v = |V^{\mathrm{C}}|$.

The *derived substrate histogram coder*, $C_{\mathrm{A},V,\mathrm{T}}$, divides the *histogram* encoding between (i) a *derived histogram* encoding, and (ii) a set of *component sub-histogram* encodings. Similarly, the *derived substrate history coders* divide the *history* encoding into (i) a *derived history* encoding, and (ii) a set of *component sub-history* encodings. The *canonical history coders* are (i) the *index history coder*, C_{H}, and (ii) the *classification history coder*, C_{G}. The *index history coder*, C_{H}, requires less *space* to encode high *entropy histories* than the *classification history coder*, C_{G}, and vice-versa. Each of the *derived history* and *component sub-histories* may be encoded with each of the *canonical history coders*. So there are four possible *derived substrate history coders*: (a) the *index derived substrate history coder* $C_{\mathrm{H},V,\mathrm{T},\mathrm{H}}$, which has *index derived* and *index components*, (b) the *classification derived substrate history coder* $C_{\mathrm{G},V,\mathrm{T},\mathrm{G}}$, which has *classification derived* and *classification components*, (c) the *specialising derived substrate history coder* $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, which has *classification derived* and *index components*, and (d) the *generalising derived substrate history coder* $C_{\mathrm{H},V,\mathrm{T},\mathrm{G}}$, which has *index derived* and *classification components*.

The *index derived substrate history coder* is constructed

$$\begin{aligned}&C_{\mathrm{H},V,\mathrm{T},\mathrm{H}} = \\ &\quad \mathrm{coderHistorySubstrateDerivedIndex}(U, X, T, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,V,X})\end{aligned}$$

The *index derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{H}}$, is similar to an *index history coder*, C_{H}. Let $H \in \mathcal{H}_{U,V,X}$ be a *history* of the *derived substrate history coder* domain. The *derived history* is $H*T$ where $H*T := \mathrm{transform}(T, H) := \{(x, P_S) : (x, S) \in H\}$ and $P = \mathrm{split}(V, \mathrm{his}(T)^{\mathrm{FS}})$. The *histogram* is $A =$

histogram(H). The *event identifiers* are encoded in *space* of spaceIds(y, z), where $y = |X|$ and $z = |H| = \text{size}(A)$. The *derived history*, $H * T$, is encoded in *fixed width space*, spaceEventsDerived$(U)(H, T) := z \ln w'$, where $W = \text{der}(T)$, $w = |W^{\text{C}}|$ and $w' = |(V^{\text{C}} * T)^{\text{F}}| = |T^{-1}| \leq w$. Then each *sub-history*, H_C, corresponding to a *component* of the *partition*, $H_C = \text{filter}((H * T)^{-1}_R, H) = \text{flip}(\text{filter}(C^{\text{S}}, \text{flip}(H))) \subseteq H$, where $(R, C) \in T^{-1}$, is encoded in a *fixed width* list. The *space* of the *sub-history* is spaceEvents$(U)(H_C) = (A * T)_R \ln |C|$. The *sub-history fixed width* lists are concatenated together into a *variable width* list.

The *history*, $H \in \mathcal{H}_{U,V,X}$, can be encoded in an intermediate tuple $T_H = ((z, Z_I), L', L) \in \mathbf{N}^2 \times \mathcal{L}(\mathbf{N}) \times \mathcal{L}(\mathbf{N})$.

There is no need to encode the *variables*, V, because these are defined by the *transform*, T, in the *derived substrate history coder* parameters. So the first element, $(z, Z_I) \in \mathbf{N}^2$, of the tuple, T_H, encodes the *event identifiers* in the same way as the *index history coder*, C_{H}, above. The *space* is spaceIds(y, z), where $y = |X|$ and $z = |H| = \text{size}(A)$.

The second element $L' \in \mathcal{L}(\mathbf{N})$ of the tuple T_H encodes the *states* of *derived events* in a list of *fixed width space*. Given order D_{S}, let M' be an enumeration of the *possible derived states*, $M' = \text{order}(D_{\text{S}}, \text{dom}(T^{-1}))$. Let $w' = |(V^{\text{C}} * T)^{\text{F}}| = |T^{-1}|$. Then $|M'| = w'$ and hence $L \in \mathcal{L}(\{1 \ldots w'\})$. Given order D_{X} and *event identifiers* $I = \text{ids}(H)$, let Q be an enumeration of the *event identifiers*, $Q = \text{order}(D_{\text{X}}, I)$. Then

$$L' = \{(Q_x, M'_R) : (x, R) \in H * T\}$$

The *space* is spaceEventsDerived$(U)(H, T)$, where spaceEventsDerived$(U) \in \mathcal{H}_U \times \mathcal{T}_{U,\text{f},1} \to \ln \mathbf{N}_{>0}$ is defined

$$\text{spaceEventsDerived}(U)(H, T) := z \ln w'$$

The *possible derived volume* is less than or equal to the *derived volume*, $w' \leq w$, so the *derived events space* is no greater than the *events space* of the *derived history*, spaceEventsDerived$(U)(H, T) \leq$ spaceEvents$(U)(H * T)$. The *possible derived volume* equals the *derived volume* if and only if the *transform* is *non-overlapped*, $\neg\text{overlap}(T) \iff w' = w$, because it is only in this case that the *transform* is *right total*, $\text{dom}(T^{-1}) = W^{\text{CS}}$. In this case, the *derived events space* equals the *events space* of the *derived history*, $\neg\text{overlap}(T) \iff \text{spaceEventsDerived}(U)(H, T) = \text{spaceEvents}(U)(H * T)$. The *possible derived volume* is less than or equal to the *underlying volume*,

$w' \leq v$, so the *derived events space* is no greater than the *events space* of the *history*, $\mathrm{spaceEventsDerived}(U)(H,T) \leq \mathrm{spaceEvents}(U)(H)$.

The last element $L \in \mathcal{L}(\mathbf{N})$ of the tuple, T_H, encodes the *underlying states* of *derived events*. Given order D_{S} and *variables* $V = \mathrm{vars}(H)$, let M be a map of enumerations indexed by *derived state*

$$M = \{(R, \mathrm{order}(D_{\mathrm{S}}, C^{\mathrm{S}})) : (R,C) \in T^{-1}\}$$

Then

$$L = \{(Q_x, M_R(S)) : (x,S) \in H,\ R = P_S\}$$

where $P = \mathrm{split}(V, \mathrm{his}(T)^{\mathrm{S}})$. Define the *space* of the list, L, encoding the *partitioned events* of *history* H as $\mathrm{spaceEventsPartition} \in \mathcal{A} \times \mathcal{T}_{\mathrm{f}} \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceEventsPartition}(A,T) := \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C|$$

The total *space* of the *index derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{H}}$, of a *history* $H \in \mathcal{H}_{U,V,X}$ is the sum of the *ids space*, *derived events space*, and *partitioned events space*

$$\begin{aligned}\mathrm{space}(C_{\mathrm{H},V,\mathrm{T},\mathrm{H}})(H) \quad &= \quad \mathrm{spaceIds}(|X|, |H|) + \\ &\quad\quad \mathrm{spaceEventsDerived}(U)(H,T) + \\ &\quad\quad \mathrm{spaceEventsPartition}(A,T)\end{aligned}$$

The *index derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{H}}$, may be compared to the *index substrate history coder*, $C_{\mathrm{H},V} \in \mathrm{coders}(\mathcal{H}_{U,V,X})$. The difference in *space* for a *history* $H \in \mathcal{H}_{U,V,X}$ is

$$\begin{aligned}&C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) \\ &\quad = \quad \mathrm{spaceEventsDerived}(U)(H,T) + \mathrm{spaceEventsPartition}(A,T) - \\ &\quad\quad\quad \mathrm{spaceEvents}(U)(H) \\ &\quad = \quad z \ln w' \ + \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C| \ - \\ &\quad\quad\quad z \ln v\end{aligned}$$

The *partitioned events space* varies against the *component size cardinality cross entropy*,

$$\begin{aligned}\mathrm{spaceEventsPartition}(A,T) \quad &:= \quad \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C| \\ &\sim \quad -\ \mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T)\end{aligned}$$

The difference varies with the *log possible derived volume* and varies against the *component size cardinality cross entropy*,

$$C^{\rm s}_{{\rm H},V,{\rm T},{\rm H}}(H) - C^{\rm s}_{{\rm H},V}(H) \quad \sim \quad \ln w' \; - \; \mathrm{entropyCross}(A * T, V^{\rm C} * T)$$

So the *space* of the *index derived substrate history coder*, $C_{{\rm H},V,{\rm T},{\rm H}}$, is minimised when (a) the *possible derived volume* is minimised, and (b) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

For example, consider the *mono-variate tri-valent singleton cardinal substrate histogram* $A = \{(\{(u,1)\},z)\}$, where $\mathrm{vars}(A) = \{u\}$, with *binary partition transform* $T = \{\{\{(u,1)\}\},\{\{(u,2)\},\{(u,3)\}\}\}^{\rm T}$ which *rolls* the *ineffective states* into a single *derived state*. The difference in *space* is

$$\begin{aligned}
C^{\rm s}_{{\rm H},V,{\rm T},{\rm H}}(H) - C^{\rm s}_{{\rm H},V}(H) &= z \ln 2 + z \ln 1 + 0 \ln 2 - z \ln 3 \\
&= z \ln 2/3 \\
&< 0
\end{aligned}$$

where $z > 0$. Equivalently, the expected logarithm of the *component cardinality* scaled by the *derived volume* fraction is

$$\begin{aligned}
C^{\rm s}_{{\rm H},V,{\rm T},{\rm H}}(H) - C^{\rm s}_{{\rm H},V}(H) &= z\frac{2}{2} \ln \left(\frac{2}{3} \times 1\right) + z\frac{0}{2} \ln \left(\frac{2}{3} \times 2\right) \\
&= z \ln 2/3
\end{aligned}$$

A *transform* T that *partitions* the *volume*, $T^{\rm P} \in \mathrm{B}(V^{\rm CS})$, into *components* having the same cardinality, $\forall C \in \mathrm{ran}(T^{-1})\ (|C| = v/w')$, and the same *size*, $\forall C \in \mathrm{ran}(T^{-1})\ (\mathrm{size}(A * C) = z/w')$, has no *space* difference

$$\begin{aligned}
C^{\rm s}_{{\rm H},V,{\rm T},{\rm H}}(H) - C^{\rm s}_{{\rm H},V}(H) &= w' \frac{z}{w'} \ln \frac{w'}{v} \frac{v}{w'} \\
&= 0
\end{aligned}$$

An example is where the *derived histogram* is a *scaled cartesian*, $A * T = \mathrm{resize}(z, W^{\rm C})$, and the *component cardinalities* equal the *derived volume* factor, $\forall C \in \mathrm{ran}(T^{-1})\ (|C| = v/w)$.

The *classification derived substrate history coder* is constructed

$$\begin{aligned}
&C_{{\rm G},V,{\rm T},{\rm G}} = \\
&\quad \mathrm{coderHistorySubstrateDerivedClassification}(U, X, T, D_{\rm S}, D_{\rm X}) \\
&\qquad\qquad\qquad \in \mathrm{coders}(\mathcal{H}_{U,V,X})
\end{aligned}$$

The *classification derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{G}}$, is similar to a *classification history coder*, C_{G}. Let $H \in \mathcal{H}_{U,V,X}$ be a *history* of the *derived substrate history coder* domain. The *derived history* is $H * T$. The *histogram* is $A = \mathrm{histogram}(H)$. The *event identifiers* are encoded in *space* of $\mathrm{spaceIds}(y, z)$. The *derived histogram*, $A * T$, is encoded in *space* of $\mathrm{spaceCountsDerived}(U)(A, T)$. The *derived history*, $H * T$, is encoded as a *classification*, $(H * T)^{-1} = \mathrm{classification}(H * T)$. The *classification* has *space* of $\mathrm{spaceClassification}(A * T)$. Then each *sub-history*, H_C, corresponding to a *component* of the *partition*, $H_C \subseteq H$, where $(R, C) \in T^{-1}$, is encoded as a *component histogram* having *space* $\mathrm{spaceCounts}(U)(A * C)$ and a *component classification* having *space* $\mathrm{spaceClassification}(A * C)$.

The *history*, $H \in \mathcal{H}_{U,V,X}$, can be encoded in an intermediate tuple $T_H = ((z, Z_I), R'_{A*T}, F'_{Q'}, L, M) \in \mathbf{N}^2 \times \mathbf{N} \times \mathbf{N} \times \mathcal{L}(\mathbf{N}) \times \mathcal{L}(\mathbf{N})$.

The first element, $(z, Z_I) \in \mathbf{N}^2$, of the tuple, T_H, encodes the *event identifiers* in the same way as the *index history coder*, C_{H}, above. The *space* is $\mathrm{spaceIds}(y, z)$, where $y = |X|$ and $z = |H| = \mathrm{size}(A)$.

The second element, $R'_{A*T} \in \mathbf{N}$, of the tuple, T_H, is the encoding of the *derived counts* as in the *derived histogram coder*, $C_{\mathrm{A},V,\mathrm{T}}$, above,

$$R'_{A*T} \in \{1 \ldots \frac{(z + w' - 1)!}{z!\ (w' - 1)!}\}$$

where $W = \mathrm{der}(T)$, $w = |W^{\mathrm{C}}|$ and $w' = |(V^{\mathrm{C}} * T)^{\mathrm{F}}| = |T^{-1}| \leq w$. The *space* is $\mathrm{spaceCountsDerived}(U)(A, T)$.

The third element, $F'_{Q'} \in \mathbf{N}$, of the tuple, T_H, is the encoding of the *classification* of the *derived history* $(H * T)^{-1}$, as in the *classification history coder*, C_{G}, above. Here Q' is the partition of *event identifiers*, $Q' = \mathrm{ran}((H*T)^{-1}) \in \mathrm{B}(\mathrm{ids}(H))$, and F' is an enumeration of the possible partitions of corresponding *component* cardinalities,

$$F' \in \mathrm{enums}(\{P : P \in \mathrm{B}(\mathrm{ids}(H)),\ \exists X \in P :\leftrightarrow: Q'\ \forall (Y, Z) \in X\ (|Y| = |Z|)\})$$

so that

$$\begin{aligned} F'_{Q'} &\in \{1 \ldots \frac{z!}{\prod_{R \in \mathrm{dom}((H*T)^{-1})} |(H * T)^{-1}_R|!}\} \\ &= \{1 \ldots \frac{z!}{\prod_{R \in (A*T)^{\mathrm{S}}} (A * T)_R!}\} \end{aligned}$$

The *space* is $\mathrm{spaceClassification}(A * T)$.

The fourth element $L \in \mathcal{L}(\mathbf{N})$ of the tuple, T_H, encodes the *counts* of *components*. For all $(R, C) \in T^{-1}$ let

$$R_R(A * C) \in \{1 \ldots \frac{((A * T)_R + |C| - 1)!}{(A * T)_R! \ (|C| - 1)!}\}$$

Then

$$L = \{(M'_R, R_R(A * C)) : (R, C) \in T^{-1}\}$$

The *space* is $\mathrm{spaceCountsPartition}(A, T)$.

The last element $M \in \mathcal{L}(\mathbf{N})$ of the tuple, T_H, encodes the *classifications* of *components*. For all $(R, C) \in T^{-1}$ let $Q_R = \mathrm{ran}(H_C^{-1}) \in \mathrm{B}(\mathrm{ids}(H_C))$ and

$$F_R \in \mathrm{enums}(\{P : P \in \mathrm{B}(\mathrm{ids}(H_C)),\ \exists X \in P :\leftrightarrow: Q_R\ \forall (Y, Z) \in X\ (|Y| = |Z|)\})$$

so that

$$F_R(Q_R) \ \in \ \{1 \ldots \frac{(A * T)_R!}{\prod_{S \in C^{\mathrm{S}}} A_S!}\}$$

in

$$M = \{(M'_R, F_R(Q_R)) : (R, C) \in T^{-1}\}$$

Define the *space* of the list, M, encoding the *partitioned classifications* of the *sub-histories* of H as $\mathrm{spaceClassificationPartition} \in \mathcal{A} \times \mathcal{T}_{\mathrm{f}} \to \ln \mathbf{N}_{>0}$

$$\begin{aligned}
\mathrm{spaceClassificationPartition}(A, T) \ &:= \ \sum_{(R,C) \in T^{-1}} \Big(\ln (A * T)_R! - \sum_{S \in C^{\mathrm{S}}} \ln A_S! \Big) \\
&= \ \sum_{(R,\cdot) \in T^{-1}} \Big(\ln (A * T)_R! \Big) - \sum_{S \in A^{\mathrm{S}}} \Big(\ln A_S! \Big)
\end{aligned}$$

which is such that

$$\begin{aligned}
\mathrm{spaceClassification}(A) \ &= \ \mathrm{spaceClassification}(A * T) + \\
&\quad\ \mathrm{spaceClassificationPartition}(A, T)
\end{aligned}$$

The total *space* of the *classification derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{G}}$, of a *history* $H \in \mathcal{H}_{U,V,X}$ is the sum of the *ids space*, *derived counts space*, *derived classification space*, *partitioned counts space*, and *partitioned classification space*

$$\begin{aligned}
\text{space}(C_{\text{G},V,\text{T},\text{G}})(H) &= \text{spaceIds}(|X|,|H|) + \\
&\quad \text{spaceCountsDerived}(U)(A,T) + \\
&\quad \text{spaceClassification}(A * T) + \\
&\quad \text{spaceCountsPartition}(A,T) + \\
&\quad \text{spaceClassificationPartition}(A,T) \\
&= \text{spaceIds}(|X|,|H|) + \\
&\quad \text{spaceCountsDerived}(U)(A,T) + \\
&\quad \text{spaceCountsPartition}(A,T) + \\
&\quad \text{spaceClassification}(A)
\end{aligned}$$

The *classification derived substrate history coder*, $C_{\text{G},V,\text{T},\text{G}}$, may be compared to the *classification substrate history coder*, $C_{\text{G},V} \in \text{coders}(\mathcal{H}_{U,V,X})$. The difference in *space* for a *history* $H \in \mathcal{H}_{U,V,X}$ equals the difference in *space* between the *derived substrate histogram coder*, $C_{\text{A},V,\text{T}}$, and the *substrate histogram coder*, $C_{\text{A},V}$, for the *histogram* $A = \text{histogram}(H)$,

$$\begin{aligned}
&C^{\text{s}}_{\text{G},V,\text{T},\text{G}}(G) - C^{\text{s}}_{\text{G},V}(G) \\
&\quad = \text{spaceCountsDerived}(U)(A,T) + \text{spaceCountsPartition}(A,T) - \\
&\quad\quad \text{spaceCounts}(U)(A) \\
&\quad = C^{\text{s}}_{\text{A},V,\text{T}}(A) - C^{\text{s}}_{\text{A},V}(A) \\
&\quad = \ln \frac{(z+w'-1)!}{z!\ (w'-1)!} + \sum_{(R,C) \in T^{-1}} \ln \frac{((A*T)_R + |C| - 1)!}{(A*T)_R!\ (|C|-1)!} - \\
&\quad\quad \ln \frac{(z+v-1)!}{z!\ (v-1)!}
\end{aligned}$$

So the *space* of the *classification derived substrate history coder*, $C_{\text{G},V,\text{T},\text{G}}$, is minimised when (a) the *possible derived volume* is minimised, (b) the *component entropy* is minimised, and (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

The *specialising derived substrate history coder* is constructed

$$\begin{aligned}
&C_{\text{G},V,\text{T},\text{H}} = \\
&\quad \text{coderHistorySubstrateDerivedSpecialising}(U,X,T,D_{\text{S}},D_{\text{X}}) \\
&\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \in \text{coders}(\mathcal{H}_{U,V,X})
\end{aligned}$$

The *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, is intermediate between a *classification history coder*, C_{G}, and an *index history coder*, C_{H}. Let $H \in \mathcal{H}_{U,V,X}$ be a *history* of the *derived substrate history coder* domain. The *derived history* is $H * T$. The *histogram* is $A = \mathrm{histogram}(H)$. The *event identifiers* are encoded in *space* of $\mathrm{spaceIds}(y, z)$. The *derived histogram*, $A * T$, is encoded in *space* of $\mathrm{spaceCountsDerived}(U)(A, T)$. The *derived history*, $H * T$, is encoded as a *classification*, $(H * T)^{-1} = \mathrm{classification}(H * T)$, having *space* of $\mathrm{spaceClassification}(A * T)$. Then each *sub-history*, H_C, corresponding to a *component* of the *partition*, $H_C \subseteq H$, where $(R, C) \in T^{-1}$, is encoded in a *fixed width* list. The *space* of the *sub-history* is $\mathrm{spaceEvents}(U)(H_C) = (A * T)_R \ln |C|$. The *sub-history fixed width* lists are concatenated together into a *variable width* list which has *space* $\mathrm{spaceEventsPartition}(A, T)$.

The *history*, $H \in \mathcal{H}_{U,V,X}$, can be encoded in an intermediate tuple $T_H = ((z, Z_I), R'_{A*T}, F'_{Q'}, L) \in \mathbf{N}^2 \times \mathbf{N} \times \mathbf{N} \times \mathcal{L}(\mathbf{N})$.

The first element, $(z, Z_I) \in \mathbf{N}^2$, of the tuple, T_H, encodes the *event identifiers* in the same way as the *index history coder*, C_{H}, above. The *space* is $\mathrm{spaceIds}(y, z)$, where $y = |X|$ and $z = |H| = \mathrm{size}(A)$.

The second element, $R'_{A*T} \in \mathbf{N}$, and the third element, $F'_{Q'} \in \mathbf{N}$, of the tuple, T_H, are encoded as in the *classification derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{G}}$, above. The *space* is

$$\mathrm{spaceCountsDerived}(U)(A, T) + \mathrm{spaceClassification}(A * T)$$

The last element $L \in \mathcal{L}(\mathbf{N})$ of the tuple, T_H, encodes the *underlying states* of *derived events* in the same way as for the *index derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{H}}$, above. The *space* is $\mathrm{spaceEventsPartition}(A, T)$.

The total *space* of the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, of a *history* $H \in \mathcal{H}_{U,V,X}$ is the sum of the *ids space*, *derived counts space*, *derived classification space*, and *partitioned events space*

$$\begin{aligned}\mathrm{space}(C_{\mathrm{G},V,\mathrm{T},\mathrm{H}})(H) \;&=\; \mathrm{spaceIds}(|X|, |H|) + \\ &\quad\; \mathrm{spaceCountsDerived}(U)(A, T) + \\ &\quad\; \mathrm{spaceClassification}(A * T) + \\ &\quad\; \mathrm{spaceEventsPartition}(A, T)\end{aligned}$$

The *space* of the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, is minimised for a *history* $H \in \mathcal{H}_{U,V,X}$ which has a single *state* S_1, $H \in$

$X \to \{S_1\}$, in the case of *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ which (i) is *non-overlapping*, $\neg\mathrm{overlap}(T)$, so the *possible derived volume* equals the *derived volume*, $w' = w$ where $w' = |T^{-1}|$, $W = \mathrm{der}(T)$, and $w = |W^{\mathrm{CS}}|$, (ii) has two *derived states*, $\{R_1, R_2\} = W^{\mathrm{CS}}$, of which only one is *effective*, $(A * T)_{R_1} = A_{S_1} = z$ and $(A * T)_{R_2} = 0$ where $A = \mathrm{histogram}(H)$ and $z = |H|$, and (iii) is such that the *component* $C_1 = \{S_1\}^{\mathrm{U}}$ corresponding to the *effective derived state*, R_1, is a singleton, $|C_1| = 1$, and the remaining *volume*, $|C_2| = v - 1$, corresponds to the *ineffective derived state*, R_2, where $\{(R_1, C_1), (R_2, C_2)\} = T^{-1}$, and $v = |V^{\mathrm{CS}}|$.

The *partitioned events space* varies against the *component size cardinality cross entropy*,

$$\begin{aligned}\mathrm{spaceEventsPartition}(A,T) &:= \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C| \\ &= z \ln v \; - \; z \times \mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T)\end{aligned}$$

The *partitioned events space* is minimised when high *size components* are low *cardinality components* and low *size components* are high *cardinality components*. In the case of single *state history*, H, and single *effective derived state transform*, T, all of the *size* is in a singleton *component*, $\mathrm{size}(A * C_1) = (A * T)_{R_1} = A_{S_1} = z$, and hence the *partitioned events space* is zero.

The *derived counts space*,

$$\mathrm{spaceCountsDerived}(U)(A,T) \;:=\; \ln \frac{(z + w' - 1)!}{z!\ (w' - 1)!}$$

is minimised for fixed *size*, z, when the *possible derived volume* is smallest, $w' = 1$. Then the *derived counts space* is zero. In that case, however, the *partitioned events space* would be $z \ln v$. In the case of the *transform*, T, the *possible derived volume* is two, $w' = 2$, the *derived counts space* is $\ln(z + 1) < z \ln v$, and so the overall *space* is smaller.

In the domain where the *size* is less than or equal to the *possible derived volume*, $z \leq w'$, the *derived counts space* varies with the *log possible derived volume*,

$$\mathrm{spaceCountsDerived}(U)(A,T) \;\sim\; \ln w'$$

In the domain where the *size* is greater than the *possible derived volume*, $z > w'$, the *derived counts space* varies with the *possible derived volume*,

$$\mathrm{spaceCountsDerived}(U)(A,T) \;\sim\; w'$$

The *derived classification space* varies with the *derived entropy*

$$\begin{aligned}
\text{spaceClassification}(A * T) &:= \ln z! - \sum_{(R,\cdot) \in T^{-1}} \ln (A * T)_R! \\
&\sim z \times \text{entropy}(A * T)
\end{aligned}$$

So the *derived classification space* is minimised when the *derived entropy* or *component size entropy*, $\text{entropy}(A * T)$, is minimised. This is the case here of single *effective derived state*, $|(A * T)^{\text{F}}| = 1$, which has zero *entropy*, $\text{entropy}(A*T) = 0$ and zero *derived classification space*, $\text{spaceClassification}(A* T) = 0$.

The *transform*, T, described above only minimises the *space* where the *history* has one *state*, $A = \{(S_1, z)\}$. If there is more than one *effective state*, $|A^{\text{F}}| > 1$, the *derived classification space* is non-zero, $\text{spaceClassification}(A * T) > 0$, and so there is a balance between the increasing *derived entropy*, $\text{entropy}(A * T)$, and the increasing *derived counts space*, $\text{spaceCountsDerived}(U)(A, T)$, caused by increasing *possible derived volume*, w'. Also the *partitioned events space*, $\text{spaceEventsPartition}(A, T)$, may be non-zero if the *components* are not all singletons.

The *space* of the *specialising derived substrate history coder*, $C_{\text{G},V,\text{T},\text{H}}$, is

$$\begin{aligned}
\text{space}(C_{\text{G},V,\text{T},\text{H}})(H) &= \text{spaceIds}(|X|, |H|) + \\
&\quad \text{spaceCountsDerived}(U)(A, T) + \\
&\quad \text{spaceClassification}(A * T) + \\
&\quad \text{spaceEventsPartition}(A, T) \\
&= \text{spaceIds}(|X|, |H|) + \\
&\quad \ln \frac{(z + w' - 1)!}{z!\ (w' - 1)!} + \\
&\quad \ln z! - \sum_{R \in (A*T)^{\text{S}}} \ln (A * T)_R! + \\
&\quad \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C|
\end{aligned}$$

The *space* of the *specialising derived substrate history coder*, $C_{\text{G},V,\text{T},\text{H}}$, varies (i) with the *possible derived volume*, w', where the *possible derived volume* is less than the *size*, $w' < z$, otherwise with the *size* scaled *log possible derived volume*, $z \ln w'$, and (ii) against the *size* scaled *component size cardinality* *relative entropy*,

$$\begin{aligned} C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) \ \sim & \\ & (w' \ : \ w' < z) \ + \ (z \ln w' \ : \ w' \geq z) \\ & - \ z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T) \end{aligned}$$

So the *space* of the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, is minimised when (a) the *possible derived volume* is minimised, (b) the *derived entropy* or *component size entropy* is minimised, and (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

Compare the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, to the *index derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{H}}$. The difference in *space* of the *history*, H, is

$$C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H)$$

In the case of *non-overlapping transform*, $\neg\mathrm{overlap}(T)$, the difference in *space* of the *history*, H, equals the difference in *space* of the *derived history*, $H * T$, between the *classification history coder*, C_{G}, and the *index history coder*, C_{H},

$$C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) = C^{\mathrm{s}}_{\mathrm{G}}(H * T) - C^{\mathrm{s}}_{\mathrm{H}}(H * T)$$

As shown in section 'Classification space', above, the *classification coder*, C_{G}, may require less *space* than the *index coder*, C_{H}, in the case of low *entropy*. So if the *derived histogram*, $A*T$, has low *entropy*, then the *specialising coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, may have smaller *space* than the *index derived coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{H}}$. For example, a *regular pluri-derived-variate transform* T_1, of *derived valency* d and *derived dimension* $n > 1$, where the *derived histogram* is *uniformly diagonalised*, has lower *entropy* than a *transform* T_2 which has *uniform cartesian congruent derived histogram*,

$$\begin{aligned} \mathrm{entropy}(A * T_1) - \mathrm{entropy}(A * T_2) \ &= \ \left(-\ln \frac{1}{d}\right) - \left(-\ln \frac{1}{d^n}\right) \\ &= \ -(n-1) \ln d \\ &< \ 0 \end{aligned}$$

The difference in *space* between the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, and the *index substrate history coder*, $C_{\mathrm{H},V}$, is

$$C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)$$

In the case of *non-overlapping transform*, $\neg\text{overlap}(T)$, this simplifies to

$$\begin{aligned}
C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) &= C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) + \\
&\quad\ C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) \\
&= C^{\mathrm{s}}_{\mathrm{G}}(H * T) - C^{\mathrm{s}}_{\mathrm{H}}(H * T) + \\
&\quad\ C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)
\end{aligned}$$

The difference in *space* between the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, and the *classification derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{G}}$, equals the difference in summed *space* between the *index history coder*, C_{H}, and the *classification history coder*, C_{G}, for each of the *components*,

$$C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{G}}(H) = \sum_{(\cdot,C) \in T^{-1}} C^{\mathrm{s}}_{\mathrm{H}}(H_C) - C^{\mathrm{s}}_{\mathrm{G}}(H_C)$$

where the *sliced-history* is defined $H_C = \{(x,S) : (x,S) \in H,\ \{S\}^{\mathrm{U}} * C \neq \emptyset\} \subseteq H$. The *space* of the *components* will be lower in the *index history coder*, C_{H}, if their *entropies* are high. This is the case, for example, if the *transform* is *ideal*, $A = A{*}T{*}T^{\dagger A}$, and the *histogram* is *completely effective*, $A^{\mathrm{F}} = V^{\mathrm{C}}$. A special case is the *frame full functional transform*, $T = \{\{u\}^{\mathrm{CS}\{\}V\mathrm{T}} : u \in V\}^{\mathrm{T}}$. In this case the difference in summed *component space* is zero.

The difference in *space* between the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, and the *classification substrate history coder*, $C_{\mathrm{G},V}$, is

$$\begin{aligned}
C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) &= C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{G}}(H) + \\
&\quad\ C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) \\
&= \sum_{(\cdot,C) \in T^{-1}} \big(C^{\mathrm{s}}_{\mathrm{H}}(H_C) - C^{\mathrm{s}}_{\mathrm{G}}(H_C)\big) + \\
&\quad\ C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A)
\end{aligned}$$

Given some *substrate history* $H \in \mathcal{H}_{U,V,X}$, the *transform* T in the parameters of the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, may be chosen to minimise the encoding *space* with respect to the *space* in either of the *canonical history coders*, (i) the *index history coder*, C_{H}, and (ii) the *classification history coder*, C_{G}. That is, the *transform* is chosen to minimise

$$\big(C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)\big) + \big(C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)\big)$$

In the case of *non-overlapping transform*, $\neg\mathrm{overlap}(T)$, this simplifies to

$$\begin{aligned}
&\big(C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)\big) + \big(C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)\big) \\
&\quad= \big(C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H)\big) + \big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)\big) + \\
&\qquad \big(C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{G}}(H)\big) + \big(C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)\big) \\
&\quad= \big(C^{\mathrm{s}}_{\mathrm{G}}(H * T) - C^{\mathrm{s}}_{\mathrm{H}}(H * T)\big) + \big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)\big) + \\
&\qquad \sum_{(\cdot,C) \in T^{-1}} \big(C^{\mathrm{s}}_{\mathrm{H}}(H_C) - C^{\mathrm{s}}_{\mathrm{G}}(H_C)\big) + \big(C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A)\big)
\end{aligned}$$

The first term, $C^{\mathrm{s}}_{\mathrm{G}}(H * T) - C^{\mathrm{s}}_{\mathrm{H}}(H * T)$, varies with the *derived entropy* or *component size entropy*,

$$C^{\mathrm{s}}_{\mathrm{G}}(H * T) - C^{\mathrm{s}}_{\mathrm{H}}(H * T) \quad \sim \quad \mathrm{entropy}(A * T)$$

The third term, $\sum_{(\cdot,C) \in T^{-1}} \big(C^{\mathrm{s}}_{\mathrm{H}}(H_C) - C^{\mathrm{s}}_{\mathrm{G}}(H_C)\big)$, varies against the *size expected component entropy*

$$\begin{aligned}
&\sum_{(\cdot,C) \in T^{-1}} \big(C^{\mathrm{s}}_{\mathrm{H}}(H_C) - C^{\mathrm{s}}_{\mathrm{G}}(H_C)\big) \\
&\qquad\qquad \sim \ - \sum_{(R,C) \in T^{-1}} (\hat{A} * T)_R \times \mathrm{entropy}(A * C) \\
&\qquad\qquad = \ - \ \mathrm{entropyComponent}(A, T)
\end{aligned}$$

The second term, $C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)$, varies with the *log possible derived volume* and varies against the *component size cardinality cross entropy*,

$$C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) \quad \sim \quad \ln w' \ - \ \mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T)$$

The fourth term, $C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A)$, varies as follows. First, in the domain where the *size* is less than or equal to the *possible derived volume*, $z \leq w' \leq v$, the difference varies with the *log possible derived volume* and against the *component size cardinality relative entropy*,

$$C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A) \quad \sim \quad \ln w' \ - \ \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

In the domain where the *size* is between the *possible derived volume* and the *volume*, $w' \leq z \leq v$, the difference varies with the *possible derived volume* and against the *size* scaled *component size cardinality relative entropy*,

$$C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A) \quad \sim \quad w' \ - z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

Last, in the domain where the *size* is greater than the *volume*, $w' \leq v < z$, the difference varies with the *possible derived volume* and against the *volume* scaled *component cardinality size relative entropy*,

$$C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A) \quad \sim \quad w' \ - v \times \mathrm{entropyRelative}(V^{\mathrm{C}} * T, A * T)$$

The *specialising-index space difference* is

$$\begin{aligned}
C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) \quad &= \quad \mathrm{spaceCountsDerived}(U)(A,T) + \\
&\quad\quad \mathrm{spaceClassification}(A * T) + \\
&\quad\quad \mathrm{spaceEventsPartition}(A,T) - \\
&\quad\quad \mathrm{spaceEvents}(U)(A,T) \\
&= \quad \ln \frac{(z + w' - 1)!}{z!\ (w' - 1)!} \\
&+ \quad \ln z! - \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A * T)_R! \\
&+ \quad \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C| \ - \\
&- \quad z \ln v
\end{aligned}$$

This varies just as the *specialising space*, $C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H)$. That is, the *specialising-index space difference* varies (i) with the *possible derived volume*, w', where the *possible derived volume* is less than the *size*, $w' < z$, otherwise with the *size* scaled *log possible derived volume*, $z \ln w'$, and (ii) against the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned}
C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) \ &\sim \\
&(w' \ : \ w' < z) \ + \ (z \ln w' \ : \ w' \geq z) \\
&- \ z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)
\end{aligned}$$

The *specialising-classification space difference* is

$$\begin{aligned}
C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) &= \mathrm{spaceCountsDerived}(U)(A,T) + \\
&\ \mathrm{spaceClassification}(A * T) + \\
&\ \mathrm{spaceEventsPartition}(A,T) - \\
&\ \mathrm{spaceCounts}(U)(A) - \\
&\ \mathrm{spaceClassification}(A) \\
&= \ln \frac{(z+w'-1)!}{z!\ (w'-1)!} \\
&+ \ln z! - \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A*T)_R! \\
&+ \sum_{(R,C) \in T^{-1}} (A*T)_R \ln |C| \\
&- \ln \frac{(z+v-1)!}{z!\ (v-1)!} \\
&- \big(\ln z! - \sum_{S \in A^{\mathrm{S}}} \ln A_S!\big)
\end{aligned}$$

So the *specialising-classification space difference* varies as

$$\begin{aligned}
C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) &\sim \ln \frac{(z+w'-1)!}{z!\ (w'-1)!} \\
&+ \sum_{(R,C) \in T^{-1}} (A*T)_R \ln |C| \\
&- \sum_{(R,C) \in T^{-1}} \big(\ln (A*T)_R! - \sum_{S \in C^{\mathrm{S}}} \ln (A*C)_S!\big)
\end{aligned}$$

The *specialising-classification space difference* varies (i) with the *possible derived volume*, w', where $w' < z$, otherwise with the *size* scaled *log possible derived volume*, $z \ln w'$, (ii) against the *size* scaled *component size cardinality cross entropy* and (iii) against the *size* scaled *size expected component entropy*,

$$\begin{aligned}
&C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) \sim \\
&\qquad (w' \ :\ w' < z) \ +\ (z \ln w' \ :\ w' \geq z) \\
&\qquad - z \times \mathrm{entropyCross}(A*T, V^{\mathrm{C}}*T) \\
&\qquad - z \times \mathrm{entropyComponent}(A,T)
\end{aligned}$$

Overall, the *specialising-canonical space difference*, $2C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, is the sum of the *specialising-index space difference*, $C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(H) -$

$C^{\rm s}_{{\rm H},V}(H)$, and the *specialising-classification space difference*, $C^{\rm s}_{{\rm G},V,{\rm T},{\rm H}}(H) - C^{\rm s}_{{\rm G},V}(H)$. The *specialising-canonical space difference* varies (i) with twice the *possible derived volume*, $2w'$, where $w' < z$, otherwise with twice the *size* scaled *log possible derived volume*, $2z \ln w'$, (ii) with the *size* scaled *derived entropy*, (iii) against twice the *size* scaled *component size cardinality cross entropy* and (iv) against the *size* scaled *size expected component entropy*,

$$\begin{aligned}
2C^{\rm s}_{{\rm G},V,{\rm T},{\rm H}}(H) - C^{\rm s}_{{\rm H},V}(H) - C^{\rm s}_{{\rm G},V}(H) \ \sim \\
&2\big((w' \ :\ w' < z)\ +\ (z \ln w' \ :\ w' \geq z)\big) \\
&+\ z \times \mathrm{entropy}(A * T) \\
&-\ 2z \times \mathrm{entropyCross}(A * T, V^{\rm C} * T) \\
&-\ z \times \mathrm{entropyComponent}(A, T)
\end{aligned}$$

So the *specialising-canonical space difference*, $2C^{\rm s}_{{\rm G},V,{\rm T},{\rm H}}(H) - C^{\rm s}_{{\rm H},V}(H) - C^{\rm s}_{{\rm G},V}(H)$, is minimised when (a) the *possible derived volume* is minimised, (b) the *derived entropy* is minimised, (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*, and (d) the *expected component entropy* is maximised.

The *canonical* term, $C^{\rm s}_{{\rm H},V}(H) + C^{\rm s}_{{\rm G},V}(H)$, is independent of the *transform*, T, so properties of the *specialising-canonical space difference*, $2C^{\rm s}_{{\rm G},V,{\rm T},{\rm H}}(H) - C^{\rm s}_{{\rm H},V}(H) - C^{\rm s}_{{\rm G},V}(H)$, are also properties of the *specialising space*, $C^{\rm s}_{{\rm G},V,{\rm T},{\rm H}}(H)$. This is because the *derived entropy*, $\mathrm{entropy}(A * T)$, and the *size expected component entropy*, $\mathrm{entropyComponent}(A, T)$, are dual.

The fourth *derived coder* is the *generalising derived substrate history coder*, constructed

$$\begin{aligned}
&C_{{\rm H},V,{\rm T},{\rm G}} = \\
&\quad \mathrm{coderHistorySubstrateDerivedGeneralising}(U, X, T, D_{\rm S}, D_{\rm X}) \\
&\qquad\qquad\qquad\qquad \in \mathrm{coders}(\mathcal{H}_{U,V,X})
\end{aligned}$$

The *generalising derived substrate history coder*, $C_{{\rm H},V,{\rm T},{\rm G}}$, is intermediate between an *index history coder*, $C_{\rm H}$ and a *classification history coder*, $C_{\rm G}$. Let $H \in \mathcal{H}_{U,V,X}$ be a *history* of the *derived substrate history coder* domain. The *derived history* is $H*T$. The *histogram* is $A = \mathrm{histogram}(H)$. The *event identifiers* are encoded in *space* of $\mathrm{spaceIds}(y, z)$. The *derived history*, $H * T$, is encoded in *fixed width space*, $\mathrm{spaceEventsDerived}(U)(H, T) := z \ln w'$, where $w' = |T^{-1}|$. Then each *sub-history*, H_C, corresponding to a *component* of the *partition*, $H_C \subseteq H$, where $(R, C) \in T^{-1}$, is encoded as a *component histogram* having *space* $\mathrm{spaceCounts}(U)(A * C)$ and a *component classification* having *space* $\text{spaceClassification}(A * C)$.

The *history*, $H \in \mathcal{H}_{U,V,X}$, can be encoded in an intermediate tuple $T_H = ((z, Z_I), L', L, M) \in \mathbf{N}^2 \times \mathcal{L}(\mathbf{N}) \times \mathcal{L}(\mathbf{N}) \times \mathcal{L}(\mathbf{N})$.

The first element, $(z, Z_I) \in \mathbf{N}^2$, of the tuple, T_H, encodes the *event identifiers* in the same way as the *index history coder*, C_{H}, above. The *space* is $\text{spaceIds}(y, z)$, where $y = |X|$ and $z = |H| = \text{size}(A)$.

The second element, $L' \in \mathcal{L}(\mathbf{N})$, of the tuple, T_H, is encoded as in the *index derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{H}}$, above. The *space* of this element is $\text{spaceEventsDerived}(U)(H, T) := z \ln w'$.

The third element, $L \in \mathcal{L}(\mathbf{N})$, and the last element, $M \in \mathcal{L}(\mathbf{N})$, of the tuple, T_H, are encoded as in the *classification derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{G}}$, above. The *space* is

$$\text{spaceCountsPartition}(A, T) + \text{spaceClassificationPartition}(A, T)$$

The total *space* of the *generalising derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{G}}$, of a *history* $H \in \mathcal{H}_{U,V,X}$ is the sum of the *ids space*, *derived events space*, *partitioned counts space*, and *partitioned classification space*

$$\begin{aligned} \text{space}(C_{\mathrm{H},V,\mathrm{T},\mathrm{G}})(H) \;=\; & \text{spaceIds}(|X|, |H|) + \\ & \text{spaceEventsDerived}(U)(H, T) + \\ & \text{spaceCountsPartition}(A, T) + \\ & \text{spaceClassificationPartition}(A, T) \end{aligned}$$

Similar to the *specialising coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, above, given some *substrate history* $H \in \mathcal{H}_{U,V,X}$, the *transform* T in the parameters of the *generalising derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{G}}$, may be chosen to minimise the encoding *space* with respect to the *space* in either of the *canonical history coders*, (i) the *classification history coder*, C_{G}, and (ii) the *index history coder*, C_{H}. That is, the *transform* is chosen to minimise

$$\big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)\big) + \big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)\big)$$

In the case of *non-overlapping transform*, $\neg\text{overlap}(T)$, this simplifies to

$$\begin{aligned}
&\big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)\big) + \big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)\big) \\
&\quad = \big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{G}}(H)\big) + \big(C^{\mathrm{s}}_{\mathrm{G},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)\big) + \\
&\qquad \big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H)\big) + \big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)\big) \\
&\quad = \big(C^{\mathrm{s}}_{\mathrm{H}}(H * T) - C^{\mathrm{s}}_{\mathrm{G}}(H * T)\big) + \big(C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A)\big) + \\
&\qquad \sum_{(\cdot,C) \in T^{-1}} \big(C^{\mathrm{s}}_{\mathrm{G}}(H_C) - C^{\mathrm{s}}_{\mathrm{H}}(H_C)\big) + \big(C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)\big)
\end{aligned}$$

The first term, $C^{\mathrm{s}}_{\mathrm{H}}(H * T) - C^{\mathrm{s}}_{\mathrm{G}}(H * T)$, equals the negative of the first term of the *specialising-canonical space difference*. The second term, $C^{\mathrm{s}}_{\mathrm{A},V,\mathrm{T}}(A) - C^{\mathrm{s}}_{\mathrm{A},V}(A)$, equals the fourth term of the *specialising-canonical space difference*. The third term, $\sum_{(\cdot,C) \in T^{-1}} \big(C^{\mathrm{s}}_{\mathrm{G}}(H_C) - C^{\mathrm{s}}_{\mathrm{H}}(H_C)\big)$, equals the negative of the third term of the *specialising-canonical space difference*. The fourth term, $C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{H}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)$, equals the second term of the *specialising-canonical space difference*.

Thus, the *generalising-canonical space difference*, $2C^{\mathrm{s}}_{\mathrm{H},V,\mathrm{T},\mathrm{G}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H)$, is minimised when (a) the *possible derived volume* is minimised, (b) the *derived entropy* or *component size entropy* is maximised, (c) the *expected component entropy* is minimised, and (d) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

The four *derived substrate history coders*, (a) the *index derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{H}}$, (b) the *classification derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{G}}$, (c) the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, and (d) the *generalising derived substrate history coder*, $C_{\mathrm{H},V,\mathrm{T},\mathrm{G}}$, all encode the *history* by means of a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ in *variables* $V = \text{und}(T)$. Now consider extending the *model* for the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, to (i) *fuds*, (ii) *decompositions*, and (iii) *fud decompositions*.

Given the *one functional definition set* $F \in \mathcal{F}_{U,1}$, such that $\text{und}(F) = V$, which is constrained such that only first *layer transforms* depend on the *substrate*, $\forall T \in F\ (\text{und}(T) \nsubseteq V \implies \text{und}(T) \cap V = \emptyset)$, a *substrate history* $H \in \mathcal{H}_{U,V,X}$ could be encoded simply by encoding each of the *transforms* separately in a *specialising coder*. The total *space* of *coder* $C(F) \in \text{coders}(\mathcal{H}_{U,V,X})$

would be

$$\begin{aligned} C(F)^{\mathrm{s}}(H) &= \sum (C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H\%V_T) - s : T \in F,\ V_T \subseteq V) \\ &+ \sum (C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H * \mathrm{dep}(F,V_T)^{\mathrm{T}}) - s : T \in F,\ V_T \nsubseteq V) \\ &+ s \end{aligned}$$

where $s = \mathrm{spaceIds}(|X|,|H|)$, $V_T = \mathrm{und}(T)$, $\mathrm{dep} = \mathrm{depends}$ and the *specialising derived substrate history coder* is constructed

$$C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T) = \mathrm{coderHistorySubstrateDerivedSpecialising}(U,X,T,D_{\mathrm{S}},D_{\mathrm{X}})$$

This method of encoding, however, means that any *fud*, F, requires at least as much *space* as its bottom *layer*, $C(F)^{\mathrm{s}}(H) \geq C(\{T : T \in F,\ V_T \subseteq V\})^{\mathrm{s}}(H)$. This is because the bottom *layer* has complete coverage of the *substrate* and so the *history* can be decoded from the decodings of the *reduced histories* of the bottom *layer coders*, $\{H\%V_T : T \in F,\ V_T \subseteq V\}$.

Consider a two *layer fud* $F = \{T_1, T_2\}$, where $V_1 = V$ and $V_2 = W_1$. A *specialising coder* of the first *layer*, $C_{\mathrm{G},V_1,\mathrm{T},\mathrm{H}}(T_1)$, encodes the *derived history*, $H * T_1$, in a *classification coder* which has a *space* of $C^{\mathrm{s}}_{\mathrm{G},W_1}(H * T_1)$ (ignoring *ids space*). In some cases, however, the first *layer derived history* may be encoded in less *space* by means of a *specialising coder* of the second *layer*, if $C_{\mathrm{G},V_2,\mathrm{T},\mathrm{H}}(T_2)^{\mathrm{s}}(H * T_1) < C^{\mathrm{s}}_{\mathrm{G},W_1}(H * T_1)$. So consider a *coder* that encodes the *layer derived history* in the *layer* above if it exists. The *space* in this case is the sum of the *ids space*, second *layer derived counts space*, second *layer derived classification space*, second *layer partitioned events space* and first *layer partitioned events space*,

$$\begin{aligned} &C_{\mathrm{G},V_1,\mathrm{T},\mathrm{H}}(T_1)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{G},W_1}(H * T_1) + C_{\mathrm{G},V_2,\mathrm{T},\mathrm{H}}(T_2)^{\mathrm{s}}(H * T_1) = \\ &\qquad \mathrm{spaceIds}(|X|,|H|) + \\ &\qquad \mathrm{spaceCountsDerived}(U)(A,T_2) + \\ &\qquad \mathrm{spaceClassification}(A * T_1 * T_2) + \\ &\qquad \mathrm{spaceEventsPartition}(A * T_1, T_2) + \\ &\qquad \mathrm{spaceEventsPartition}(A, T_1) \end{aligned}$$

The difference in *space* is

$$\begin{aligned} &C_{\mathrm{G},V_2,\mathrm{T},\mathrm{H}}(T_2)^{\mathrm{s}}(H * T_1) - C^{\mathrm{s}}_{\mathrm{G},W_1}(H * T_1) = \\ &\qquad \sum_{(\cdot,C) \in T_2^{-1}} \big(C^{\mathrm{s}}_{\mathrm{H}}((H * T_1)_C) - C^{\mathrm{s}}_{\mathrm{G}}((H * T_1)_C)\big) + \\ &\qquad C_{\mathrm{A},V_2,\mathrm{T}}(T_2)^{\mathrm{s}}(A * T_1) - C^{\mathrm{s}}_{\mathrm{A},V_2}(A * T_1) \end{aligned}$$

which is sometimes negative.

The *specialising fud substrate history coder* is constructed

$C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F) =$
$\quad \mathrm{coderHistorySubstrateFudSpecialising}(U,X,F,D_{\mathrm{S}},D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,V,X})$

The total *space* of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}$, of a *history* $H \in \mathcal{H}_{U,V,X}$ is the sum of the *specialising derived substrate history coder space* for each *transform* less the sum of the *classification history coder space* for each *transform* which has *derived variables* that are not in the *fud derived variables*,

$$\begin{aligned}
&\mathrm{space}(C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F))(H) = \\
&\qquad \sum (C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H\%V_T) - s : T \in F,\ V_T \subseteq V) \\
&\quad + \ \sum (C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H * \mathrm{dep}(F,V_T)^{\mathrm{T}}) - s : T \in F,\ V_T \not\subseteq V) \\
&\quad - \ \sum (C^{\mathrm{s}}_{\mathrm{G},W_T}(H * \mathrm{dep}(F,W_T)^{\mathrm{T}}) - s : T \in F,\ W_T \cap \mathrm{der}(F) = \emptyset) \\
&\quad + \ s
\end{aligned}$$

This definition can be generalised to allow for *fuds* having *transforms* higher than the first *layer* that have *underlying variables* in the *substrate*, $\mathrm{und}(T) \cap V \notin \{\emptyset, \mathrm{und}(T)\}$,

$$\begin{aligned}
&\mathrm{space}(C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F))(H) = \\
&\qquad \sum (C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H_F\%V_T) - s : T \in F) \\
&\quad - \ \sum (C^{\mathrm{s}}_{\mathrm{G},W_T}(H_F\%W_T) - s : T \in F,\ W_T \cap \mathrm{der}(F) = \emptyset) \\
&\quad + \ s
\end{aligned}$$

where the *history*, H_F, is the *expanded history* to $\mathrm{vars}(F)$,

$$H_F \ = \ \{(x,S) : (x,R) \in H,\ (S,\cdot) \in \prod_{(X,\cdot) \in F} X,\ R \subseteq S\}$$

Note that, given a *fud* that has *underlying variables* that are a proper subset of the *substrate*, $\mathrm{und}(F) \subset V$, the *fud* can be *expanded* to the remaining *variables* $L = V \setminus \mathrm{und}(F)$ by adding a *unary partition transform*, $\{L^{\mathrm{CS}}\}^{\mathrm{T}}$, which adds *space* of $C^{\mathrm{s}}_{\mathrm{H},L}(H\%L) - s$.

Note also that, given a *fud* that has a *transform* $T \in F$ having *derived variables* that are partially a subset of the *fud derived variables*, $W_T \cap \mathrm{der}(F) \subset$

W_T, the *fud* can be altered to save the *classification space* of the remaining *variables* $L_T = W_T \cap \mathrm{der}(F)$ by adding a *self partition transform*, $L_T^{\mathrm{CS}\{\}\mathrm{T}}$, which subtracts *space* of $C^{\mathrm{s}}_{\mathrm{G},W_T}(H_F \% W_T) - C^{\mathrm{s}}_{\mathrm{G},L_T}(H_F \% L_T)$.

In the *law-like* case where the *fud* has a *top transform*, $\exists T \in F\ (W_T = \mathrm{der}(F))$, the *space* is

$$\begin{aligned}
&\mathrm{space}(C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F))(H) = \\
&\qquad\qquad \mathrm{spaceIds}(|X|, |H|) + \\
&\qquad\qquad \mathrm{spaceCountsDerived}(U)(A, F^{\mathrm{T}}) + \\
&\qquad\qquad \mathrm{spaceClassification}(A * F^{\mathrm{T}}) + \\
&\qquad\qquad \sum_{T \in F} \mathrm{spaceEventsPartition}(A * \mathrm{dep}(F, V_T)^{\mathrm{T}}, T)
\end{aligned}$$

Let w' be the *possible derived volume* of the *transform* of the *fud*, $w' = |(F^{\mathrm{T}})^{-1}|$. The *space* of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}$, varies (i) with the *possible fud derived volume*, w', where the *possible fud derived volume* is less than the *size*, $w' < z$, otherwise with the *size* scaled *log possible fud derived volume*, $z \ln w'$, (ii) with the *size* scaled *transform fud derived entropy* and (iii) against the sum of the *size* scaled *component size cardinality cross entropies* of the *transforms* of the *fud*,

$$\begin{aligned}
C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H)\ \sim&\\
&(w'\ :\ w' < z)\ +\ (z \ln w'\ :\ w' \geq z)\\
&+ z \times \mathrm{entropy}(A * F^{\mathrm{T}})\\
&- z \times \sum_{T \in F} \mathrm{entropyCross}(A * \mathrm{dep}(F, W_T)^{\mathrm{T}}, V_T^{\mathrm{C}} * T)
\end{aligned}$$

So the *space* of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}$, is minimised when (a) the *possible fud derived volume* is minimised, (b) the *derived entropy* or *component size entropy* of the *fud transform* is minimised, and (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for each of the *fud transforms*.

Consider a *transform* T added to the *top* of a *fud* F. As mentioned above, in some cases the *specialising space* of the *transform*, T, is less than the *classification space* of its *underlying*. The change in the *specialising fud coder space* equals the difference between the *specialising space* and the *underlying classification space* of the *transform*,

$$\begin{aligned}
C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F \cup \{T\})^{\mathrm{s}}(H) - C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H)\ =&\\
&C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H * F^{\mathrm{T}}) - C^{\mathrm{s}}_{\mathrm{G},V_T}(H * F^{\mathrm{T}})
\end{aligned}$$

Recursing for all *transforms*, conjecture that the *specialising-classification space difference* for the *fud* varies with the sum of the *specialising-classification space differences* for the *transforms*,

$$C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) \sim \sum_{T \in F} \big(C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H * \mathrm{dep}(F,V_T)^{\mathrm{T}}) - C^{\mathrm{s}}_{\mathrm{G},V_T}(H * \mathrm{dep}(F,V_T)^{\mathrm{T}})\big)$$

Conjecture that it is also the case that the *specialising-index space difference* for the *fud* varies with the sum of the *specialising-index space differences* for the *transforms*,

$$C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) \sim \sum_{T \in F} \big(C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H * \mathrm{dep}(F,V_T)^{\mathrm{T}}) - C^{\mathrm{s}}_{\mathrm{H},V_T}(H * \mathrm{dep}(F,V_T)^{\mathrm{T}})\big)$$

Together the *specialising-canonical space difference* for the *fud* varies with the sum of the *specialising-canonical space differences* for the *transforms*,

$$\begin{aligned}2C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) \sim \\ \sum_{T \in F} \big(2C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H * \mathrm{dep}(F,V_T)^{\mathrm{T}}) \\ -C^{\mathrm{s}}_{\mathrm{H},V_T}(H * \mathrm{dep}(F,V_T)^{\mathrm{T}}) - C^{\mathrm{s}}_{\mathrm{G},V_T}(H * \mathrm{dep}(F,V_T)^{\mathrm{T}})\big)\end{aligned}$$

The *specialising-canonical space difference* varies (i) with twice the total *possible derived volume* of the *transforms*, where the *possible derived volumes* are less than the *size*, otherwise with twice the total *size* scaled *log possible derived volume*, (ii) with the sum of the *size* scaled *derived entropies*, (iii) against twice the sum of the *size* scaled *component size cardinality cross entropies* and (iv) against the sum of the *size* scaled *size expected component entropies*,

$$\begin{aligned}2C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) \sim \\ \sum_{T \in F} 2\big((w'_T \ : \ w'_T < z) \ + \ (z \ln w'_T \ : \ w'_T \geq z)\big) \\ + \sum_{T \in F} z \times \mathrm{entropy}(A * T_F) \\ - \sum_{T \in F} 2z \times \mathrm{entropyCross}(A * T_F, V_T^{\mathrm{C}} * T) \\ - \sum_{T \in F} z \times \mathrm{entropyComponent}(A * \mathrm{dep}(F,V_T)^{\mathrm{T}}, T)\end{aligned}$$

where $w'_T = |T^{-1}|$ and $T_F = \mathrm{dep}(F, W_T)^{\mathrm{T}}$. So the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, is minimised when (a) the total *possible derived volume* is minimised, (b) the total *derived entropy* is minimised, (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for each *transform*, and (d) the total *expected component entropy* is maximised. The *canonical* terms, $C^{\mathrm{s}}_{\mathrm{H},V}(H)$ and $C^{\mathrm{s}}_{\mathrm{G},V}(H)$, are independent of the *model*, so these properties are also the properties of the *specialising derived substrate history coder space*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H)$.

If it is the case that the *space* of the *fud*, F, is less than that of any proper *subfud*,

$$\forall G \subset F\ (\mathrm{und}(G) = \mathrm{und}(F) \implies C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) < C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(G)^{\mathrm{s}}(H))$$

then the *specialising-classification space* differences are always negative,

$$\begin{aligned}&\forall T \in F\ (V_T \cap V = \emptyset \implies \\ &\quad C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H * \mathrm{dep}(F, V_T)^{\mathrm{T}}) - C^{\mathrm{s}}_{\mathrm{G},V_T}(H * \mathrm{dep}(F, V_T)^{\mathrm{T}}) < 0)\end{aligned}$$

Let it also be the case that (i) the *fud*, F, consists of a single *transform* in each *layer*, $\forall i \in \{1 \ldots l\}\ (F_i = \{T_i\})$ where $l = \mathrm{layer}(F, \mathrm{der}(F))$ and $F_i = \{T : T \in F,\ \mathrm{layer}(F, W_T) = i\}$, (ii) the *fud*, F, is a *linear fud* which is such that the *underlying variables* of each *layer* are the *derived variables* of the *layer* immediately below, $\forall i \in \{2 \ldots l\}\ (V_i = W_{i-1})$ where $V_i = \mathrm{und}(F_i)$ and $W_i = \mathrm{der}(F_i)$, and (iii) the *size* is greater than the *derived volume* of each of the intermediate *layers*, $\forall i \in \{1 \ldots l\}\ (z > |W_i^{\mathrm{C}}|)$. Let the cumulative *fud* $F_{\{1 \ldots i\}} = \bigcup_{j \in \{1 \ldots i\}} F_j = \mathrm{dep}(F, W_i)$. Then conjecture that, in general, (i) the *derived entropy* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (\mathrm{entropy}(A * F^{\mathrm{T}}_{\{1 \ldots i\}}) < \mathrm{entropy}(A * F^{\mathrm{T}}_{\{1 \ldots i-1\}}))$$

(ii) the *possible derived volume* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (|W_i^{\mathrm{C}}| < |W_{i-1}^{\mathrm{C}}|)$$

(iii) the *expected component entropy* increases up the *layers*,

$$\begin{aligned}&\forall i \in \{2 \ldots l\} \\ &\quad (\mathrm{entropyComponent}(A, F^{\mathrm{T}}_{\{1 \ldots i\}})\ >\ \mathrm{entropyComponent}(A, F^{\mathrm{T}}_{\{1 \ldots i-1\}}))\end{aligned}$$

and (iv) the *component size cardinality cross entropy* increases up the *layers*,

$$\begin{aligned}&\forall i \in \{2 \ldots l\} \\ &\quad (\mathrm{entropyCross}(A * F^{\mathrm{T}}_{\{1 \ldots i\}}, V^{\mathrm{C}} * F^{\mathrm{T}}_{\{1 \ldots i\}})\ > \\ &\qquad\qquad \mathrm{entropyCross}(A * F^{\mathrm{T}}_{\{1 \ldots i-1\}}, V^{\mathrm{C}} * F^{\mathrm{T}}_{\{1 \ldots i-1\}}))\end{aligned}$$

Terms (i) and (iv) together are equivalent to the *component size cardinality relative entropy* increasing up the *layers*,

$$\begin{aligned}&\forall i \in \{2 \dots l\}\\ &\quad(\text{entropyRelative}(A * F^{\mathrm{T}}_{\{1\dots i\}}, V^{\mathrm{C}} * F^{\mathrm{T}}_{\{1\dots i\}}) \ > \\ &\qquad\qquad \text{entropyRelative}(A * F^{\mathrm{T}}_{\{1\dots i-1\}}, V^{\mathrm{C}} * F^{\mathrm{T}}_{\{1\dots i-1\}}))\end{aligned}$$

Therefore to minimise *specialising fud space*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H)$, by varying the *fud*, in the case where the first *layer* has complete coverage of the *substrate*, $V_1 = V$, it is sufficient to (a) maximise the *component size cardinality cross entropy* and the *expected component entropy*, and (b) minimise the *derived entropy* and *possible derived volume*, for each *layer* in sequence from the first *layer* upwards. Thus the optimisation of a *fud* without a *layer* limit may be made computable by building the *fud layer* by *layer*, minimising the *specialising space* at each step, until the addition of a *layer* fails to reduce the *specialising space*.

The *space* of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H)$, is related to the *space* of the *specialising derived substrate history coder* parameterised by the *transform* of the *fud*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(F^{\mathrm{T}})^{\mathrm{s}}(H)$. In some cases the *spaces* are equal. For example, consider the case of a single *layer fud* of *unary partition transforms*, $\forall T \in F\ ((V_T \subseteq V) \wedge (T = \{V_T^{\mathrm{CS}}\}^{\mathrm{T}}))$, when applied to a *scaled cartesian histogram*, $A = V^{\mathrm{CS}} \times \{z/v\}$ where $z = |H|$ and $v = |V^{\mathrm{C}}|$. There is no *derived history classification space*,

$$\begin{aligned}C^{\mathrm{s}}_{\mathrm{G}}(H * F^{\mathrm{T}}) - s &= \sum_{T \in F} (C^{\mathrm{s}}_{\mathrm{G}}(H * T) - s)\\ &= 0\end{aligned}$$

and the *partitioned events spaces* of the *underlying components* are equal,

$$\begin{aligned}\text{spaceEventsPartition}(A, F^{\mathrm{T}}) &= \sum_{T \in F} \text{spaceEventsPartition}(A, T)\\ \sum_{(R,C) \in (F^{\mathrm{T}})^{-1}} (A * F^{\mathrm{T}})_R \ln |C| &= \sum_{T \in F} \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C|\\ &= z \ln v\end{aligned}$$

so the *specialising spaces* are equal, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) = C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(F^{\mathrm{T}})^{\mathrm{s}}(H)$.

In the *law-like* case where the *fud* has a *top transform*, $\exists T \in F\ (W_T =$

der(F)), the *space* difference is just the difference in *partitioned events space*,

$$C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) - C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(F^{\mathrm{T}})^{\mathrm{s}}(H) = \\ \sum_{T \in F} \mathrm{spaceEventsPartition}(A * \mathrm{dep}(F,V_T)^{\mathrm{T}}, T) \\ - \mathrm{spaceEventsPartition}(A, F^{\mathrm{T}})$$

which is the *size* scaled difference in *component size cardinality cross entropies*,

$$C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) - C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(F^{\mathrm{T}})^{\mathrm{s}}(H) = \\ z \times \mathrm{entropyCross}(A * F^{\mathrm{T}}, V^{\mathrm{C}} * F^{\mathrm{T}}) \\ - z \times \sum_{T \in F} \mathrm{entropyCross}(A * \mathrm{dep}(F,W_T)^{\mathrm{T}}, V_T^{\mathrm{C}} * T)$$

Now consider extending the *model* for the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, to *decompositions*. Let a *distinct decomposition* of *one functional transforms* $D \in \mathcal{D}_U = \mathcal{D} \cap \mathrm{trees}(\mathcal{S}_U \times \mathcal{T}_{U,\mathrm{f},1})$ be such that the *fud* of each path of the *application* tree has complete coverage of the *substrate*,

$$\forall L \in \mathrm{paths}(D^*)\ (\bigcup_{(\cdot,(T,\cdot)) \in L} V_T = V)$$

where $V_T = \mathrm{und}(T)$.

The *specialising decomposition substrate history coder* is constructed

$$C_{\mathrm{G},V,\mathrm{D},\mathrm{H}}(D) = \\ \mathrm{coderHistorySubstrateDecompSpecialising}(U,X,D,D_{\mathrm{S}},D_{\mathrm{X}}) \\ \in \mathrm{coders}(\mathcal{H}_{U,V,X})$$

The total *space* of the *specialising decomposition substrate history coder*, $C_{\mathrm{G},V,\mathrm{D},\mathrm{H}}$, of a *history* $H \in \mathcal{H}_{U,V,X}$ is the sum of the *specialising derived substrate history coder space* for each *transform* for each *slice*,

$$\mathrm{space}(C_{\mathrm{G},V,\mathrm{D},\mathrm{H}}(D))(H) = \\ \sum(C_{\mathrm{G},V_T,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H_C\%V_T) - s : (C,T) \in \mathrm{cont}(D)) \\ + \ s$$

where $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$, the *sliced-history* is defined $H_C = \{(x,S) : (x,S) \in H,\ \{S\}^{\mathrm{U}} * C \neq \emptyset\}$, and $s = \mathrm{spaceIds}(|X|,|H|)$.

Similarly, consider extending the *model* for the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}$, to *fud decompositions.* Let a *distinct fud decomposition* of *one functional definition sets* $D \in \mathcal{D}_{\mathrm{F},U} = \mathcal{D}_{\mathrm{F}} \cap \mathrm{trees}(\mathcal{S}_U \times \mathcal{F}_{U,1})$ be such that the *fud* of each path of the *application* tree has complete coverage of the *substrate*,

$$\forall L \in \mathrm{paths}(D^*)\ (\bigcup_{(\cdot,(F,\cdot)) \in L} V_F = V)$$

where $V_F = \mathrm{und}(F)$.

The *specialising fud decomposition substrate history coder* is constructed

$$\begin{aligned} &C_{\mathrm{G},V,\mathrm{D},\mathrm{F},\mathrm{H}}(D) = \\ &\quad \mathrm{coderHistorySubstrateFudDecompSpecialising}(U, X, D, D_{\mathrm{S}}, D_{\mathrm{X}}) \\ &\qquad\qquad \in \mathrm{coders}(\mathcal{H}_{U,V,X}) \end{aligned}$$

The total *space* of the *specialising fud decomposition substrate history coder*, $C_{\mathrm{G},V,\mathrm{D},\mathrm{F},\mathrm{H}}$, of a *history* $H \in \mathcal{H}_{U,V,X}$ is the sum of the *specialising fud substrate history coder space* for each *fud* for each *slice*,

$$\begin{aligned} &\mathrm{space}(C_{\mathrm{G},V,\mathrm{D},\mathrm{F},\mathrm{H}}(D))(H) = \\ &\quad \sum (C_{\mathrm{G},V_F,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H_C \% V_F) - s_C : (C, F) \in \mathrm{cont}(D)) \\ &\quad +\ s \end{aligned}$$

where $s_C = \mathrm{spaceIds}(|X|, |H_C|)$.

Note that, whereas the optimisation of a *fud* without a *layer* limit may be made computable by building the *fud layer* by *layer*, this method of optimisation of a *fud decomposition* merely constructs a *decomposition* that has singleton *fuds* with singleton *underlying variables*, $\forall F \in \mathrm{fuds}(D)\ (|F| = 1 \wedge |\mathrm{und}(F)| = 1)$. This is because a *decomposition fud* $F \in \mathrm{fuds}(D)$ is not constrained to cover the *substrate*, $V_F \subseteq V$. Only the union of the *fuds* of the *application* paths must have complete coverage. So the first *layer* of a *fud*, F, is not constrained at least to partition the *substrate*, $\{V_T : T \in F_1\} \in \mathrm{B}(V)$, and the minimisation of *specialising space* is uninteresting. A way to address this would be to limit the *transforms* of the *fud* to have a minimum *underlying dimension*, $\forall T \in F\ (|V_T| \geq \mathrm{kmin})$ where $\mathrm{kmin} \in \mathbf{N}_{>1}$.

The *derived history coders* discussed above, such as the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}} \in \mathrm{coders}(\mathcal{H}_{U,X})$, are *substrate history coders* in that they are restricted to the *histories* in the *underlying variables* of the *transform*, $\mathcal{H}_{U,V,X}$, where $V = \mathrm{und}(T)$. Now consider how the *derived substrate history coders* may be generalised to the unrestricted *history coder domain* where the *histories* may be in any of the *system variables*, $\mathcal{H}_{U,X} = \bigcup\{X \to V^{\mathrm{CS}} : V \subseteq \mathrm{vars}(U)\}$.

The *expanded specialising derived history coder* $C_{\mathrm{G,T,H}}(T) \in \mathrm{coders}(\mathcal{H}_{U,X})$ is derived from the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T,H}}$. It *expands* the *transform* to the *history variables*, V_H, where the set of *history variables* is a superset of the *underlying variables*, $V = \mathrm{und}(T)$, and otherwise defaults to an *index coder*,

$$\begin{aligned} C_{\mathrm{G,T,H}}(T)^{\mathrm{s}}(H) &= (C_{\mathrm{G},V_H,\mathrm{T,H}}(T^{\mathrm{PV}_H\mathrm{T}})^{\mathrm{s}}(H) + s_{|V_H|} \ : \ V_H \supseteq V) \ + \\ &\quad\ (C^{\mathrm{s}}_{\mathrm{H}}(H) \ : \ V_H \not\supseteq V) \end{aligned}$$

where $s_n = \mathrm{spaceVariables}(U)(n)$. Note that the *expansion* is equivalent to adding *index space* for the additional *variables*, $C_{\mathrm{G},V_H,\mathrm{T,H}}(T^{\mathrm{PV}_H\mathrm{T}})^{\mathrm{s}}(H) = C_{\mathrm{G},V,\mathrm{T,H}}(T)^{\mathrm{s}}(H\%V) + C^{\mathrm{s}}_{\mathrm{H},V}(H \ \% \ (V_H \setminus V))$. The *expansion* always adds at least *canonical space*, $\mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}(H \ \% \ (V_H \setminus V)), C^{\mathrm{s}}_{\mathrm{G}}(H \ \% \ (V_H \setminus V)))$.

Similarly the *expanded specialising fud history coder* $C_{\mathrm{G,F,H}}(F) \in \mathrm{coders}(\mathcal{H}_{U,X})$ is derived from the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F,H}}$,

$$\begin{aligned} C_{\mathrm{G,F,H}}(F)^{\mathrm{s}}(H) &= (C_{\mathrm{G},V_H,\mathrm{F,H}}(F^{V_H})^{\mathrm{s}}(H) + s_{|V_H|} \ : \ V_H \supseteq V) \ + \\ &\quad\ (C^{\mathrm{s}}_{\mathrm{H}}(H) \ : \ V_H \not\supseteq V) \end{aligned}$$

where F^V is the *expansion* that adds a *unary transform* in the remaining *underlying variables*, $F \cup \{\{(V \setminus \mathrm{und}(F))^{\mathrm{CS}}\}^{\mathrm{T}}\}$. Again, the *expansion* is equivalent to adding *index space* for the additional *variables*.

Lastly the *expanded specialising fud decomposition history coder* $C_{\mathrm{G,D,F,H}}(D) \in \mathrm{coders}(\mathcal{H}_{U,X})$ is derived from the *specialising fud decomposition substrate history coder*, $C_{\mathrm{G},V,\mathrm{D,F,H}}$,

$$\begin{aligned} C_{\mathrm{G,D,F,H}}(D)^{\mathrm{s}}(H) &= (C_{\mathrm{G},V_H,\mathrm{D,F,H}}(D^{V_H})^{\mathrm{s}}(H) + s_{|V_H|} \ : \ V_H \supseteq V) \ + \\ &\quad\ (C^{\mathrm{s}}_{\mathrm{H}}(H) \ : \ V_H \not\supseteq V) \end{aligned}$$

where D^V is the *expansion* that adds a *unary transform* in the remaining *underlying variables* to the leaf *fuds* in the *decomposition* tree such that the *fud* of each path of the *application* tree has complete coverage of the *substrate*,

$$\forall L \in \mathrm{paths}(D^*) \ (\bigcup_{(\cdot,(F,\cdot)) \in L} V_F = V)$$

where $V_F = \mathrm{und}(F)$.

Given a *system* U and *event identifiers* X, a *history coder domain probability function* $P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is defined as *entropic* with respect to *history coder* $C \in \text{coders}(\mathcal{H}_{U,X})$ if the *coder* is an *entropy coder*. See appendix 'Coders and entropy' for the definition of the *entropy coder*. The *coder* is an *entropy history coder* if and only if the *space* of a *history* equals the negative logarithm of the non-zero probability, $\forall H \in \mathcal{H}_{U,X}\ (P_H > 0 \implies C^{\text{s}}(H) = -\ln P_H)$. Then the expected *space* of the *coder* equals the *entropy* of the *history probability function*.

$$\begin{aligned}
\text{expected}(P)(C^{\text{s}}) &= \sum_{H \in \mathcal{H}_{U,X}} P_H \times C^{\text{s}}(H) \\
&= -\sum(P_H \ln P_H : H \in \mathcal{H}_{U,X},\ P_H > 0) \\
&= \text{entropy}(P)
\end{aligned}$$

An *entropy coder* has the smallest expected *space* of all *coders* given the *probability function*.

Note that *entropy history coders* should be distinguished from the theoretical *variable-width history coder* C_{E} that encodes its *states* in an *entropy coder*, described in appendix 'Entropy encoding of states'. The *variable-width history coder*, C_{E}, can only encode the subset of *histories* for which there exists a *states coder* that is an *entropy coder* with respect to the normalised *histogram*. That is, $\exists C \in \text{coders}(A^{\text{FS}})\ \forall R \in A^{\text{FS}}\ (C^{\text{s}}(R) = -\ln \hat{A}_R)$ where $H \in \mathcal{H}_{U,X}$ and $A = \text{histogram}(H)$. Even for this subset of the *histories* for which an *entropy states coder*, C, can be constructed, the *space* of the *variable-width history coder* is always greater than or equal to the *space* of the *classification coder*, $C^{\text{s}}_{\text{E}}(H) \geq C^{\text{s}}_{\text{G}}(H)$.

Similar to the definition of *entropic history probability functions*, a *history coder domain probability function* $P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is defined as *structured* with respect to *derived history coder* $C \in \text{coders}(\mathcal{H}_{U,X})$, if the expected *space* of the *derived history coder* is less than the expected lesser *space* of the *canonical history coders*, (i) *index history coder*, C_{H}, and (ii) *classification history coder*, C_{G},

$$\text{expected}(P)(C^{\text{s}}) < \text{expected}(P)(\text{minimum}(C^{\text{s}}_{\text{H}}, C^{\text{s}}_{\text{G}}))$$

where $\text{minimum}(C^{\text{s}}_{\text{H}}, C^{\text{s}}_{\text{G}}) \in \mathcal{H}_{U,X} \to \ln \mathbf{N}_{>0}$.

The *degree of structure* is defined $\text{structure}(U, X) \in ((\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}) \times$

$\text{coders}(\mathcal{H}_{U,X}) \to \mathbf{Q} \ln \mathbf{Q}_{>0} / \ln \mathbf{Q}_{>0}$ as

$$\text{structure}(U,X)(P,C) := \frac{\text{canonical}(U,X)(P) - \text{expected}(P)(C^{\text{s}})}{\text{canonical}(U,X)(P) - \text{entropy}(P)}$$

where $\text{canonical}(U,X) \in ((\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}) \to \mathbf{Q}_{\geq 0} \ln \mathbf{N}_{>0}$ is defined

$$\text{canonical}(U,X)(P) := \text{expected}(P)(\text{minimum}(C^{\text{s}}_{\text{H}}, C^{\text{s}}_{\text{G}}))$$

The *degree of structure* is undefined if the *canonical coders* are already *entropic*, $\text{canonical}(U,X)(P) = \text{entropy}(P)$. The *degree of structure* is defined for all *history coders*, not just *derived history coders*.

Define the *compression* of *coder* C with respect to *probability function* P as a synonym for the *degree of structure* of *probability function* P with respect to the *coder* C.

The *degree of structure* is always less than or equal to one,

$$\forall P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}\ (\text{structure}(U,X)(P,C) \leq 1)$$

If the *degree of structure* equals one, $\text{structure}(U,X)(P,C) = 1$, the *coder*, C, is an *entropy coder* of the *probability function*, P, $\text{expected}(P)(C^{\text{s}}) = \text{entropy}(P)$.

If the *degree of structure* less than or equal to zero, $\text{structure}(U,X)(P,C) \leq 0$, the *probability function*, P, is *structureless* with respect to the *coder*, C, or, equivalently, the *coder*, C, is *non-compressing* with respect to the *probability function*, P. For example, the theoretical *variable-width history coder*, C_{E}, is *non-compressing* with respect to all *probability functions* for which it can be defined, because the *space* is always greater than or equal to the *space* of the *classification coder*, $C^{\text{s}}_{\text{E}}(H) \geq C^{\text{s}}_{\text{G}}(H)$.

Structured history probability functions are less strongly constrained than *entropic history probability functions* because *entropy coders* have least expected *space*, $0 < \text{structure}(U,X)(P,C) \leq 1$.

A *history coder* $C_{\text{min(H,G)}}$ of the lesser *space* of the *canonical history coders* can be implemented with a flag to indicate which of the *canonical coders* was chosen. The *space* is $C^{\text{s}}_{\text{min(H,G)}}(H) = \text{minimum}(C^{\text{s}}_{\text{H}}(H), C^{\text{s}}_{\text{G}}(H)) + \ln 2$. The *lesser canonical history coder*, $C_{\text{min(H,G)}}$, is necessarily *structureless*,

$$\forall P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}\ (\text{structure}(U,X)(P, C_{\text{min(H,G)}}) < 0)$$

because of the additional *space* of the flag.

Conjecture that there is no *coder* such that the uniform *history probability function*, $\hat{\mathcal{H}}_{U,X} = \mathcal{H}_{U,X} \times \{1/|\mathcal{H}_{U,X}|\} \in \mathcal{P}$, has *structure*,

$$\forall C \in \text{coders}(\mathcal{H}_{U,X})\ (\text{structure}(U,X)(\hat{\mathcal{H}}_{U,X}, C) < 0)$$

where $\text{canonical}(U,X)(\hat{\mathcal{H}}_{U,X}) \neq \text{entropy}(\hat{\mathcal{H}}_{U,X})$.

The *degree of structure* has two arguments, (i) the *probability function* $P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$, and (ii) the *coder* $C \in \text{coders}(\mathcal{H}_{U,X})$. The function can be viewed as (i) a measure of the *structure* of the *histories* of the *probability function*, P, with respect to a fixed *coder*, C, or (ii) a measure of the *compression*, or *canonical-entropic* relative *space*, of the *coder*, C, given a *probability function*, P.

3.19 Computation time and representation space

The set of *computers*, computers, is a type class that formalises computation *time* and representation *space*. Define the application of a *computer*, $\text{apply} \in \text{computers} \to (\mathcal{X} \to \mathcal{Y})$. Define the shorthand $I^* := \text{apply}(I)$. Define the domain of the application, $\text{domain} \in \text{computers} \to \text{P}(\mathcal{X})$, and the range of the application, $\text{range} \in \text{computers} \to \text{P}(\mathcal{Y})$, such that $\forall I \in \text{computers}\ (I^* \in \text{domain}(I) \to \text{range}(I))$ and $\forall I \in \text{computers}\ (\text{dom}(I^*) = \text{domain}(I))$. The computation or application *time* is defined as $\text{time} \in \text{computers} \to (\mathcal{X} \to \mathbf{N}_{>0})$. Define the shorthand $I^{\text{t}} := \text{time}(I)$. The representation *space* is defined as $\text{space} \in \text{computers} \to (\mathcal{X} \to \ln \mathbf{N}_{>0})$. Define the shorthand $I^{\text{s}} := \text{space}(I)$. See appendix 'Computers' for a more formal definition.

3.19.1 Computation of histograms

The computation of various operations on *histograms* depends on the concrete representation or encoding. Consider an *array histogram representation*. Let $A \in \mathcal{A}_U$ and $V = \text{vars}(A)$. Let *complete histogram* $A' = A + A^{\text{CZ}}$. The array is implemented in a list $L_A \in \mathcal{L}(\mathbf{Q}_{\geq 0})$ such that $\exists X \in L_A \leftrightarrow A'\ \forall((i,d),(S,c)) \in X\ (d = c)$ and $|X| = |L_A| = |A'| = v$ where $v = \text{volume}(U)(V)$. The operation to set a *count* in the *histogram* is implemented with the *list setter* on the positive rationals, $I_{\text{L,s}} = \text{listSetter}(\mathbf{Q}_{\geq 0})$. The operation to access a *count* is implemented with the corresponding *list getter*, $I_{\text{L,g}} = \text{listGetter}(\mathbf{Q}_{\geq 0})$. The *space* complexity of the list accessors is the length of the list, which is the *volume*, v.

In order to index the array, define an encoding of the *state*, $\text{index}(U,V) \in \text{enum}(V^{\text{CS}}) \subset \mathcal{S}_U \leftrightarrow \mathbf{N}_{>0}$, which is such that $\forall S \in V^{\text{CS}}\ (\text{index}(U,V)(S) \in \{1 \ldots v\})$. Then L_A is such that $\forall S \in V^{\text{CS}}\ (L_A(\text{index}(U,V)(S)) = A'_S)$. The lookup, $I^*_{\text{L,g}}(L_A, \text{index}(U,V)(S)) \in \mathbf{Q}_{\geq 0}$, has overall *time* complexity equal to that of the index computation, because the *time* complexity of the *list getter* is constant.

If the *values* of V are constrained such that $\forall w \in V\ (U_w = \{0 \ldots |U_w| - 1\})$, then a *state* $S \in V^{\text{CS}}$ maps to a tuple of *values* in $\mathbf{N}^n$, where *dimension* $n = |V|$, which is bounded by a tuple of *valencies* in $\mathbf{N}^n_{>0}$. The index function can be defined to map from the tuple of *values* to the *volume*, $\{1 \ldots v\}$, by application of the tuple of *valencies*. Let $M \in \text{L}(V)$ be the bijective map between positions in the tuples and the *variables*. M is the inverse of some enumeration of the *variables*, $\text{flip}(M) \in \text{enums}(V)$. The tuple of *values* is $\{(i, S_w) : i \in \{1 \ldots n\},\ w = M_i\} \in \mathcal{L}(\mathbf{N})$ and the tuple of *valencies* is $\{(i, |U_w|) : i \in \{1 \ldots n\},\ w = M_i\} \in \mathcal{L}(\mathbf{N}_{>\mathbf{0}})$. The index function can be defined $\text{index}(U,V)(S) = \text{index}(U,S,M) + 1$, and then recursively as

$$\begin{aligned} \text{index}(U,S,M) &:= \text{index}(U,S,\text{tail}(M)) * |U(M_1)| + S(M_1) \\ \text{index}(U,S,\emptyset) &:= 0 \end{aligned}$$

The index method is similar to the *encode* method of *coders*.

A concrete implementation of the index function depends on the representation of the *state* S. Consider the case where the *state* is represented as a list or tuple of *variable-value* pairs, $K \in \mathcal{L}(V \times \mathbf{N})$ such that $\text{set}(K) = S$ and $|K| = |S|$. That is, $\text{flip}(K) \in \text{enums}(S)$. If K is ordered in the same order as M, $\forall i \in \{1 \ldots n\}\ \Diamond(w, \cdot) = K_i\ (M_i = w)$, then the lists can be zipped together and the *time* complexity of the index computation is n. Let N be the tuple of *valencies*, $N = \{(i, |U_w|) : i \in \{1 \ldots n\},\ w = M_i\}$. Consider the *ordered list state representation*. Let $I_{\text{S,o}} = \text{stateOrderedIndexer}(U,V,M,N) \in \text{computers}$, $\text{domain}(I_{\text{S,o}}) = \{\text{flip}(Q) :\ S \in V^{\text{CS}},\ Q \in \text{enums}(S),\ \forall((w,\cdot),i) \in Q\ (M_i = w)\}$, $\text{range}(I_{\text{S,o}}) = \{1 \ldots |V^{\text{CS}}|\}$ and $\text{apply}(I_{\text{S,o}})(K) = \text{in}(N,L,n)+1$ where $L = \{(i,u) : (i,(\cdot,u)) \in K\}$, $\text{in}(N,L,i) := \text{in}(N,L,i-1) * N_i + L_i$ and $\text{in}(N,L,0) := 0$. Then $I^{\text{t}}_{\text{S,o}}(K) > nI^{\text{t}}_{\times}(1,1) + nI^{\text{t}}_{+}(0,0)$ and

$$\exists m \in \mathbf{N}_{>0}\ (I^{\text{t}}_{\text{S,o}} \in \text{O}(\{(K,n) : K \in \text{domain}(I_{\text{S,o}}),\ n = |K|\}, m))$$

Another implementation of the index function is of the *unordered list state representation* where $K \in \mathcal{L}(V \times \mathbf{N})$ is not ordered with M. Let $I_{\text{S,u}} = \text{stateUnorderedIndexer}(U,V,M,N) \in \text{computers}$. The domain is a superset of the ordered case, $\text{domain}(I_{\text{S,u}}) = \{\text{flip}(Q) :\ S \in V^{\text{CS}},\ Q \in \text{enums}(S)\}$. In the unordered case the *time* complexity of the index computation, $I^{\rm t}_{\rm S,u}$, is n^2 because the *indexer* must find each *variable* within the unordered *state* list by searching the entire list.

A third implementation of the index function is where the *variables* of the *state* are constrained to be natural numbers, $V \subset \mathbf{N}$, then the *state*, $S \in \mathbf{N} \to \mathbf{N}$, has a *binary map state representation*, $B \in \mathcal{B}(\mathrm{ran}(S))$, where $B = \mathrm{mapBinary}(S)$. Let $I_{\rm S,B} = \mathrm{stateMapBinaryIndexer}(U, V, M, N) \in \mathrm{computers}$. Let $\mathrm{domain}(I_{\rm S,B}) = \{\mathrm{mapBinary}(S) : S \in V^{\rm CS}\}$, $\mathrm{range}(I_{\rm S,B}) = \{1 \ldots |V^{\rm CS}|\}$ and $\mathrm{apply}(I_{\rm S,B})(B) = \mathrm{in}(M, N, B, n){+}1$ where $\mathrm{in}(M, N, B, i) := \mathrm{in}(M, N, B, i-1) * N_i + \mathrm{find}(B, M_i)$ and $\mathrm{in}(M, N, B, 0) := 0$. In this case the *time* complexity, $I^{\rm t}_{\rm S,B}$, is $n \ln n$, which intermediate between the ordered, $I^{\rm t}_{\rm S,o}$, and unordered, $I^{\rm t}_{\rm S,u}$, cases.

Each of the *indexers*, $I_{\rm S,o}$, $I_{\rm S,B}$ and $I_{\rm S,u}$, has a corresponding *inverse indexer*, $J_{\rm S,o}$, $J_{\rm S,B}$ and $J_{\rm S,u}$. These recursively take the modulus and perform integral division of the given index to obtain the *representation* of the *state*. For example, $\mathrm{domain}(J_{\rm S,o}) = \{1 \ldots |V^{\rm CS}|\}$ and $\mathrm{apply}(J_{\rm S,o})(k) = \mathrm{un}(M, N, n, k)$ where $\mathrm{un}(M, N, i, k) := ((M_i, k\%N_i), \mathrm{un}(M, N, i-1, k/N_i))$ and $\mathrm{un}(M, N, 0, k) := \emptyset$. Modulus and divide are the natural number operators. The method is similar to the *decode method* of *fixed width coders*. Note that the *inverse indexer* $J_{\rm S,u}$ is simply equal to $J_{\rm S,o}$ and therefore not a true *inverse computer* of $I_{\rm S,u}$ because the order of the *variable-value* list is not preserved, $\exists K \in \mathrm{domain}(I_{\rm S,u})\ (J^*_{\rm S,u}(I^*_{\rm S,u}(K)) \neq K)$, although $\forall K \in \mathrm{domain}(I_{\rm S,u})\ (\mathrm{set}(J^*_{\rm S,u}(I^*_{\rm S,u}(K))) = \mathrm{set}(K))$. The *time* complexity of *inverse indexers* $J_{\rm S,o}$ and $J_{\rm S,B}$ is the same as that of the corresponding *indexer*, n and $n \ln n$ respectively. The *time* complexity of *inverse indexer* $J_{\rm S,u}$ is that of $J_{\rm S,o}$, n. Note that these complexities suggest that it may require less computation to apply the *inverse index* to an index k than to find the *state* S indexed by k in a binary map $\mathbf{N} \to \mathcal{S}$. That is, $J^{\rm t}_{\rm S,o}(k) < I^{\rm t}_{\rm B,g}((B, k))$ where $B \in \mathcal{B}(V^{\rm CS})$, $S = \mathrm{find}(B, k)$ and $I_{\rm B,g} = \mathrm{mapBinaryGetter}(V^{\rm CS})$. If the binary map has cardinality $|\mathrm{function}(B)|$ equal to the *volume* $|V^{\rm CS}|$ then the *time* complexity is $\ln v$ where $v = |V^{\rm CS}|$. This complexity is greater than n, at least in the case of *regular* V of *valency* d when $\ln v = n \ln d$.

Given an *ordered list state representation* K where $\mathrm{set}(K) = S$, the *indexer*, $I_{\rm S,o}$, has *time* complexity n. So the overall *time* complexity of an *ordered list state representation* index operation on an *array histogram representation*, $I^*_{\rm L,g}(L_A, I^*_{\rm S,o}(K))$, is n.

Having considered the *array histogram representation* and three types of *indexers* of *state representations* that index the array, consider a *binary map histogram representation*. The binary map $B_A \in \mathcal{B}(\mathbf{Q}_{\geq 0})$ is such that $\exists X \in \mathrm{function}(B_A) \leftrightarrow A\ \forall((i,d),(S,c)) \in X\ (d = c)$ and $|\mathrm{function}(B_A)| = |A|$. The operation to set a *count* in the *histogram* is implemented with the *binary map setter* on the positive rationals, $I_{\mathrm{B,s}} = \mathrm{mapBinarySetter}(\mathbf{Q}_{\geq 0})$. The operation to access a *count* is implemented with the corresponding *binary map getter*, $I_{\mathrm{B,g}} = \mathrm{mapBinaryGetter}(\mathbf{Q}_{\geq 0})$. The *space* complexity of the *binary map accessors* is $|A| \ln |A|$. The *space* complexity of the *binary map histogram representation* is greater than the *array histogram representation* if the *histogram* is *complete*, $A^{\mathrm{U}} = A^{\mathrm{C}}$, and so $v \log_2 v \geq v$ where the *volume* $v > 1$. The *space* of the *binary map histogram representation* is least for the *effective equivalent*, $\mathrm{trim}(A) \equiv A$, because $|A^{\mathrm{F}}| \leq |A|$.

Given an index function, $\mathrm{index}(U,V) \in \mathrm{enum}(V^{\mathrm{CS}}) \subset \mathcal{S}_U \leftrightarrow \mathbf{N}_{>0}$, then B_A is such that $\forall(S,c) \in A\ (\mathrm{find}(B_A, \mathrm{index}(U,V)(S)) = c)$. The lookup, $I^*_{\mathrm{B,g}}(B_A, \mathrm{index}(U,V)(S)) \in \mathbf{Q}_{\geq 0}$, has *time* complexity of the greater of the *binary map getter*, $\ln |A|$, and that of the index computation. The implementation of the index function may be one of the *indexers* (above) depending on the *representation* of the *state*. Given an *ordered list state representation* $K \in \mathcal{L}(S)$ such that $\mathrm{flip}(K) \in \mathrm{enums}(S)$, the *indexer*, $I_{\mathrm{S,o}}$, has *time* complexity n. So the overall *time* complexity of an *ordered list state representation* index operation on a *binary map histogram representation*, $I^*_{\mathrm{B,g}}(B_A, I^*_{\mathrm{S,o}}(K))$, is $\ln v$ where $v = |A^{\mathrm{C}}|$. In the case of a *regular histogram* A of *valency* d, the *time* complexity of the operation is $n \ln d$. This *time* complexity of a *binary map histogram representation*, $n \ln d$, is greater than that for the *array histogram representation*, n. The *array histogram representation* may require less *time*, but the array must be of cardinality equal to the *volume*, and so its *space* complexity is v, which may be greater than that of the *effective binary map histogram representation*, $|A^{\mathrm{F}}| \ln |A^{\mathrm{F}}|$.

Closely related to the *binary map histogram representation*, having a function indexed by the natural numbers, are the *poset binary map histogram representations*, which are indexed by *state representations* and which implement the find operation using corresponding *comparators* rather than *indexers*. For example the poset binary map of *histogram* A on the *ordered list state representation* $B_A \in \mathrm{mapBinaryPosets}(\mathrm{domain}(I_{\mathrm{S,o}}), \mathbf{Q}_{\geq 0})$, where $I_{\mathrm{S,o}} = \mathrm{stateOrderedIndexer}(U,V,M,N)$ and $V = \mathrm{vars}(A)$, is such that $\{(\mathrm{set}(K), c) : (K,c) \in \mathrm{function}(B_A)\} = A$. The accessor operations to set and get a *count* in the *histogram* are implemented with a *poset binary map* *setter*, $I_{\mathrm{B,P,s}} = \mathrm{mapBinaryPosetSetter}(I_{\pm}, \mathrm{domain}(I_{\mathrm{S,o}}), \mathbf{Q}_{\geq 0})$ and a *poset binary map getter*, $I_{\mathrm{B,P,g}} = \mathrm{mapBinaryPosetGetter}(I_{\pm}, \mathrm{domain}(I_{\mathrm{S,o}}), \mathbf{Q}_{\geq 0})$ given some *ordered list state comparator* $I_{\pm}$ such that $\mathrm{domain}(I_{\pm}) = \mathrm{domain}(I_{\mathrm{S,o}}) \times \mathrm{domain}(I_{\mathrm{S,o}})$. The *ordered list state comparator* compares the *values* of a pair of *ordered list states* in sequence, $I_{\pm}^{*}((K, J)) \in \{-1, 0, 1\}$ where $K, J \in \mathrm{domain}(I_{\mathrm{S,o}})$. The *comparators* do not have *inverse computers* unlike the *indexers*. Assuming the self comparison has the greatest *time*, the *time complexity* of the *ordered list state comparator* is the same as for the *ordered list state indexer*, n, where $n = |V|$. The lookup, $I_{\mathrm{B,P,g}}^{*}(B_A, K) \in \mathbf{Q}_{\geq 0}$, has *time* complexity of $n \ln v$ where $v = \mathrm{volume}(U)(V)$. In the case of a *regular histogram* A of *valency* d, the *time* complexity of the operation is $n^2 \ln d$.

As well as *array* and *binary map representations* of *histograms* there are *list histogram representations*. For example, consider the *list representation* of *histogram* A in *variables* $V = \mathrm{vars}(A)$ in *system* U where each *state* is an *ordered list state representation* of *variable-values*, $P_A \in \mathcal{L}(\{(\mathrm{ord}(U, D)(S), c) : (S, c) \in A\}) \subset \mathcal{L}(\mathcal{L}(\mathcal{V}_U \times \mathcal{W}_U) \times \mathbf{Q}_{\geq 0})$ where $D \in \mathrm{enums}(\mathrm{vars}(U))$, and $\mathrm{ord}(U, D)(S) \in \mathcal{L}(S)$. In this case, $\{((\mathrm{set}(L), c), i) : (i, (L, c)) \in P_A\} \in \mathrm{enums}(A)$. The $\mathrm{ord}(U, D)$ function implies some *variables* tuple, $\exists M \in \mathcal{L}(V)\ (\mathrm{flip}(M) \in \mathrm{enums}(V) \wedge (\forall S \in V^{\mathrm{CS}}\ (\{(i, w) : (i, (w, \cdot)) \in \mathrm{ord}(U, D)(S)\} = M)))$. Note that the order D is implied if the *variables* are natural numbers, $V \subset \mathbf{N}$. If the *values* are also natural numbers, for example if $\forall w \in V\ (U_w = \{1 \ldots |U_w|\})$, then $P_A \in \mathcal{L}(\mathcal{L}(\mathbf{N} \times \mathbf{N}) \times \mathbf{Q}_{\geq 0})$.

Another example of a *list representation* of *histograms* is where *variables* are natural numbers, $V \subset \mathbf{N}$, and the *states* are *binary map state representations*, $Q_A \in \mathcal{L}(\mathcal{B}(\mathcal{W}_U) \times \mathbf{Q}_{\geq 0})$ which is such that $\{(\mathrm{function}(B), c) : (B, c) \in Q_A\} = A$. If the *values* are also natural numbers, then $Q_A \in \mathcal{L}(\mathcal{B}(\mathbf{N}) \times \mathbf{Q}_{\geq 0})$.

A third example of *list representation* of *histograms* is where the *state* is represented by an index, $R_A \in \mathcal{L}(\{(\mathrm{index}(U, V)(S), c) : (S, c) \in A\}) \subset \mathcal{L}(\mathbf{N} \times \mathbf{Q}_{\geq 0})$. Here the *index state representation* is implemented with one of the *indexers* (above). The *histogram* is recovered from the *representation* by means of the corresponding *inverse indexer*. For example, in the case of the *ordered list state representation indexer*, $\mathrm{flip}(R_A) \in \mathrm{enums}(\{(I_{\mathrm{S,o}}^{*}(\mathrm{ord}(U, D)(S)), c) : (S, c) \in A\})$ and $\{(\mathrm{set}(J_{\mathrm{S,o}}^{*}(k)), c) : (i, (k, c)) \in R_A\} = A$. This example of a *list representation* is related to the *array representation* of *histograms*, $\mathrm{set}(R_A) \in \mathcal{L}(\mathbf{Q}_{\geq 0})$, in the special case where A is *complete*, $A^{\mathrm{U}} = A^{\mathrm{C}}$. The *list representation* is also related to the *binary map representation* of *histograms*, $\mathrm{set}(R_A) = \mathrm{function}(B)$ where $B \in \mathcal{B}(\mathbf{Q}_{\geq 0})$.

All of these *list histogram representations* are such that the cardinality of the list is the same as that of the *histogram*, $|P_A| = |Q_A| = |R_A| = |A|$. Unlike the *array representation*, these *list representations* are not subject to the constraint that the cardinality be equal to the *volume*, $|A^{\mathrm{C}}|$. In fact, there are *list representations* which are not constrained to be in a bijective mapping to the *histogram*, but which equal the given *histogram* after a summation. For example, $X_A \in \mathcal{L}(\{\mathrm{ord}(U,D)(S) : S \in V^{\mathrm{CS}}\} \times \mathbf{Q}_{\geq 0})$ such that $\sum(\{\{(\mathrm{set}(L),c)\} : (i,(L,c)) \in X_A) \equiv A$. In this case $\exists X_A \in \mathcal{L}(\mathcal{L}(\mathcal{V}_U \times \mathcal{W}_U) \times \mathbf{Q}_{\geq 0})$ $(\{(\mathrm{set}(L),c) : (i,(L,c)) \in X_A\} \notin \mathcal{S}_U \to \mathbf{Q}_{\geq 0})$. An example is that of a *history* $H \in \mathcal{H}$ which is such that $\mathrm{ids}(H) = \{1 \ldots |H|\}$, $\mathrm{histogram}(H) = A$, $|\mathrm{states}(H)| < |H|$ and $X_A = \{(i,(\mathrm{ord}(U,D)(S),1)) : (i,S) \in H\}$. Then $|X_A| = |H| > |A|$. Both the *time* and *space* complexities of *list histogram representations* is the length of the list, which is at least $|A|$. While the *space* complexity of *list histogram representations* may be less than that of the *binary map representations*, the *time* complexity may be greater.

Given a *system* U and set of *variables* $V \subseteq \mathcal{V}_U$, consider the computation *time* to calculate the *cartesian* set of *states*, $\mathrm{cartesian}(U)(V) \subset \mathcal{S}_U$. Let $I_{\mathrm{C}} = \mathrm{cartesianer}(U) \in \mathrm{computers}$. Then $\mathrm{domain}(I_{\mathrm{C}}) = \mathrm{P}(\mathcal{V}_U)$, $\mathrm{range}(I_{\mathrm{C}}) = \{Q : Q \subset \mathcal{S}_U,\ \forall S \in Q\ (\mathrm{vars}(S) = \mathrm{vars}(Q))\}$ and $\mathrm{apply}(I_{\mathrm{C}})(V) = \mathrm{cartesian}(U)(V)$. If the *cartesianer* is implemented using a *binary map histogram representation* on *ordered list state representations*, then the *time* complexity is $v \ln v$

$$\exists m \in \mathbf{N}_{>0}\ (I_{\mathrm{C}}^{\mathrm{t}} \in \mathrm{O}(\{(V, v \ln v) : V \in \mathrm{domain}(I_{\mathrm{C}}),\ v = |V^{\mathrm{C}}|\}, m))$$

If V is *regular* having *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$ then $v = d^n$ and the *time* complexity is $d^n \ln d^n = nd^n \ln d$.

Given a *histogram* $A \in \mathcal{A}$ and set of *variables* $V \subseteq \mathcal{V}$, consider the computation *time* to *reduce* the *histogram*, $A\%V$. Let $I_{\%} = \mathrm{reducer} \in \mathrm{computers}$. Then $\mathrm{domain}(I_{\%}) = \mathrm{P}(\mathcal{V}) \times \mathcal{A}$, $\mathrm{range}(I_{\%}) = \mathcal{A}$ and $\mathrm{apply}(I_{\%})((V,A)) = A\%V$. If the *reducer* is implemented using a *binary map histogram representation* on *ordered list state representations*, then the *reducer* is constrained to *system* U. That is, $A \in \mathcal{A}_U$ and $V \subseteq \mathrm{vars}(U)$. The *time* complexity is defined

$$\begin{gathered}\exists m \in \mathbf{N}_{>0}\ (I_{\%}^{\mathrm{t}} \in \mathrm{O}(\{((V,A), \mathrm{maximum}(y \ln y, ny)) : \\ (V,A) \in \mathrm{domain}(I_{\%}),\ y = |A|,\ n = |V|\}, m))\end{gathered}$$

Note that $|A| \geq |A\%V|$ and so $|A|$ has the greater complexity. In the case where A is *reduced* to a *scalar* the *reducer* must compute at least $|A| - 1$

additions, $I^{\rm t}_{\%}((\emptyset,A)) > (|A|-1)I^{\rm t}_{+}((0,0))$. If A is *regular* in *system* U having *dimension* $n = |\mathrm{vars}(A)|$ and *valency* $\{d\} = \{|U_w| : w \in \mathrm{vars}(A)\}$ and A is *complete*, $A^{\rm U} = A^{\rm C}$, then $y = d^n$ and the *time* complexity is at most $d^n \ln d^n = nd^n \ln d$.

Given a pair of *histograms* $A, B \in \mathcal{A}$, consider the computation *time* of the *multiplication*, $A * B$. Let $I_* = \mathrm{multiplicationer} \in \mathrm{computers}$. Then $\mathrm{domain}(I_*) = \mathcal{A} \times \mathcal{A}$, $\mathrm{range}(I_*) = \mathcal{A}$ and $\mathrm{apply}(I_*)((A,B)) = A * B$. If the *multiplicationer* is implemented using a *binary map histogram representation* on *ordered list state representations* in *system* U, then

$$\exists m \in \mathbf{N}_{>0}\ (I^{\rm t}_* \in \mathrm{O}(\{((A,B),\mathrm{maximum}(xy\ln xy, \mathrm{maximum}(n_A x, n_B y))) : A,B \in \mathcal{A},\ x = |A|,\ y = |B|,\ n_A = |\mathrm{vars}(A)|,\ n_B = |\mathrm{vars}(B)|\}, m))$$

In the case where B is a *scalar* the *multiplicationer* must compute at least $|A|$ multiplications, $I^{\rm t}_*((A,\mathrm{scalar}(\mathrm{size}(B)))) > |A|I^{\rm t}_{\times}((1,1))$ where $I_\times = \mathrm{multiplier}$. In the case where the *variables* of A and B do not intersect, $\mathrm{vars}(A) \cap \mathrm{vars}(B) = \emptyset$, the *multiplicationer* must compute at least $|A||B|$ multiplications, $I^{\rm t}_*((A,B)) > |A||B|I^{\rm t}_{\times}((1,1))$. If A and B are *complete* and have *volumes* $v = |A^{\rm C}|$ and $w = |B^{\rm C}|$, then the *time* complexity is at most $vw \ln vw$.

A special case of a *multiplicationer* is the *one functional multiplicationer* where the second *histogram* B is the *histogram* of some *one functional transform* T. That is, $\exists T \in \mathcal{T}_{U,\mathrm{f},1}\ (\mathrm{his}(T) = B)$. The domain is also constrained such that the *underlying variables* of the *transform* T are a subset of the *variables* of the first *histogram* A, $\mathrm{und}(T) \subseteq \mathrm{vars}(A)$, and the *derived variables* of T are disjoint with the *variables* of A, $\mathrm{der}(T) \cap \mathrm{vars}(A) = \emptyset$. Then the *multiplication* can be thought of as adding *variables* to A without changing the cardinality or the *counts*. That is, $\mathrm{vars}(A*B) = V \cup W = V \cup \mathrm{der}(T)$ and $\{(S\%V, c) : (S,c) \in A*B\} = A$ where $V = \mathrm{vars}(A)$ and $W = \mathrm{vars}(B)$. There exists a mapping $Q = \{(S, S \cup R) : S \in A^{\rm S},\ R \in B^{\rm S},\ |S \cap R| = |V \cap W|\} \in V^{\rm CS} \to (V \cup W)^{\rm CS}$ which is such that $|Q| = |A|$. The *one functional multiplicationer* need only remap the *states* without modifying the *counts* however they are represented. No multiplications nor additions are computed. Let $I_{*1} = \mathrm{multiplicationOneFunctionaler} \in \mathrm{computers}$. Then $\mathrm{domain}(I_{*1}) = \{(A,\mathrm{his}(T)) : U \in \mathcal{U},\ A \in \mathcal{A}_U,\ T \in \mathcal{T}_{U,\mathrm{f},1},\ \mathrm{und}(T) \subseteq \mathrm{vars}(A),\ \mathrm{der}(T) \cap \mathrm{vars}(A) = \emptyset\}$, $\mathrm{range}(I_{*1}) = \mathcal{A}$ and $\mathrm{apply}(I_{*1})((A,B)) = A * B$. If the *histograms* A and B are represented with a *binary map histogram representation* on *ordered list state representations* then $I^{\rm t}_{*1}((A,B)) > |A| \ln |B|$ and $I^{\rm t}_{*1}((A,B)) > |A| \ln |A|$. The *volume* of A is greater than or equal to that of the *underlying* of T, $|V^{\rm C}| \geq |(V \cap W)^{\rm C}|$, so the *time complexity* is at most $v \ln v$ where $v = |A^{\rm C}|$.

3.19.2 Computation of the application of a transform

The application of the *transform* $T \in \mathcal{T}_U$ to a *histogram* $A \in \mathcal{A}_U$ in *system* U can be implemented by applying a *multiplicationer* followed by a *reducer*, $A * T = I^*_{\%}((W, I^*_*((A, X))))$ where $(X, W) = T$. In the case when the *histograms* are represented with a *binary map histogram representation* on *ordered list state representations*, this method has *time* complexity, $vw \ln vw$, where $v = \text{volume}(U)(\text{vars}(A))$ and $w = \text{volume}(U)(\text{vars}(T))$. The computation does not assume that the *transform* is *functional* and hence is less efficient than an implementation using a *one functional multiplicationer*.

If the *transform* is constrained to be *one functional*, $T \in \mathcal{T}_{U,\text{f},1}$, and such that the *underlying variables* of T are a subset of the *histogram* A, $\text{und}(T) \subseteq \text{vars}(A)$, and the *derived variables* of T are disjoint with the *variables* of A, $\text{der}(T) \cap \text{vars}(A) = \emptyset$, then the application can be implemented by applying a *one functional multiplicationer* followed by a *reducer*, $A * T = I^*_{\%}((W, I^*_{*1}((A, X))))$. Let $I_{*\text{T}} = \text{transformer} \in \text{computers}$, $\text{domain}(I_{*\text{T}}) = \{(T, A) : U \in \mathcal{U},\ A \in \mathcal{A}_U,\ T \in \mathcal{T}_{U,\text{f},1},\ \text{und}(T) \subseteq \text{vars}(A),\ \text{der}(T) \cap \text{vars}(A) = \emptyset\}$, $\text{range}(I_{*\text{T}}) = \text{range}(I_{\%})$ and

$$\text{apply}(I_{*\text{T}})((T, A)) = \text{transform}(T, A) = I^*_{\%}((W, I^*_{*1}((A, X))))$$

where $(X, W) = T$. Then $I^{\text{t}}_{*T}((T, A)) > |A| \ln |X|$ and $I^{\text{t}}_{*T}((T, A)) > |A| \ln |A|$. The overall *time* complexity is $v \ln v$ where $v = \text{volume}(U)(V)$ and $V = \text{vars}(A)$. If A is a *regular histogram* of *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$ then $v = d^n$ and the *time* complexity is $d^n \ln d^n$.

3.19.3 Computation of functional definition sets

A *functional definition set* F is a set of *unit functional transforms*, $F \in \mathcal{F} \subset \text{P}(\mathcal{T}_{\text{U}} \cap \mathcal{T}_{\text{f}})$. The *equivalent transform*, F^{T}, is defined $\text{transform}(F) := (\prod \text{his}(F)\ \%\ (V\ \cup\ W), W) \in \mathcal{T}_{\text{U}} \cap \mathcal{T}_{\text{f}}$, where $W = \text{der}(F)$ and $V = \text{und}(F)$. Let $L \in \mathcal{L}(\mathcal{A})$ be a list of the *histograms* of a non-empty *fud* F such that the inverse is an enumeration of the *histograms*, $\text{flip}(L) \in \text{enums}(\text{his}(F))$. Then the product can be computed by application of a *multiplicationer* recursing on L. Let $\text{mul}(L) := I^*_*((\text{mul}(\text{tail}(L)), L_1))$ and $\text{mul}(\{(1, A)\}) := A$. Let $I_{\text{F,T}} = \text{fudsTransformer} \in \text{computers}$. Then $\text{range}(I_{\text{F,T}}) = \mathcal{T}_{\text{U}} \cap \mathcal{T}_{\text{f}}$, $\text{domain}(I_{\text{F,T}}) = \mathcal{F} \setminus \{\emptyset\}$ and $\text{apply}(I_{\text{F,T}})(F) = (I^*_{\%}((V\ \cup\ W, \text{mul}(L))), W) = F^{\text{T}}$ where $\text{flip}(L) \in \text{enums}(\text{his}(F))$. The cardinality of the intermediate *histogram* $|\prod \text{set}(L_{\{1 \ldots i\}})|$, and hence the computation *times* of subsequent *multiplications*, $I^{\text{t}}_*((\prod \text{set}(L_{\{1 \ldots i\}}), L_{i+1}))$, depends on the order in which the *multiplications* take place in L. If the *fud transformer* is implemented where

L is chosen such that the intermediate *equivalent transform* is always *functional*, $\forall i \in \{1 \ldots |L|\}\ ((X_{\{1 \ldots i\}}, \mathrm{vars}(X_{\{1 \ldots i\}}) \setminus V) \in \mathcal{T}_{\mathrm{f}})$, where $X_{\{1 \ldots i\}} = \prod \mathrm{set}(L_{\{1 \ldots i\}})$, then the *time* complexity of the *fud transformer* is at most the *time* complexity of the *reducer*, $v \ln u$ where $v = \mathrm{volume}(U)(V)$ and $u = \mathrm{volume}(U)(\mathrm{vars}(F))$. If L is chosen arbitrarily then the *time* complexity of the *fud transformer* is at most $u \ln u$.

The *partition set transformer* $I_{\mathrm{N,T}} = \mathrm{partitionSetsTransformer} \in \mathrm{computers}$ is a variation of the *fud transformer* which has a domain of *partition sets*, $\mathrm{domain}(I_{\mathrm{N,T}}) = \mathrm{P}(\mathcal{R}) \setminus \{\emptyset\}$. The application promotes each *partition* to a *transform* and the computes the *equivalent fud*, $\mathrm{apply}(I_{\mathrm{N,T}})(Q) = \{P^{\mathrm{T}} :\in Q\}^{\mathrm{T}}$. The *time* complexity of the *partition set transformer* is the same as the *fud transformer*, $v \ln y$.

Direct application of *functional definition sets* to *histograms*, $\mathrm{apply} \in \mathcal{F} \times \mathcal{A} \to \mathcal{A}$, reduces computation *time* and *space* by navigating through the *fud reducing* any *non-derived variables* as soon as possible. If the *fud* is constrained to be *one functional*, $F \in \mathcal{F}_{U,1} \subset \mathrm{P}(\mathcal{T}_{U,\mathrm{f},1})$, then the implementation may use the *one functional multiplicationer* rather than the *multiplicationer* and the *reduction* can optionally be left to the end without increasing the cardinality of the cumulative *histogram* product. The *one functional multiplicationer* must be applied in sequence such that the intermediate *equivalent transform* is always *functional*. Let $M \in \mathcal{L}(\mathcal{T}_{U,\mathrm{f},1})$ be a list of the *transforms* of F such that $\mathrm{flip}(M) \in \mathrm{enums}(F)$ and such that $\mathrm{und}(M_1) \subseteq V$ and $\mathrm{der}(M_1) \cap V = \emptyset$ where $V = \mathrm{vars}(A)$ and $\forall i \in \{2 \ldots |M|\}\ \Diamond Q = \mathrm{vars}(\mathrm{set}(M_{\{1 \ldots i-1\}}))\ (\mathrm{und}(M_i) \subseteq Q \wedge \mathrm{der}(M_i) \cap Q = \emptyset)$. Let $\mathrm{mul1}(A, M) := I^{*}_{*1}((\mathrm{mul1}(A, \mathrm{tail}(M)), \mathrm{his}(M_1)))$ and $\mathrm{mul1}(A, \emptyset) := A$.

First define the application without *reduction*. Let $I_{*\mathrm{X}} = \mathrm{applier} \in \mathrm{computers}$ in *system* U. Then $\mathrm{domain}(I_{*\mathrm{X}}) = \{(F, A) : U \in \mathcal{U},\ A \in \mathcal{A}_U,\ F \in \mathcal{F}_{U,1},\ \mathrm{und}(F) \subseteq \mathrm{vars}(A),\ \mathrm{vars}(F) \setminus \mathrm{und}(F) \cap \mathrm{vars}(A) = \emptyset\}$, $\mathrm{range}(I_{*\mathrm{X}}) = \mathcal{A}$ and $I^{*}_{*\mathrm{X}}((F, A)) = \mathrm{mul1}(A, \mathrm{reverse}(M))$. Then $I^{\mathrm{t}}_{*\mathrm{X}}((F, A)) > r|A|$ where $r = |\mathrm{vars}(F)|$, and $I^{\mathrm{t}}_{*\mathrm{X}}((F, A)) > f|A| \ln |A|$ where $f = |F|$.

Now define the application with *reduction*. Let $I_{*\mathrm{F}} = \mathrm{fuder} \in \mathrm{computers}$ in *system* U. Then $\mathrm{domain}(I_{*\mathrm{F}}) = \mathrm{domain}(I_{*\mathrm{X}})$, $\mathrm{range}(I_{*\mathrm{F}}) = \mathrm{range}(I_{\%})$ and $I^{*}_{*\mathrm{F}}((F, A)) = \mathrm{apply}(F, A) = I^{*}_{\%}((W, I^{*}_{*\mathrm{X}}((F, A))))$ where $W = \mathrm{der}(F)$.

3.19.4 Computation of independent

Given a *histogram* A consider the computation *time* to calculate its *independent* A^{X}. Let $I_{\mathrm{X}} = \mathrm{independenter} \in \mathrm{computers}$. Let $\mathrm{domain}(I_{\mathrm{X}}) = \mathrm{range}(I_{\mathrm{X}}) = \mathcal{A}$, and $\mathrm{apply}(I_{\mathrm{X}})(A) = A^{\mathrm{X}}$. Consider *non-zero histogram* $A \in \mathcal{A}$ having *size* $z = \mathrm{size}(A)$, *variables* $V = \mathrm{vars}(A)$, and *dimension* $n = |V|$. In order to calculate A^{X} the *independenter* must calculate $(n-1)$ additions for each *effective state*, A^{F}, to construct the *reductions* $R = \{(v, A\%\{v\}) : v \in V\} = \{(v, \sum(\{(S\%\{v\}, c)\} : (S,c) \in A,\ c > 0)) : v \in V\}$. Then the *independenter* must calculate n multiplications for each *state* of the *effective cartesian sub-volume* A^{XF}, $A^{\mathrm{X}} = \{(S, \prod(R_v(S\%\{v\}) : v \in V)/z^{n-1}) : S \in \mathrm{states}(\prod\{R_v^{\mathrm{F}} : v \in V\})\}$. Thus $I_{\mathrm{X}}^{\mathrm{t}}(A) > |A^{\mathrm{F}}|(n-1)I_{+}^{\mathrm{t}}(0,0) + |A^{\mathrm{XF}}|nI_{\times}^{\mathrm{t}}(1,1)$ where $I_{+} = \mathrm{adder}$ and $I_{\times} = \mathrm{multiplier}$. Implementing the *histograms* with an *array histogram representation* on *ordered list state representations*

$$\exists m \in \mathbf{N}_{>0}\ (I_{\mathrm{X}}^{\mathrm{t}} \in \mathrm{O}(\{(A, ny) : A \in \mathcal{A},\ y = |A^{\mathrm{XF}}|,\ n = |\mathrm{vars}(A)|\}, m))$$

In other words, the *time* to calculate the *independent histogram* A^{X} is of complexity of ny where y is the *effective independent cartesian sub-volume*. If A is a *regular histogram* in a *system* U of *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_v| : v \in V\}$ for which the *independent* is *completely effective*, $A^{\mathrm{XF}} = A^{\mathrm{C}}$, then y is the *volume*, $|V^{\mathrm{C}}|$ and the *time* complexity is nd^n.

If the *independent* is *completely effective*, $A^{\mathrm{XF}} = A^{\mathrm{C}}$, then the *space* complexity of an *array histogram representation*, v, is less than the *space* complexity of a *binary map histogram representation*, $v \ln v$.

4 Alignment

4.1 Definition

The *alignment* of a *histogram* A is defined, $\mathrm{alignment} \in \mathcal{A} \to \mathbf{R}$

$$\mathrm{alignment}(A) := \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A_S^{\mathrm{X}}$$

where the unit-translated gamma function is defined $(\Gamma_!) \in \mathbf{R} \to \mathbf{R}$ as $\Gamma_! x = \Gamma(x+1)$ which is such that $\forall i \in \mathbf{N}\ (\Gamma_! i = i!)$. The *alignment* of the *empty histogram* is defined as zero, $\mathrm{alignment}(\emptyset) := 0$. The first term of the expression, $\sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S$, is called the *non-independent* term. The second term, $\sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A_S^{\mathrm{X}}$, is the *independent* term.

In the case where $A, A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$ then

$$\mathrm{alignment}(A) = \sum_{S \in A^{\mathrm{S}}} \ln A_S! - \sum_{S \in A^{\mathrm{XS}}} \ln A^{\mathrm{X}}_S! \ \in \ \ln \mathbf{Q}_{>0}$$

In the following, the *alignment* function will sometimes be abbreviated, $\mathrm{algn} = \mathrm{alignment}$.

The *alignments* of *equivalent histograms* are equal, $\forall A, B \in \mathcal{A}\ (A \equiv B \implies \mathrm{alignment}(A) = \mathrm{alignment}(B))$.

For each *histogram* $A \in \mathcal{A}_U$ in *system* U there exists a set of *cardinal substrate permutations* which are *non-literal frame* mappings $X \in \mathcal{V} \leftrightarrow (\mathbf{N}_{>0} \times (\mathcal{W} \leftrightarrow \mathbf{N}_{>0}))$ such that the *reframed histogram* is a *cardinal substrate histogram*, $\mathrm{reframe}(X, A) \in \mathcal{A}_{\mathrm{c}}$. The *alignment* of each of the *cardinal substrate histograms* under *permutation* equals the *alignment* of the *histogram*, $\mathrm{alignment}(\mathrm{reframe}(X, A)) = \mathrm{alignment}(A)$. There are $|V|! \prod_{w \in V} |U_w|!$ *cardinal substrate permutations* of *histogram* A. These *cardinal substrate histograms* are isomorphic with respect to *alignment*.

The *alignment* of an *independent histogram*, $A = A^{\mathrm{X}}$, is zero. The *alignment* of each of the following is zero because in each of these cases the *histogram* is *equivalent* to its *independent histogram*, $A \equiv A^{\mathrm{X}}$: (i) *zero histograms*, $\mathrm{size}(A) = 0$, (ii) *scalar histograms*, $\mathrm{vars}(A) = \emptyset$, (iii) *mono-variate histograms*, $|\mathrm{vars}(A)| = 1$, (iv) *uniform cartesian histograms*, $A = \mathrm{scalar}(q) * A^{\mathrm{C}}$ where $q \in \mathbf{Q}_{\geq 0}$, (v) *uniform full planar histograms*, (vi) *uniform linear histograms*, (vii) *singleton histograms*, $|A| = 1$, and (viii) *uniform cartesian sub-volumes*, $A = \mathrm{scalar}(q) * A^{\mathrm{XF}}$.

Conjecture that the *alignment* of an *integral histogram* is zero if and only if it is *independent*, $\forall A \in \mathcal{A}_{\mathrm{i}}\ (A = A^{\mathrm{X}} \iff \mathrm{algn}(A) = 0)$. The set of *integral iso-independents* of *histogram* A of *size* z and *variables* V in *system* U is $Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \subset \mathcal{A}_{U,\mathrm{i},V,z}$. It contains at most one *independent histogram*, $|\{B : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}),\ B = B^{\mathrm{X}}\}| \leq 1$, which, if it exists, has zero *alignment*, $\mathrm{algn}(A^{\mathrm{X}}) = 0$. The other *histograms* have *alignment* not equal zero, $\forall B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})\ (B \neq A^{\mathrm{X}} \implies \mathrm{algn}(B) \neq 0)$.

4.2 Derivation

The *generalised multinomial probability distribution* $\hat{Q}_{\mathrm{m},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is defined

$$\hat{Q}_{\mathrm{m},U}(E,z) := \{(A, \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

where $(E,z) \in \mathcal{A}_U \times \mathbf{N}$, $z_E = \mathrm{size}(E) > 0$, $V = \mathrm{vars}(E)$ and E is *complete*, $E^{\mathrm{U}} = E^{\mathrm{C}}$. The domain of the *generalised multinomial probability distribution* is the finite *integral congruent support*

$$\mathrm{dom}(\hat{Q}_{\mathrm{m},U}(E,z)) = \mathcal{A}_{U,\mathrm{i},V,z} = \{A : A \in \mathcal{A}_U \cap \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{U}} = V^{\mathrm{C}},\ \mathrm{size}(A) = z\}$$

The *generalised multinomial probability distribution* can be further generalised to a probability density function by means of the gamma function. First generalise the *support*. Let the set of *complete congruent histograms* in *system* U, of *variables* V and *size* z be

$$\mathcal{A}_{U,V,z} = \{A : A \in \mathcal{A}_U,\ A^{\mathrm{U}} = V^{\mathrm{C}},\ \mathrm{size}(A) = z\}$$

The *complete congruent histograms* is an infinite superset of the *integral congruent support*, $\mathcal{A}_{U,\mathrm{i},V,z} \subset \mathcal{A}_{U,V,z}$.

Let E be a *complete histogram*, $E^{\mathrm{U}} = E^{\mathrm{C}}$, of non-zero *size*, $\mathrm{size}(E) > 0$. Let the *multinomial probability density function*, $\mathrm{mpdf}(U) \in \mathcal{A}_U \times \mathbf{Q}_{\geq 0} \to (\mathcal{A}_U \to \mathbf{R}_{\geq 0})$ be such that $\mathrm{mpdf}(U)(E,z)$ is a real valued function of the *complete congruent histograms* of *size* z and *variables* equal to those of E, defined $\mathrm{mpdf}(U)(E,z) \in \mathcal{A}_{U,V,z} \to \mathbf{R}_{\geq 0}$ as

$$\mathrm{mpdf}(U)(E,z) := \{(A, \frac{\Gamma_! z}{\prod_{S \in A^{\mathrm{S}}} \Gamma_! A_S} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S}) : A \in \mathcal{A}_{U,V,z}\}$$

The *generalised multinomial probability distribution* is a subset of the *multinomial probability density function*

$$\hat{Q}_{\mathrm{m},U}(E,z) \subset \mathrm{mpdf}(U)(E,z)$$

The infinite domain of the *multinomial probability density function*, which is the set of *congruent histograms*, $\mathrm{dom}(\mathrm{mpdf}(U)(E,z)) = \mathcal{A}_{U,V,z}$, corresponds to the finite domain of the *generalised multinomial probability distribution*, which is the *integral congruent support*, $\mathrm{dom}(\hat{Q}_{\mathrm{m},U}(E,z)) = \mathcal{A}_{U,\mathrm{i},V,z}$.

The *scaled distribution histogram* $M = \text{scalar}(z/z_E) * E$ is in the domain of the *multinomial probability density function*, $M \in \text{dom}(\text{mpdf}(U)(E,z)) = \mathcal{A}_{U,V,z}$. The *mean* of the *generalised multinomial probability distribution* equals the *scaled distribution histogram*, $\text{mean}(\hat{Q}_{\text{m},U}(E,z)) = M = \text{scalar}(z/z_E) * E$ where the *draw size* z is integral. Thus the *mean* of the *generalised multinomial probability distribution* is in the domain of the *multinomial probability density function* even if the *mean* is not itself *integral*, $M \notin \mathcal{A}_\text{i}$.

The *independent histogram* of each of the *histograms* in the domain of the *multinomial probability density function* are also in the domain because the *independent histogram* is *congruent*, $\forall A \in \mathcal{A}_{U,V,z}\ (A^\text{X} \in \mathcal{A}_{U,V,z})$.

The *scaled uniform histogram* $\text{resize}(z, V^\text{C})$ is in the domain of the *multinomial probability density function*, $\text{scalar}(z/v) * V^\text{C} \in \mathcal{A}_{U,V,z}$, where $v = |V^\text{C}|$.

Similarly to the *generalised multinomial probability distribution*, the *multinomial probability density function* can be approximated by means of the Stirling approximation

$$\text{mpdf}(U)(E,z)(A) = \Gamma_! z \prod_{S \in A^\text{S}} \frac{\hat{E}_S^{A_S}}{\Gamma_! A_S} \approx \prod_{S \in A^\text{FS}} \left(\frac{\hat{E}_S}{\hat{A}_S}\right)^{A_S}$$

where $\hat{E} = \text{resize}(1, E)$ and $\hat{A} = \text{resize}(1, A)$. The approximation is best for high *entropy sample histograms* for which the *multinomial coefficient* is largest.

Compare this approximation to the same term for a *scaled draw size* kz and *scaled sample histogram* $\text{scalar}(k) * A$ where $k \in \mathbf{Q}_{>0}$

$$\begin{aligned}
\text{mpdf}(U)(E,kz)(\text{scalar}(k) * A) &= \Gamma_! kz \prod_{S \in A^\text{S}} \frac{\hat{E}_S^{kA_S}}{\Gamma_! kA_S} \\
&\approx \left(\prod_{S \in A^\text{FS}} \left(\frac{\hat{E}_S}{\hat{A}_S}\right)^{A_S}\right)^k \\
&= (\text{mpdf}(U)(E,z)(A))^k
\end{aligned}$$

The parameterised *multinomial probability density function*, $\text{mpdf}(U)(E,z)$, does not have a continuous domain. Rather, its domain is countably infinite, $\mathcal{A}_{U,V,z} \leftrightarrow \mathbf{N}$, consisting as it does of *histograms* which are rational valued functions of finite domain. The integration of the *multinomial probability density function*, which is the cumulative density function, when summed over the whole domain equals one. The integration is defined here in terms of the *scaled multinomial probability density function*. The *scaled complete integral congruent histograms* equals the *complete congruent histograms* in the limit

$$\lim_{k \to \infty} \{A/Z_k : A \in \mathcal{A}_{U,\mathrm{i},V,kz}\} = \mathcal{A}_{U,V,z}$$

where $k \in \mathbf{N}_{>0}$ and $Z_k = \mathrm{scalar}(k)$. Therefore define the integration

$$\int_{A \in \mathcal{A}_{U,V,z}} \mathrm{mpdf}(U)(E,z)(A)\ dA = \lim_{k \to \infty} \sum (\mathrm{mpdf}(U)(E,kz)(Z_k * A) : A \in \mathcal{A}_{U,V,z},\ Z_k * A \in \mathcal{A}_\mathrm{i})$$

In the case where the *distribution histogram* is *integral*, $E \in \mathcal{A}_\mathrm{i}$ and the *draw size* z is integral and non-zero, $z \in \mathbf{N}_{>0}$

$$\begin{aligned}\sum (\mathrm{mpdf}(U)(E,kz)(Z_k * A) : A \in \mathcal{A}_{U,V,z},\ Z_k * A \in \mathcal{A}_\mathrm{i}) &= \\ \sum (\mathrm{mpdf}(U)(E,kz)(A) : A \in \mathcal{A}_{U,\mathrm{i},V,kz}) &= \\ \sum (\hat{Q}_{\mathrm{m},U}(E,kz)(A) : A \in \mathcal{A}_{U,\mathrm{i},V,kz}) &= 1\end{aligned}$$

Thus the integration can be approximated by a finite summation for some large k, with the approximation improving as k is increased. A further approximation is to take the k-th power of the *unscaled density*

$$\int_{A \in \mathcal{A}_{U,V,z}} \mathrm{mpdf}(U)(E,z)(A)\ dA \approx \lim_{k \to \infty} \sum ((\mathrm{mpdf}(U)(E,z)(A))^k : A \in \mathcal{A}_{U,V,z},\ Z_k * A \in \mathcal{A}_\mathrm{i})$$

When $k = z^{n-1}$, where $n = |V|$, all of the *independent histograms* of the *integral sample histograms* are approximated

$$\{A^\mathrm{X} : A \in \mathcal{A}_{U,\mathrm{i},V,z}\} \subset \{A/\mathrm{scalar}(z^{n-1}) : A \in \mathcal{A}_{U,\mathrm{i},V,z^n}\}$$

because $\{\mathrm{scalar}(z^{n-1}) * A^\mathrm{X} : A \in \mathcal{A}_{U,\mathrm{i},V,z}\} \subset \mathcal{A}_\mathrm{i}$.

The definition of the unit-translated gamma function, $(\Gamma_!) \in \mathbf{R} \to \mathbf{R}$, is such that the minimum value of positive real arguments is less than one, approximately $(0.4616, 0.8856)$. In fact, $\forall x \in \mathbf{R}\ (0 < x < 1 \implies 0! > \Gamma_! x < 1!)$.

However, all *integral histograms* have *multinomial probability density function* less than or equal to one

$$\forall A \in \mathcal{A}_{U,V,z}\ (A \in \mathcal{A}_{\mathrm{i}} \implies \mathrm{mpdf}(U)(E,z)(A) = \hat{Q}_{\mathrm{m},U}(E,z)(A) \leq 1)$$

where E is *integral* and *complete*.

The condition for an *independent histogram* A^{X} to contain at least one fractional *count*, $\mathrm{minr}(\mathrm{trim}(A^{\mathrm{X}})) < 1$, is

$$\prod_{v \in V} \mathrm{minr}(\mathrm{trim}(A)\%\{v\}) < z^{n-1}$$

where $V = \mathrm{vars}(A)$, $n = |V|$, $n > 0$ and $z = \mathrm{size}(A) > 0$. For some *histograms* the *independent* term of the *alignment* expression is negative, $\sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S < 0$. For example, let $A = \mathrm{resize}(0.4616v, V^{\mathrm{C}})$ where $v = |V^{\mathrm{C}}| \geq 1$, then

$$\sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S = v \ln \Gamma_! 0.4616 \approx -0.1215v$$

If the *sample histogram* is *integral*, $A \in \mathcal{A}_{\mathrm{i}}$, and the *independent* term is negative, $\sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S < 0$, then the *alignment* must be positive, $\mathrm{algn}(A) \geq 0$.

Consider the *complete integral congruent support sample histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$. In the case where (a) the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and therefore also in the *integral congruent support*, $A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},V,z}$, and (b) the *distribution histogram* E is as *effective* as the *independent*, $E^{\mathrm{F}} \geq A^{\mathrm{XF}}$, then the *generalised multinomial probability* of the *sample histogram*, $\hat{Q}_{\mathrm{m},U}(E,z)(A)$, may be decomposed into (i) the *independent multinomial probability* and (ii) *relative dependent multinomial probability*

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = \hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}}) \times \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})}$$

The *relative dependent multinomial probability* is

$$\begin{aligned}
\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})} &= \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S} / \frac{z!}{\prod_{S \in A^{\mathrm{XS}}} A_S!} \prod_{S \in A^{\mathrm{XS}}} \left(\frac{E_S}{z_E}\right)^{A^{\mathrm{X}}_S} \\
&= \prod_{S \in A^{\mathrm{XS}}} \frac{A^{\mathrm{X}}_S!}{A_S!} \frac{E_S^{A_S}}{E_S^{A^{\mathrm{X}}_S}}
\end{aligned}$$

In some cases the *relative dependent multinomial probability* may be greater than 1, $\hat{Q}_{\mathrm{m},U}(E,z)(A)/\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}}) > 1$. Therefore *relative probability* is not strictly speaking a probability per se.

In the case where the *sample histogram* is *independent*, $A = A^{\mathrm{X}}$, the *relative dependent multinomial probability* is 1

$$\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})}{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})} = 1$$

The *relative dependent multinomial probability* may be generalised to cases where the *independent* is not *integral*, $A^{\mathrm{X}} \notin \mathcal{A}_{\mathrm{i}}$, and therefore not in the finite *integral congruent support*, $A^{\mathrm{X}} \notin \mathcal{A}_{U,\mathrm{i},V,z}$, by considering the *multinomial probability density function*, $\mathrm{mpdf}(U)(E,z)$. Here the *sample histogram* need not be *integral* either, but is a *complete congruent histogram*, $A \in \mathcal{A}_{U,V,z}$. Decompose the *multinomial probability density* into (i) the *independent multinomial probability density* and (ii) *relative dependent multinomial probability density*

$$\mathrm{mpdf}(U)(E,z)(A) = \mathrm{mpdf}(U)(E,z)(A^{\mathrm{X}}) \times \frac{\mathrm{mpdf}(U)(E,z)(A)}{\mathrm{mpdf}(U)(E,z)(A^{\mathrm{X}})}$$

Again, the *distribution histogram* must be as *effective* as the *independent*, $E^{\mathrm{F}} \geq A^{\mathrm{XF}}$, so that the *relative independent multinomial probability density* is non-zero, $\mathrm{mpdf}(U)(E,z)(A^{\mathrm{X}}) > 0$. The *relative dependent multinomial probability density* is

$$\begin{aligned}
&\frac{\mathrm{mpdf}(U)(E,z)(A)}{\mathrm{mpdf}(U)(E,z)(A^{\mathrm{X}})} \\
&= \frac{\Gamma_! z}{\prod_{S \in A^{\mathrm{S}}} \Gamma_! A_S} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S} \Big/ \frac{\Gamma_! z}{\prod_{S \in A^{\mathrm{XS}}} \Gamma_! A_S} \prod_{S \in A^{\mathrm{XS}}} \left(\frac{E_S}{z_E}\right)^{A^{\mathrm{X}}_S} \\
&= \prod_{S \in A^{\mathrm{XS}}} \frac{\Gamma_! A^{\mathrm{X}}_S}{\Gamma_! A_S} \frac{E_S^{A_S}}{E_S^{A^{\mathrm{X}}_S}}
\end{aligned}$$

The negative logarithm *relative dependent multinomial probability density* is

$$\begin{aligned}
&-\ln \frac{\mathrm{mpdf}(U)(E,z)(A)}{\mathrm{mpdf}(U)(E,z)(A^{\mathrm{X}})} \\
&= \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S - \sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln E_S \\
&= \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S - \sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln E_S
\end{aligned}$$

because

$$\sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A_S = \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S$$

In the case where the *sample histogram* is *independent*, $A = A^{\mathrm{X}}$, the negative logarithm *relative dependent multinomial probability density* is 0

$$-\ln \frac{\mathrm{mpdf}(U)(E,z)(A^{\mathrm{X}})}{\mathrm{mpdf}(U)(E,z)(A^{\mathrm{X}})} = 0$$

In the special case where the *distribution histogram*, E, from which the *sample histogram* is *drawn*, is a *scaled uniform cartesian distribution histogram* $E = \mathrm{resize}(z_E, V^{\mathrm{C}})$, then the negative logarithm *relative dependent multinomial probability density* simplifies to

$$\begin{aligned}
&-\ln \frac{\mathrm{mpdf}(U)(\mathrm{resize}(z_E, V^{\mathrm{C}}), z)(A)}{\mathrm{mpdf}(U)(\mathrm{resize}(z_E, V^{\mathrm{C}}), z)(A^{\mathrm{X}})} \\
&= \quad -\ln \frac{\mathrm{mpdf}(U)(V^{\mathrm{C}}, z)(A)}{\mathrm{mpdf}(U)(V^{\mathrm{C}}, z)(A^{\mathrm{X}})} = \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S
\end{aligned}$$

because

$$\mathrm{mpdf}(U)(\mathrm{resize}(z_E, V^{\mathrm{C}}), z) = \mathrm{mpdf}(U)(V^{\mathrm{C}}, z)$$

and

$$\sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln V^{\mathrm{C}}_S = \ln \frac{1}{v} \sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) = 0$$

where *volume* $v = |V^{\mathrm{C}}|$. Thus, in this case the negative logarithm *relative dependent multinomial probability density* does not depend on the *distribution histogram*.

In fact, the negative logarithm *relative dependent multinomial probability density* simplifies to the same expression under the weaker constraint that the *distribution histogram* is *independent*, $E = E^{\mathrm{X}}$,

$$-\ln \frac{\mathrm{mpdf}(U)(E^{\mathrm{X}}, z)(A)}{\mathrm{mpdf}(U)(E^{\mathrm{X}}, z)(A^{\mathrm{X}})} = \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S$$

To prove this it is necessary to show that

$$\sum_{S \in A^{\mathrm{S}}} A_S \ln \hat{E}^{\mathrm{X}}_S = \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln \hat{E}^{\mathrm{X}}_S$$

where $\hat{E} = \text{resize}(1, E) \in \mathcal{P}$. Define $\ln \in \mathcal{A} \to (\mathcal{S} \to \ln \mathbf{Q}_{>0})$ as $\ln(A) := \{(S, \ln A_S) : S \in A^{\text{FS}}\}$, in

$$\begin{aligned}
\sum_{S \in A^{\text{S}}} A_S \ln \hat{E}^{\text{X}}_S &= \sum_{S \in A^{\text{S}}} A_S \ \ln(\hat{E}^{\text{X}})(S) \\
&= \sum_{S \in A^{\text{S}}} A_S \sum_{v \in V} \ln(\hat{E}^{\text{X}}\%\{v\})(S\%\{v\}) \\
&= \sum_{v \in V} \sum_{S \in A^{\text{S}}} A_S \ \ln(\hat{E}^{\text{X}}\%\{v\})(S\%\{v\}) \\
&= \sum_{v \in V} \sum_{R \in (A\%\{v\})^{\text{S}}} A\%\{v\}(R) \ \ln(\hat{E}^{\text{X}}\%\{v\})(R) \\
&= \sum_{v \in V} \sum_{R \in (A^{\text{X}}\%\{v\})^{\text{S}}} A^{\text{X}}\%\{v\}(R) \ \ln(\hat{E}^{\text{X}}\%\{v\})(R) \\
&= \sum_{S \in A^{\text{XS}}} A^{\text{X}}_S \ln \hat{E}^{\text{X}}_S
\end{aligned}$$

In this case, where the *distribution histogram* is *independent*, $E = E^{\text{X}}$, the negative logarithm *relative dependent multinomial probability density* does not depend on the *distribution histogram*.

The *alignment* may be derived from the negative logarithm *relative dependent multinomial probability density* in the case when the *distribution histogram* is *independent*, $E = E^{\text{X}}$,

$$\begin{aligned}
-\ln \frac{\text{mpdf}(U)(E^{\text{X}}, z)(A)}{\text{mpdf}(U)(E^{\text{X}}, z)(A^{\text{X}})} &= \sum_{S \in A^{\text{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\text{XS}}} \ln \Gamma_! A^{\text{X}}_S \\
&= \text{alignment}(A)
\end{aligned}$$

That is, the *alignment* is the negative logarithm *independently-distributed relative dependent multinomial probability density*.

The *alignment*, alignment(A), does not depend on the *distribution histogram*,

$$\forall E \in \mathcal{A}_{U,V,z_E} \ (E^{\text{XF}} \geq A^{\text{XF}} \implies -\ln \frac{\text{mpdf}(U)(E^{\text{X}}, z)(A)}{\text{mpdf}(U)(E^{\text{X}}, z)(A^{\text{X}})} = \text{alignment}(A))$$

The *alignment*, alignment(A), does not depend on the *completeness* or otherwise of the *sample histogram*, A,

$$\forall U \in \mathcal{U} \ \forall A, B \in \mathcal{A}_U \ (A \equiv B \implies \text{alignment}(A) = \text{alignment}(B))$$

If the *distribution histogram* equals the *independent sample*, $E = A^{\mathrm{X}}$, then the negative logarithm *independent-sample-distributed relative dependent multinomial probability density* equals the *alignment*

$$-\ln \frac{\mathrm{mpdf}(U)(A^{\mathrm{X}}, z)(A)}{\mathrm{mpdf}(U)(A^{\mathrm{X}}, z)(A^{\mathrm{X}})} \quad = \quad \mathrm{alignment}(A)$$

If the *distribution histogram* equals the *cartesian histogram*, $E = V^{\mathrm{C}}$, then the negative logarithm *cartesian-distributed relative dependent multinomial probability density* equals the *alignment*

$$-\ln \frac{\mathrm{mpdf}(U)(V^{\mathrm{C}}, z)(A)}{\mathrm{mpdf}(U)(V^{\mathrm{C}}, z)(A^{\mathrm{X}})} \quad = \quad \mathrm{alignment}(A)$$

In the case where both the *sample histogram* and the *independent sample* are *integral*, $A, A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},V,z}$, and the *distribution histogram* is *independent*, $E = E^{\mathrm{X}}$, and sufficiently *effective*, $E^{\mathrm{F}} \geq A^{\mathrm{XF}}$, then the negative logarithm *independently-distributed relative dependent multinomial probability* equals the *alignment* expressed in terms of factorials,

$$\begin{aligned} -\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A^{\mathrm{X}})} \quad &= \quad \sum_{S \in A^{\mathrm{S}}} \ln A_S! - \sum_{S \in A^{\mathrm{XS}}} \ln A^{\mathrm{X}}_S! \\ &= \quad \mathrm{alignment}(A) \end{aligned}$$

In this case the *alignment* is the difference of the logarithms of the *multinomial coefficients* of the *independent histogram* and the *sample histogram*

$$\mathrm{alignment}(A) = \ln \frac{z!}{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!} - \ln \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!}$$

In this case, also, it is conjectured below ('Minimum alignment') that the *alignment* is always positive,

$$(A, A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},V,z}) \wedge (E = E^{\mathrm{X}}) \implies \mathrm{alignment}(A) \geq 0$$

and so the *relative dependent multinomial probability* must be less than or equal to one,

$$0 < \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A^{\mathrm{X}})} \leq 1$$

Therefore in this case *relative probability* is a probability per se.

In the case where both the *sample histogram* and the *independent sample* are *integral*, $A, A^{\rm X} \in \mathcal{A}_{U,\mathrm{i},V,z}$, and the *distribution histogram* is equal to the *scaled independent*, $E = \mathrm{resize}(z_E, A^{\rm X})$, then the *independent* is the *modal mean*, $A^{\rm X} = \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(A^{\rm X}, z)) \in \mathrm{modes}(\hat{Q}_{\mathrm{m},U}(A^{\rm X}, z))$. In this case the *alignment* is the negative logarithm *modal-independently-distributed relative dependent multinomial probability*,

$$\mathrm{alignment}(A) = -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\rm X}, z)(A)}{\hat{Q}_{\mathrm{m},U}(A^{\rm X}, z)(A^{\rm X})}$$

In the cases where the *distribution histogram* is not *independent*, $E \neq E^{\rm X}$, then the *alignment* may not be equal to the negative logarithm *relative dependent multinomial probability density*. The difference is the *mis-alignment* $\sum((A_S - A^{\rm X}_S) \ln E_S : S \in A^{\rm XS})$. That is,

$$-\ln \frac{\mathrm{mpdf}(U)(E, z)(A)}{\mathrm{mpdf}(U)(E, z)(A^{\rm X})} \quad = \quad \mathrm{alignment}(A) - \sum_{S \in A^{\rm XS}} (A_S - A^{\rm X}_S) \ln E_S$$

The *mis-alignment* does depend on the *distribution histogram*, but not its *size*, z_E

$$\forall q \in \mathbf{Q}_{>0}\ (\sum_{S \in A^{\rm XS}} (A_S - A^{\rm X}_S) \ln(Z_q * E)(S) = \sum_{S \in A^{\rm XS}} (A_S - A^{\rm X}_S) \ln E_S)$$

where $Z_q = \mathrm{scalar}(q)$.

In the derivations of *alignment* above, the starting point has been to take a *distribution histogram* E and then to show that if the *distribution histogram* is *independent*, $E = E^{\rm X}$, then the *alignment* can be derived from it but does not depend on it. In fact, the *alignment* can be derived without reference to a *distribution histogram* at all. Consider the *classification coder* of *histories*

$$C_{\rm G} = \mathrm{coderClassification}(U, X, D_{\rm V}, D_{\rm S}, D_{\rm X}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

The *coder domain* is the finite set of *histories* $\mathcal{H}_{U,X} \subset \mathcal{H}_U$ in *system* U where the domains of the *histories* are restricted to a finite subset of the *event identifiers* $X \subset \mathcal{X}$. Consider the *history* $H \in \mathcal{H}_{U,X}$ of *size* $z = |H|$ and *variables* $V = \mathrm{vars}(H)$. The total *space* of a *classification coder* of a *history* H is the sum of the *variables space*, *ids space*, *histogram counts space* and *events classification space*

$$\begin{aligned}\mathrm{space}(C_{\rm G})(H) \quad = \quad & \mathrm{spaceVariables}(U)(|V|) + \\ & \mathrm{spaceIds}(|X|, z) + \\ & \mathrm{spaceCounts}(U)(A) + \\ & \mathrm{spaceClassification}(A)\end{aligned}$$

where $A = \text{histogram}(H)$ and *events classification space* is defined

$$\text{spaceClassification}(A) := \ln z! - \sum_{S \in A^{\mathrm{S}}} \ln A_S!$$

In the case where the *independent histogram*, A^{X}, is in the *histograms* of the *coder domain*, $A^{\mathrm{X}} \in \{\text{histogram}(G) : G \subseteq \mathcal{H}_{U,X}\}$, and is therefore *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, then the *classification coder space* of *history* H may be decomposed into (i) the *independent classification coder space* and (ii) *relative dependent classification coder space*

$$\begin{aligned}\text{space}(C_{\mathrm{G}})(H) \quad = \quad & \text{spaceVariables}(U)(|V|) + \\ & \text{spaceIds}(|X|, z) + \\ & \text{spaceCounts}(U)(A^{\mathrm{X}}) + \\ & \text{spaceClassification}(A^{\mathrm{X}}) + \\ & (\text{spaceClassification}(A) - \\ & \text{spaceClassification}(A^{\mathrm{X}}))\end{aligned}$$

In this case the negative *relative dependent classification coder space* equals the *alignment*

$$\begin{aligned}& -(\text{spaceClassification}(A) - \text{spaceClassification}(A^{\mathrm{X}})) \\ & = \sum_{S \in A^{\mathrm{S}}} \ln A_S! - \sum_{S \in A^{\mathrm{XS}}} \ln A^{\mathrm{X}}_S! = \text{alignment}(A)\end{aligned}$$

In the case where the *sample histogram* is *independent*, $A = A^{\mathrm{X}}$, the negative *relative dependent classification coder space* is 0

$$-(\text{spaceClassification}(A^{\mathrm{X}}) - \text{spaceClassification}(A^{\mathrm{X}})) = 0$$

Although there is no mention of a *distribution histogram* in this derivation, it may be viewed as defaulting to the *cartesian histogram*, V^{C}.

4.3 Alignment of types of histogram

4.3.1 Diagonal alignment

The *alignment* of a *uniform full diagonal regular histogram* of *size* z, *dimension* n and *valency* d is

$$d \ln \Gamma_! \frac{z}{d} - d^n \ln \Gamma_! \frac{z}{d^n}$$

A *uniform full diagonal regular histogram* is maximally *diagonal*. If the *size* is scaled by *valency*, $z = dx$ where x is the *size per value*, then the *non-independent* term scales with *valency* for given *size per value*, $d \ln \Gamma_!(z/d) = d \ln \Gamma_! x$.

Applying Stirling's approximation, the *alignment* approximates to $z(n-1)\ln d$. Thus *diagonal alignment* increases linearly with increasing *size* or *dimension*, and logarithmically with *valency*. Note that the first term decreases with increasing *valency* but the second term decreases more rapidly.

A *uniform full anti-diagonal histogram* is *planar* and hence its *alignment* is zero.

4.3.2 Crown alignment

The *alignment* of a *uniform full crown histogram* of *size* z and *dimension* $n > 0$ is

$$n \ln \Gamma_! \frac{z}{n} - \sum_{k \in \{0 \ldots n\}} \frac{n!}{k!(n-k)!} \ln \Gamma_! \frac{(n-1)^k z}{n^n}$$

A *uniform full crown histogram* is maximally *orthogonal*. If the *size* is scaled by *dimension*, $z = nx$ where x is the *size per dimension*, then the *non-independent* term scales with *dimension* for given *size per dimension*, $n \ln \Gamma_!(z/n) = n \ln \Gamma_! x$.

The *alignment* increases with *size* and *dimension*. *Valencies* greater than two have the same *alignment* because in each *effective state* each *variable* has exactly one of two *values*, $\forall w \in V\ (|A^{\mathrm{F}}\%\{w\}| = 2)$. Thus the *cartesian sub-volume* of the *independent* is *effectively bi-valent*, $|A^{\mathrm{XF}}| = 2^n$.

4.3.3 Axial alignment

The *alignment* of a *uniform full axial regular histogram* missing the *pivot* of *size* z, *dimension* n and *valency* d is

$$\begin{aligned} & b \ln \Gamma_! \frac{z}{b} - \sum_{k \in \{0 \ldots n\}} \frac{n!}{k!(n-k)!} (d-1)^k \ln \Gamma_! \frac{(n-1)^{n-k} z}{n^n (d-1)^k} \\ = \ & b \ln \Gamma_! \frac{z}{b} - \sum_{k \in \{0 \ldots n\}} \frac{n!}{k!(n-k)!} (d-1)^k \ln \Gamma_! \frac{(b-(d-1))^{n-k} z}{b^n} \end{aligned}$$

where the cardinality of *effective states* is $b = n(d-1)$.

The *alignment* of a *uniform full axial regular histogram* with a *pivot* is

$$c \ln \Gamma_! \frac{z}{c} - \sum_{k \in \{0 \ldots n\}} \frac{n!}{k!(n-k)!}(d-1)^k \ln \Gamma_! \frac{(c-(d-1))^{n-k} z}{c^n}$$

where the cardinality of *effective states* is $c = b + 1$ and the cardinality of *non-pivot effective states* is $b = n(d-1)$.

In the case of *uniform full axial regular histogram* missing the *pivot*, of *size* $(1-p)z$, plus a *singleton* of the *pivot state*, of *size* pz, where the *pivot* fraction is $p \in \mathbf{Q}$ such that $0 \leq p \leq 1$, the *alignment* is

$$b \ln \Gamma_! \frac{qz}{b} + \ln \Gamma_! pz - \sum_{k \in \{0 \ldots n\}} \frac{n!}{k!(n-k)!}(d-1)^k \ln \Gamma_! \frac{q^k (n-q)^{n-k} z}{n^n (d-1)^k}$$

where the cardinality of *effective states* is $c = b + 1$, the cardinality of *non-pivot effective states* is $b = n(d-1)$ and the *non-pivot* fraction is $q = 1 - p$. The *alignment* varies between (i) zero where the *pivot* fraction is one, $p = 1$, and the *histogram* is *effectively singleton*, (ii) the *alignment* of the with-*pivot* case where the *pivot* fraction $p = 1/c$, and (iii) the *alignment* of the without-*pivot* case where the *pivot* fraction is zero, $p = 0$.

A *uniform full axial regular histogram* missing the *pivot* is part *diagonal* and part *orthogonal*. If the *size* is scaled by *dimension* times the *non-null valency*, $z = n(d-1)x$ where x is the *size per dimension per non-null value*, then the *non-independent* term scales with *dimension* times the *non-null valency* for given *size per dimension per non-null value*, $b \ln \Gamma_!(z/b) = n(d-1) \ln \Gamma_! x$. If the *histogram* is *bi-valent*, $d = 2$, then the *uniform full axial regular histogram* missing the *pivot* is a *uniform full crown histogram* and therefore *orthogonal*. Otherwise the *uniform full axial regular histogram* missing the *pivot* is partly *diagonal*. In the case of *uniform full axial regular histogram* with non-zero *pivot count*, the *histogram* is partly *singleton*.

Axial histograms are intermediate between *diagonal*, *orthogonal* and *singleton*. The smaller the *valency*, the more *orthogonal*. The larger the *pivot count*, the more *singleton*.

4.3.4 Skeletal alignment

A *uniform full regular skeleton histogram* A can be defined such that the *variables* $V = \text{vars}(A)$ map to the *derived variables* of the *nullable transform*

D^{T} of a *well behaved decomposition* $D \in \mathcal{D}_{\mathrm{w},U}$, $V \leftrightarrow \mathrm{der}(D^{\mathrm{T}})$.

Given *valency* $d \in \mathbf{N}_{>0}$, let $Q \in \mathrm{trees}(\mathcal{V} \times \{1 \ldots d\})$ be a tree of depth $l \in \mathbf{N}_{>0}$ such that (i) $\forall L \in \mathrm{paths}(Q)\ (|\mathrm{dom}(\mathrm{set}(L))| = |L| = l)$ and (ii) $\forall X \in \{Q\} \cup \mathrm{ran}(\mathrm{nodes}(Q))\ (X \neq \emptyset \implies (|X| = d) \wedge (|\mathrm{dom}(\mathrm{dom}(X))| = 1) \wedge (\mathrm{ran}(\mathrm{dom}(X)) = \{1 \ldots d\}))$. Then a *skeletal histogram* of *size* z can be constructed $A = \mathrm{resize}(z, \{S \cup ((V \setminus \mathrm{vars}(S)) \times \{\mathrm{null}\}) : L \in \mathrm{paths}(Q),\ S = \mathrm{set}(L)\}^{\mathrm{U}}) \in \mathcal{A}$ where $V = \mathrm{dom}(\mathrm{elements}(Q))$ and the *null value* is $\mathrm{null} \in \mathcal{W}$. Then (i) $\mathrm{skeletal}(A)$, (ii) $\mathrm{size}(A) = z$, (iii) $\mathrm{ran}(A) = \{z/d^l\}$ and (iv) $\forall w, u \in V\ (\mathrm{axial}(A\%\{w,u\}))$. The *dimension* is $n = |V| = \sum d^{i-1} : i \in \{1 \ldots l\}$. The *volume* is $v = |A^{\mathrm{C}}| = d(d+1)^{n-1}$. The *effective states* $b = |A^{\mathrm{F}}| = d^l$.

The *alignment* of *uniform full regular skeleton histogram* A is

$$b \ln \Gamma_! \frac{z}{b} - \sum \left(\frac{n!}{p\ (n-m)!}\ d^m\ \ln \Gamma_! \Big(\frac{z}{d^m} \prod_{i \in \{1 \ldots l\}} \frac{(d^{i-1}-1)^{d^{i-1}-K_i}}{d^{(i-1)d^{i-1}}} \Big) : K \in \prod_{i \in \{1 \ldots l\}} \{i\} \times \{1 \ldots d^{i-1}\},\ m = \sum_{i \in \{1 \ldots l\}} K_i,\ p = \prod_{i \in \{1 \ldots l\}} K_i! \right)$$

If the depth, l, is constrained such that the *counts* of the *histogram*, A, are at least one, $l = \lfloor \ln z / \ln d \rfloor$, the *skeleton alignment* is minimised at integral *valency* $d = 2$. That is, where the *regular skeleton* tree, Q, is a binary tree.

4.3.5 Pivoted alignment

The *alignment* of a *uniform full pivoted regular histogram* of *size* z, *dimension* n and *valency* d is

$$b \ln \Gamma_! \frac{z}{b} - \sum_{k \in \{0 \ldots n\}} \binom{n}{k} (d-1)^k \ln \Gamma_! \frac{(d-1)^{k(n-1)} z}{b^n}$$

where the cardinality of *effective states* is $b = (d-1)^n + 1$. When the *regular histogram* is *bi-valent*, $d = 2$, the *histogram* is *diagonalised*. So the cardinality of *effective states* is two, $b = d$, and the *alignment* equals the *diagonal alignment*.

A *uniform full pivoted regular histogram* is roughly *volumar*. If the *size* is scaled by *volume*, $z = d^n x$ where x is the *size per volume*, then the *non-independent* term approximately scales with *volume* for given *size per volume*,

$b \ln \Gamma_! (z/b) \approx d^n \ln \Gamma_! x$.

The *alignment* of a *full pivoted regular histogram* that has *uniform non-pivot states* and a *pivot state* of *size* pz, where the *pivot* fraction is $p \in \{x : x \in \mathbf{Q},\ 0 \leq x \leq 1\}$, is

$$\ln \Gamma_! pz + (d-1)^n \ln \Gamma_! qz - \sum_{k \in \{0 \ldots n\}} \binom{n}{k} (d-1)^k \ln \Gamma_! (d-1)^{k(n-1)} q^k p^{n-k} z$$

where the *non-pivot* fraction of a single *non-pivot state* is $q = (1-p)/(d-1)^n$.

The *alignment* of a *uniform full anti-pivoted regular histogram* of *size* z, *dimension* n and *valency* d is

$$b \ln \Gamma_! \frac{z}{b} - \sum_{k \in \{0 \ldots n\}} \binom{n}{k} (d-1)^k \ln \Gamma_! \frac{(d^{n-1} - (d-1)^{n-1})^k (d^{n-1} - 1)^{n-k} z}{b^n}$$

where the cardinality of *effective states* is $b = d^n - ((d-1)^n + 1)$. When the *regular histogram* is *bi-valent*, $d = 2$, the *histogram* is *complement diagonal*. When the *regular histogram* is *bi-variate*, $n = 2$, the *histogram* is *axial* missing the *pivot*.

A *uniform full anti-pivoted regular histogram* is also roughly *volumar*.

4.4 Scaled alignment

The *alignment* of a *scaled histogram* $Z_k * A$, where $k \in \mathbf{Q}_{\geq 0}$ and $Z_k = \text{scalar}(k)$, weakly approximates to the scaled *alignment* of the *histogram*, $\text{alignment}(Z_k * A) \approx k \times \text{alignment}(A)$

$$\begin{aligned} \text{alignment}(Z_k * A) &= -\ln \frac{\text{mpdf}(U)(E^{\mathrm{X}}, kz)(Z_k * A)}{\text{mpdf}(U)(E^{\mathrm{X}}, kz)(Z_k * A^{\mathrm{X}})} \\ &\approx -\ln \left(\frac{\text{mpdf}(U)(E^{\mathrm{X}}, z)(A)}{\text{mpdf}(U)(E^{\mathrm{X}}, z)(A^{\mathrm{X}})} \right)^k \\ &= k \times \text{alignment}(A) \end{aligned}$$

Thus scale analysis suggests that *alignment* has the units of *size*.

The *scaled multinomial probability density function*

$$\text{mpdf}(U)(E^{\mathrm{X}}, kz)(Z_k * A) \approx (\text{mpdf}(U)(E^{\mathrm{X}}, z)(A))^k$$

approximates best for high *entropy sample histograms* for which the *multinomial coefficient* is largest. The *entropy* of the *independent histogram* is often greater than that of the *histogram*, $\mathrm{entropy}(A^{\mathrm{X}}) \geq \mathrm{entropy}(A)$ (see 'Minimum alignment' below), so the approximation is best for low *alignments*.

4.5 Minimum alignment

If a *histogram* is *integral*, $A \in \mathcal{A}_{\mathrm{i}}$, and the *independent histogram*, A^{X}, is such that all of the *counts* are fractional, $\forall c \in \mathrm{ran}(A^{\mathrm{X}})\ (0 \leq c \leq 1)$, then the gamma functions of the *counts* are such that $\forall c \in \mathrm{ran}(A^{\mathrm{X}})\ (0! \geq \Gamma_! c \leq 1!)$. Hence $\sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S \leq 0$ and so the *alignment* is positive, $\mathrm{algn}(A) \geq 0$. Note that in this case $z \leq v$ where $z = \mathrm{size}(A)$ and $v = |A^{\mathrm{C}}|$.

If the *independent histogram* is *integral scaled uniform cartesian*, $A^{\mathrm{X}} = Z_{z/v} * V^{\mathrm{C}} \in \mathcal{A}_{\mathrm{i}}$ where $Z_x = \mathrm{scalar}(x)$, $V = \mathrm{vars}(A)$, $z = \mathrm{size}(A)$ and $v = |V^{\mathrm{C}}|$ such that $z/v \in \mathbf{N}_{>0}$, then its *multinomial coefficient*

$$\frac{z!}{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!} = \frac{z!}{((z/v)!)^v}$$

is maximised. The *classification space* is the logarithm of the *multinomial coefficient* and so it is maximised too

$$0 \leq \mathrm{spaceClassification}(A) \leq \mathrm{spaceClassification}(A^{\mathrm{X}}) = \ln z! - v \ln(z/v)!$$

See 'Classification space' above for a discussion showing that the *classification space* is maximised at maximum *entropy*. Thus in this case the *alignment* is positive

$$\mathrm{algn}(A) = \mathrm{spaceClassification}(A^{\mathrm{X}}) - \mathrm{spaceClassification}(A) \geq 0$$

Note that there exists a non-singleton set of *integral iso-independents* of *integral scaled uniform cartesian independent*, $|Y^{-1}_{U,\mathrm{i},V,z}(Z_{z/v} * V^{\mathrm{C}})| > 1$, if the *histogram* is *pluri-variate*, $|V| > 1$, and each *variable* is *pluri-valent*, $\forall w \in V\ (|U_w| > 1)$, for all $z/v \in \mathbf{N}_{>0}$. Of the *integral iso-independents* set only the *independent histogram* has zero *alignment*, $Z_{z/v} * V^{\mathrm{C}} \in Y^{-1}_{U,\mathrm{i},V,z}(Z_{z/v} * V^{\mathrm{C}})$ and $\mathrm{algn}(Z_{z/v} * V^{\mathrm{C}}) = 0$. The others have *alignment* greater than zero, $\forall B \in Y^{-1}_{U,\mathrm{i},V,z}(Z_{z/v} * V^{\mathrm{C}})\ (B \neq Z_{z/v} * V^{\mathrm{C}} \implies \mathrm{algn}(B) > 0)$.

The *minimum alignment conjecture* states that if the *independent histogram* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, then the *minimum alignment* is conjectured to be zero,

$$\forall A \in \mathcal{A}\ (A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies \mathrm{algn}(A) \geq 0)$$

or $\mathrm{minr}(\{(A, \mathrm{algn}(A)) : A \in \mathcal{A},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\}) \geq 0$. This is a consequence of the *integral mean multinomial probability distribution conjecture* which states that if the *mean* of the *multinomial probability distribution* is *integral* then it is also *modal*, $\mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E,z)) \in \mathcal{A}_i \implies \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E,z)) \in \mathrm{modes}(\hat{Q}_{\mathrm{m},U}(E,z))$. Thus for *complete integral independent histogram* $(A^{\mathrm{XU}} = A^{\mathrm{C}}) \wedge (A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}) \implies A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},V,z}$ and $A^{\mathrm{X}} = \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)) \implies A^{\mathrm{X}} \in \mathrm{maxd}(\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z))$. Thus

$$\forall A \in \mathcal{A}_{U,\mathrm{i},V,z}\ (A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies$$
$$\frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{A^{\mathrm{X}}_S}{z}\right)^{A_S} \leq \frac{z!}{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!} \prod_{S \in A^{\mathrm{XS}}} \left(\frac{A^{\mathrm{X}}_S}{z}\right)^{A^{\mathrm{X}}_S})$$

where $z = \mathrm{size}(A) > 0$. The *distribution histogram*, A^{X}, is *independent* so

$$\forall A \in \mathcal{A}_{U,\mathrm{i},V,z}\ (A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \leq \frac{z!}{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!})$$

Alignment does not depend on *completeness*, $\mathrm{algn}(A) = \mathrm{algn}(A + A^{\mathrm{CZ}}) = \mathrm{algn}(A * A^{\mathrm{F}})$, thus *integral independent histogram* implies positive *alignment*, $\forall A \in \mathcal{A}\ (A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies \mathrm{algn}(A) \geq 0)$.

Moreover, within the degree to which Stirling's approximation holds, the *minimum alignment* is zero even for *non-integral independent histogram*. Utilising the identity

$$\sum_{S \in N^{\mathrm{FS}}} N_S \ln N^{\mathrm{X}}_S = \sum_{S \in N^{\mathrm{XFS}}} N^{\mathrm{X}}_S \ln N^{\mathrm{X}}_S$$

where $N \in \mathcal{A} \cap \mathcal{P}$, and noting that the *relative entropy* between the *sample histogram* and its *independent histogram* is greater than or equal to zero by Gibbs' inequality, then

$$\begin{aligned}
\mathrm{alignment}(A) &:= \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S \\
&\approx z \sum_{S \in N^{\mathrm{FS}}} N_S \ln N_S - z \sum_{S \in N^{\mathrm{XFS}}} N^{\mathrm{X}}_S \ln N^{\mathrm{X}}_S \\
&= z \sum_{S \in N^{\mathrm{FS}}} N_S \ln N_S - z \sum_{S \in N^{\mathrm{FS}}} N_S \ln N^{\mathrm{X}}_S \\
&= z \sum_{S \in N^{\mathrm{FS}}} N_S \ln \frac{N_S}{N^{\mathrm{X}}_S} \\
&= z \times \mathrm{entropyRelative}(N, N^{\mathrm{X}}) \\
&\geq 0
\end{aligned}$$

where $z = \text{size}(A) > 0$ and $N = \text{resize}(1, A) \in \mathcal{P}$. *Alignment* is approximately equal to the scaled difference between the *entropy* of the *independent histogram* and the *entropy* of the *histogram*, which in this case is the *scaled relative entropy* between *histogram* and *independent*, or *scaled mutual entropy*

$$\begin{aligned} \text{alignment}(A) &\approx z \times \text{entropy}(A^{\mathrm{X}}) - z \times \text{entropy}(A) \\ &= z \times \text{entropyRelative}(A, A^{\mathrm{X}}) \end{aligned}$$

Thus $\text{entropy}(A^{\mathrm{X}}) \geq \text{entropy}(A)$.

Similar logic was used above to show that the log of the *generalised multinomial probability distribution*, $\ln \circ \hat{Q}_{\mathrm{m},U}(E, z) \in \mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{R}$, can be approximated by the negative *relative entropy* between the *sample histogram* and the *distribution histogram* by means of the Stirling approximation

$$\hat{Q}_{\mathrm{m},U}(E, z)(A) = z! \prod_{S \in A^{\mathrm{S}}} \frac{P_S^{A_S}}{A_S!} \approx \prod_{S \in A^{\mathrm{FS}}} \left(\frac{P_S}{N_S}\right)^{A_S}$$

where $P = \text{resize}(1, E)$ and $N = \text{resize}(1, A)$. So

$$\ln \text{mpdf}(U)(E, z)(A) \approx \sum_{S \in A^{\mathrm{S}}} A_S \ln \frac{P_S}{N_S} = -z \sum_{S \in N^{\mathrm{FS}}} N_S \ln \frac{N_S}{P_S}$$

where $A^{\mathrm{F}} \leq E^{\mathrm{F}}$. Let the *distribution histogram* equal the *independent histogram*, $E = A^{\mathrm{X}}$, then

$$\begin{aligned} \ln \text{mpdf}(U)(A^{\mathrm{X}}, z)(A) &\approx -z \sum_{S \in N^{\mathrm{FS}}} N_S \ln \frac{N_S}{N_S^{\mathrm{X}}} \\ &\approx -\text{alignment}(A) \\ &= \ln \frac{\text{mpdf}(U)(A^{\mathrm{X}}, z)(A)}{\text{mpdf}(U)(A^{\mathrm{X}}, z)(A^{\mathrm{X}})} \end{aligned}$$

which implies that $\text{mpdf}(U)(A^{\mathrm{X}}, z)(A^{\mathrm{X}}) \approx 1$ within the Stirling approximation and that the *alignment* is positive.

As shown above, the gamma function is log convex and hence the expected logarithm of the factorial of the *counts* of the *states* of the *sample histograms* is greater than or equal to the logarithm of the factorial of the *counts* of the *states* of the *mean histogram* by Jensen's inequality

$$\forall S \in V^{\mathrm{CS}}\ (\text{expected}(\hat{Q}_{\mathrm{m},U}(E, z))(\{(A, \ln A_S!) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \geq \ln \Gamma_! M_S)$$

where the *mean histogram* is $M = \text{mean}(\hat{Q}_{\text{m},U}(E, z))$. Consider a *draw* of *size* z from *independent distribution histogram* E^{X}. Let

$$Y = \{(S, \text{expected}(\hat{Q}_{\text{m},U}(E^{\text{X}}, z))(\{(A, \ln A_S!) : A \in \mathcal{A}_{U,\text{i},V,z}\})) : S \in V^{\text{CS}}\}$$

and $X = \{(S, \Gamma_!^{-1}(e^y)) : (S, y) \in Y\} \in \mathcal{S}_U \to \mathbf{R}$. Let X' be a *histogram*, which is a rational valued function, that approximates closely to X, which is a real valued function, $X' \approx X$. Let $z' = \text{size}(X')$. Conjecture that $z' \geq z$. Conjecture that $\text{scalar}(z/z') * X' \approx \text{scalar}(z/z_E) * E^{\text{X}}$, but that $\text{algn}(\text{scalar}(z/z') * X') \geq 0$. That is, even where the *alignment* is small, the log convexity tends to make it positive.

However, *alignment* is negative for some *histograms*. For example, the *alignment* of a *uniform full anti-pivoted regular histogram* A of *size* $z = 1000$, *dimension* $n = 2$ and *valency* $d = 101$ is

$$\text{algn}(A) = \\ b \ln \Gamma_! \frac{z}{b} - \sum_{k \in \{0 \ldots n\}} \frac{n!}{k!(n-k)!}(d-1)^k \ln \Gamma_! \frac{z(n-1)^k}{n^n (d-1)^k} \approx -277.52$$

where the cardinality of *effective states* is $b = d^n - ((d-1)^n + 1) = 200$. Note that the *independent* of the example *uniform full anti-pivoted regular histogram* is *non-integral*, $A^{\text{X}} \notin \mathcal{A}_{\text{i}}$. The *independent* cannot be the *distribution histogram* of a *generalised multinomial probability distribution*. The *independent* is not maximal in the *multinomial probability density function* parameterised by the *independent*, $\text{mpdf}(U)(A^{\text{X}}, z)(A^{\text{X}}) < \text{mpdf}(U)(A^{\text{X}}, z)(A)$.

Conjecture that there exist *non-integral histograms* which are not *independent* but have zero *alignment*, $\exists A \in \mathcal{A}\ (A \neq A^{\text{X}} \wedge \text{algn}(A) = 0)$. However, although the example above, of a *uniform full anti-pivoted regular histogram* of *dimension* $n = 2$ and *valency* $d = 101$, has a real solution of $z \approx 3850.38 \in \mathbf{R}$ such that *alignment* equals zero, there may not be a rational solution *size* $z \in \mathbf{Q}$.

The *minimum alignment* is sometimes negative, depending on the *geometry* of the *variables* V in *system* U, and the *size* z. However, conjecture that the *expected alignment* is always positive if the *sample histograms* are *drawn* from the *generalised multinomial probability distribution* where the *distribution* *histogram* is *independent*, $E = E^{\mathrm{X}}$. The *expected exponential alignment* is

$$\begin{aligned}
&\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z))(\{(A,\exp(\mathrm{algn}(A))) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \\
&= \sum_{A \in \mathcal{A}_{U,\mathrm{i},V,z}} \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E^{\mathrm{X}}_S}{z_E}\right)^{A_S} \frac{\mathrm{mpdf}(U)(E^{\mathrm{X}},z)(A^{\mathrm{X}})}{\mathrm{mpdf}(U)(E^{\mathrm{X}},z)(A)} \\
&= \sum_{A \in \mathcal{A}_{U,\mathrm{i},V,z}} \frac{z!}{\prod_{S \in A^{\mathrm{XS}}} \Gamma_! A^{\mathrm{X}}_S} \prod_{S \in A^{\mathrm{XS}}} \left(\frac{E^{\mathrm{X}}_S}{z_E}\right)^{A^{\mathrm{X}}_S} \\
&= \sum_{A \in \mathcal{A}_{U,\mathrm{i},V,z}} \mathrm{mpdf}(U)(E^{\mathrm{X}},z)(A^{\mathrm{X}}) \\
&= \mathrm{expected}(\hat{Q}_{\mathrm{u},U}(V,z))(\{(A,\mathrm{mpdf}(U)(E^{\mathrm{X}},z)(A^{\mathrm{X}})) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \times |\mathcal{A}_{U,\mathrm{i},V,z}| \\
&\geq \mathrm{expected}(\hat{Q}_{\mathrm{u},U}(V,z))(\{(A,\mathrm{mpdf}(U)(E^{\mathrm{X}},z)(A)) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \times |\mathcal{A}_{U,\mathrm{i},V,z}| \\
&= \mathrm{sum}(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)) = 1
\end{aligned}$$

where $\hat{Q}_{\mathrm{u},U}(V,z)$ is the *uniform probability distribution* and $\exp \in \mathbf{R} \to \mathbf{R}$ is the exponential function. The *entropy* of the *independent histogram* is greater than or equal to the *entropy* of the *histogram*, $\mathrm{entropy}(A^{\mathrm{X}}) \geq \mathrm{entropy}(A)$. So the probability of *drawing* the *independent histogram*, A^{X}, from an *independent distribution*, E^{X}, is expected in the *uniform probability distribution*, $\hat{Q}_{\mathrm{u},U}(V,z)$, to be greater than or equal to that of the *sample histogram*, A. If the *expected exponential alignment* is conjectured to be greater than or equal to 1 then conjecture that the *expected alignment* is positive

$$\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z))(\{(A,\mathrm{algn}(A)) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \geq 0$$

However, note that Jensen's inequality implies that the *expected alignment* is only less than or equal to the logarithm of the *expected exponential alignment*

$$\begin{aligned}
&\exp(\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z))(\{(A,\mathrm{algn}(A)) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\})) \\
&\leq \mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z))(\{(A,\exp(\mathrm{algn}(A))) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\})
\end{aligned}$$

In the case where $E = V^{\mathrm{C}}$ then the sum must be less than the maximum *entropy histogram*, $A = \mathrm{resize}(z,V^{\mathrm{C}})$

$$\begin{aligned}
&\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}},z))(\{(A,\exp(\mathrm{algn}(A))) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \\
&\leq \sum_{A \in \mathcal{A}_{U,\mathrm{i},V,z}} \frac{z!}{\prod_{S \in V^{\mathrm{CS}}} \Gamma_!(z/v)} \prod_{S \in V^{\mathrm{CS}}} \left(\frac{1}{v}\right)^{z/v} = \frac{(z+v-1)!}{(v-1)!} \frac{1}{((z/v)!)^v} \frac{1}{v^z}
\end{aligned}$$

where $v = |V^{\mathrm{C}}|$. Hence, by Jensen's inequality

$$\begin{aligned}&\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}}, z))(\{(A, \mathrm{algn}(A)) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \\ &\leq \ \ln(z+v-1)! - \ln(v-1)! - v\ln(z/v)! - z\ln v\end{aligned}$$

If $z \gg v$ then this approximates to $v \ln(z/v)$. Therefore conjecture that *expected alignment* varies as the *volume* for constant *size* greater than the *volume*

$$\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z))(\{(A, \mathrm{algn}(A)) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \sim v$$

If $z \ll v$ then the expression above approximates to $z \ln(v/z)$. Conjecture that *expected alignment* varies as the logarithm of the *volume* for constant *size* less than the *volume*

$$\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z))(\{(A, \mathrm{algn}(A)) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \sim \ln v$$

The *partially independent set* R_A of *histogram* A of *variables* $V = \mathrm{vars}(A)$ is the set of *partially independent histograms*

$$R_A = \{Z_A * \prod\{\frac{A}{Z_A} \ \% \ C : C \in P\} : P \in \mathrm{B}(V)\}$$

where $Z_A = \mathrm{scalar}(\mathrm{size}(A))$. The *independent* is a member, $A^{\mathrm{X}} \in R_A$. The *alignment* of the *partially independent histograms* is such that

$$\forall B \in R_A \ (0 \leq \mathrm{algn}(B) \leq \mathrm{algn}(A))$$

in the case where $\mathrm{algn}(A) \geq 0$.

4.6 Maximum alignment

The maximum *alignment* of a *histogram* A is conjectured to occur when the *histogram* is both *uniform*, $|\mathrm{ran}(\mathrm{trim}(A))| = 1$, and *fully diagonalised*, $\mathrm{diagonalFull}(U)(A)$. The set of *congruent maximum alignment histograms* for a set of *variables* V in *system* U can be calculated explicitly. There is a *uniform histogram* of *size* z for each of the subsets of the *cartesian states* having maximum cardinality and which are such that the elements have zero mutual *incidence*, $\{\mathrm{resize}(z, A) : A \in \mathrm{P}(V^{\mathrm{C}}),\ \mathrm{diagonalFull}(U)(A)\}$.

The maximum *alignment* of a *regular histogram* A with *variables* V in *system* U and *size* $z = \mathrm{size}(A)$, *dimension* $n = |V|$ and *valency* d, where $\{d\} = \{|U_v| : v \in V\}$, is

$$d \ln \Gamma_! \frac{z}{d} - d^n \ln \Gamma_! \frac{z}{d^n}$$

If a *histogram* is not *regular*, its *maximum alignment* is that of a *regular histogram* of the same *size* z and *dimension* n having *valency* d equal to the minimum *valency*, $d = \mathrm{minr}(\{(v, |U_v|) : v \in V\})$. This *regular histogram* is *effectively congruent* to the *independent histogram's cartesian sub-volume*, $|A^{\mathrm{XF}}| = d^n$. Define alignmentMaximum$(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathbf{Q}_{\geq 0} \to \mathbf{R}$

$$\mathrm{alignmentMaximum}(U)(V, z) := d \ln \Gamma_! \frac{z}{d} - d^n \ln \Gamma_! \frac{z}{d^n}$$

where $n = |V|$ and $d = \mathrm{minr}(\{(v, |U_v|) : v \in V\})$. alignmentMaximum is undefined for *scalars*, $V = \emptyset$.

If a *uniform diagonalised histogram* is not *fully diagonalised* its *alignment* is equal to the *maximum alignment* of the *regular histogram* having the same *dimension* n and *valency* d equal to the *effective valency* of the *independent histogram's cartesian sub-volume*, $d = |A^{\mathrm{XF}}|^{1/n}$.

Applying Stirling's approximation, the maximum *alignment* approximates to $z(n-1) \ln d$. So maximum *alignment* increases with increasing *size*, *dimension* and *valency*. For a given set of *variables* V, the *alignment* is of the same complexity as the *size*, alignment $\in \mathrm{O}(\mathrm{size}, m)$ where the multiplier m depends on the *dimension* and *valencies* of V.

In some special cases, the application of Stirling's approximation is exact, $\mathrm{alignmentMaximum}(U)(V, z) = z(n-1) \ln d$. *Zero histograms*, $z = 0$, and *mono-variate histograms*, $n = 1$, have zero *alignment* and zero maximum *alignment* in the approximation. *Pluri-variate singleton histograms* satisfy the requirement for zero *incidence*, but the maximum *alignment* is zero because they are *effectively mono-valent*, $\ln d = 0$.

Consider constructing a single *derived variable* w of *valency* equal to the *effective* cardinality of a *diagonalised pluri-variate histogram* A in a *system* U such that $U_w = \mathrm{states}(A^{\mathrm{F}})$ then we can calculate a *mono-variate histogram* $B = \{(\{(w, S)\}, c) : (S, c) \in \mathrm{trim}(A)\}$ that represents the *diagonal* of A. The *events classification space* of the *derived histogram* spaceClassification(B) is maximised when the *alignment* of the *diagonalised histogram* is maximised. In other words, the *entropy* of the *diagonal* is maximised when *alignment* is maximised. Let D be the set of *diagonalised histograms* of *size* z and *variables* V in a *system* U, $D \in \mathrm{P}(\mathcal{A}_U)$

$$D = \{A : A \in \mathcal{A}_U,\ \mathrm{vars}(A) = V,\ \mathrm{size}(A) = z,\ \mathrm{diagonal}(A)\}$$

then

$$\begin{aligned}\text{maxd}(\{(A, \text{algn}(A)) : A \in D\}) &= \text{maxd}(\{(A, \text{spCl}(A)) : A \in D\}) \\ &= \text{maxd}(\{(A, \text{entropy}(A)) : A \in D\})\end{aligned}$$

where algn = alignment and spCl = spaceClassification. This is counter-intuitive because a glance at the definition of *alignment* would suggest that maximum *alignment* would increase with decreasing *events classification space*. But as the maximum is approached the sensitivity to the *independent events classification space* becomes more important. For example a *singleton* which has a low *events classification space*, but also has an equally low *independent events classification space* and hence zero *alignment*.

Consider the *effective states* $\text{states}(A^{\text{F}})$ of a *diagonalised histogram* that is not necessarily at maximum *alignment*. The subsets of the *states* which are *incident* on each of the *states* on the *diagonal* form a set of exclusive *singleton histograms* and thus *independent histograms*

$$\forall A \in \mathcal{A}\,(\text{diagonal}(A) \implies (\forall S \in A^{\text{FS}}\,\Diamond B = A\backslash\text{incidence}(A, S, 0)\,(B = B^{\text{X}})))$$

In other words, the partition of the *volume* of *diagonalised histogram* A, $\{A \setminus \text{incidence}(A, S, 0) : S \in A^{\text{FS}}\} \in \text{B}(A^{\text{CS}})$, which has the *components* corresponding to each of these *singletons*, consists of a set of *independent histograms*.

4.7 Dependent alignment

Given a *substrate histogram* $A \in \mathcal{A}_{U,V,z}$, the *independent*, $A^{\text{X}} \in \mathcal{A}_{U,V,z}$, is conjectured in section 'Likely histograms', above, to be the *maximum likelihood estimate* of the sum of the *generalised multinomial probabilities* of the *integral iso-independents* of the *histogram*, A,

$$\{A^{\text{X}}\} = \text{maxd}(\{(D, \sum(Q_{\text{m},U}(D, z)(B) : B \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}}))) : D \in \mathcal{A}_{U,V,z}\})$$

where the *integral iso-independents* is

$$Y^{-1}_{U,\text{i},V,z}(A^{\text{X}}) \;=\; \{B : B \in \mathcal{A}_{U,\text{i},V,z},\ B^{\text{X}} = A^{\text{X}}\}$$

The corresponding *dependent histogram*, $A^{\text{Y}} \in \mathcal{A}_{U,V,z}$, is defined

$$\{A^{\text{Y}}\} = \text{maxd}(\{(D, \frac{Q_{\text{m},U}(D, z)(A)}{\sum Q_{\text{m},U}(D, z)(B) : B \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})}) : D \in \mathcal{A}_{U,V,z}\})$$

Alignment may be defined as the negative logarithm *independent-sample-distributed relative dependent multinomial probability density*

$$\text{algn}(A) \quad = \quad -\ln \frac{\text{mpdf}(U)(A^{\text{X}}, z)(A)}{\text{mpdf}(U)(A^{\text{X}}, z)(A^{\text{X}})}$$

which is the *independent-distributed-relative multinomial space*,

$$\text{algn}(A) \quad = \quad \text{spaceRelative}(A^{\text{X}})(A)$$

where the *distribution-relative multinomial space* is defined, in section 'Likely histograms', above, as

$$\text{spaceRelative}(E)(A) \quad := \quad -\ln \frac{\text{mpdf}(U)(E, z)(A)}{\text{mpdf}(U)(E, z)(E)}$$

In the case where both the *histogram* and *independent* are *integral*, $A, A^{\text{X}} \in \mathcal{A}_{\text{i}}$, the *independent-distributed-relative multinomial space* is

$$\text{spaceRelative}(A^{\text{X}})(A) \quad := \quad -\ln \frac{Q_{\text{m},U}(A^{\text{X}}, z)(A)}{Q_{\text{m},U}(A^{\text{X}}, z)(A^{\text{X}})}$$

The *independent-distributed-relative multinomial space* of the *independent* is zero,

$$\text{spaceRelative}(A^{\text{X}})(A^{\text{X}}) = \text{algn}(A^{\text{X}}) = 0$$

In section 'Likely histograms', above, it is conjectured that the logarithm of the *maximum conditional probability* with respect to the *dependent analogue* varies with the *relative space* with respect to the *independent analogue*, which in this case is the *alignment*,

$$\begin{aligned}
\ln \frac{Q_{\text{m},U}(A^{\text{Y}}, z)(A)}{\sum Q_{\text{m},U}(A^{\text{Y}}, z)(B) : B \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})} \quad &\sim \quad -\ln \frac{Q_{\text{m},U}(A^{\text{X}}, z)(A)}{Q_{\text{m},U}(A^{\text{X}}, z)(A^{\text{X}})} \\
&= \quad \text{spaceRelative}(A^{\text{X}})(A) \\
&= \quad \text{algn}(A)
\end{aligned}$$

In the case where the *histogram* and *independent* are *integral*, $A, A^{\text{X}} \in \mathcal{A}_{\text{i}}$, the *independent-distributed-relative multinomial space* is conjectured to be greater than or equal to zero and less than or equal to the *independent-distributed-relative multinomial space* of the *dependent*,

$$0 \leq \text{spaceRelative}(A^{\text{X}})(A) \leq \text{spaceRelative}(A^{\text{X}})(A^{\text{Y}})$$

In the case where the *independent* of the *dependent* equals the *independent*, $A^{\mathrm{YX}} = A^{\mathrm{X}}$, then the inequality is

$$0 \leq \mathrm{algn}(A) \leq \mathrm{algn}(A^{\mathrm{Y}})$$

This is consistent with the *entropies*,

$$\mathrm{entropy}(A^{\mathrm{X}}) \geq \mathrm{entropy}(A) \geq \mathrm{entropy}(A^{\mathrm{Y}})$$

Conjecture that if the *histogram* is at *maximum alignment* the *dependent* equals the *histogram*,

$$\mathrm{algn}(A) = \mathrm{alignmentMaximum}(U)(V, z) \implies A^{\mathrm{Y}} = A$$

4.8 Capacity and Alignment density

Functions on the *geometry*, called *capacities*, of a *histogram* A can be formalised in terms of the *effective states* A^{FS}. Let this set of functions be defined as capacities $\subset \mathrm{P}(\mathcal{S}) \to \mathbf{R}_{>0}$ such that $\forall K \in$ capacities $\forall Q \in \mathrm{dom}(K)\ \forall S, T \in Q\ (\mathrm{vars}(S) = \mathrm{vars}(T))$. The application of *capacity* function K to the empty set, $K(\emptyset)$, is undefined.

Given some *capacity* function, $K \in$ capacities, the *alignment density* of *non-zero histogram* A is defined as $\mathrm{alignment}(A)/K(A^{\mathrm{FS}}) \in \mathbf{R}$. The *alignment density* of a *zero histogram*, $\mathrm{size}(A) = 0$, is undefined.

The *unit capacity* function is constant 1. Define capacityUnit $\in$ capacities as $\mathrm{capacityUnit}(Q) := 1$. The *alignment unit density* equals the *alignment* $\mathrm{alignment}(A)/\mathrm{capacityUnit}(A^{\mathrm{FS}}) = \mathrm{alignment}(A)$.

The *effective capacity* is the cardinality of the *effective histogram*. Define capacityEffective $\in$ capacities as $\mathrm{capacityEffective}(Q) := |Q|$.

The *volume capacity* is the *volume* of the *variables*. Define $\mathrm{capacityVolume}(U) \in$ capacities as $\mathrm{capacityVolume}(U)(Q) := v$ where $v = \mathrm{volume}(U)(\mathrm{vars}(Q))$, and $\mathrm{vars} \in \mathrm{P}(\mathcal{S}) \to \mathrm{P}(\mathcal{V})$ is defined as $\mathrm{vars}(Q) = \bigcup\{\mathrm{vars}(S) : S \in Q\}$.

The *valency capacity* is the geometrical mean of the *valencies* of the *variables*. Define $\mathrm{capacityValency}(U) \in$ capacities in *system* U as

$$\mathrm{capacityValency}(U)(Q) := v^{1/n}$$

where $n = |\mathrm{vars}(Q)|$. The *alignment valency density* is

$$\frac{\mathrm{alignment}(A)}{\mathrm{capacityValency}(U)(A^{\mathrm{FS}})}$$

The *effective valency capacity* is similar to the *valency capacity* except that the interesting *volume* is the cardinality of the *cartesian sub-volume* of the *effective independent*. Define capacityValencyEffective$(U) \in$ capacities in *system* U as

$$\mathrm{capacityValencyEffective}(U)(Q) := |Q^{\mathrm{UXF}}|^{1/n}$$

The *diagonal capacity* is the cardinality of the *values* of the shortest *variable*. Define capacityDiagonal$(U) \in$ capacities as capacityDiagonal$(U)(Q) :=$ minr$(\{|U_w| : w \in \mathrm{vars}(Q)\})$.

The *maximum alignment* of a *histogram* of *size* z and *variables* V approximates to $z(n-1)\ln d$ where $d = \mathrm{minr}(\{|U_w| : w \in V\})$ and $n = |V|$. The *aligned capacity* is the log of the *valency capacity* scaled by $n-1$. Define capacityAligned$(U) \in$ capacities as

$$\mathrm{capacityAligned}(U)(Q) := (1 - 1/n)\ln v$$

The *alignment aligned density* at *maximum alignment* of a *regular volume*, $\{d\} = \{|U_w| : w \in V\}$, is approximately independent of *geometry*

$$\frac{\mathrm{alignmentMaximum}(U)(V,z)}{\mathrm{capacityAligned}(U)(V^{\mathrm{CS}})} \approx \frac{z(n-1)\ln d}{(1-1/n)\ln v} = z$$

4.9 Alignment and independent histograms

The *multinomial probability density* of an *independent histogram* A^{X} of *size* z and *variables* V *drawn* from an *independent distribution* E^{X} is approximately equal to the product of the *multinomial probability densities* of the *reduced independent histogram* $A^{\mathrm{X}}\%\{w\}$, where $w \in V$, *drawn* from the *reduced independent distribution* $E^{\mathrm{X}}\%\{w\}$

$$\mathrm{mpdf}(U)(E^{\mathrm{X}}, z)(A^{\mathrm{X}}) \approx \prod_{w \in V} \mathrm{mpdf}(U)(E^{\mathrm{X}}\%\{w\}, z)(A^{\mathrm{X}}\%\{w\})$$

To see this (i) use the notation $A_w = A\%\{w\}$ and $S_w = S\%\{w\}$, (ii) let $P^{\mathrm{X}} = \mathrm{resize}(1, E^{\mathrm{X}})$, (iii) note that

$$\sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln P^{\mathrm{X}}_S = \sum_{w \in V} \sum_{R \in A^{\mathrm{XS}}_w} A^{\mathrm{X}}_{w,R} \ln P^{\mathrm{X}}_{w,R}$$

which follows from the proof that $\sum_{S \in A^{\mathrm{S}}} A_S \ln P^{\mathrm{X}}_S = \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln P^{\mathrm{X}}_S$ above, and (iv) apply Stirling's approximation,

$$\begin{aligned}
&\ln \mathrm{mpdf}(U)(E^{\mathrm{X}}, z)(A^{\mathrm{X}}) \\
&= \ln \Gamma_! z - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S + \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln P^{\mathrm{X}}_S \\
&\approx z \ln z - \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln A^{\mathrm{X}}_S + \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln P^{\mathrm{X}}_S \\
&= \sum_{w \in V} \left(z \ln z - \sum_{R \in A^{\mathrm{XS}}_w} A^{\mathrm{X}}_{w,R} \ln A^{\mathrm{X}}_{w,R} + \sum_{R \in A^{\mathrm{XS}}_w} A^{\mathrm{X}}_{w,R} \ln P^{\mathrm{X}}_{w,R} \right) \\
&\approx \sum_{w \in V} \left(\ln \Gamma_! z - \sum_{R \in A^{\mathrm{XS}}_w} \ln \Gamma_! A^{\mathrm{X}}_{w,R} + \sum_{R \in A^{\mathrm{XS}}_w} A^{\mathrm{X}}_{w,R} \ln P^{\mathrm{X}}_{w,R} \right) \\
&= \ln \prod_{w \in V} \mathrm{mpdf}(U)(E^{\mathrm{X}}\%\{w\}, z)(A^{\mathrm{X}}\%\{w\})
\end{aligned}$$

It can be shown, however, that the *multinomial probability density* of an *independent histogram* is not exactly equal to product of the *multinomial probability densities*, but only approximately equal. For example, in the case of a *regular scaled cartesian histogram* of *size* z *dimension* $n = |V|$ and *valency* d, $A^{\mathrm{X}} = \mathrm{scalar}(z/d^n) * V^{\mathrm{C}}$ *drawn* from V^{C}

$$\begin{aligned}
&\ln \frac{\prod_{w \in V} \mathrm{mpdf}(U)(E^{\mathrm{X}}\%\{w\}, z)(A^{\mathrm{X}}\%\{w\})}{\mathrm{mpdf}(U)(E^{\mathrm{X}}, z)(A^{\mathrm{X}})} \\
&= n \ln \Gamma_! z - nd \ln \Gamma_! \frac{z}{d} + nz \ln \frac{1}{d} \\
&\quad - \ln \Gamma_! z + d^n \ln \Gamma_! \frac{z}{d^n} - z \ln \frac{1}{d^n} \\
&= n \ln \Gamma_! z - nd \ln \Gamma_! \frac{z}{d} - \ln \Gamma_! z + d^n \ln \Gamma_! \frac{z}{d^n} \neq 0 \\
&\approx nz \ln d - z \ln d^n = 0
\end{aligned}$$

This approximation of the *multinomial probability density* of an *independent histogram* is related to the method used by the *dimensional classification coder* of *histories*, $C_{\mathrm{G,n}}$, to encode the *reduced classifications*. This is a special case where the *distribution histogram* is *cartesian*, $E^{\mathrm{X}} = V^{\mathrm{C}}$

$$\begin{aligned}
\prod_{w \in V} \mathrm{mpdf}(U)(\{w\}^{\mathrm{C}}, z)(A^{\mathrm{X}}\%\{w\}) &= n \ln \Gamma_! z - \sum_{w \in V} \sum_{R \in A^{\mathrm{XS}}_w} \ln \Gamma_! A^{\mathrm{X}}_{w,R} \\
&= \sum_{w \in V} \mathrm{spaceClassification}(A^{\mathrm{X}}\%\{w\})
\end{aligned}$$

There is an analogy between the *size scaling* of *alignment* and what may be called *dimension scaling* of the logarithm of the *multinomial probability density* of an *independent histogram*. Choose one of the *variables* $w \in V$ then

$$\begin{aligned}
\text{alignment}(Z_k * A) &\approx -\ln \left(\frac{\text{mpdf}(U)(E^{\text{X}}, z)(A)}{\text{mpdf}(U)(E^{\text{X}}, z)(A^{\text{X}})} \right)^k \\
&= k \ \text{alignment}(A) \\
\ln \text{mpdf}(U)(E^{\text{X}}, z)(A^{\text{X}}) &\approx \ln(\text{mpdf}(U)(E^{\text{X}}\%\{w\}, z)(A^{\text{X}}\%\{w\}))^n \\
&= n \ \ln \text{mpdf}(U)(E^{\text{X}}\%\{w\}, z)(A^{\text{X}}\%\{w\})
\end{aligned}$$

where $k \in \mathbf{Q}_{\geq 0}$, $Z_k = \text{scalar}(k)$ and $n = |V|$.

4.10 Independent cartesian sub-volume

The *effective states* of the *independent histogram* form a *cartesian sub-volume*, that is, $A^{\text{XF}} = \prod\{(A\%\{v\})^{\text{F}} : v \in V\}$. This *volume* is less than the whole *volume* if there are *volumes incident* on the *reduced variables* such that the *incident states* have *zero count*. The union of this set of *zero sub-volumes* forms the complement of the *independent cartesian sub-volume*, $\bigcup\{C^{\text{U}} : v \in V,\ w \in U_v,\ S = \{(v, w)\},\ B = A + A^{\text{CZ}},\ C = B \setminus \text{incidence}(B, S, 0),\ \text{size}(C) = 0\} = A^{\text{C}} \setminus A^{\text{XF}}$. The *alignment* of the *cartesian sub-volume* is the same as that for the whole *volume*, $\text{alignment}(A * A^{\text{XF}}) = \text{alignment}(A)$.

A *unit functional transform* that represents the *cartesian sub-volume slice* having *derived variables* of truncated *valency* can be constructed. The *transformed histogram* has the same *alignment*. Let $P = \{(v, (A\%\{v\})^{\text{FS}}) : v \in V\}$, such that $\forall v \in V\ ((P_v, P_v) \in U)$ and *transform* $T = (\{S \cup \{(P_v, (v, w)) : (v, w) \in S\} : S \in \prod \text{ran}(P)\}^{\text{U}}, \text{ran}(P))$. Then $\text{alignment}(A * T) = \text{alignment}(A)$.

4.11 Mis-alignment

The negative *relative dependent space* of *sample histogram* A *drawn* from *distribution histogram* E equals the *alignment* minus the *mis-alignment*

$$\text{alignment}(A) - \sum_{S \in A^{\text{XS}}} (A_S - A^{\text{X}}_S) \ln E_S$$

where the *sample histogram* is as *complete* as the *independent*, $A^{\text{U}} \geq A^{\text{XU}}$ and the *distribution histogram* is as *effective* as the *independent*, $E^{\text{F}} \geq A^{\text{XU}}$.

In terms of probability density, the negative *relative dependent space* is the negative logarithm *relative dependent multinomial probability density*

$$-\ln \frac{\mathrm{mpdf}(U)(E,z)(A)}{\mathrm{mpdf}(U)(E,z)(A^{\mathrm{X}})} \quad = \quad \mathrm{alignment}(A) - \sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln E_S$$

The *mis-alignment* does depend on the *distribution histogram*, but not its *size*, z_E

$$\forall q \in \mathbf{Q}_{>0}\ (\sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln(Z_q * E)(S) = \sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln E_S)$$

where $Z_q = \mathrm{scalar}(q)$.

The *mis-alignment* is zero if the *sample histogram* is *independent*, $A = A^{\mathrm{X}}$,

$$\sum_{S \in A^{\mathrm{XS}}} (A^{\mathrm{X}}_S - A^{\mathrm{X}}_S) \ln E_S = 0$$

The *mis-alignment* is zero if the *distribution histogram* is *independent*, $E = E^{\mathrm{X}}$,

$$\sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln E^{\mathrm{X}}_S = 0$$

Now consider some examples of the *mis-alignment* for *non-independent distribution histograms*, $E \neq E^{\mathrm{X}}$. First consider the case where the *distribution histogram* E is equal to the *sample histogram* A, $E = A$, and where the *sample* is not *independent*, $A \neq A^{\mathrm{X}}$, and is as *effective* as its *independent* $A^{\mathrm{F}} = A^{\mathrm{XF}}$. In this case the *alignment* of the *distribution distribution* equals the *sample histogram*, $\mathrm{algn}(E) = \mathrm{algn}(A)$, because $E = A$. The *mis-alignment* is approximately greater than or equal to the *alignment* of the *sample histogram*, $\mathrm{algn}(A)$. So the negative *relative dependent space* is approximately less than or equal to zero. Applying Gibbs' inequality and then Stirling's approximation,

$$\begin{aligned}
\sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln A_S &= \sum_{S \in A^{\mathrm{S}}} A_S \ln A_S - \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln A_S \\
&> \sum_{S \in A^{\mathrm{S}}} A_S \ln A_S - \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln A^{\mathrm{X}}_S \\
&\approx \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S \\
&= \mathrm{algn}(A)
\end{aligned}$$

In terms of probability density, the *probability* of drawing *histogram* A from *distribution histogram* A is approximately higher than the *probability* of drawing *histogram* A^{X}

$$-\ln \frac{\mathrm{mpdf}(U)(A,z)(A)}{\mathrm{mpdf}(U)(A,z)(A^{\mathrm{X}})} \leq 0$$

and so

$$\mathrm{mpdf}(U)(A,z)(A) \geq \mathrm{mpdf}(U)(A,z)(A^{\mathrm{X}})$$

When *mis-alignment* is positive the *distribution distribution*, E, is said to be *aligned with* the *sample histogram* A.

A similar case is where the *distribution histogram* is a ratio of the *sample* and the *independent sample*, $E = \mathrm{resize}(z, A/A^{\mathrm{X}})$, where $A \neq A^{\mathrm{X}}$, $A^{\mathrm{F}} \geq A^{\mathrm{XF}}$ and $z = \mathrm{size}(A)$. The *alignment* of the *distribution histogram* is less than the *alignment* of the *sample*, $\mathrm{algn}(E) < \mathrm{algn}(A)$. The *mis-alignment* is positive, but less than the *alignment* of the *sample*, $\mathrm{algn}(A)$,

$$\begin{aligned}\sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln \frac{A_S}{A^{\mathrm{X}}_S} &= \sum_{S \in A^{\mathrm{XS}}} A_S \ln \frac{A_S}{A^{\mathrm{X}}_S} + \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln \frac{A^{\mathrm{X}}_S}{A_S} \\ &> 0\end{aligned}$$

Again the inequality is Gibbs' inequality. The *distribution distribution* A/A^{X} is *aligned with* the *sample histogram* A.

A symmetrical example is where the *distribution histogram* is a scaled ratio of the *independent sample* and the *sample*, $E = \mathrm{resize}(z, A^{\mathrm{X}}/A)$, where $A \neq A^{\mathrm{X}}$ and $A^{\mathrm{F}} \geq A^{\mathrm{XF}}$. The *alignment* of the *distribution histogram* is still less than the *alignment* of the *sample*, $\mathrm{algn}(E) < \mathrm{algn}(A)$, but the *mis-alignment* is negative

$$\begin{aligned}\sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln \frac{A^{\mathrm{X}}_S}{A_S} &= -\sum_{S \in A^{\mathrm{XS}}} A_S \ln \frac{A_S}{A^{\mathrm{X}}_S} - \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln \frac{A^{\mathrm{X}}_S}{A_S} \\ &> 0\end{aligned}$$

The *distribution distribution* A^{X}/A is *aligned against* the *sample histogram* A.

A fourth example of a *distribution histogram* E derived from the *sample histogram* A is where E is the *independent additive complement* defined $E = \mathrm{resize}(z, (A - A^{\mathrm{X}})^{\mathrm{F}} * A^{\mathrm{X}} + (A^{\mathrm{X}} - A)^{\mathrm{F}} * A)$, where $A^{\mathrm{F}} \geq A^{\mathrm{XF}}$. In this case, the *distribution histogram*, E, may be more *aligned* than the *sample*, $\mathrm{algn}(E) > \mathrm{algn}(A)$ but the *mis-alignment* is negative

$$\begin{aligned}
&\sum_{S \in A^{\mathrm{XS}}} A_S \ln E_S - \sum_{S \in A^{\mathrm{S}}} A^{\mathrm{X}}_S \ln E_S \\
&= \sum_{S \in (A^{\mathrm{X}}-A)^{\mathrm{FS}}} A_S \ln A_S + \sum_{S \in (A-A^{\mathrm{XS}})^{\mathrm{F}}} A_S \ln A^{\mathrm{X}}_S - \\
&\quad (\sum_{S \in (A^{\mathrm{X}}-A)^{\mathrm{FS}}} A^{\mathrm{X}}_S \ln A_S + \sum_{S \in (A-A^{\mathrm{X}})^{\mathrm{FS}}} A^{\mathrm{X}}_S \ln A^{\mathrm{X}}_S) \\
&= -\sum_{S \in (A^{\mathrm{X}}-A)^{\mathrm{FS}}} (A^{\mathrm{X}}_S - A_S) \ln A_S - \sum_{S \in (A-A^{\mathrm{X}})^{\mathrm{FS}}} (A_S - A^{\mathrm{X}}_S) \ln A^{\mathrm{X}}_S \\
&< 0
\end{aligned}$$

In this case, the *distribution distribution* E is *aligned against* the *sample histogram* A.

The examples above are of *aligned distribution histograms* that are either *aligned with* the *sample histogram* or *aligned against* the *sample histogram*. Another case is where the *distribution histogram* is *orthogonally aligned* to the *sample histogram*. In this case the *distribution histogram alignment* is non-zero, $E \neq E^{\mathrm{X}}$, and the *sample histogram alignment* is non-zero, $A \neq A^{\mathrm{X}}$, but the *mis-alignment* is zero,

$$\sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln E_S = 0$$

If the *size* of the *distribution histogram* is much larger than the *size* of the *sample histogram*, $z_E \gg z$, and such that each *distribution histogram count* is greater than the corresponding *count* of both the *sample* and *independent sample*, $E > A$ and $E > A^{\mathrm{X}}$, then there is a approximate condition for *orthogonal alignment*. First approximate the *mis-alignment* as an *alignment* *delta*

$$\begin{aligned}
&\sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln E_S \\
&= \sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln E_S - \sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}}_S) \ln E^{\mathrm{X}}_S \\
&\approx \sum_{S \in E^{\mathrm{S}}} (E + A - A^{\mathrm{X}})_S \ln E_S - \sum_{S \in E^{\mathrm{XS}}} (E^{\mathrm{X}} + A - A^{\mathrm{X}})_S \ln E^{\mathrm{X}}_S - \mathrm{algn}(E) \\
&\approx \sum_{S \in E^{\mathrm{S}}} (E + A - A^{\mathrm{X}})_S \ln E_S - \sum_{S \in E^{\mathrm{XS}}} (E + A - A^{\mathrm{X}})^{\mathrm{X}}_S \ln E^{\mathrm{X}}_S - \mathrm{algn}(E) \\
&\approx \sum_{S \in E^{\mathrm{S}}} (E + A - A^{\mathrm{X}})_S \ln (E + A - A^{\mathrm{X}})_S - \\
&\quad \sum_{S \in E^{\mathrm{XS}}} (E + A - A^{\mathrm{X}})^{\mathrm{X}}_S \ln (E + A - A^{\mathrm{X}})^{\mathrm{X}}_S - \mathrm{algn}(E) \\
&\approx \mathrm{algn}(E + A - A^{\mathrm{X}}) - \mathrm{algn}(E)
\end{aligned}$$

The *perturbed distribution histogram*, $E + A - A^{\mathrm{X}}$, is the *distribution histogram*, E, plus the *delta*, (A^{X}, A). The approximate condition for *orthogonal alignment* is thus $\mathrm{algn}(E + A - A^{\mathrm{X}}) \approx \mathrm{algn}(E)$. A symmetrical argument yields a similar approximate condition $\mathrm{algn}(E + A^{\mathrm{X}} - A) \approx \mathrm{algn}(E)$. Together the condition is $\mathrm{algn}(E + A - A^{\mathrm{X}}) - \mathrm{algn}(E + A^{\mathrm{X}} - A) \approx 0$. The appropriate degree of approximation can be guaged by calculating the *alignment delta* of the *perturbed distribution histogram* relative to itself after scaling

$$\mathrm{algn}(E + Z_{z/z_E} * E - Z_{z/z_E} * E^{\mathrm{X}}) - \mathrm{algn}(E + Z_{z/z_E} * E^{\mathrm{X}} - Z_{z/z_E} * E)$$

where $Z_q = \mathrm{scalar}(q)$.

The constraints on E to make the *perturbed distribution histogram alignment delta*, $\mathrm{algn}(E + A - A^{\mathrm{X}}) - \mathrm{algn}(E + A^{\mathrm{X}} - A)$, a reasonable limit on the *orthogonal alignment* condition, are similar to those that make *multinomial distributions* approximations to *historical distributions*. That is, that the *generalised multinomial probability distribution* approximates to the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{m},U}(E, z) \approx \hat{Q}_{\mathrm{h},U}(E, z)$, where $z \ll \mathrm{minr}(E)$, $E \in \mathcal{A}_{\mathrm{i}}$ and $E^{\mathrm{F}} = E^{\mathrm{C}}$.

4.12 Alignment of partially independent

A *histogram* A of *variables* $V = \text{vars}(A)$ is said to be *partially independent* in a partition of the *variables* $P \in \text{B}(V)$ if

$$A = Z_A * \prod_{K \in P} \frac{A}{Z_A} \%K$$

where $Z_A = \text{scalar}(\text{size}(A))$. Conjecture that the *alignment* of the *histogram* equals the sum of the *alignments* of the *reductions*

$$\text{algn}(A) = \sum_{K \in P} \text{algn}(A\%K)$$

The components of the partition are said to be *independent* of eachother. So, for example, given $K, J \in P$ then $\text{algn}(A * \{K^{\text{CS}\{\}\text{T}}, J^{\text{CS}\{\}\text{T}}\}^{\text{T}}) = 0$.

A *histogram* may be *contracted* by removing the largest subset of *independent variables*. Let $\{J\} = \text{mind}(\{(K, |K|) : K \subseteq V,\ A = A\%K * \prod_{w \in V \setminus K} (A/Z_A)\%\{w\}\})$, then $\text{algn}(A) = \text{algn}(A\%J)$. Note that there is exactly one *contraction*. If the *histogram* is *independent*, $A = A^{\text{X}}$, then it *contracts* to a *scalar*, $A\%\emptyset$.

4.13 Alignment of axial reductions

Let *non-scalar non-singleton trimmed histogram* $A = \text{trim}(A)$ have *variables* $V = \text{vars}(A)$. Let P_A be a non-unary partition of the *trimmed histogram*, $P_A \in \text{B}(A)$ and $|P_A| > 1$. Let P_V be a partition of the *variables*, $P_V \in \text{B}(V)$, of the same cardinality, $|P_V| = |P_A|$. Consider the total bijection $Q \in P_A :\leftrightarrow: P_V$. If the map, Q, is (i) such that there exists a *pivot state* $X = \bigcup\{(A * C\ \%\ (V \setminus K))^{\text{S}} : (C, K) \in Q\} \in \text{cartesian}(U)(V)$ for *implied system* $U = \text{implied}(A)$, and (ii) such that all of the *sliced reductions* are *diagonalised*, $\forall (C, K) \in Q\ (\text{diagonal}(A * C\ \%\ K))$, then for all selections $M \in P_A :\leftrightarrow V$ of Q, $\forall C \in P_A\ (M_C \in Q_C)$, there exists a *reduced histogram*

$$B_M = \sum_{(C,w) \in M} (A * C\ \%\ \{w\}) * (X\ \%\ (\text{ran}(M) \setminus \{w\}))$$

which is *axial*, $\text{axial}(B_M)$. Conjecture that for all such selections, M, of such maps, Q, the *alignment* of the *histogram* equals the *alignment* of the *axial reduction* plus the sum of the *alignments* of the *sliced diagonalised reductions*

$$\text{algn}(A) = \text{algn}(B_M) + \sum_{(C,K) \in Q} \text{algn}(A * C\ \%\ K)$$

The *sliced diagonalised reductions*, $\{A * C \ \% \ K : (C,K) \in Q\}$, are said to be *axially independent* of eachother and *axially independent* of the *axial reduced histogram*, B_M, because the *alignments* sum together similarly to the *alignments* of a *partially independent histogram*, see section 'Alignment of partially independent' above. The *pivot state*, X, need not be *effective*, only that it is a member of the *cartesian* set of *states*, $X \in V^{\mathrm{CS}}$. If $X \notin A^{\mathrm{S}}$ then the *axial* is missing the *pivot*.

4.14 Alignment and conditional probability

Consider the *complete integral congruent support sample histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ *drawn with replacement* from *distribution histogram* $E \in \mathcal{A}_{U,V,z_E}$. In the case where the *distribution histogram* E is as *effective* as the *independent*, $E^{\mathrm{F}} \geq A^{\mathrm{XF}}$, then the *generalised multinomial probability* of the *sample histogram*, $\hat{Q}_{\mathrm{m},U}(E,z)(A)$, may be decomposed into (i) the *iso-independents multinomial probability* and (ii) *iso-independent conditional dependent multinomial probability*

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B) \times \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B)}$$

where the *generalised multinomial probability distribution* $\hat{Q}_{\mathrm{m},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is defined

$$\hat{Q}_{\mathrm{m},U}(E,z) := \{(A, \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

and the *integral iso-independent* function, $Y_{U,\mathrm{i},V,z} \in \mathcal{A}_{U,\mathrm{i},V,z} \to \mathcal{A}_{U,V,z}$, is defined

$$Y_{U,\mathrm{i},V,z} = \{(B, B^{\mathrm{X}}) : B \in \mathcal{A}_{U,\mathrm{i},V,z}\} \subset Y_{U,V,z} \subset \text{independent}$$

Compare the *conditional dependent* to the *relative dependent*. In the case of the *relative dependent* the *generalised multinomial probability* is decomposed into (i) the *independent multinomial probability* and (ii) *relative dependent multinomial probability*

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = \hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}}) \times \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})}$$

Unlike in the *relative dependent* case, where the *independent histogram* must be *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, in the *conditional dependent* case there is no need for the *independent histogram* to be *integral* because the *integral iso-independents*, $Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$, is non-empty regardless.

Defined in terms of the *generalised multinomial probability*, the *generalised iso-independent conditional multinomial probability distribution*, $\hat{Q}_{\mathrm{m,y},U}$, is

$$\hat{Q}_{\mathrm{m,y},U}(E,z) \quad = \quad \mathrm{normalise}(\{(A, \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\})$$

So

$$\hat{Q}_{\mathrm{m,y},U}(E,z)(A) = \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|} \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B)}$$

and the *generalised multinomial probability* may be decomposed

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B) \times |\mathrm{ran}(Y_{U,\mathrm{i},V,z})| \hat{Q}_{\mathrm{m,y},U}(E,z)(A)$$

The cardinality of the components of the partition of $\mathcal{A}_{U,\mathrm{i},V,z}$ is the normalisation factor,

$$|\mathrm{ran}(Y_{U,\mathrm{i},V,z})| = \prod_{w \in V} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!}$$

The *relative dependent multinomial probability* equals the *iso-independent conditional dependent multinomial probability* if the *iso-independents* set is a singleton containing the *independent*, for example if the *histogram* is *mono-variate*, $|V| = 1$. In this case, however, the *sample* must be *independent*, $Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) = \{A^{\mathrm{X}}\} \implies A = A^{\mathrm{X}}$, and therefore the *probability* is 1,

$$\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})}{\sum_{B \in \{A^{\mathrm{X}}\}} \hat{Q}_{\mathrm{m},U}(E,z)(B)} = \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})}{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})} = 1$$

and the *generalised iso-independent conditional multinomial probability* does not depend on A^{X}

$$\hat{Q}_{\mathrm{m,y},U}(E,z)(A^{\mathrm{X}}) = \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|}$$

The *iso-independent conditional dependent multinomial probability* is greater than 0 and less than or equal to 1

$$0 < \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B)} \leq 1$$

because

$$0 < \hat{Q}_{\mathrm{m},U}(E,z)(A) \leq \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B) \leq 1$$

The *iso-independent conditional dependent multinomial probability* only equals 1 if the *sample* is *independent*

$$\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B)} = 1 \implies A = A^{\mathrm{X}}$$

because a *non-independent sample* has more than one *integral iso-independents*, $A \neq A^{\mathrm{X}} \implies |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| > 1$.

In some cases the *relative probability* may be greater than one,

$$\exists E, A \in \mathcal{A}\ (\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})} > 1)$$

and hence *relative probability* is not strictly speaking a probability per se. In the *conditional dependent* case, however, the *conditional probability* is always between zero and one, yielding a *probability function*,

$$\{(C, \frac{\hat{Q}_{\mathrm{m},U}(E,z)(C)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B)}) : C \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})\} \in \mathcal{P}$$

Therefore *conditional probability* is a probability proper.

The *iso-independent conditional dependent multinomial probability* may be generalised to a *probability density*. Instead of *drawing* an *integral sample histogram* from the finite *integral congruent support*, $\mathcal{A}_{U,\mathrm{i},V,z}$, the *sample histogram* is *drawn* from the infinite *complete congruent histograms*, $A \in \mathcal{A}_{U,V,z}$. The *iso-independent conditional dependent multinomial probability density* given the infinite *iso-independents* is

$$\frac{\mathrm{mpdf}(U)(E,z)(A)}{\int_{B \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})} \mathrm{mpdf}(U)(E,z)(B)\ dB}$$

which is defined if the *distribution histogram* E is as *effective* as the *independent sample*, $E^{\mathrm{F}} \geq A^{\mathrm{XF}}$.

The *iso-independent conditional dependent multinomial probability density* is greater than 0 and less than or equal to 1

$$0 < \frac{\mathrm{mpdf}(U)(E,z)(A)}{\int_{B \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})} \mathrm{mpdf}(U)(E,z)(B)\ dB} \leq 1$$

because

$$0 < \mathrm{mpdf}(U)(E,z)(A) \leq \int_{B \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})} \mathrm{mpdf}(U)(E,z)(B)\ dB \leq 1$$

The *iso-independent conditional dependent multinomial probability* tends to the *iso-independent conditional dependent multinomial probability density* as the *size* increases

$$\lim_{k \to \infty} \frac{\hat{Q}_{\mathrm{m},U}(E,kz)(Z_k * A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,kz}(Z_k * A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,kz)(B)} = \frac{\mathrm{mpdf}(U)(E,z)(A)}{\int_{B \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})} \mathrm{mpdf}(U)(E,z)(B)\ dB}$$

where $Z_k = \mathrm{scalar}(k)$. This is because (i) either the finite *integral iso-independents* becomes a larger subset of the *iso-independents* as the *size* increases, $Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \subset Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, or (ii) both are singletons, $Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) = Y^{-1}_{U,V,z}(A^{\mathrm{X}}) = \{A^{\mathrm{X}}\}$.

Consider the case where the *distribution histogram* is *independent*, $E = E^{\mathrm{X}}$, as well as sufficiently *effective*, $E^{\mathrm{XF}} \geq A^{\mathrm{XF}}$. The negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability* is

$$\begin{aligned}
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B)} : E^{\mathrm{XF}} \geq A^{\mathrm{XF}}\right) \\
&= -\ln \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A) \ + \ \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B) \\
&= -\ln \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E^{\mathrm{X}}_S}{z_E}\right)^{A_S} \ + \ \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{z!}{\prod_{S \in B^{\mathrm{S}}} B_S!} \prod_{S \in B^{\mathrm{S}}} \left(\frac{E^{\mathrm{X}}_S}{z_E}\right)^{B_S} \\
&= \sum_{S \in A^{\mathrm{S}}} \ln A_S! \ + \ \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!}
\end{aligned}$$

because

$$\forall B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \left(\sum_{S \in B^{\mathrm{S}}} B_S \ln E^{\mathrm{X}}_S = \sum_{S \in B^{\mathrm{XS}}} B^{\mathrm{X}}_S \ln E^{\mathrm{X}}_S = \sum_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S \ln E^{\mathrm{X}}_S = \sum_{S \in A^{\mathrm{S}}} A_S \ln E^{\mathrm{X}}_S\right)$$

As in the case of the negative logarithm *independently-distributed relative dependent multinomial probability density* of the *sample*, which is the *alignment*,

$$\sum_{S \in A^{\mathrm{S}}} \ln A_S! - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S = \mathrm{alignment}(A)$$

the negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability*,

$$\sum_{S \in A^{\mathrm{S}}} \ln A_S! \;+\; \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!}$$

does not depend on the *distribution histogram*, E, so long as the *distribution histogram* is sufficiently *effective*, $E^{\mathrm{F}} \geq A^{\mathrm{XF}}$, and *independent*, $E = E^{\mathrm{X}}$.

Let *integral congruent delta* $(D, I) \in \mathcal{A}_{\mathrm{i}} \times \mathcal{A}_{\mathrm{i}}$ be such that its *perturbation*, $A-D+I$, is *iso-independence conserving*, $A-D+I \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, so that $(A-D+I)^{\mathrm{X}} = A^{\mathrm{X}}$. The change in negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability* because of the *application* of *delta*, (D, I), is

$$\begin{aligned}
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A-D+I)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}((A-D+I)^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)}\right) - \\
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)}\right) \\
= &\left(\sum_{S \in (A-D+I)^{\mathrm{S}}} \ln (A-D+I)_S! \;+\; \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}((A-D+I)^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!}\right) - \\
&\left(\sum_{S \in A^{\mathrm{S}}} \ln A_S! \;+\; \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!}\right) \\
= &\sum_{S \in (A-D+I)^{\mathrm{S}}} \ln (A-D+I)_S! - \sum_{S \in A^{\mathrm{S}}} \ln A_S!
\end{aligned}$$

This difference equals the difference in *alignments*, $\mathrm{algn}(A-D+I) - \mathrm{algn}(A)$, because the *independent perturbation*, $(A-D+I)^{\mathrm{X}}$, and the *independent sample*, A^{X}, are equal.

The *idealisation* of a *histogram* given an *effective transform*, $A * T * T^{\dagger A}$, is in the *iso-independents*, $A * T * T^{\dagger A} \in Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, because the *independent* of the *idealisation* equals the *independent histogram*, $(A * T * T^{\dagger A})^{\mathrm{X}} = A^{\mathrm{X}}$. In the case where the *idealisation* is *integral*, $A * T * T^{\dagger A} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, there is a corresponding *iso-independence conserving delta*, $A * T * T^{\dagger A} = A - D + I$. The change in negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability* because of the *integral idealisation* of the *sample histogram* is

$$\begin{aligned}
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A * T * T^{\dagger A})}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}((A*T*T^{\dagger A})^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)} \right) - \\
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)} \right) \\
= \quad &\left(\sum_{S \in (A*T*T^{\dagger A})^{\mathrm{S}}} \ln (A * T * T^{\dagger A})_S! \; + \; \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}((A*T*T^{\dagger A})^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} \right) - \\
&\left(\sum_{S \in A^{\mathrm{S}}} \ln A_S! \; + \; \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} \right) \\
= \quad &\sum_{S \in (A*T*T^{\dagger A})^{\mathrm{S}}} \ln (A * T * T^{\dagger A})_S! - \sum_{S \in A^{\mathrm{S}}} \ln A_S!
\end{aligned}$$

This difference equals the difference in *alignments*, $\mathrm{algn}(A*T*T^{\dagger A}) - \mathrm{algn}(A)$, because the *independent idealisation*, $(A * T * T^{\dagger A})^{\mathrm{X}}$, and the *independent sample*, A^{X}, are equal.

In the case where the *independent histogram* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, then the *independent histogram* is in the *iso-independents*, $A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and the negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability* can be rearranged in terms of the *alignment*,

$$\sum_{S \in A^{\mathrm{S}}} \ln A_S! + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} = \mathrm{algn}(A) + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}$$

The *minimum alignment conjecture*, defined above in 'Minimum alignment', states that the *alignment* is conjectured to be always greater than or equal to zero where the *independent* is *integral*, $\forall B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})\ (\mathrm{algn}(B) \geq 0)$, and hence

$$\forall B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})\ (\frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!} \leq 1)$$

and so

$$0 \leq \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!} \leq \ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|$$

and so

$$\mathrm{algn}(A) \leq \sum_{S \in A^{\mathrm{S}}} \ln A_S! + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} \leq \mathrm{algn}(A) + \ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|$$

The negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability* where the *independent* is *integral* is such that

$$\begin{aligned}
& \mathrm{algn}(A) \\
\leq\ & \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)} : E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \right) \\
\leq\ & \mathrm{algn}(A) + \ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|
\end{aligned}$$

The negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability* where the *independent* is *integral* equals the *alignment* only if the *sample histogram* is *independent* and the *iso-independents* is a singleton,

$$\sum_{S \in A^{\mathrm{XS}}} \ln A^{\mathrm{X}}_S! + \ln \sum_{B \in \{A^{\mathrm{X}}\}} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} = \mathrm{algn}(A^{\mathrm{X}}) = 0$$

Therefore the *alignment* is always an underestimate of the negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability* where the *independent* is *integral* and the *sample* is *non-independent*

$$A \neq A^{\mathrm{X}} \implies \mathrm{algn}(A) < \sum_{S \in A^{\mathrm{S}}} \ln A_S! \ +\ \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!}$$

The cardinality of the *integral iso-independents* must be less than or equal to the cardinality of the *integral congruent support*,

$$|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| \leq |\mathcal{A}_{U,\mathrm{i},V,z}| = \frac{(z+v-1)!}{z!\ (v-1)!}$$

where $v = |V^{\mathrm{C}}|$. Thus $\ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| < \overline{v} \ln z$ if $z > v$. So

$$\sum_{S \in A^{\mathrm{S}}} \ln A_S! + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} < \mathrm{algn}(A) + \overline{v} \ln z$$

Compare this to *maximum alignment*, $\mathrm{alignmentMaximum}(U)(V, z)$, which for large *size*, $z \gg v$, approximates to $z(n-1)\ln d$ for a *regular histogram* of *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$. Therefore, in some cases the difference between the *alignment* and the negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability* is less than the *alignment*, $\ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| < \overline{v} \ln z < \mathrm{alignment}(A)$. That is, in some cases

$$\mathrm{algn}(A) \leq \sum_{S \in A^{\mathrm{S}}} \ln A_S! \ + \ \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} \leq 2 \times \mathrm{algn}(A)$$

In the case where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, the negative logarithm *independently-distributed relative dependent multinomial probability*, which is the *alignment*, can be expressed in terms of *multinomial coefficients*

$$\mathrm{alignment}(A) = \ln \left(\frac{z!}{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!} \right) - \ln \left(\frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \right)$$

In all cases, the negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability* can be expressed in terms of *multinomial coefficients*

$$\sum_{S \in A^{\mathrm{S}}} \ln A_S! + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} =$$

$$\ln \left(\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{z!}{\prod_{S \in B^{\mathrm{S}}} B_S!} \right) - \ln \left(\frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \right)$$

In the case where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, the negative logarithm *independently-distributed iso-independent conditional dependent multi-*

nomial probability,

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)} : E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \right)$$

$$= \left(\mathrm{algn}(A) + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!} \right) \in \ln \mathbf{Q}_{>0}$$

may be abbreviated to the *alignment-bounded iso-independent space*.

The difference between the *alignment-bounded iso-independent space* and the *alignment* is the *alignment-bounded iso-independent error*

$$\ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}$$

The numerator in the *alignment-bounded iso-transform error* expression is derived from the *independent* term of the *alignment*, $\sum_{S \in A^{\mathrm{XS}}} \ln A^{\mathrm{X}}_S!$, which varies against the *entropy* of the *independent histogram*, $\mathrm{entropy}(A^{\mathrm{X}})$. In the case of *uniform independent histogram* of *size* z and *volume* v where $z/v \in \mathbf{N}_{>0}$, the *independent* term is $v \ln(z/v)! \approx z \ln(z/v)$. So the *alignment-bounded iso-transform error* with respect to the numerator varies with the *size*, z, and varies against the logarithm of the *volume*, $\ln v$. The *independent histogram*, A^{X}, tends to be more *uniform* at higher *alignments*.

The *alignment-bounded iso-independent error* varies with the cardinality of the *integral iso-independents*, $|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|$. As shown above, the average cardinality of the *integral iso-independents* is

$$\frac{|\mathcal{A}_{U,\mathrm{i},V,z}|}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|} = \frac{(z+v-1)!}{z!\ (v-1)!} / \prod_{w \in V} \frac{(z+|U_w|-1)!}{z!\ (|U_w|-1)!}$$

The average cardinality of the *integral iso-independents* varies with both *size*, z, and *volume*, v. Hence the *error* varies with both *size*, z, and *volume*, v.

In the case where the *size* is greater than the *volume*, $z > v$, the logarithm of the average cardinality is less than $\overline{v} \ln z$. In this case the negative contribution to the variation between the *error* and the *volume* from the numerator, $\ln v$, is outweighed by the positive contribution from the summation, $\overline{v}$. Hence, in the case where $z > v$, the *error* varies with both *size*, z, and *volume*, v.

For a given *volume*, v, the average cardinality of the *integral iso-independents* varies with the entropy of the *valencies*, $\mathrm{entropy}(\{(w, |U_w|) : w \in V\})$. Hence the *error* also varies with *valency* entropy. Thus the *error* tends to increase with *dimension*, $n = |V|$. *Regular histograms* tend to have higher *error* than *irregular*.

It is conjectured above that the cardinality of the *integral iso-independents* corresponding to A^{X} varies with the *entropy* of the *independent*, A^{X},

$$\ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| \quad \sim \quad z \times \mathrm{entropy}(A^{\mathrm{X}})$$

Therefore the *alignment-bounded iso-independent error* also varies with the *entropy* of the *independent*, $\mathrm{entropy}(A^{\mathrm{X}})$.

The ratio of the *alignment-bounded iso-independent error* to the *alignment* is

$$\left(\ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!} \right) / \mathrm{algn}(A)$$

where the *histogram* is not *independent*, $A \neq A^{\mathrm{X}} \implies \mathrm{algn}(A) > 0$.

As the *alignment* increases to *maximum alignment*, $\mathrm{alignmentMaximum}(U)(V, z) \approx z \ln v$ where $z \gg v$, the ratio decreases, $\overline{v} \ln z / z \ln v$.

On the other hand, as noted above, the *alignment* approximates to the difference in *entropy* between the *independent* and the *sample histogram*, $\mathrm{algn}(A) \approx z \times \mathrm{entropy}(A^{\mathrm{X}}) - z \times \mathrm{entropy}(A)$. Hence increases in *alignment* imply increases in the *entropy* of the *independent* to some degree. So there is a tendency to increase the ratio of the *alignment-bounded iso-independent error* to the *alignment* at higher *alignments* due to the *independent entropy* which partly counteracts the tendency to decrease the ratio at higher *alignments* due to the *size*.

In the case where the *alignment* is approximately equal to the *expected alignment*, it is conjectured above ('Minimum alignment') that *expected alignment* varies as the *volume* for constant *size* greater than the *volume*

$$\mathrm{expected}(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z))(\{(A, \mathrm{algn}(A)) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}) \sim v$$

So in the case of *expected alignment* the *alignment-bounded iso-independent error* tends to be greater than the *alignment* and the ratio is greater than one, $\overline{v} \ln z/v$.

The change in *alignment-bounded iso-independent space* because of the *application* of *iso-independence conserving delta*, $A - D + I \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, is equal to the change in *alignment*

$$\begin{aligned}
&\left(\mathrm{algn}(A-D+I) + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}((A-D+I)^{\mathrm{X}})} \frac{\prod_{S \in (A-D+I)^{\mathrm{XS}}} (A-D+I)^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}\right) - \\
&\left(\mathrm{algn}(A) + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}\right) \\
= \quad & \mathrm{algn}(A-D+I) - \mathrm{algn}(A)
\end{aligned}$$

A special case is the *integral idealisation*, $A * T * T^{\dagger A} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, where the change in *alignment-bounded iso-independent space* because of the *integral idealisation* of the *sample histogram* is

$$\begin{aligned}
&\left(\mathrm{algn}(A * T * T^{\dagger A}) + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}((A * T * T^{\dagger A})^{\mathrm{X}})} \frac{\prod_{S \in (A * T * T^{\dagger A})^{\mathrm{XS}}} (A * T * T^{\dagger A})^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}\right) - \\
&\left(\mathrm{algn}(A) + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}\right) \\
= \quad & \mathrm{algn}(A * T * T^{\dagger A}) - \mathrm{algn}(A)
\end{aligned}$$

Similarly, consider the *integral idealisations* of two *transforms* T_1 and T_2, where $A * T_1 * T_1^{\dagger A},\ A * T_2 * T_2^{\dagger A} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$. In this case the change in *alignment-bounded iso-independent space* between the two *integral idealisations* of the *sample histogram* is

$$\mathrm{algn}(A * T_2 * T_2^{\dagger A}) - \mathrm{algn}(A * T_1 * T_1^{\dagger A})$$

4.15 Transform alignment

Let the set $\mathcal{O}_{U,z} \subset \mathcal{A}_U \times \mathcal{T}_{U,\mathrm{f},1}$ be the set of pairs of (i) *histograms* of non-zero *size* $z > 0$ such that the *independent histogram* is *completely effective* and (ii) *one functional transforms* having *underlying variables* equal to the *histogram variables*

$$\mathcal{O}_{U,z} = \{(A,T) : A \in \mathcal{A}_U,\ \text{size}(A) = z,\ A^{\text{XF}} = A^{\text{C}},\ T \in \mathcal{T}_{U,\text{f},1},\ \text{und}(T) = \text{vars}(A)\}$$

So $\forall A \in \text{dom}(\mathcal{O}_{U,z})\ (\text{size}(A) = z)$, $\forall A \in \text{dom}(\mathcal{O}_{U,z})\ (A^{\text{XF}} = A^{\text{C}})$ and $\forall (A,T) \in \mathcal{O}_{U,z}\ (\text{und}(T) = \text{vars}(A))$. For any given $A \in \text{dom}(\mathcal{O}_{U,z})$ of *variables* $V = \text{vars}(A)$ the set $\{T : (B,T) \in \mathcal{O}_{U,z},\ B = A\}$ is a superset of the finite *substrate transforms set* in V, $\mathcal{T}_{U,V} \subset \text{ran}(\text{filter}(\{A\}, \mathcal{O}_{U,z}))$.

Let $(A,T) \in \mathcal{O}_{U,z}$. The application of the *transform* T to its corresponding *histogram* A is called the *derived histogram* $A*T$. In this context, A is called the *underlying histogram*. The application to create the *derived histogram* is *size conserving*, $\forall (A,T) \in \mathcal{O}_{U,z}\ (\text{size}(A*T) = z)$. If the *transform* is not *right total*, $|T^{-1}| < |W^{\text{C}}|$, then the *derived histogram* is always *incompletely effective*, $(X\%W)^{\text{F}} < W^{\text{C}} \implies (A*T)^{\text{F}} < (A*T)^{\text{C}}$, where $(X,W) = T$. If the *transform* is *right total*, then the *volume* of the *derived histogram* must be less than or equal to that of the *underlying*, $(X\%W)^{\text{F}} = W^{\text{C}} \implies |V^{\text{C}}| \geq |W^{\text{C}}|$ and so $|A^{\text{C}}| \geq |(A*T)^{\text{C}}|$, where $(X,W) = T$ and $V = \text{und}(T)$.

The *idealisation* and the *neutralisation* are both *size conserving*, $\forall (A,T) \in \mathcal{O}_{U,z}\ (\text{size}(A*T*T^{\dagger A}) = z)$ and $\forall (A,T) \in \mathcal{O}_{U,z}\ (\text{size}(A*T*T^{\odot A^{\text{X}}}) = z)$. The *idealisation* and the *neutralisation* are both at least as *effective* as the *underlying*, $(A*T*T^{\dagger A})^{\text{F}} \geq A^{\text{F}}$ and $(A*T*T^{\odot A^{\text{X}}})^{\text{F}} \geq A^{\text{F}}$. A *surrealisation* is *size-conserving* if the *derived histogram* is as *effective* as the *abstract histogram*, $\forall (A,T) \in \mathcal{O}_{U,z}\ ((A*T)^{\text{F}} = (A*T)^{\text{XF}} \implies \text{size}((A*T)^{\text{X}}*T^{\odot A}) = z)$. Otherwise the *size* of the *surrealisation* is less than the *size* of the *histogram*, $\forall (A,T) \in \mathcal{O}_{U,z}\ ((A*T)^{\text{F}} < (A*T)^{\text{XF}} \implies \text{size}((A*T)^{\text{X}}*T^{\odot A}) < z)$. A *contentisation* is *size-conserving* if the *derived histogram* is as *effective* as the *formal histogram*, $\forall (A,T) \in \mathcal{O}_{U,z}\ ((A*T)^{\text{F}} \geq (A^{\text{X}}*T)^{\text{F}} \implies \text{size}(A^{\text{X}}*T*T^{\odot A}) = z)$. Otherwise the *size* of the *contentisation* is less than the *size* of the *histogram*, $\forall (A,T) \in \mathcal{O}_{U,z}\ ((A*T)^{\text{F}} < (A^{\text{X}}*T)^{\text{F}} \implies \text{size}(A^{\text{X}}*T*T^{\odot A}) < z)$. The *surrealisation* and the *contentisation* are no more *effective* than the *underlying*, $((A*T)^{\text{X}}*T^{\odot A})^{\text{F}} \leq A^{\text{F}}$ and $(A^{\text{X}}*T*T^{\odot A})^{\text{F}} \leq A^{\text{F}}$.

The *alignment* of the *derived histogram* is called the *derived alignment*,

$$\text{algn}(A*T)$$

where $(A,T) \in \mathcal{O}_{U,z}$ and algn = alignment.

The *alignment* of the *underlying histogram* is called the *underlying alignment*,

$$\mathrm{algn}(A)$$

The *formal alignment* is the *derived alignment* of the *independent histogram*,

$$\mathrm{algn}(A^{\mathrm{X}} * T)$$

The *derived alignment* relative to the *formal alignment* is called the *content alignment*

$$\mathrm{algn}(A * T) - \mathrm{algn}(A^{\mathrm{X}} * T)$$

The *idealisation alignment* is

$$\mathrm{algn}(A * T * T^{\dagger A})$$

The *surrealisation alignment* is

$$\mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})$$

The *midisation pseudo-alignment* is the *histogram alignment* less the *surrealisation alignment* less the *idealisation alignment*

$$\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A})$$

The *derived variables* of a *transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ are *non-overlapping* if there exists an *equivalent transform* of a *fud* $F \in \mathcal{F}_{U,1}$ which is *non-overlapping*, $\exists F \in \mathcal{F}_{U,1}\ ((F^{\mathrm{T}} = T) \wedge \neg\mathrm{overlap}(F))$. There exists *histogram-transform* pairs $(A,T) \in \mathcal{O}_{U,z}$ for which the *transform*, T, is *non-overlapping*, $\exists (A,T) \in \mathcal{O}_{U,z}\ (\neg\mathrm{overlap}(T))$.

Consider the *histogram-transform* pair $(A,T) \in \mathcal{O}_{U,z}$. The *transform* T is *effectively non-overlapping* with respect to *histogram* A if there exists an equivalent *fud* $F \in \mathcal{F}_{U,1}$ which is *non-overlapping* in the application to the *effective histogram*, $\exists F \in \mathcal{F}_{U,1}\ ((A^{\mathrm{F}} * F^{\mathrm{T}} = A^{\mathrm{F}} * T) \wedge \neg\mathrm{overlap}(F))$. A *transform* T that is *overlapping*, but is *effectively non-overlapping* with respect to *histogram* A, must be *effectively overlapping* with respect to the *independent underlying* A^{X} because the *independent underlying* is *completely effective* $A^{\mathrm{XF}} = A^{\mathrm{C}}$.

Define the *degree of overlap* of $T \in \mathcal{T}_{U,\mathrm{f},1}$ as $\mathrm{algn}(V_z^{\mathrm{C}} * T)$ relative to *size* z where $V = \mathrm{und}(T)$ and the *scaled cartesian* is $V_z^{\mathrm{C}} = \mathrm{resize}(z, V^{\mathrm{C}})$. Define $\mathrm{alignmentOverlap}(U) \in \mathcal{T}_{U,\mathrm{f},1} \times \mathbf{Q}_{>0} \to \mathbf{R}$ as

$$\mathrm{alignmentOverlap}(U)(T,z) := \mathrm{algn}(V_z^{\mathrm{C}} * T)$$

If a *transform* T is *non-overlapping* its application to a *complete uniform histogram* leaves the *derived variables independent* of each other, and so the *derived alignment* is zero and the *degree of overlap* is zero, $\neg\text{overlap}(T) \implies V^{\text{C}} * T = (V^{\text{C}} * T)^{\text{X}}$ and hence $\text{alignmentOverlap}(U)(T, z) = 0$.

Derived variables $x, y \in W$ are said to be *tautological* if their *partitions* are equal, $\text{partition}((X\%(V \cup \{x\}), \{x\})) = \text{partition}((X\%(V \cup \{y\}), \{y\}))$, where $(X, W) = T$ and $V = \text{und}(T)$. A *transform* is *tautologically overlapped* if all of its *derived variables* are *tautological*. A *tautology* is always *overlapped*, $\forall T \in \mathcal{T}_{\text{f}} \cap \mathcal{T}_{\text{U}}\ (\text{tautology}(T) \implies \text{overlap}(T))$. A *tautologically overlapped transform* has a high *degree of overlap*, $\text{algn}(V_z^{\text{C}} * T)$, because $V^{\text{C}} * T$ is *regular* and *fully diagonalised* in the *underlying cartesian*, $\text{diagonalFull}(U)(V^{\text{C}} * T)$. None of the *transforms* in the *substrate transforms set*, $\mathcal{T}_{U,V}$, contain *tautologically derived variables* because each *derived variable* corresponds to a different *partition*, $\forall (X, W) \in \mathcal{T}_{U,V}\ (|\{\text{partition}((X\%(V \cup \{w\}), \{w\})) : w \in W\}| = |W|)$. Thus none are *tautologically overlapped* $\forall T \in \mathcal{T}_{U,V}\ (\neg\text{tautology}(T))$.

There is a case of a *tautologically overlapping transform* that has *maximum degree of overlap*, $\text{algnOver}(U)(T, z) = \text{algnMax}(U)(W, z)$ where $T \in \mathcal{T}_{U,\text{f},1}$, $W = \text{der}(T)$, $\text{algnOver} = \text{alignmentOverlap}$, $\text{algnMax} = \text{alignmentMaximum}$, and T is such that $\text{tautology}(T)$. This is the case where the *derived variables* are each *frames* of the same *underlying variable* v. For example, let $V = \{v\}$, and $\forall w \in W\ (U_w = U_v)$ and $\text{his}(T) = \{\{(v, u)\} \cup \{(w, u) : w \in W\} : u \in U_v\}^{\text{U}}$. All the *derived variables* are *tautologically aligned*. The *degree of overlap* of T is equal to *maximum alignment*, $\text{algn}(\text{resize}(z, \{v\}^{\text{C}}) * T) = \text{algnMax}(U)(W, z)$. This example is of a *literal frame* where the *values* are shared, $U_w = U_v$, but *non-literal frames* are also *tautologically aligned*. The *non-literal frame transforms* can be constructed using bijective maps, $\forall w \in W\ \exists M_w \in U_v \leftrightarrow U_w\ (|M_w| = |U_v|)$ and $\text{his}(T) = \{\{(v, u)\} \cup \{(w, M_w(u)) : w \in W\} : u \in U_v\}^{\text{U}}$.

Another case of *tautologically overlapping transform* $T \in \mathcal{T}_{U,\text{f},1}$ that has *maximum degree of overlap* is such that each of the *tautologically aligned derived variables* enumerates the *cartesian underlying states* and are therefore *self-partitions*. For example, $\forall w \in W\ (U_w = \{\{S\} : S \in V^{\text{CS}}\})$ and $\text{his}(T) = \{S \cup \{(w, \{S\}) : w \in W\} : S \in V^{\text{CS}}\}^{\text{U}}$ where $W = \text{der}(T)$ and $V = \text{und}(T)$. The *degree of overlap* of T is equal to *maximum alignment*, $\text{algnOver}(U)(T, z) = \text{algnMax}(U)(W, z)$. This approximately scales the *underlying maximum alignment* $(m-1)n/(n-1) \times \text{algnMax}(U)(V, z)$ where $m = |W|$ and $n = |V|$ and the *underlying histogram* is *regular*,

$\mathrm{algnMax}(U)(V,z) \approx z(n-1)\ln d$, where d is the *regular valency*, $|V^{\mathrm{CS}}| = d^n$.

Consider the *histogram-transform* pair (A,T) where $A \in \mathcal{A}_U$, $T \in \mathcal{T}_{U,\mathrm{f},1}$, $\mathrm{und}(T) = \mathrm{vars}(A)$ and the T is *tautologically aligned*, $\mathrm{algnOver}(U)(T,z) = \mathrm{algnMax}(U)(W,z)$. If the *derived alignment* is less than *maximum alignment*, $\mathrm{algn}(A*T) < \mathrm{algnMax}(U)(W,z)$, then the *content alignment* may be negative, $\mathrm{algn}(A^{\mathrm{X}}*T) > \mathrm{algn}(A*T)$. A^{X} is more *uniform* than A in the sense that the *entropy* of A^{X} is approximately greater than or equal to the *entropy* of A, $\mathrm{entropy}(A^{\mathrm{X}}) \geq \mathrm{entropy}(A)$, if the *minimum alignment conjecture* is true. Hence the *diagonalised formal histogram*, $\mathrm{diagonal}(A^{\mathrm{X}}*T)$, is sometimes more *uniform* along the *diagonal* than the *derived histogram*, $\mathrm{entropy}(A^{\mathrm{X}}*T\%\{w\}) \geq \mathrm{entropy}(A*T\%\{w\})$ where $w \in W$. Therefore the *formal alignment* $\mathrm{algn}(A^{\mathrm{X}}*T)$ may be closer to *maximum alignment* $\mathrm{algnMax}(U)(W,z)$ than the *derived alignment* $\mathrm{algn}(A*T)$.

There exist *histogram-transform* pairs, $(A,T) \in \mathcal{O}_{U,z}$, such that the *derived alignment* equals the *underlying alignment*, $\mathrm{algn}(A*T) = \mathrm{algn}(A)$. An example is the *value full functional transform* T for which the *derived histogram* is a *non-literal reframe* of the *underlying histogram*. That is, let $M \in (V \leftrightarrow W) \to (\mathcal{U} \leftrightarrow \mathcal{U})$, such that $\mathrm{dom}(M) \in V \cdot W$ and $\forall (v,w) \in \mathrm{dom}(M)\ (M_{v,w} \in U_v \cdot U_w)$ and $\mathrm{his}(T) = \{S \cup \{(w, M_{v,w}(u)) : (v,u) \in S,\ w = \mathrm{dom}(M)(v)\} : S \in V^{\mathrm{CS}}\}^{\mathrm{U}}$, where $V = \mathrm{und}(T)$ and $W = \mathrm{der}(T)$. For example, the *self non-overlapping substrate self-cartesian value full functional transform*, $T = \{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}}$. The *content alignment* of the *non-overlapping transform* also equals the *alignment* of the *underlying histogram*, $\mathrm{algn}(A*T) - \mathrm{algn}(A^{\mathrm{X}}*T) = \mathrm{algn}(A)$ because $\mathrm{algn}(A^{\mathrm{X}}*T) = \mathrm{algn}(A^{\mathrm{X}}) = 0$.

A *derived histogram* can be *independent* even though the *underlying histogram* is not, $\exists (A,T) \in \mathcal{O}_{U,z}\ ((A*T = (A*T)^{\mathrm{X}}) \wedge (A \neq A^{\mathrm{X}}))$. This implies that the *derived alignment* is sometimes less than the *underlying alignment*, $\exists (A,T) \in \mathcal{O}_{U,z}\ (\mathrm{algn}(A*T) < \mathrm{algn}(A))$. Examples of *independent derived histograms* include *singletons*, *mono-variate* and *effectively mono-valent*. If the *transform* is a *unary partition transform*, $|\mathrm{inverse}(T)| = 1$, for example $T = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, then the *derived histogram* is necessarily a *singleton*. *Transforms* having one *derived variable*, $|\mathrm{der}(T)| = 1$, imply a *mono-variate derived histogram*.

A *transform* that is *non-overlapping*,$\neg\mathrm{overlap}(T)$, and such that the *derived variables* partition the *underlying variables* of a *partially independent underlying histogram*, must be *independent*, $\mathrm{algn}(A*T) = 0$. The *independent* *underlying histogram* A^{X} is a *partially independent histogram* by definition and so it follows that for *non-overlapping transforms* the *formal histogram* is *independent*, $\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$, hence *formal alignment* is zero, $\neg\mathrm{overlap}(T) \implies \mathrm{algn}(A^{\mathrm{X}} * T) = 0$.

If a *transform* is *non-overlapping*, $\neg\mathrm{overlap}(T)$, then it has an *equivalent fud*, $F^{\mathrm{T}} = T$, such that the *transforms* have a single *derived variable*, $\forall R \in F\ (|\mathrm{der}(R)| = 1)$, and the *underlying* of the *transforms* partition the *underlying variables* V, $\{\mathrm{und}(\mathrm{depends}(F, \{w\})) : w \in \mathrm{der}(T)\} \in \mathrm{B}(\mathrm{und}(T))$. Therefore the *derived dimension* must be less than or equal to the *underlying dimension*, $\forall T \in \mathcal{T}_{U,V}\ (\neg\mathrm{overlap}(T) \implies |W| \leq |V|)$ where $V = \mathrm{und}(T)$ and $W = \mathrm{der}(T)$. Each of the *transforms* is *one functional*, $F \subset \mathcal{T}_{U,\mathrm{f},1}$, and each is *right total* because the *derived variables* are *partition variables* and hence the *transform* T must be *right total*, $\forall T \in \mathcal{T}_{U,V}\ (\neg\mathrm{overlap}(T) \implies (X\%W)^{\mathrm{F}} = W^{\mathrm{C}})$ where $(X, W) = T$. Therefore the *derived volume* must be less than or equal to the *underlying volume*, $\forall T \in \mathcal{T}_{U,V}\ (\neg\mathrm{overlap}(T) \implies |W^{\mathrm{C}}| \leq |V^{\mathrm{C}}|)$.

If a *histogram* $A \in \mathrm{dom}(\mathcal{O}_{U,z})$ is *irregular*, $|\{|U_v| : v \in V\}| > 1$ where $V = \mathrm{vars}(A)$, then it must be less than *maximally aligned*, $\mathrm{algn}(A) < \mathrm{algnMax}(U)(V, z)$. The *independent histogram* A^{X} is constrained to be *completely effective*, $A^{\mathrm{XF}} = A^{\mathrm{C}}$ and so A cannot be *diagonalised*.

The *alignment* of a *derived histogram* may be less than the *alignment* of the *underlying histogram* even when neither is *independent*, $\exists (A, T) \in \mathcal{O}_{U,z}\ (((A * T) \neq (A * T)^{\mathrm{X}}) \wedge (A \neq A^{\mathrm{X}}) \wedge \mathrm{algn}(A * T) < \mathrm{algn}(A))$. This can be shown to be true by first showing that if a *transform* is *non-overlapping*, then the *maximum alignment* of the *derived histogram* must be less than or equal to the *maximum underlying alignment*, $\neg\mathrm{overlap}(T) \implies \mathrm{algnMax}(U)(W, z) \leq \mathrm{algnMax}(U)(V, z)$ where $V = \mathrm{und}(T)$ and $W = \mathrm{der}(T)$. The *derived volume* must be less than or equal to the *underlying volume* $|W^{\mathrm{C}}| \leq |V^{\mathrm{C}}|$. If an *underlying histogram* is *maximally aligned*, $\mathrm{algn}(A) = \mathrm{algnMax}(U)(V, z)$, then it must be *regular* because the *independent histogram* A^{X} is constrained to be *completely effective*, $A \in \mathrm{dom}(\mathcal{O}_{U,z})$. The largest *maximum derived alignment* occurs when the *derived histogram* is *regular*. Let d_V be the *underlying valency* such that $d_V^{|V|} = |V^{\mathrm{C}}|$. Let d_W be the *derived valency* such that $d_W^{|W|} = |W^{\mathrm{C}}|$. Approximate the *maximal alignments*, $\mathrm{algnMax}(U)(V, z) \approx z(|V| - 1) \ln d_V$ and $\mathrm{algnMax}(U)(W, z) \approx z(|W| - 1) \ln d_W$. Then $|W^{\mathrm{C}}| \leq |V^{\mathrm{C}}| \implies d_W^{|W|} \leq d_V^{|V|} \implies |W| \ln d_W \leq |V| \ln d_V$. But $|W| \leq |V|$ hence $(|W| - 1) \ln d_W \leq (|V| - 1) \ln d_V \implies \mathrm{algnMax}(U)(W, z) \leq \mathrm{algnMax}(U)(V, z)$. If both A and $A*T$ are *maximally aligned* and $|W^{\mathrm{C}}| < |V^{\mathrm{C}}|$ then $\mathrm{algn}(A*T) < \mathrm{algn}(A)$.

If the *transform* is not *right total*, the *derived histogram* is *pluri-variate* and each of the *reductions* of the *derived histogram* is *complete*, then the *derived histogram* must be *aligned*, $((X\%W)^{\mathrm{F}} < W^{\mathrm{C}}) \wedge (|W| > 1) \wedge (\forall w \in W\ (((A * T)\%\{w\})^{\mathrm{F}} = \{w\}^{\mathrm{C}}) \implies \mathrm{algn}(A * T) > 0$, where $(X, W) = T$. This is the case even if the *underlying histogram* is *independent*, $((X\%W)^{\mathrm{F}} < W^{\mathrm{C}}) \wedge (|W| > 1) \wedge (\forall w \in W\ (((A^{\mathrm{X}}*T)\%\{w\})^{\mathrm{F}} = \{w\}^{\mathrm{C}}) \implies \mathrm{algn}(A^{\mathrm{X}}*T) > 0$.

The *derived alignment*, $\mathrm{algn}(A*T)$, of a *histogram-transform* pair $(A, T) \in \mathcal{O}_{U,z}$ constructed from a partition of the *variables* $Y \in \mathrm{B}(V)$, where $V = \mathrm{und}(T) = \mathrm{vars}(A)$, such that the *derived variables* map to components and the *values* enumerate the *cartesian states* of the components, $\forall w \in W\ \exists K \in Y\ (|U_w| = |K^{\mathrm{C}}|)$ where $W = \mathrm{der}(T)$, is less than or equal to the *underlying alignment*, $\mathrm{algn}(A * T) \leq \mathrm{algn}(A)$. For example, $\forall Y \in \mathrm{B}(V)\ (\mathrm{algn}(A * \{K^{\mathrm{CS}\{\}} : K \in Y\}^{\mathrm{T}}) \leq \mathrm{algn}(A))$.

This may be shown for the partition of the *variables* $Y = \{K, V \setminus K\}$ and *transform* $T = \{K^{\mathrm{CS}\{\}}, (V \setminus K)^{\mathrm{CS}\{\}}\}^{\mathrm{T}}$. The entropy of the *partially independent histogram* corresponding to the *independent* term of the *alignment* is such that

$$\begin{aligned}
\sum_{S \in W^{\mathrm{XS}}} \ln \Gamma_!(A * T)^{\mathrm{X}}_S &= \sum_{S \in V^{\mathrm{XS}}} \ln \Gamma_!(\frac{1}{Z_A} * (A\%K) * (A\%(V \setminus K)))_S \\
&\geq \sum_{S \in V^{\mathrm{XS}}} \ln \Gamma_!(\frac{1}{Z_A} * (A\%K) * (A\%(V \setminus K))^{\mathrm{X}})_S \\
&\geq \sum_{S \in V^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S
\end{aligned}$$

The *dependent* term of the *alignment* is unchanged

$$\sum_{S \in W^{\mathrm{XS}}} \ln \Gamma_!(A * T)_S = \sum_{S \in V^{\mathrm{XS}}} \ln \Gamma_! A_S$$

Hence the upper bound of the *underlying alignment*,

$$\mathrm{algn}(A * T) \leq \mathrm{algn}(A)$$

Similarly, the *derived alignment* function and the parent partition relation are monotonic,

$$\forall Y_1, Y_2 \in \mathrm{B}(V)\ (\mathrm{parent}(Y_1, Y_2) \implies \\ \mathrm{algn}(A * \{K^{\mathrm{CS}\{\}} : K \in Y_1\}^{\mathrm{T}}) \leq \mathrm{algn}(A * \{K^{\mathrm{CS}\{\}} : K \in Y_2\}^{\mathrm{T}}))$$

The *derived alignment*, $\mathrm{algn}(A * T)$, and the *derived alignment valency-density*, $\mathrm{algn}(A * T)/\mathrm{capacityValency}(U)((A * T)^{\mathrm{FS}}) = \mathrm{algn}(A * T)/w^{1/m}$, where *derived variables* $W = \mathrm{der}(T)$, *derived dimension* $m = |W| = |Y|$, *volume* $w = |W^{\mathrm{C}}| = |V^{\mathrm{C}}|$, and the *valency capacity* is defined in section 'Capacity and Alignment density', above, are therefore also monotonic. For example, $\forall Y \in \mathrm{B}(V)\ \lozenge T = \{K^{\mathrm{CS}\{\}} : K \in Y\}^{\mathrm{T}}\ \lozenge W = \mathrm{der}(T)\ (\mathrm{algn}(A * T)/|W^{\mathrm{C}}|^{1/|W|} \leq \mathrm{algn}(A)/|V^{\mathrm{C}}|^{1/|V|})$. This is because for any *derived dimension*, m, the *valency capacity* is constant

$$(\prod_{K \in Y} |K^{\mathrm{C}}|)^{1/m} = |V^{\mathrm{C}}|^{1/m} = w^{1/m}$$

and the *valency capacity* of a parent partition of the *variables* is greater than that of the child partition where the *volume* is greater than one, $|V^{\mathrm{C}}| > 1$,

$$\forall Y_1, Y_2 \in \mathrm{B}(V)\ (\mathrm{parent}(Y_1, Y_2) \implies |V^{\mathrm{C}}|^{1/|Y_1|} > |V^{\mathrm{C}}|^{1/|Y_2|})$$

Conjecture that the *alignments* of the *abstract converse actions*, which depend on the *derived histogram*, $A * T$, and the *abstract histogram*, $(A * T)^{\mathrm{X}}$, are constrained to be less than or equal to the *alignment* of the *histogram*, in the case where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, given the *minimum alignment conjecture*. Conjecture that the *idealisation alignment* is always less than or equal to the *alignment* of the *histogram*, where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$,

$$\forall (A, T) \in \mathcal{O}_{U,z}\ (A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies \mathrm{algn}(A * T * T^{\dagger A}) \leq \mathrm{algn}(A))$$

The *non-idealisation alignment* is defined as the difference, $\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A})$.

Conjecture that the *surrealisation alignment* is always less than or equal to the *alignment* of the *histogram*, where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$,

$$\forall (A, T) \in \mathcal{O}_{U,z}\ (A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) \leq \mathrm{algn}(A))$$

This is the case whether the *surrealisation* is *effective* or not, $(A * T)^{\mathrm{F}} \leq (A * T)^{\mathrm{XF}}$. If it is *ineffective*, $(A * T)^{\mathrm{F}} < (A * T)^{\mathrm{XF}}$, then the *size* of the *surrealisation* is less, $\text{size}((A * T)^{\text{X}} * T^{\odot A}) < z$, and the *alignment* is correspondingly less. The *non-surrealisation alignment* is defined as the difference, $\text{algn}(A) - \text{algn}((A * T)^{\text{X}} * T^{\odot A})$.

If the *abstract converse actions* are *integral* then the *alignments* must be positive, given the *minimum alignment conjecture*

$$\forall (A, T) \in \mathcal{O}_{U,z}\ (A * T * T^{\dagger A} \in \mathcal{A}_{\text{i}} \implies 0 \leq \text{algn}(A * T * T^{\dagger A}))$$

and

$$\forall (A, T) \in \mathcal{O}_{U,z}\ ((A * T)^{\text{X}} * T^{\odot A} \in \mathcal{A}_{\text{i}} \implies 0 \leq \text{algn}((A * T)^{\text{X}} * T^{\odot A}))$$

Consider the *histogram-transform* pair $(A, T) \in \mathcal{O}_{U,z}$ where the *derived histogram* is as *effective* as the *formal histogram*, $(A * T)^{\text{F}} \geq (A^{\text{X}} * T)^{\text{F}}$, so that the *contentisation* is *size-conserving*, $\text{size}(A^{\text{X}} * T * T^{\odot A}) = \text{size}(A)$.

In section 'Likely histograms', above, it is conjectured that the *midisation entropy* varies as the *entropy* of the *histogram* less the *entropies* of the *liftisation* and the *surrealisation*,

$$\begin{aligned} \text{entropy}(A^{\text{M}(T)}) \ \sim\ \text{entropy}(A) &- \text{entropy}(A^{\text{K}(T)}) \\ &- \text{entropy}((A * T)^{\text{X}} * T^{\odot A}) \end{aligned}$$

and that insofar as the *idealisation entropy* approximates to the *liftisation entropy*, $\text{entropy}(A*T*T^{\dagger A}) \approx \text{entropy}(A^{\text{K}(T)})$, the *midisation entropy* varies as the *histogram entropy* less the *entropies* of the *idealisation* and the *surrealisation*,

$$\begin{aligned} \text{entropy}(A^{\text{M}(T)}) \ \sim\ \text{entropy}(A) &- \text{entropy}(A * T * T^{\dagger A}) \\ &- \text{entropy}((A * T)^{\text{X}} * T^{\odot A}) \end{aligned}$$

Alignment is approximately equal to the scaled difference between the *independent entropy* and the *histogram entropy*,

$$\text{algn}(A) \ \approx\ z \times \text{entropy}(A^{\text{X}}) - z \times \text{entropy}(A)$$

where $z = \text{size}(A)$. So, insofar as the *midisation independent entropy* approximates to the *independent entropy*, $\text{entropy}(A^{\text{M}(T)\text{X}}) \approx \text{entropy}(A^{\text{X}})$, and the *surrealisation independent entropy* approximates to the *independent entropy*, $\text{entropy}(((A * T)^{\text{X}} * T^{\odot A})^{\text{X}}) \approx \text{entropy}(A^{\text{X}})$, the *midisation alignment* varies as the *histogram alignment* less the *alignments* of the *idealisation* and the *surrealisation*,

$$\mathrm{algn}(A^{\mathrm{M}(T)}) \quad \sim \quad \mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})$$

The computable right hand side is called the *midisation pseudo-alignment* to distinguish it from the usually incomputable *midisation alignment*.

The *midisation pseudo-alignment* is not necessarily positive, but it is always greater than or equal to the greater of $-\mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})$ and $-\mathrm{algn}(A * T * T^{\dagger A})$, because $\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) \geq 0$ and $\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) \geq 0$. If the *abstract converse actions* are *integral*, $A * T * T^{\dagger A} \in \mathcal{A}_{\mathrm{i}}$ and $(A * T)^{\mathrm{X}} * T^{\odot A} \in \mathcal{A}_{\mathrm{i}}$, then *midisation pseudo-alignment* must be less than or equal to the *histogram alignment*, $\mathrm{algn}(A)$.

If the *transform* is a *unary partition transform* $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$ then the *idealisation* equals the *independent*, $A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A} \equiv A^{\mathrm{X}}$, and the *surrealisation* equals the *histogram*, $(A * T_{\mathrm{u}})^{\mathrm{X}} * T_{\mathrm{u}}^{\odot A} \equiv A$, so the *midisation pseudo-alignment* is zero, $\mathrm{algn}(A) - \mathrm{algn}((A * T_{\mathrm{u}})^{\mathrm{X}} * T_{\mathrm{u}}^{\odot A}) - \mathrm{algn}(A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A}) = 0$.

If the *transform* is a *full functional transform*, for example a *value full functional transform* $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}}$, then the *idealisation* equals the *histogram*, $A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger A} \equiv A$, and the *surrealisation* equals the *independent* if the *histogram* is as *effective* as the *independent*, $A^{\mathrm{F}} = A^{\mathrm{XF}} \implies (A * T_{\mathrm{s}})^{\mathrm{X}} * T_{\mathrm{s}}^{\odot A} \equiv A^{\mathrm{X}}$, so the *midisation pseudo-alignment* is zero, $\mathrm{algn}(A) - \mathrm{algn}((A * T_{\mathrm{s}})^{\mathrm{X}} * T_{\mathrm{s}}^{\odot A}) - \mathrm{algn}(A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger A}) = 0$.

The *midisation alignment* may also be expressed in terms of the *contentisation alignment*. In section 'Likely histograms', above, it is conjectured that the *midisation entropy* varies as the difference between the *entropies* of the *contentisation* and the *surrealisation* less the *entropy* of the *independent*,

$$\begin{aligned}\mathrm{entropy}(A^{\mathrm{M}(T)}) \quad \sim \quad \mathrm{entropy}(A^{\mathrm{X}} * T * T^{\odot A}) \; &- \; \mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\odot A}) \\ &- \; \mathrm{entropy}(A^{\mathrm{X}})\end{aligned}$$

Insofar as the *contentisation independent entropy* approximates to the *independent entropy*, $\mathrm{entropy}((A^{\mathrm{X}} * T * T^{\odot A})^{\mathrm{X}}) \approx \mathrm{entropy}(A^{\mathrm{X}})$, the *midisation alignment* varies as the difference between the *alignments* of the *contentisation* and the *surrealisation*,

$$\mathrm{algn}(A^{\mathrm{M}(T)}) \quad \sim \quad \mathrm{algn}(A^{\mathrm{X}} * T * T^{\odot A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})$$

As shown in section 'Likely histograms', above, if the *histogram*, A, is a given, and the *formal* is constrained to be *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, so that the *contentisation* equals the doubly-*independent formal independent converse action*, $A^{\mathrm{X}} * T * T^{\odot A} = (A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\odot A}$, then as the *midisation entropy*, $\mathrm{entropy}(A^{\mathrm{M}(T)})$, is minimised the *contentisation entropy* decreases to equal the *surrealisation entropy*, and *formal* tends to equal the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$. Similarly, as the *midisation alignment*, $\mathrm{algn}(A^{\mathrm{M}(T)})$, is maximised the *contentisation alignment* increases to equal the *surrealisation alignment*, and so *midisation alignment* maximisation also tends to *formal-abstract equivalence*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$.

The discussion of *midisation entropy* in 'Likely histograms', above, goes on to conjecture that there exists a *mid substrate transform* $T_{\mathrm{m}} \in \mathcal{T}_{U,V}$ which is neither *self* nor *unary*, $T_{\mathrm{m}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$, where the *formal* is *independent* and the *midisation entropy* is minimised,

$$T_{\mathrm{m}} \in \mathrm{mind}(\{(T, \mathrm{entropy}(A^{\mathrm{M}(T)})) : T \in \mathcal{T}_{U,V},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\})$$

At the *mid transform* the *formal* tends to the *abstract*, $A^{\mathrm{X}} * T_{\mathrm{m}} \approx (A * T_{\mathrm{m}})^{\mathrm{X}}$, and the *mid component size cardinality relative entropy* is small,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V^{\mathrm{C}} * T_{\mathrm{m}}) \quad \approx \quad 0$$

Conjecture that an approximation to the *mid transform* may also be obtained by a maximisation of the *midisation pseudo-alignment*,

$$\begin{aligned} T_{\mathrm{m}} \quad \in \quad & \mathrm{maxd}(\{(T, \mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})) : \\ & \qquad T \in \mathcal{T}_{U,V},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}) \end{aligned}$$

The discussion in 'Likely histograms' then goes on to show that subsequent minimisation of the *idealisation entropy*, where the *mid idealisation* is *integral*, $A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A} \in \mathcal{A}_{\mathrm{i}}$, tends to increase the *mid component size cardinality relative entropy*,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V^{\mathrm{C}} * T_{\mathrm{m}}) \quad \sim \quad -\ \mathrm{entropy}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A})$$

Conjecture, therefore, that subsequent maximisation of the *idealisation alignment* also tends to increase the *relative entropy*,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V^{\mathrm{C}} * T_{\mathrm{m}}) \quad \sim \quad \mathrm{algn}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A})$$

Consider the *histogram-transform* pair $(A, T) \in \mathcal{O}_{U,z}$ where the *derived histogram* is as *effective* as the *formal histogram*, $(A * T)^{\mathrm{F}} \geq (A^{\mathrm{X}} * T)^{\mathrm{F}}$, so that the *contentisation* is *size-conserving*, $\mathrm{size}(A^{\mathrm{X}} * T * T^{\odot A}) = \mathrm{size}(A)$. The *idealisation histogram*, $A * T * T^{\dagger A}$, can be defined as the *summation* of its *independent components*

$$A * T * T^{\dagger A} \equiv \sum((A * C)^{\mathrm{X}} : (R, C) \in T^{-1})$$

where T^{-1}. Hence conjecture that the *non-idealisation alignment*, $\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A})$, varies as the sum of the *alignments* of the *components*

$$\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) \sim \sum(\mathrm{algn}(A * C) : (R, C) \in T^{-1})$$

So, the *non-idealisation alignment* varies as the sum of the *alignments* of the *derivedly re-sized components*

$$\begin{aligned} &\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) \sim \\ &\quad \sum(\mathrm{algn}(Z_{D_R/D_R} * A * C) : (R, C) \in T^{-1},\ D_R > 0) \end{aligned}$$

where $D = A * T$ and $Z_x = \mathrm{scalar}(x)$. Similarly, the *surrealisation alignment*, $\mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})$, varies as the sum of the *alignments* of the *abstractly re-sized components*

$$\begin{aligned} &\mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) \sim \\ &\quad \sum(\mathrm{algn}(Z_{D^{\mathrm{X}}_R/D_R} * A * C) : (R, C) \in T^{-1},\ D_R > 0) \end{aligned}$$

Therefore the *midisation pseudo-alignment*, $\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A})$, varies as the difference

$$\begin{aligned} &\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A}) \sim \\ &\quad \sum(\mathrm{algn}(Z_{D_R/D_R} * A * C) - \mathrm{algn}(Z_{D^{\mathrm{X}}_R/D_R} * A * C) : (R, C) \in T^{-1},\ D_R > 0) \end{aligned}$$

Stated roughly in terms of the most *effective states* of a highly *diagonalised derived histogram*, the *midisation pseudo-alignment* increases as *on-diagonal component alignments*, $\mathrm{algn}(A * C)$ where $(R, C) \in T^{-1}$ and $R \in ((A * T) - (A * T)^{\mathrm{X}})^{\mathrm{FS}}$, exceed *off-diagonal abstractly re-sized component alignments*, $\mathrm{algn}(Z_{D^{\mathrm{X}}_R/D_R} * A * C)$ where $(R, C) \in T^{-1}$ and $R \in ((A * T)^{\mathrm{X}} - (A * T))^{\mathrm{FS}}$. Conjecture that the *midisation pseudo-alignment* (i) increases with the *derived alignment*, increasing the difference between the *on-diagonal* and *off-diagonal component sizes*, but (ii) decreases with the length of the *diagonal*, because long *diagonals* decrease the *component sizes*. That is,

$$\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A}) \sim \mathrm{algn}(A * T)/w^{1/m}$$

where $m = |W|$, $w = |W^{\mathrm{C}}|$ and $W = \mathrm{der}(T)$. Here the length of the *diagonal*, $w^{1/m}$, is approximated to the geometric average of the *valencies* of the *derived variables*, $w^{1/m} = (\prod_{u \in W} |U_u|)^{1/m}$. In other words, the *midisation pseudo-alignment* varies with the *derived alignment valency density*,

$$\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A}) \sim \mathrm{algn}(A * T)/\mathrm{capacityValency}(U)((A * T)^{\mathrm{FS}})$$

where *valency capacity* is defined in section 'Capacity and Alignment density', above, $\mathrm{capacityValency}(U) \in \mathrm{capacities}$ in *system* U as

$$\mathrm{capacityValency}(U)(Q) := x^{1/p}$$

where $x = \mathrm{volume}(U)(\mathrm{vars}(Q))$ and $p = |\mathrm{vars}(Q)|$.

Although the *idealisation alignment*, $\mathrm{algn}(A * T * T^{\dagger A})$, varies with the *derived alignment*, $\mathrm{algn}(A * T)$, this is less the case in the region of short *diagonals*. The *idealisation alignment* is a convex, positive-gradient function of *derived histogram diagonal length*. The convexity increases with *derived alignment* for given *diagonal length* because the *on-diagonal components* have greater *size* and the *alignment* of the *components* is a convex function of *component size*. That is, the *alignment* only scales linearly with *size*, $\mathrm{algn}(\mathrm{scalar}(k) * A) \approx k \times \mathrm{algn}(A)$, at high *entropies* and low *alignments*, as discussed in 'Scaled alignment', above. For example, the sum of the *alignments* of the *components* of a *uniformly diagonalised regular derived histogram* of *valency* 2 is expected to be greater than for *valency* 3 in the same *underlying histogram*, $\sum_{(\cdot, C_1) \in T_1^{-1}} \mathrm{algn}(A * C_1) > \sum_{(\cdot, C_2) \in T_2^{-1}} \mathrm{algn}(A * C_2)$ where $|(A * T_1)^{\mathrm{F}}| = 2$ and $|(A * T_2)^{\mathrm{F}}| = 3$. So the effect of rendering the *compenents independent*, when the *independent converse*, $T^{\dagger A}$, is applied, is greater in the *idealisation* at shorter *valencies*.

The *surrealisation alignment*, $\mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})$, varies against the *derived alignment*, $\mathrm{algn}(A * T)$, because the application of the *actual converse*, $T^{\odot A}$, to the *derived*, $A * T$, is the constant *underlying histogram*, $A * T * T^{\odot A} = A$. The *surrealisation alignment* is a convex, negative-gradient function of *derived histogram diagonal length*. The convexity increases with *derived alignment* for given *diagonal length* because *ineffective off-diagonal components* are still *ineffective* after *abstract re-sizing*, $\mathrm{size}(A * C) = 0 \implies \mathrm{algn}(Z_{D_R^{\mathrm{X}}/D_R} * A * C) = 0$, where $(R, C) \in T^{-1}$ and $R \notin (A * T)^{\mathrm{FS}}$.

Thus the *midisation pseudo-alignment*, $\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) -$

$\mathrm{algn}(A * T * T^{\dagger A})$, which depends on the *alignments* of *histograms* in the *underlying variables*, varies with the *derived alignment valency density*, $\mathrm{algn}(A * T)/w^{1/m}$, which depends only on the *geometry* and *alignment* of the *derived histogram* in the *derived variables*.

Consider the example of a *histogram-transform* pair $(A, T) \in \mathcal{O}_{U,z+y}$, where the *histogram* is a *regular cardinal histogram*, $A \in \mathcal{A}_{\mathrm{c}}$, of *dimension* n and *valency* d that consists of a *diagonal histogram* of *size* z *plus* a *cartesian histogram* of *size* y,

$$\begin{aligned} A &= \{(S, z/d) : S \in \{\{(x, u) : x \in \{1 \dots n\}\} : u \in \{1 \dots d\}\}\} + \\ &\quad \{(S, y/d^n) : S \in \prod \{\{(x, u) : u \in \{1 \dots d\}\} : x \in \{1 \dots n\}\}\} \end{aligned}$$

and the *transform*, T, has a *derived variable* for every *underlying variable* having *derived values* which roll up adjacent d/c *underlying values*, where c is the *derived valency*,

$$\begin{aligned} T = (\{(S \cup \{(x+n, (u/c)+1) : (x, u) \in S\}, 1) : \\ S \in \prod \{\{(x, u) : u \in \{1 \dots d\}\} : x \in \{1 \dots n\}\}\},\ \{n+1 \dots n+n\}) \end{aligned}$$

where $(/) \in \mathbf{N} \times \mathbf{N}_{>0} \to \mathbf{N}$.

Let $q_0 = y$, $q_1 = z + y$, $q_{\mathrm{c}} = c^{n-1}z + y$, $q_{\mathrm{d}} = d^{n-1}z + y$, and $r = d/c$. The *histogram alignment* is

$$\mathrm{algn}(A) \;=\; d \ln \frac{q_{\mathrm{d}}}{d^n}! + (d^n - d) \ln \frac{q_0}{d^n}! - d^n \ln \frac{q_1}{d^n}!$$

The *derived alignment* is

$$\mathrm{algn}(A * T) = c \ln \frac{q_{\mathrm{c}}}{c^n}! + (c^n - c) \ln \frac{q_0}{c^n}! - c^n \ln \frac{q_1}{c^n}!$$

The *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, so the *formal alignment* is zero

$$\mathrm{algn}(A^{\mathrm{X}} * T) = 0$$

The *neutralisation* equals the *idealisation*, $A * T * T^{\odot A^{\mathrm{X}}} = A * T * T^{\dagger A}$. The *idealisation alignment* is

$$\begin{aligned} \mathrm{algn}(A * T * T^{\dagger A}) &= r^n c \ln \frac{q_{\mathrm{c}}}{d^n}! + r^n (c^n - c) \ln \frac{q_0}{d^n}! - d^n \ln \frac{q_1}{r^n c^n}! \\ &= r^n c \ln \frac{q_{\mathrm{c}}}{r^n c^n}! + r^n (c^n - c) \ln \frac{q_0}{r^n c^n}! - r^n c^n \ln \frac{q_1}{r^n c^n}! \end{aligned}$$

Each of the terms of the expression for the *derived alignment*, $\mathrm{algn}(A * T)$, has a corresponding term in the expression for the *idealisation alignment*, $\mathrm{algn}(A * T * T^{\dagger A})$, scaled by r^n and such that the argument to the factorial is inversely scaled by r^n. Thus the *idealisation alignment* is approximately less than the *derived alignment*, $\mathrm{algn}(A*T*T^{\dagger A}) \approx \mathrm{algn}(A*T)$, and $\mathrm{algn}(A * T * T^{\dagger A}) \leq \mathrm{algn}(A * T)$.

When $c = 1$ both the *derived alignment* and the *idealisation alignment* are zero, $c = 1 \implies \mathrm{algn}(A*T) = \mathrm{algn}(A*T*T^{\dagger A}) = 0$. This is the case, for example, if the *transform* is a *unary partition transform* $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$. When $c = d$ both the *derived alignment* and the *idealisation alignment* equal the *histogram alignment*, $c = d \implies \mathrm{algn}(A*T) = \mathrm{algn}(A*T*T^{\dagger A}) = \mathrm{algn}(A)$. This is the case, for example, if the *transform* is a *value full functional transform* $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}}$. The partial derivative of the *idealisation alignment* with respect to c is positive where $1 \leq c \leq d$,

$$\frac{\partial}{\partial c}\left(\frac{d^n}{c^{n-1}} \ln \frac{q_{\mathrm{c}}}{d^n}! - \frac{1}{c^{n-1}} \ln \frac{q_0}{d^n}!\right) \geq 0$$

The *contentisation* equals the *surrealisation*, $A^{\mathrm{X}} * T * T^{\odot A} = (A*T)^{\mathrm{X}} * T^{\odot A}$. The *surrealisation alignment* is

$$\mathrm{algn}((A*T)^{\mathrm{X}} * T^{\odot A}) \;=\; d \ln \frac{q_{\mathrm{d}}}{d^n}\frac{q_1}{q_{\mathrm{c}}}! + (r^n c - d) \ln \frac{q_0}{d^n}\frac{q_1}{q_{\mathrm{c}}}! - r^n c \ln \frac{q_1}{d^n}!$$

When $c = 1$ the *surrealisation alignment* equals the *histogram alignment*, $c = 1 \implies \mathrm{algn}((A*T)^{\mathrm{X}} * T^{\odot A}) = \mathrm{algn}(A)$. When $c = d$ the *surrealisation alignment* is zero, $c = d \implies \mathrm{algn}((A*T)^{\mathrm{X}} * T^{\odot A}) = 0$. The partial derivative of the *surrealisation alignment* with respect to c is negative where $1 < c < d$,

$$\frac{\partial}{\partial c}\left(d \ln \frac{q_{\mathrm{d}}}{d^n}\frac{q_1}{q_{\mathrm{c}}}! + (\frac{d^n}{c^{n-1}} - d) \ln \frac{q_0}{d^n}\frac{q_1}{q_{\mathrm{c}}}! - \frac{d^n}{c^{n-1}} \ln \frac{q_1}{d^n}!\right) \leq 0$$

The *midisation pseudo-alignment* is

$$\begin{aligned}
&\mathrm{algn}(A) - \mathrm{algn}((A*T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A}) = \\
&\quad \left(d \ln \frac{q_{\mathrm{d}}}{d^n}! + (r^n c - d) \ln \frac{q_0}{d^n}! + r^n c \ln \frac{q_1}{d^n}!\right) - \\
&\quad \left(d \ln \frac{q_{\mathrm{d}}}{d^n}\frac{q_1}{q_{\mathrm{c}}}! + (r^n c - d) \ln \frac{q_0}{d^n}\frac{q_1}{q_{\mathrm{c}}}! + r^n c \ln \frac{q_{\mathrm{c}}}{d^n}!\right)
\end{aligned}$$

When $c = 1$ or $c = d$ the *midisation pseudo-alignment* is zero, $(c = 1) \wedge (c = d) \implies \mathrm{algn}(A) - \mathrm{algn}((A*T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A}) = 0$.

Elsewhere the *midisation pseudo-alignment* is greater than zero, $1 < c < d \implies \text{algn}(A) - \text{algn}((A*T)^{\text{X}} * T^{\odot A}) - \text{algn}(A*T*T^{\dagger A}) > 0$, therefore the *midisation pseudo-alignment* as a function of c contains a maximum away from the boundaries.

Having shown that the *midisation pseudo-alignment* varies with *derived alignment valency density*

$$\text{algn}(A) - \text{algn}((A*T)^{\text{X}} * T^{\odot A}) - \text{algn}(A*T*T^{\dagger A}) \sim \text{algn}(A*T)/w^{1/m}$$

conjecture that the approximation to the *mid transform* obtained by a maximisation of the *midisation pseudo-alignment*, where the *formal* is *independent*,

$$\begin{aligned} T_{\text{m}} \in\ & \text{maxd}(\{(T, \text{algn}(A) - \text{algn}(A*T*T^{\dagger A}) - \text{algn}((A*T)^{\text{X}} * T^{\odot A})) : \\ & \qquad T \in \mathcal{T}_{U,V},\ A^{\text{X}} * T = (A^{\text{X}} * T)^{\text{X}}\}) \end{aligned}$$

also has high *derived alignment*, $\text{algn}(A * T_{\text{m}})$, and hence low *derived entropy*, $\text{entropy}(A * T_{\text{m}})$. Although the *mid derived entropy* decreases, the *mid derived relative entropy* is small, $\text{entropyRelative}(A * T_{\text{m}}, V^{\text{C}} * T_{\text{m}}) \approx 0$, because the *formal* tends to the *abstract*, $A^{\text{X}} * T_{\text{m}} \approx (A*T_{\text{m}})^{\text{X}}$. That is, while a maximisation of the *midisation pseudo-alignment*, where the *formal* is *independent*, tends to decrease the *derived entropy*, the *relative entropy* also tends to decrease. A subsequent maximisation of the *mid integral idealisation alignment*, $\text{algn}(A*T_{\text{m}}*T_{\text{m}}^{\dagger A})$, which tends to lengthen *derived diagonals* towards *full functional*, is necessary to recover the *relative entropy*.

Given a *histogram-transform* pair $(A, T) \in \mathcal{O}_{U,z}$, having *variables* $V = \text{vars}(A)$, *size* $z = \text{size}(A)$ and *volume* $v = |V^{\text{C}}|$, it is conjectured in section 'Likely histograms', above, that the *maximum likelihood estimate* for the *integral iso-deriveds* is the *naturalisation*, $A * T * T^{\dagger}$,

$$\begin{aligned} &\{A * T * T^{\dagger}\} = \\ &\qquad \text{maxd}(\{(D, \textstyle\sum(Q_{\text{m},U}(D,z)(B) : B \in D^{-1}_{U,\text{i},T,z}(A*T))) : D \in \mathcal{A}_{U,V,z}\}) \end{aligned}$$

The corresponding *dependent analogue* is the *derived-dependent*, $A^{\text{D}(T)} \in \mathcal{A}_{U,V,z}$, defined

$$\{A^{\text{D}(T)}\} = \text{maxd}(\{(D, \frac{Q_{\text{m},U}(D,z)(A)}{\sum Q_{\text{m},U}(D,z)(B) : B \in D^{-1}_{U,\text{i},T,z}(A*T)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *naturalisation-distributed-relative multinomial space* is

$$\text{spaceRelative}(A * T * T^{\dagger})(A) \quad := \quad -\ln \frac{\text{mpdf}(U)(A * T * T^{\dagger}, z)(A)}{\text{mpdf}(U)(A * T * T^{\dagger}, z)(A * T * T^{\dagger})}$$

where the *distribution-relative multinomial space* is defined, in section 'Likely histograms', above, as

$$\text{spaceRelative}(E)(A) \quad := \quad -\ln \frac{\text{mpdf}(U)(E, z)(A)}{\text{mpdf}(U)(E, z)(E)}$$

The *naturalisation* is computable, so a rational approximation to the *naturalisation distributed-relative multinomial space* is computable.

In the case where both the *histogram* and *naturalisation* are *integral*, $A,\ A * T * T^{\dagger} \in \mathcal{A}_{\mathrm{i}}$, the *naturalisation-distributed-relative multinomial space* is

$$\text{spaceRelative}(A * T * T^{\dagger})(A) \quad := \quad -\ln \frac{Q_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A)}{Q_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})}$$

The *naturalisation-distributed-relative multinomial space* of the *naturalisation* is zero,

$$\text{spaceRelative}(A * T * T^{\dagger})(A * T * T^{\dagger}) = 0$$

In the case where the *histogram* and *naturalisation* are *integral*, $A,\ A*T*T^{\dagger} \in \mathcal{A}_{\mathrm{i}}$, the *naturalisation-distributed-relative multinomial space* is conjectured to be greater than or equal to zero, and less than or equal to the *naturalisation-distributed-relative multinomial space* of the *derived-dependent*,

$$0 \leq \text{spaceRelative}(A * T * T^{\dagger})(A) \leq \text{spaceRelative}(A * T * T^{\dagger})(A^{\mathrm{D}(T)})$$

This is consistent with the *entropies*,

$$\text{entropy}(A * T * T^{\dagger}) \geq \text{entropy}(A) \geq \text{entropy}(A^{\mathrm{D}(T)})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *naturalisation* equals the *histogram*, $A * T_{\mathrm{f}} * T_{\mathrm{f}}^{\dagger} = A$, and so the *naturalisation-distributed-relative multinomial space* equals zero,

$$\text{spaceRelative}(A * T_{\mathrm{f}} * T_{\mathrm{f}}^{\dagger})(A) \quad = \quad 0$$

At the other extreme where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *naturalisation* equals the *scaled normalised cartesian*, $A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger} = \text{scalar}(z/v) * V^{\mathrm{C}}$, and so the *naturalisation-distributed-relative multinomial space* simplifies to

$$\text{spaceRelative}(A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger})(A) = \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - v \ln \Gamma_!(z/v)$$

The *naturalisation-derived* is equal to the *derived*, $(A * T * T^{\dagger}) * T = A * T$, so the *naturalisation-derived alignment* equals the *derived alignment*,

$$\mathrm{algn}((A * T * T^{\dagger}) * T) = \mathrm{algn}(A * T)$$

Of the *integral iso-deriveds*, only the *naturalisation* has zero *naturalisation-distributed-relative multinomial space*,

$$\forall B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)\ (B \neq A*T*T^{\dagger} \implies \mathrm{spaceRelative}(A*T*T^{\dagger})(B) > 0)$$

Insofar as the *naturalisation* is approximately equal to the *independent*, $A * T * T^{\dagger} \approx A^{\mathrm{X}}$, then the *naturalisation-distributed-relative multinomial space* approximates to

$$\begin{aligned}\mathrm{spaceRelative}(A * T * T^{\dagger})(A) &\approx \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! (A * T * T^{\dagger})_S \\ &\approx z \times \mathrm{entropy}(A * T * T^{\dagger}) - z \times \mathrm{entropy}(A)\end{aligned}$$

The difference is the *mis-naturalisation-distributed-relative multinomial space*,

$$\sum_{S \in A^{\mathrm{XS}}} (A_S - (A * T * T^{\dagger})_S) \ln (A * T * T^{\dagger})_S$$

The degree to which the *integral iso-derived* is *aligned-like* is the *iso independence*,

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)\ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|D^{-1}_{U,\mathrm{i},T,z}(A * T)\ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

As the *iso-independence* of an *iso-set* increases, the *independent analogue* tends to the *independent*. In this case the *naturalisation*, $A*T*T^{\dagger}$, tends to the *independent*, A^{X}, and so the *relative space*, $\mathrm{spaceRelative}(A*T*T^{\dagger})(A)$, tends to the *alignment*, $\mathrm{algn}(A)$. That is, as the *integral iso-deriveds* becomes more *aligned-like*, the *relative space* becomes less dependent on the *transform*, T.

The *iso-set* is *law-like*, so in the case where the *dependent analogue* is in the *iso-set*, $A^{\mathrm{D}(T)} \in D^{-1}_{U,T,z}(A * T)$, the *dependent derived* equals the *derived*, $A^{\mathrm{D}(T)} * T = A * T$, and the difference in *relative space* between the *histogram* and the *dependent* must be in the *relative spaces* of the *components*,

$$\begin{aligned}&A^{\mathrm{D}(T)} \in D^{-1}_{U,T,z}(A * T) \implies \\ &\quad \sum_{(\cdot,C) \in T^{-1}} \mathrm{spaceRelative}(A * T * T^{\dagger} * C)(A * C) \\ &\leq \quad \sum_{(\cdot,C) \in T^{-1}} \mathrm{spaceRelative}(A * T * T^{\dagger} * C)(A^{\mathrm{D}(T)} * C)\end{aligned}$$

and so

$$A^{\mathrm{D}(T)} \in D^{-1}_{U,T,z}(A * T) \implies \sum_{(\cdot,C) \in T^{-1}} \sum_{S \in C^{\mathrm{S}}} \ln \Gamma_! A_S \leq \sum_{(\cdot,C) \in T^{-1}} \sum_{S \in C^{\mathrm{S}}} \ln \Gamma_! A^{\mathrm{D}(T)}_S$$

or

$$A^{\mathrm{D}(T)} \in D^{-1}_{U,T,z}(A * T) \implies \sum_{S \in V^{\mathrm{CS}}} \ln \Gamma_! A_S \leq \sum_{S \in V^{\mathrm{CS}}} \ln \Gamma_! A^{\mathrm{D}(T)}_S$$

In 'Iso-sets', above, the cardinality of the set of *integral iso-deriveds* is the product of the weak compositions of the *components*,

$$|D^{-1}_{U,\mathrm{i},T,z}(A * T)| = \prod_{(R,C) \in T^{-1}} \frac{((A * T)_R + |C| - 1)!}{(A * T)_R! \ (|C| - 1)!}$$

It is shown in 'Integral iso-sets and entropy', above, that the *integral iso-deriveds log-cardinality* varies against the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned} &\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \sim \\ &\quad - \big((z + v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\ &\qquad\qquad -z \times \mathrm{entropy}(A * T) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)\big) \end{aligned}$$

In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *integral iso-deriveds log-cardinality* varies against the *size* scaled *component size cardinality relative entropy*,

$$\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \sim -z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

In the domain where the *size* is greater than the *volume*, $z > v$, the *integral iso-deriveds log-cardinality* varies against the *volume* scaled *component cardinality size relative entropy*,

$$\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \sim -v \times \mathrm{entropyRelative}(V^{\mathrm{C}} * T, A * T)$$

The *relative entropy* is the *cross entropy* minus the *component entropy*, so in the case where the *size* is less than or equal to the *volume*, $z \leq v$, the *iso-derived log-cardinality* varies against the *component size cardinality cross entropy* and varies with the *derived entropy* or *component size entropy*,

$$\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \sim -z \times \mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T)$$

and

$$\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \sim z \times \mathrm{entropy}(A * T)$$

Given a *histogram-transform* pair $(A, T) \in \mathcal{O}_{U,z}$, it is conjectured above that the *maximum likelihood estimate* for the *integral iso-idealisations* is the *idealisation*, $A * T * T^{\dagger A}$,

$$\{A * T * T^{\dagger A}\} = \\ \text{maxd}(\{(D, \sum(Q_{\text{m},U}(D, z)(B) : B \in \text{isoi}(U)(T, A))) : D \in \mathcal{A}_{U,V,z}\})$$

where $\text{isoi}(U)(T, A) := Y^{-1}_{U,\text{i},T,\dagger,z}(A * T * T^{\dagger A})$.

The corresponding *dependent analogue* is the *idealisation-dependent*, $A^{\dagger(T)} \in \mathcal{A}_{U,V,z}$, defined

$$\{A^{\dagger(T)}\} = \\ \text{maxd}(\{(D, \frac{Q_{\text{m},U}(D, z)(A)}{\sum Q_{\text{m},U}(D, z)(B) : B \in \text{isoi}(U)(T, A)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *idealisation-distributed-relative multinomial space* is

$$\text{spaceRelative}(A * T * T^{\dagger A})(A) \quad := \quad -\ln \frac{\text{mpdf}(U)(A * T * T^{\dagger A}, z)(A)}{\text{mpdf}(U)(A * T * T^{\dagger A}, z)(A * T * T^{\dagger A})}$$

The *idealisation* is computable, so a rational approximation to the *idealisation-distributed-relative multinomial space* is computable.

In the case where both the *histogram* and *idealisation* are *integral*, $A,\ A * T * T^{\dagger A} \in \mathcal{A}_\text{i}$, the *idealisation-distributed-relative multinomial space* is

$$\text{spaceRelative}(A * T * T^{\dagger A})(A) \quad := \quad -\ln \frac{Q_{\text{m},U}(A * T * T^{\dagger A}, z)(A)}{Q_{\text{m},U}(A * T * T^{\dagger A}, z)(A * T * T^{\dagger A})}$$

The *idealisation-distributed-relative multinomial space* of the *idealisation* is zero,

$$\text{spaceRelative}(A * T * T^{\dagger A})(A * T * T^{\dagger A}) = 0$$

In the case where the *histogram* and *idealisation* are *integral*, $A,\ A*T*T^{\dagger A} \in \mathcal{A}_\text{i}$, the *idealisation-distributed-relative multinomial space* is conjectured to be greater than or equal to zero, and less than or equal to the *idealisation-distributed-relative multinomial space* of the *idealisation-dependent*,

$$0 \leq \text{spaceRelative}(A * T * T^{\dagger A})(A) \leq \text{spaceRelative}(A * T * T^{\dagger A})(A^{\dagger(T)})$$

This is consistent with the *entropies*,

$$\text{entropy}(A * T * T^{\dagger A}) \geq \text{entropy}(A) \geq \text{entropy}(A^{\dagger(T)})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *idealisation* equals the *histogram*, $A * T_{\mathrm{f}} * T_{\mathrm{f}}^{\dagger A} = A$, and so the *idealisation* equals zero,

$$\mathrm{spaceRelative}(A * T_{\mathrm{f}} * T_{\mathrm{f}}^{\dagger A})(A) \quad = \quad 0$$

At the other extreme where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *idealisation* equals the *independent*, $A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A} = A^{\mathrm{X}}$, and so the *idealisation-distributed-relative multinomial space* equals the *alignment*

$$\mathrm{spaceRelative}(A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A})(A) = \mathrm{algn}(A)$$

Conjecture that the *relative space* of the *histogram* with respect to the *idealisation* is less than or equal to that with respect to the *independent*,

$$\mathrm{spaceRelative}(A * T * T^{\dagger A})(A) \leq \mathrm{spaceRelative}(A^{\mathrm{X}})(A) = \mathrm{algn}(A)$$

and similarly for the *idealisation-dependent*

$$\mathrm{spaceRelative}(A * T * T^{\dagger A})(A^{\dagger(T)}) \leq \mathrm{spaceRelative}(A^{\mathrm{X}})(A^{\dagger(T)}) = \mathrm{algn}(A^{\dagger(T)})$$

because the *idealisation entropy* is less than or equal to the *independent entropy*, $\mathrm{entropy}(A * T * T^{\dagger A}) \leq \mathrm{entropy}(A^{\mathrm{X}})$.

The *idealisation-derived* is equal to the *derived*, $(A * T * T^{\dagger A}) * T = A * T$, so the *idealisation-derived alignment* equals the *derived alignment*,

$$\mathrm{algn}((A * T * T^{\dagger A}) * T) = \mathrm{algn}(A * T)$$

Of the *integral iso-idealisations*, only the *idealisation* has zero *idealisation-distributed-relative multinomial space*,

$$\begin{aligned} &\forall B \in Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A}) \\ &\quad (B \neq A * T * T^{\dagger A} \implies \mathrm{spaceRelative}(A * T * T^{\dagger A})(B) > 0) \end{aligned}$$

Insofar as the *transform* approximates to *unary*, $T \approx T_{\mathrm{u}}$, the *idealisation* approximates to the *independent*, $A * T * T^{\dagger A} \approx A^{\mathrm{X}}$, and the *idealisation-distributed-relative multinomial space* approximates to

$$\begin{aligned} \mathrm{spaceRelative}(A * T * T^{\dagger A})(A) \quad &\approx \quad \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! (A * T * T^{\dagger A})_S \\ &\approx \quad z \times \mathrm{entropy}(A * T * T^{\dagger A}) - z \times \mathrm{entropy}(A) \end{aligned}$$

The difference is the *mis-idealisation-distributed-relative multinomial space*,

$$\sum_{S \in A^{\mathrm{XS}}} (A_S - (A * T * T^{\dagger A})_S) \ln(A * T * T^{\dagger A})_S$$

The *iso-idealisations* is a subset of the *iso-independents*, $Y^{-1}_{U,T,\dagger,z}(A*T*T^{\dagger A}) \subseteq Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, so the *iso-independence* is the fraction of their *integral* cardinalities,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

In some cases the *iso-independence* of the *iso-idealisations* is greater than or equal to the *iso-independence* of the *iso-derived*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} \geq \frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T) \cap Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|D^{-1}_{U,\mathrm{i},T,z}(A * T) \cup Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

As the *iso-independence* increases, the *transform* becomes more *unary*, the *idealisation*, $A*T*T^{\dagger A}$, tends to the *independent*, A^{X}, and the *relative space*, $\mathrm{spaceRelative}(A * T * T^{\dagger A})(A)$, tends to the *alignment*, $\mathrm{algn}(A)$.

The *iso-set* is *law-like*, so in the case where the *dependent analogue* is in the *iso-set*, $A^{\dagger(T)} \in D^{-1}_{U,T,z}(A * T)$, the *dependent derived* equals the *derived*, $A^{\dagger(T)} * T = A * T$, and the difference in *relative space* between the *histogram* and the *dependent* must be in the *relative spaces* of the *components*,

$$\begin{aligned}
&A^{\dagger(T)} \in D^{-1}_{U,T,z}(A * T) \implies \\
&\quad \sum_{(\cdot,C) \in T^{-1}} \mathrm{spaceRelative}(A * T * T^{\dagger A} * C)(A * C) \\
&\quad \leq \sum_{(\cdot,C) \in T^{-1}} \mathrm{spaceRelative}(A * T * T^{\dagger A} * C)(A^{\dagger(T)} * C)
\end{aligned}$$

So, in the case of the *idealisation-dependent*, the *component alignments* must be greater than or equal to the *component alignments* of the *histogram*,

$$\begin{aligned}
&A^{\dagger(T)} \in D^{-1}_{U,T,z}(A * T) \implies \\
&\quad \sum_{(\cdot,C) \in T^{-1}} \mathrm{algn}(A * C) \leq \sum_{(\cdot,C) \in T^{-1}} \mathrm{algn}(A^{\dagger(T)} * C)
\end{aligned}$$

The *iso-derivedence*, or degree of *law-likeness*, of the *iso-idealisations* is

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|} \leq 1$$

As the *iso-derivedence* increases, the difference between the *relative spaces* of the *dependents*, $\mathrm{spaceRelative}(A * T * T^{\dagger})(A^{\mathrm{D}(T)}) - \mathrm{spaceRelative}(A * T * T^{\dagger A})(A^{\dagger(T)})$, decreases.

Given a *histogram-transform* pair $(A,T) \in \mathcal{O}_{U,z}$, it is conjectured in section 'Likely histograms', above, that the *maximum likelihood estimate* for the *integral iso-abstracts* is the *naturalised abstract*, $(A*T)^{\mathrm{X}}*T^{\dagger}$,

$$\begin{aligned}&\{(A*T)^{\mathrm{X}}*T^{\dagger}\} = \\ &\quad \mathrm{maxd}(\{(D, \sum (Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}}))) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The corresponding *dependent analogue* is the *abstract-dependent*, $A^{\mathrm{W}(T)} \in \mathcal{A}_{U,V,z}$, defined

$$\begin{aligned}&\{A^{\mathrm{W}(T)}\} = \\ &\quad \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})}) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *naturalised abstract distributed relative multinomial space* is

$$\begin{aligned}&\mathrm{spaceRelative}((A*T)^{\mathrm{X}}*T^{\dagger})(A) := \\ &\qquad -\ln \frac{\mathrm{mpdf}(U)((A*T)^{\mathrm{X}}*T^{\dagger},z)(A)}{\mathrm{mpdf}(U)((A*T)^{\mathrm{X}}*T^{\dagger},z)((A*T)^{\mathrm{X}}*T^{\dagger})}\end{aligned}$$

The *naturalised abstract* is computable, so a rational approximation to the *naturalised abstract distributed relative multinomial space* is computable.

In the case where both the *histogram* and *naturalised abstract* are *integral*, A, $(A*T)^{\mathrm{X}}*T^{\dagger} \in \mathcal{A}_{\mathrm{i}}$, the *naturalised abstract distributed relative multinomial space* is

$$\begin{aligned}&\mathrm{spaceRelative}((A*T)^{\mathrm{X}}*T^{\dagger})(A) := \\ &\qquad -\ln \frac{Q_{\mathrm{m},U}((A*T)^{\mathrm{X}}*T^{\dagger},z)(A)}{Q_{\mathrm{m},U}((A*T)^{\mathrm{X}}*T^{\dagger},z)((A*T)^{\mathrm{X}}*T^{\dagger})}\end{aligned}$$

The *naturalised abstract distributed relative multinomial space* of the *naturalised abstract* is zero,

$$\mathrm{spaceRelative}((A*T)^{\mathrm{X}}*T^{\dagger})((A*T)^{\mathrm{X}}*T^{\dagger}) = 0$$

In the case where the *histogram* and *naturalised abstract* are *integral*, A, $(A*T)^{\mathrm{X}}*T^{\dagger} \in \mathcal{A}_{\mathrm{i}}$, the *naturalised abstract distributed relative multinomial space* is conjectured to be greater than or equal to zero, and less than or equal to the *naturalised abstract distributed relative multinomial space* of the *abstract-dependent*,

$$0 \leq \mathrm{spaceRelative}((A*T)^{\mathrm{X}}*T^{\dagger})(A) \leq \mathrm{spaceRelative}((A*T)^{\mathrm{X}}*T^{\dagger})(A^{\mathrm{W}(T)})$$

This is consistent with the *entropies*,

$$\mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\dagger}) \geq \mathrm{entropy}(A) \geq \mathrm{entropy}(A^{\mathrm{W}(T)})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, the *naturalised abstract* equals the *independent*, $(A * T_{\mathrm{f}})^{\mathrm{X}} * T_{\mathrm{f}}^{\dagger} = A^{\mathrm{X}}$, and so the *naturalised abstract distributed relative multinomial space* equals the *aligment*,

$$\mathrm{spaceRelative}((A * T_{\mathrm{f}})^{\mathrm{X}} * T_{\mathrm{f}}^{\dagger})(A) \quad = \quad \mathrm{algn}(A)$$

At the other extreme where the *transform* is *unary*, $T = T_{\mathrm{u}}$, the *naturalised abstract* equals the *scaled normalised cartesian*, $(A * T_{\mathrm{u}})^{\mathrm{X}} * T_{\mathrm{u}}^{\dagger} = \mathrm{scalar}(z/v) * V^{\mathrm{C}}$, and so the *naturalised abstract distributed relative multinomial space* simplifies to

$$\mathrm{spaceRelative}((A * T_{\mathrm{u}})^{\mathrm{X}} * T_{\mathrm{u}}^{\dagger})(A) = \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - v \ln \Gamma_!(z/v)$$

Of the *integral iso-abstracts*, only the *naturalised abstract* has zero *naturalised abstract distributed relative multinomial space*,

$$\begin{aligned}
&\forall B \in Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}}) \\
&\qquad (B \neq (A * T)^{\mathrm{X}} * T^{\dagger} \implies \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(B) > 0)
\end{aligned}$$

Insofar as the *naturalised abstract* is approximately equal to the *independent*, $(A * T)^{\mathrm{X}} * T^{\dagger} \approx A^{\mathrm{X}}$, then the *naturalised abstract distributed relative multinomial space* approximates to

$$\begin{aligned}
&\mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A) \\
&\qquad \approx \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_!((A * T)^{\mathrm{X}} * T^{\dagger})_S \\
&\qquad \approx z \times \mathrm{entropy}((A * T)^{\mathrm{X}} * T^{\dagger}) - z \times \mathrm{entropy}(A)
\end{aligned}$$

The difference is the *mis-naturalised abstract distributed relative multinomial space*,

$$\sum_{S \in A^{\mathrm{XS}}} (A_S - ((A * T)^{\mathrm{X}} * T^{\dagger})_S) \ln((A * T)^{\mathrm{X}} * T^{\dagger})_S$$

The degree to which the set of *integral iso-abstracts* is *aligned-like* is the *iso-independence*,

$$\frac{|Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}}) \ \cap \ Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})|}{|Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}}) \ \cup \ Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})|}$$

As the *iso-independence* increases, the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, tends to the *independent*, A^{X}, and the *relative space*, $\mathrm{spaceRelative}((A*T)^{\mathrm{X}} * T^{\dagger})(A)$, tends to the *alignment*, $\mathrm{algn}(A)$. That is, as the *integral iso-abstracts* becomes more *aligned-like*, the *relative space* becomes less dependent on the *transform*, T.

The *independent analogue* is the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$. The *derived alignment* of the *independent analogue* is zero, $\mathrm{algn}((A*T)^{\mathrm{X}} * T^{\dagger} * T) = \mathrm{algn}((A*T)^{\mathrm{X}}) = 0$. The set of *iso-abstracts* is *entity-like* so the *derived*, $A*T$, and the *dependent derived*, $A^{\mathrm{W}(T)} * T$, are not necessarily equal to each other and nor are they necessarily equal to the *abstract*, $(A * T)^{\mathrm{X}}$. Conjecture that the relation between the *relative spaces*,

$$\begin{aligned} 0 &= \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})((A * T)^{\mathrm{X}} * T^{\dagger}) \\ &\leq \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A) \\ &\leq \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A^{\mathrm{W}(T)}) \end{aligned}$$

can be *lifted*,

$$\begin{aligned} 0 &= \mathrm{spaceRelative}((A * T)^{\mathrm{X}})((A * T)^{\mathrm{X}}) \\ &\leq \mathrm{spaceRelative}((A * T)^{\mathrm{X}})(A * T) \\ &\leq \mathrm{spaceRelative}((A * T)^{\mathrm{X}})(A^{\mathrm{W}(T)} * T) \end{aligned}$$

and so conjecture that the *dependent analogue derived alignment* is greater than or equal to the *derived alignment* which in turn is greater than or equal to the *independent analogue derived alignment*,

$$0 = \mathrm{algn}((A * T)^{\mathrm{X}}) \leq \mathrm{algn}(A * T) \leq \mathrm{algn}(A^{\mathrm{W}(T)} * T)$$

The *iso-derivedence* of the *iso-abstracts* equals the *iso-abstractence* of the *iso-deriveds*,

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}$$

As the *iso-derivedence* and *iso-abstractence* increases, the difference between the *relative spaces* of the *dependents*, $\mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A^{\mathrm{W}(T)}) - \mathrm{spaceRelative}(A * T * T^{\dagger})(A^{\mathrm{D}(T)})$, decreases.

In ‘Iso-sets’, above, in the case where the *derived* is *independent*, $A * T = (A * T)^{\mathrm{X}}$, the cardinality of the set of *integral iso-abstracts* equals the cardinality of the set of *integral iso-deriveds*,

$$\begin{aligned}|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| &= |D^{-1}_{U,\mathrm{i},T,z}((A * T)^{\mathrm{X}})| \\ &= \prod_{(R,C) \in T^{-1}} \frac{((A * T)^{\mathrm{X}}_R + |C| - 1)!}{(A * T)^{\mathrm{X}}_R!\ (|C| - 1)!}\end{aligned}$$

and so in this case the *integral iso-abstracts log-cardinality* is approximately proportional to the negative *abstract size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}&\ln |Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| \\ &\quad\approx\ (z + v)\ln(z + v)\ -\ z \ln z\ -\ v \ln v \\ &\qquad\quad -\ \big((z + v) \times \mathrm{entropy}((A * T)^{\mathrm{X}} + V^{\mathrm{C}} * T) \\ &\qquad\qquad\qquad -z \times \mathrm{entropy}((A * T)^{\mathrm{X}})\ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)\big)\end{aligned}$$

In the case where the *size* is less than or equal to the *volume*, $z \leq v$, the *log iso-abstract cardinality* varies against the *size* scaled *component size cardinality relative abstract entropy*,

$$\ln |Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| \ \sim\ -z \times \mathrm{entropyRelative}((A * T)^{\mathrm{X}}, V^{\mathrm{C}} * T)$$

and where the *size* is greater than the *volume*, $z > v$, the *log iso-abstract cardinality* varies against the *volume* scaled *component cardinality size relative abstract entropy*,

$$\ln |Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| \ \sim\ -v \times \mathrm{entropyRelative}(V^{\mathrm{C}} * T, (A * T)^{\mathrm{X}})$$

The *relative entropy* is the *cross entropy* minus the *component entropy*, so in the case where the *size* is less than or equal to the *volume*, $z \leq v$, the *log iso-abstract cardinality* varies against the *component size cardinality cross abstract entropy* and varies with the *abstract entropy*,

$$\ln |Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| \ \sim\ -z \times \mathrm{entropyCross}((A * T)^{\mathrm{X}}, V^{\mathrm{C}} * T)$$

and

$$\ln |Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})| \ \sim\ z \times \mathrm{entropy}((\hat{A} * T)^{\mathrm{X}})$$

In this case where the *size* is less than or equal to the *volume*, $z \leq v$, the *log iso-derivedence* of the *iso-abstracts* varies with the difference between the *derived entropy* and the *abstract entropy*, and so varies against the *derived alignment*,

$$\begin{aligned}
\ln \frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} &\sim z \times \mathrm{entropy}(A * T) - z \times \mathrm{entropy}((\hat{A} * T)^{\mathrm{X}}) \\
&\approx -\ \mathrm{spaceRelative}((A * T)^{\mathrm{X}})(A * T) \\
&= -\ \mathrm{algn}(A * T)
\end{aligned}$$

That is, the fraction of the *entity-like histograms* that are also *law-like* decreases as the *derived alignment* increases. In the case where the *derived* is *independent*, the *derived alignment* is minimised, $A * T = (A * T)^{\mathrm{X}} \implies \mathrm{algn}(A * T)$, and the *iso-derivedence* is maximised,

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}((A * T)^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} = 1$$

Given a *histogram-transform* pair $(A, T) \in \mathcal{O}_{U,z}$, the *transform-independent*, $A^{\mathrm{X}(T)} \in \mathcal{A}_{U,V,z}$, is defined in section 'Likely histograms', above, as

$$\{A^{\mathrm{X}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A))) : D \in \mathcal{A}_{U,V,z}\})$$

where the *integral iso-transform-independents* is abbreviated

$$\begin{aligned}
\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) &= Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \\
&= \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}
\end{aligned}$$

The corresponding *dependent analogue* is the *transform-dependent*, $A^{\mathrm{Y}(T)} \in \mathcal{A}_{U,V,z}$, defined

$$\{A^{\mathrm{Y}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D, z)(A)}{\sum Q_{\mathrm{m},U}(D, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *transform-independent-distributed-relative multinomial space* is

$$\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) := -\ln \frac{\mathrm{mpdf}(U)(A^{\mathrm{X}(T)}, z)(A)}{\mathrm{mpdf}(U)(A^{\mathrm{X}(T)}, z)(A^{\mathrm{X}(T)})}$$

The *transform-independent* is sometimes not computable, so the *transform-independent-distributed-relative multinomial space* is sometimes not computable.

In the case where both the *histogram* and *transform-independent* are *integral*, $A, A^{\mathrm{X}(T)} \in \mathcal{A}_{\mathrm{i}}$, the *transform-independent-distributed-relative multinomial space* is

$$\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) \quad := \quad -\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(A)}{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)}, z)(A^{\mathrm{X}(T)})}$$

The *transform-independent-distributed-relative multinomial space* of the *transform independent* is zero,

$$\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{X}(T)}) = 0$$

In the case where the *histogram* is *integral*, $A \in \mathcal{A}_{\mathrm{i}}$, and the *transform-independent* is an *integral iso-transform-independent*, $A^{\mathrm{X}(T)} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, the *transform-independent-distributed-relative multinomial space* is conjectured to be greater than or equal to zero and less than or equal to the *transform-independent-distributed-relative multinomial space* of the *transform-dependent*,

$$0 \leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) \leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{Y}(T)})$$

This is consistent with the *entropies*,

$$\mathrm{entropy}(A^{\mathrm{X}(T)}) \geq \mathrm{entropy}(A) \geq \mathrm{entropy}(A^{\mathrm{Y}(T)})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, where $T_{\mathrm{f}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$, the *transform-independent* equals the *independent*, $A^{\mathrm{X}(T_{\mathrm{f}})} = A^{\mathrm{X}}$, and so the *transform-independent-distributed-relative multinomial space* equals the *alignment*,

$$\begin{aligned}\mathrm{spaceRelative}(A^{\mathrm{X}(T_{\mathrm{f}})})(A) &= \mathrm{algn}(A) \\ &= \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S\end{aligned}$$

At the other extreme where the *transform* is *unary*, $T = T_{\mathrm{u}}$, where $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$, the *transform-independent* equals the *scaled normalised cartesian*, $A^{\mathrm{X}(T_{\mathrm{u}})} = \mathrm{scalar}(z/v) * V^{\mathrm{C}}$, and so the *transform-independent-distributed-relative multinomial space* simplifies to

$$\mathrm{spaceRelative}(A^{\mathrm{X}(T_{\mathrm{u}})})(A) = \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - v \ln \Gamma_!(z/v)$$

Conjecture that if the *transform-independent* is an *integral iso-transform-independent*, $A^{\mathrm{X}(T)} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, the *relative space* of the *histogram* with respect to the *transform-independent* is greater than or equal to that with respect to the *independent*,

$$\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) \geq \mathrm{spaceRelative}(A^{\mathrm{X}})(A) = \mathrm{algn}(A)$$

and similarly for the *transform-dependent*

$$\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{Y}(T)}) \geq \mathrm{spaceRelative}(A^{\mathrm{X}})(A^{\mathrm{Y}(T)}) = \mathrm{algn}(A^{\mathrm{Y}(T)})$$

because the *transform-independent entropy* is greater than or equal to the *independent entropy*, $\mathrm{entropy}(A^{\mathrm{X}(T)}) \geq \mathrm{entropy}(A^{\mathrm{X}})$.

Of the *integral iso-independents*, only the *independent* has zero *alignment*,

$$\forall B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})\ (B \neq A^{\mathrm{X}} \implies \mathrm{algn}(B) > 0)$$

Similarly, of the *integral iso-transform-independents*, only the *transform-independent* has zero *transform-independent-distributed-relative multinomial space*,

$$\forall B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}{*}T),(A{*}T)^{\mathrm{X}}))\ (B \neq A^{\mathrm{X}(T)} \implies \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(B) > 0)$$

where the *transform-independent* is an *integral iso-transform-independent*, $A^{\mathrm{X}(T)} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$.

Insofar as the *transform* approximates to *full functional*, $T \approx T_{\mathrm{f}}$, the *transform-independent* approximates to the *independent*, $A^{\mathrm{X}(T)} \approx A^{\mathrm{X}}$, and the *transform-independent-distributed-relative multinomial space* approximates to

$$\begin{aligned}\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) &\approx \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}(T)}_S \\ &\approx z \times \mathrm{entropy}(A^{\mathrm{X}(T)}) - z \times \mathrm{entropy}(A)\end{aligned}$$

The difference is the *mis-transform-independent-distributed-relative multinomial space*,

$$\sum_{S \in A^{\mathrm{XS}}} (A_S - A^{\mathrm{X}(T)}_S) \ln A^{\mathrm{X}(T)}_S$$

The degree to which the set of *integral iso-transform-independents* is *aligned-like* is the *iso-independence*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \ \cap \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \ \cup \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

As the *iso-independence* increases, the *transform-independent*, $A^{\mathrm{X}(T)}$, tends to the *independent*, A^{X}, and the *relative space*, $\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A)$, tends to the *alignment*, $\mathrm{algn}(A)$. That is, as the *integral iso-transform-independents* becomes more *aligned-like*, the *relative space* becomes less dependent on the *transform*, T.

The set of *iso-transform-independents* may be more *law-like* then the set of *iso-abstracts*, depending on the *iso-derivedence*, but is still *entity-like*. So the *derived*, $A*T$, and the *dependent derived*, $A^{\mathrm{Y}(T)}*T$, are not necessarily equal to each other and nor are they necessarily equal to the *abstract*, $(A*T)^{\mathrm{X}}$. The set of *iso-transform-independents* is the intersection of the set of *iso-formals* and the set of *iso-abstracts*. The *independent analogue* of the *iso-abstracts* is the *naturalised abstract*, $(A*T)^{\mathrm{X}}*T^{\dagger}$, which has zero *derived alignment*, $\mathrm{algn}((A*T)^{\mathrm{X}}*T^{\dagger}*T) = \mathrm{algn}((A*T)^{\mathrm{X}}) = 0$. So the *derived alignment* of the *independent analogue* of the *iso-transform-independents*, $\mathrm{algn}(A^{\mathrm{X}(T)}*T)$, is conjectured to be less than or equal to the *derived alignment*, $\mathrm{algn}(A*T)$, which in turn is conjectured to be less than or equal to the *dependent analogue derived alignment*, $\mathrm{algn}(A^{\mathrm{Y}(T)}*T)$. The relation between the *relative spaces*,

$$\begin{aligned} 0 &= \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{X}(T)}) \\ &\leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) \\ &\leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{Y}(T)}) \end{aligned}$$

can be *lifted*,

$$\begin{aligned} 0 &= \mathrm{spaceRelative}(A^{\mathrm{X}(T)}*T)(A^{\mathrm{X}(T)}*T) \\ &\leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)}*T)(A*T) \\ &\leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)}*T)(A^{\mathrm{Y}(T)}*T) \end{aligned}$$

so conjecture that

$$\mathrm{algn}(A^{\mathrm{X}(T)}*T) \leq \mathrm{algn}(A*T) \leq \mathrm{algn}(A^{\mathrm{Y}(T)}*T)$$

Similarly, the *independent analogue* of the *iso-formals* is the *naturalised formal*, $A^{\mathrm{X}}*T*T^{\dagger}$, which is *formal*, $A^{\mathrm{X}}*T*T^{\dagger}*T = A^{\mathrm{X}}*T$, and so the *derived alignment* is equal to the *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}}*T*T^{\dagger}*T) = \mathrm{algn}(A^{\mathrm{X}}*T)$. Conjecture that the *formal alignment* of the *naturalised formal* is greater than or equal to its *derived alignment*, $\mathrm{algn}((A^{\mathrm{X}}*T*T^{\dagger})^{\mathrm{X}}*T) \geq \mathrm{algn}(A^{\mathrm{X}}*T*T^{\dagger}*T)$. So conjecture that the *formal alignment* of the *naturalised formal* is greater than or equal to the *formal alignment* of the *histogram*, $\mathrm{algn}((A^{\mathrm{X}}*T*T^{\dagger})^{\mathrm{X}}*T) \geq \mathrm{algn}(A^{\mathrm{X}}*T)$. The *transform-dependent*, is near the *histogram*, $A^{\mathrm{Y}(T)} \sim A$, only in as much as it is far from the *transform-independent*, $A^{\mathrm{Y}(T)} \nsim A^{\mathrm{X}(T)}$, so conjecture that the *formal alignment* of the *independent analogue* of the *iso-transform-independents*, $\mathrm{algn}(A^{\mathrm{X}(T)\mathrm{X}} * T)$, is greater than or equal to the *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * T)$, which in turn is greater than or equal to the *dependent analogue formal alignment*, $\mathrm{algn}(A^{\mathrm{Y}(T)\mathrm{X}} * T)$,

$$\mathrm{algn}(A^{\mathrm{X}(T)\mathrm{X}} * T) \geq \mathrm{algn}(A^{\mathrm{X}} * T) \geq \mathrm{algn}(A^{\mathrm{Y}(T)\mathrm{X}} * T)$$

That is, the *dependent analogue derived alignment* is greater than or equal to the *derived alignment*, $\mathrm{algn}(A^{\mathrm{Y}(T)} * T) \geq \mathrm{algn}(A * T)$, but the *dependent analogue formal alignment* is less than or equal to the *formal alignment*, $\mathrm{algn}(A^{\mathrm{Y}(T)\mathrm{X}} * T) \leq \mathrm{algn}(A^{\mathrm{X}} * T)$.

The *iso-abstractence* of the *iso-transform-independents* is

$$\frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \ \leq \ 1$$

As the *iso-abstractence* increases, the *iso-transform-independents* becomes more *entity-like* and the difference between the *relative spaces* of the *dependents*, $\mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A^{\mathrm{W}(T)}) - \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{Y}(T)})$, decreases.

The *iso-derivedence* of the *iso-transform-independents* is

$$\frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \ \cap \ D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \ \cup \ D^{-1}_{U,\mathrm{i},T,z}(A * T)|}$$

As the *iso-derivedence* increases, the *iso-transform-independents* becomes more *law-like* and the difference between the *relative spaces* of the *dependents*, $\mathrm{spaceRelative}((A*T)^{\mathrm{X}}*T^{\dagger})(A^{\mathrm{Y}(T)}) - \mathrm{spaceRelative}(A*T*T^{\dagger})(A^{\mathrm{D}(T)})$, decreases, in the case where the *formal* equals the *abstract*, $A^{\mathrm{X}}*T = (A*T)^{\mathrm{X}}$.

Given a *histogram-transform* pair $(A, T) \in \mathcal{O}_{U,z}$, the *partition-independent*, $A^{\mathrm{P}(T)} \in \mathcal{A}_{U,V,z}$, is defined in section 'Likely histograms', above, as

$$\{A^{\mathrm{P}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D, z)(B) : B \in \mathrm{isop}(U)(T, A))) : D \in \mathcal{A}_{U,V,z}\})$$

where the *integral iso-partition-independents* is abbreviated

$$\mathrm{isop}(U)(T, A) \ := \ Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

and the *iso-partition-independents* is such that

$$\begin{aligned}
&Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \\
&\quad = \ \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ (B^{\mathrm{X}} * T)^{\mathrm{X}} = (A^{\mathrm{X}} * T)^{\mathrm{X}},\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}
\end{aligned}$$

The corresponding *dependent analogue* is the *partition-dependent*, $A^{\mathrm{R}(T)} \in \mathcal{A}_{U,V,z}$, defined

$$\{A^{\mathrm{R}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *iso-partition-independents* is intermediate between the *iso-transform-independents* and the *iso-abstracts*,

$$\begin{aligned}
&\quad Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T) \\
&\subseteq \ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \\
&\subseteq \ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})
\end{aligned}$$

All three *iso-sets* are subsets of the *iso-abstracts*, $Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})$. The *iso-formal-independents*, $Y^{-1}_{U,T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})$, is a superset of the *iso-formals*, $Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}}*T)$. So the properties of the *likely histograms* of the *iso-partition-independents* are also intermediate between the *likely histograms* of the *iso-transform-independents* and the *likely histograms* of the *iso-abstracts*.

The *partition-independent-distributed-relative multinomial space* is

$$\mathrm{spaceRelative}(A^{\mathrm{P}(T)})(A) \ := \ -\ln \frac{\mathrm{mpdf}(U)(A^{\mathrm{P}(T)},z)(A)}{\mathrm{mpdf}(U)(A^{\mathrm{P}(T)},z)(A^{\mathrm{P}(T)})}$$

The *partition-independent-distributed-relative multinomial space* of the *partition independent* is zero,

$$\mathrm{spaceRelative}(A^{\mathrm{P}(T)})(A^{\mathrm{P}(T)}) = 0$$

In the case where the *histogram* is *integral*, $A \in \mathcal{A}_{\mathrm{i}}$, and the *partition-independent* is an *integral iso-partition-independent*, $A^{\mathrm{P}(T)} \in \mathrm{isop}(U)(T,A)$, the *partition-independent-distributed-relative multinomial space* is conjectured to be greater than or equal to zero and less than or equal to the *partition-independent-distributed-relative multinomial space* of the *partition-dependent*,

$$0 \leq \mathrm{spaceRelative}(A^{\mathrm{P}(T)})(A) \leq \mathrm{spaceRelative}(A^{\mathrm{P}(T)})(A^{\mathrm{R}(T)})$$

This is consistent with the *entropies*,

$$\text{entropy}(A^{\mathrm{P}(T)}) \geq \text{entropy}(A) \geq \text{entropy}(A^{\mathrm{R}(T)})$$

In the case where the *transform* is *full functional*, $T = T_{\mathrm{f}}$, where $T_{\mathrm{f}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$, the *partition-independent* equals the *independent*, $A^{\mathrm{P}(T_{\mathrm{f}})} = A^{\mathrm{X}}$, and so the *partition-independent-distributed-relative multinomial space* equals the *alignment*,

$$\begin{aligned}
\text{spaceRelative}(A^{\mathrm{P}(T_{\mathrm{f}})})(A) &= \text{algn}(A) \\
&= \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S
\end{aligned}$$

At the other extreme where the *transform* is *unary*, $T = T_{\mathrm{u}}$, where $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,V}$, the *partition-independent* equals the *scaled normalised cartesian*, $A^{\mathrm{P}(T_{\mathrm{u}})} = \text{scalar}(z/v) * V^{\mathrm{C}}$, and so the *partition-independent-distributed-relative multinomial space* simplifies to

$$\text{spaceRelative}(A^{\mathrm{P}(T_{\mathrm{u}})})(A) = \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - v \ln \Gamma_!(z/v)$$

Conjecture that if the *partition-independent* is *integral*, $A^{\mathrm{P}(T)} \in \text{isop}(U)(T, A)$, the *relative space* of the *histogram* with respect to the *partition-independent* is greater than or equal to that with respect to the *independent*,

$$\text{spaceRelative}(A^{\mathrm{P}(T)})(A) \geq \text{spaceRelative}(A^{\mathrm{X}})(A) = \text{algn}(A)$$

and similarly for the *partition-dependent*

$$\text{spaceRelative}(A^{\mathrm{P}(T)})(A^{\mathrm{R}(T)}) \geq \text{spaceRelative}(A^{\mathrm{X}})(A^{\mathrm{R}(T)}) = \text{algn}(A^{\mathrm{R}(T)})$$

because the *partition-independent entropy* is greater than or equal to the *independent entropy*, $\text{entropy}(A^{\mathrm{P}(T)}) \geq \text{entropy}(A^{\mathrm{X}})$.

Of the *integral iso-partition-independents*, only the *partition-independent* has zero *partition-independent-distributed-relative multinomial space*,

$$\forall B \in \text{isop}(U)(T, A)\ (B \neq A^{\mathrm{P}(T)} \implies \text{spaceRelative}(A^{\mathrm{P}(T)})(B) > 0)$$

where the *partition-independent* is *integral*, $A^{\mathrm{P}(T)} \in \text{isop}(U)(T, A)$.

Insofar as the *transform* approximates to *full functional*, $T \approx T_{\mathrm{f}}$, the *partition-independent* approximates to the *independent*, $A^{\mathrm{P}(T)} \approx A^{\mathrm{X}}$, and the *partition-independent-distributed-relative multinomial space* approximates to

$$\begin{aligned}
\mathrm{spaceRelative}(A^{\mathrm{P}(T)})(A) &\approx \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A_S^{\mathrm{P}(T)} \\
&\approx z \times \mathrm{entropy}(A^{\mathrm{P}(T)}) - z \times \mathrm{entropy}(A)
\end{aligned}$$

The difference is the *mis-partition-independent-distributed-relative multinomial space*,

$$\sum_{S \in A^{\mathrm{XS}}} (A_S - A_S^{\mathrm{P}(T)}) \ln A_S^{\mathrm{P}(T)}$$

Just as for the *iso-abstracts* and the *iso-transform-independents*, above, it is conjectured that, because the *independent analogue* of the *iso-abstracts* is the *naturalised abstract*, $(A*T)^{\mathrm{X}}*T^{\dagger}$, which has zero *derived alignment*, $\mathrm{algn}((A*T)^{\mathrm{X}}*T^{\dagger}*T) = \mathrm{algn}((A*T)^{\mathrm{X}}) = 0$, the *derived alignment* of the *independent analogue* of the *iso-partition-independents*, $\mathrm{algn}(A^{\mathrm{P}(T)}*T)$, is conjectured to be less than or equal to the *derived alignment*, $\mathrm{algn}(A*T)$, which in turn is conjectured to be less than or equal to the *dependent analogue derived alignment*, $\mathrm{algn}(A^{\mathrm{R}(T)}*T)$. So the relation between the *relative spaces*,

$$\begin{aligned}
0 &= \mathrm{spaceRelative}(A^{\mathrm{P}(T)})(A^{\mathrm{P}(T)}) \\
&\leq \mathrm{spaceRelative}(A^{\mathrm{P}(T)})(A) \\
&\leq \mathrm{spaceRelative}(A^{\mathrm{P}(T)})(A^{\mathrm{R}(T)})
\end{aligned}$$

can be *lifted* to *derived alignment*, depending on the *derived iso-independence*,

$$\mathrm{algn}(A^{\mathrm{P}(T)}*T) \leq \mathrm{algn}(A*T) \leq \mathrm{algn}(A^{\mathrm{R}(T)}*T)$$

Although the properties of the *likely histograms* of the *iso-partition independents* with respect to *derived alignment* are similar to those of the *likely histograms* of both the *iso-abstracts* and the *iso-transform-independents*, because all three *iso-sets* are subsets of the *iso-abstracts*, the properties with respect to *formal alignment* are not similar. The set of *iso-abstracts* is conditional on neither the *formal*, $A^{\mathrm{X}}*T$, nor the *formal independent*, $(A^{\mathrm{X}}*T)^{\mathrm{X}}$, so the *abstract dependent*, $A^{\mathrm{W}(T)}$, is neutral with respect to the *formal* and the *formal independent*, and nothing can be said of its *formal alignment*, $\mathrm{algn}(A^{\mathrm{W}(T)\mathrm{X}}*T)$. Indeed in some cases the *abstract dependent* may be purely *formal*, $A^{\mathrm{W}(T)}*T = A^{\mathrm{W}(T)\mathrm{X}}*T \implies \mathrm{algn}(A^{\mathrm{W}(T)}*T) = \mathrm{algn}(A^{\mathrm{W}(T)\mathrm{X}}*T)$. This contrasts with the set of *iso-transform-independents* which is conditional on the *formal*. It is conjectured above that the *formal alignment* of the *independent analogue* of the *iso-transform-independents*, $\mathrm{algn}(A^{\mathrm{X}(T)\mathrm{X}}*T)$, is greater than or equal to the *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * T)$, which in turn is greater than or equal to the *dependent analogue formal alignment*, $\mathrm{algn}(A^{\mathrm{Y}(T)\mathrm{X}} * T)$,

$$\mathrm{algn}(A^{\mathrm{X}(T)\mathrm{X}} * T) \geq \mathrm{algn}(A^{\mathrm{X}} * T) \geq \mathrm{algn}(A^{\mathrm{Y}(T)\mathrm{X}} * T)$$

Now consider the *formal alignment* of the *likely histograms* of the *iso-partition-independents*. The *independent analogue* of the *iso-partition-independents*, $A^{\mathrm{P}(T)}$, is intermediate between (i) the *independent analogue* of the *iso-formal-independents*, which is the *naturalised formal independent*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\dagger}$, and (ii) the *independent analogue* of the *iso-abstracts*, which is the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$. Conjecture that the triply-*independent formal alignment* of the *naturalised formal independent* is less than or equal to the doubly-*independent formal alignment* of the *naturalised abstract* which in turn is less than or equal to the singly-*independent formal alignment* of the *histogram*,

$$\mathrm{algn}(((A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\dagger})^{\mathrm{X}} * T) \ \leq\ \mathrm{algn}(((A * T)^{\mathrm{X}} * T^{\dagger})^{\mathrm{X}} * T) \ \leq\ \mathrm{algn}(A^{\mathrm{X}} * T)$$

So the *formal alignment* of the *partition-independent* is less than or equal to the *formal alignment* of the *histogram*,

$$\mathrm{algn}(A^{\mathrm{P}(T)\mathrm{X}} * T) \ \leq\ \mathrm{algn}(A^{\mathrm{X}} * T)$$

The *partition-dependent* varies against the *partition-independent*, $A^{\mathrm{R}(T)} \nsim A^{\mathrm{P}(T)}$, so conjecture that

$$\mathrm{algn}(A^{\mathrm{P}(T)\mathrm{X}} * T) \ \leq\ \mathrm{algn}(A^{\mathrm{X}} * T) \ \leq\ \mathrm{algn}(A^{\mathrm{R}(T)\mathrm{X}} * T)$$

That is, the *dependent analogue derived alignment* is greater than or equal to the *derived alignment*, $\mathrm{algn}(A^{\mathrm{R}(T)} * T) \geq \mathrm{algn}(A * T)$, and the *dependent analogue formal alignment* is greater than or equal to the *formal alignment*, $\mathrm{algn}(A^{\mathrm{R}(T)\mathrm{X}} * T) \geq \mathrm{algn}(A^{\mathrm{X}} * T)$.

The direction of the *formal alignment* inequality is opposite to that of the *likely histograms* of the *iso-transform-independents*. So conjecture that the *partition-dependent formal alignment* is greater than or equal to the *transform-dependent formal alignment*

$$\mathrm{algn}(A^{\mathrm{R}(T)\mathrm{X}} * T) \ \geq\ \mathrm{algn}(A^{\mathrm{X}} * T) \ \geq\ \mathrm{algn}(A^{\mathrm{Y}(T)\mathrm{X}} * T)$$

4.16 Rolled alignment

A *roll*, defined above, $R \in \mathrm{rolls} \subset \mathcal{S} \to \mathcal{S}$, in *variables* V and *system* U, is a *state* valued function of *state*, $R \in V^{\mathrm{CS}} \to V^{\mathrm{CS}}$. The application of a *roll* R in *variables* V to a *histogram* A having the same *variables* is $\mathrm{roll} \in \mathrm{rolls} \times \mathcal{A} \to \mathcal{A}$

$$\mathrm{roll}(R, A) := \sum_{S \in A^{\mathrm{S}} \setminus \mathrm{dom}(R)} \{(S, A_S)\} + \sum_{S \in A^{\mathrm{S}} \cap \mathrm{dom}(R)} \{(R_S, A_S)\}$$

Define $(*) \in \mathcal{A} \times \mathrm{rolls} \to \mathcal{A}$ as $A * R := \mathrm{roll}(R, A)$. Define the *identity roll* $\mathrm{id}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{rolls}$ as $\mathrm{id}(U)(V) := \{(S, S) : S \in V^{\mathrm{CS}}\}$.

A *roll* $R \in \mathrm{rolls}$ having *variables* V can converted to a *partition transform* $P^{\mathrm{T}} \in \mathcal{T}_{U,V}$ on the *partition* $P \in \mathcal{R}_U$ of the *cartesian states* of the *variables*, $P \in \mathrm{B}(V^{\mathrm{CS}})$, implied by the functional inverse, $P = \mathrm{ran}(\mathrm{inverse}(R \circ \mathrm{id}(U)(V)))$. This *transform* has a single *derived variable*, $|\mathrm{der}(P^{\mathrm{T}})| = 1$, and therefore the *derived histogram* is *independent*, $A * P^{\mathrm{T}} = (A * P^{\mathrm{T}})^{\mathrm{X}}$, when applied to some *underlying histogram* A in *variables* V. Hence, the *derived alignment* is zero, $\mathrm{alignment}(A * P^{\mathrm{T}}) = 0$.

The *transform* of a *roll* is defined $\mathrm{transform}(U) \in \mathrm{rolls} \to \mathcal{T}_{U,\mathrm{f},1}$ as

$$\begin{aligned}&\mathrm{transform}(U)(R) := \\ &\quad \{P^{\mathrm{T}} : v \in V,\ P = \mathrm{ran}(\mathrm{inverse}(\{(S_1, S_2\%\{v\}) : (S_1, S_2) \in R'\}))\}^{\mathrm{T}}\end{aligned}$$

where $V = \mathrm{vars}(R)$ and $R' = R \circ \mathrm{id}(U)(V)$ is the given *roll* stuffed with the *identity roll*. Define $R^{\mathrm{T}} := \mathrm{transform}(U)(R)$ where the *system* U is implicit.

The *alignment* of the *derived histogram* in *variables* $W = \mathrm{der}(R^{\mathrm{T}})$ is equal to the *alignment* in *variables* $V = \mathrm{vars}(R) = \mathrm{und}(R^{\mathrm{T}})$ of the *rolled histogram*

$$\mathrm{alignment}(A * R^{\mathrm{T}}) = \mathrm{alignment}(A * R)$$

This is the case because (i) the *sizes* are equal, $\mathrm{size}(A*R^{\mathrm{T}}) = \mathrm{size}(A*R)$, (ii) there is a map between *states*, $V^{\mathrm{CS}} \leftrightarrow W^{\mathrm{CS}}$, such that the *counts* are equal, $\exists M \in V^{\mathrm{CS}} \leftrightarrow W^{\mathrm{CS}}\ (|M| = |W^{\mathrm{CS}}| \wedge (\forall (S,T) \in M\ ((A * R)_S = (A * R^{\mathrm{T}})_T)))$ and hence

$$\sum_{S \in V^{\mathrm{CS}}} \ln \Gamma_!(A * R)_S = \sum_{T \in W^{\mathrm{CS}}} \ln \Gamma_!(A * R^{\mathrm{T}})_T$$

(iii) there exists a surjective mapping between the *variables*, $V \to W$, such that the *reductions* are equal, $\exists N \in V \to W\ (\mathrm{dom}(N) = V \wedge \mathrm{ran}(N) = W \wedge (\ \forall (v, w) \in N\ \exists Q \in \{v\}^{\mathrm{CS}} \leftrightarrow \{w\}^{\mathrm{CS}}\ (|Q| = |U_w| \wedge (\forall (S,T) \in Q\ (((A * R)\%\{v\})_S = ((A * R^{\mathrm{T}})\%\{w\})_T)))))$, so

$$\sum_{S \in V^{\mathrm{CS}}} \ln \Gamma_!(A * R)^{\mathrm{X}}_S = \sum_{T \in W^{\mathrm{CS}}} \ln \Gamma_!(A * R^{\mathrm{T}})^{\mathrm{X}}_T$$

A *value roll*, defined above, is equivalent to a special case of a *roll* $R \in V^{\mathrm{CS}} \to V^{\mathrm{CS}}$ on *variables* V in *system* U. Define the set of *value rolls* $\mathrm{rollValues}(U) \subset \mathrm{P}(\mathcal{V}_U) \times \mathcal{V}_U \times \mathcal{W}_U \times \mathcal{W}_U$ such that $\forall (V,v,s,t) \in \mathrm{rollValues}(U)\ (v \in V \wedge s,t \in U_v)$. Define $\mathrm{roll}(U) \in \mathrm{rollValues}(U) \to (\mathrm{rolls} \cap (\mathcal{S}_U \to \mathcal{S}_U))$. Define $(V,v,s,t)^{\mathrm{R}} := \mathrm{roll}(U)((V,v,s,t))$.

The *independent* of the application of a *value roll* (V,v,s,t) to a *histogram* A is equal to the application of a *value roll* to the *independent histogram*

$$(A * (V,v,s,t)^{\mathrm{R}})^{\mathrm{X}} = A^{\mathrm{X}} * (V,v,s,t)^{\mathrm{R}}$$

Thus the application of a *value roll* to an *independent histogram* is also *independent*, $A^{\mathrm{X}} * (V,v,s,t)^{\mathrm{R}} = (A^{\mathrm{X}} * (V,v,s,t)^{\mathrm{R}})^{\mathrm{X}}$. The *alignment* of the *rolled independent* is zero, $\mathrm{alignment}(A^{\mathrm{X}} * (V,v,s,t)^{\mathrm{R}}) = 0$.

$\mathcal{J}_{U,V}$ is defined above as the set of lists of *value rolls* in *variables* V and *system* U, $\mathcal{J}_{U,V} = \{L : L \in \mathcal{L}(\mathrm{rollValues}(U)),\ (\forall (W,v,s,t) \in \mathrm{set}(L)\ (W = V))\}$. The list of *value rolls* $J \in \mathcal{J}_{U,V}$ can be converted into a list of *rolls*. Define $J^{\mathrm{R}} := \mathrm{map}(\mathrm{roll}(U), J)$. *Rolled independent histograms* remain *independent* $A^{\mathrm{X}} * J^{\mathrm{R}} = (A^{\mathrm{X}} * J^{\mathrm{R}})^{\mathrm{X}}$.

The *transform* of a list of *value rolls* $J \in \mathcal{J}_{U,V}$ is defined above, $\mathrm{transform}(U) \in \mathcal{L}(\mathrm{rollValues}(U)) \to \mathcal{T}_{U,\mathrm{f},1}$. Define $J^{\mathrm{T}} := \mathrm{transform}(U)(J)$. The *value roll list transform* equals the *roll list transform*, $J^{\mathrm{T}} = J^{\mathrm{RT}}$.

The *derived alignment* equals the *alignment* of the *rolled histogram*

$$\mathrm{alignment}(A * J^{\mathrm{T}}) = \mathrm{alignment}(A * J^{\mathrm{RT}}) = \mathrm{alignment}(A * J^{\mathrm{R}})$$

The *alignment* of a *rolled independent histogram* is zero, $\mathrm{alignment}(A^{\mathrm{X}} * J^{\mathrm{T}}) = 0$ because $A^{\mathrm{X}} * J^{\mathrm{T}} = (A^{\mathrm{X}} * J^{\mathrm{T}})^{\mathrm{X}}$. The *value roll list transform* is *non-overlapping*, $\neg\mathrm{overlap}(J^{\mathrm{T}})$, so the *degree of overlap is zero*, $\mathrm{alignment}(V^{\mathrm{C}} * J^{\mathrm{T}}) = 0$.

4.17 Decomposition alignment

In section 'Decompositions', above, the *nullable transform*, D^{T}, of a *well behaved distinct decomposition* $D \in \mathcal{D}_{\mathrm{w},U}$ was defined such that the *derived variables*, $\mathrm{der}(D^{\mathrm{T}})$, consist of the union of (i) the *root frame variables*, $\{\{w\}^{\mathrm{CS}\{\}} : w \in W_{\mathrm{r}}\}$, of the *derived variables* $W_{\mathrm{r}} = \mathrm{der}(T_{\mathrm{r}})$ of the *root transform* $\{((\emptyset, T_{\mathrm{r}}), \cdot)\} = D$, and (ii) the *nullable variables*, $\mathrm{dom}(\mathrm{nullables}(U)(D))$. The *nullable transform derived variables*, $\mathrm{der}(D^{\mathrm{T}})$, originate in the *transform* *derived variables*, $\mathrm{der}(G)$, where $G = \mathrm{transforms}(D)$. So $\mathrm{originals}(U)(D) \in \mathrm{der}(D^{\mathrm{T}}) \to \mathrm{der}(G)$. The *derived alignment* given *histogram* $A \in \mathcal{A}$, having $\mathrm{vars}(A) \supseteq \mathrm{und}(D)$, is $\mathrm{algn}(A * D^{\mathrm{T}})$.

In some cases the *derived alignment*, $\mathrm{algn}(A * D^{\mathrm{T}})$, may not be practicably computable. The *derived volume*, $|N^{\mathrm{C}}|$ where $N = \mathrm{der}(D^{\mathrm{T}})$, may be larger than the *underlying volume*, $|V^{\mathrm{C}}|$ where $V = \mathrm{und}(D)$. Certainly the *volume* of the *nullable transform derived variables* may be greater than the *volume* of the *transform derived variables*, $|N^{\mathrm{C}}| \geq |W^{\mathrm{C}}|$ where $W = \mathrm{der}(G)$. For example, the application to *cartesian* may not be *completely effective*, $(V^{\mathrm{C}} * D^{\mathrm{T}})^{\mathrm{F}} < N^{\mathrm{C}}$, if (i) the *root transform*, T_{r}, is *overlapping*, $\mathrm{overlap}(T_{\mathrm{r}})$, or (ii) there is more than one path, $|\mathrm{path}(D)| > 1$, and no *transform symmetries*, and hence *overlapping* contingently applied *nullable variables*, $\mathrm{nullables}(U)(D) \neq \emptyset$. However, it is conjectured that, given certain conditions, there is a calculation of the *content alignment*, $\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}})$, that does not require the computation of the *derived histogram*, $A * D^{\mathrm{T}}$.

Let $D \in \mathcal{D}_{\mathrm{w},U}$ be a *well behaved decomposition* having no *variable symmetries*, $\{(w, (S,T)) : (S,T) \in \mathrm{elements}(D),\ w \in \mathrm{der}(T)\} \in \mathrm{der}(G) \to \mathrm{elements}(D)$, where $G = \mathrm{transforms}(D)$. First, consider the case where the *transforms* of the *decomposition*, D, are all *contingently diagonalised* with respect to *histogram* A,

$$\forall (C,T) \in \mathrm{cont}(D)\ (\mathrm{diagonal}(A * C * T))$$

where $\mathrm{cont}(D) := \mathrm{elements}(\mathrm{contingents}(D))$. Conjecture that in the *contingently diagonalised decomposition* case the *derived alignment*, $\mathrm{algn}(A * D^{\mathrm{T}})$, equals (i) the *derived alignment* of the *skeletal contingent reduction* plus (ii) the sum of the *alignments* of the *diagonalised derived histograms* of the *decomposition transforms*. Let the *skeletal contingent reduction* be $D' \in \mathrm{reductions}(A, D)$ which is such that $\mathrm{skeletal}(A * D'^{\mathrm{T}})$. Then

$$\mathrm{algn}(A * D^{\mathrm{T}}) = \mathrm{algn}(A * D'^{\mathrm{T}}) + \sum_{(C,T) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * T)$$

To show this, consider the *fud* F of children *transforms* of a *decomposition transform* $T \in G$, where $G = \mathrm{transforms}(D)$, $F = \mathrm{ran}(\mathrm{dom}(E))$ and $((\cdot, T), E) \in \mathrm{nodes}(D)$. Let $N = \{\mathrm{flip}(\mathrm{originals}(U)(D))(w) : w \in \mathrm{der}(F)\}$ be the set of *nullable variables* of the *nullable transform* which correspond to the *originating derived variables*, because the *originating* map is bijective, $\mathrm{originals}(U)(D) \in \mathrm{der}(D^{\mathrm{T}}) \leftrightarrow \mathrm{der}(G)$, where there are no *variable symmetries*. Then the *reduced derived histogram* of the *nullable variables* corresponding to F is $B = A * D^{\mathrm{T}} \ \% \ N$, and its *alignment* is $\mathrm{algn}(B)$. This *histogram*, B, has an *axial reduction* as defined in the section 'Alignment of axial reductions', above. Let $T' \in \mathrm{transforms}(D')$ be the *reduced transform* in the *contingently diagonalised decomposition*, D', that corresponds to T in the *decomposition* D. The corresponding *skeletal contingent reduction derived histogram* is $B' = A * D'^{\mathrm{T}} \ \% \ N'$ where $F' = \mathrm{ran}(\mathrm{dom}(E'))$ and $((\cdot, T'), E') \in \mathrm{nodes}(D')$. $N' \subseteq \mathrm{der}(D'^{\mathrm{T}})$ is the set of *nullable variables* of F' corresponding to N. B' is an *axial reduction* of B. That is, there exists a *pivot state*, $N_* \in N^{\mathrm{CS}}$ where $N_* \subseteq \mathrm{nullables}(U)(D)$ and $\mathrm{dom}(N_*) = N$, such that the implied set of *slices* are *diagonalised*, $\forall (C,R) \in \mathrm{cont}(D)\ (R \in F \implies \mathrm{diagonal}(A * C * R))$. It is conjectured that the *alignment* of the *histogram* B equals (i) the *alignment* of the *axial reduction*, B', plus (ii) the sum of the *alignments* of the *sliced diagonalised reductions*

$$\mathrm{algn}(B) = \mathrm{algn}(B') + \sum(\mathrm{algn}(A * C * R) : (C,R) \in \mathrm{cont}(D),\ R \in F)$$

The *contingent slice diagonalisations*, $\{A*C*R : (C,R) \in \mathrm{cont}(D),\ R \in F\}$, are *axially independent* of eachother and *axially independent* of the *axial reduction*, B', although, of course, they do not have the same *variables*. That is, they are *axially independent* in the sense defined in the section 'Alignment of axial reductions', above, where the *slices* partition the *underlying variables*.

Then conjecture that the same separation of the *alignment*, $\mathrm{algn}(B)$, of the particular children *transforms*, F, of a *decomposition transform* T, into components of (i) *axial alignment*, $\mathrm{algn}(B')$, and (ii) sum *diagonal alignments*, $\sum(\mathrm{algn}(A*C*R) : (C,R) \in \mathrm{cont}(D),\ R \in F)$, can be extended to the whole *decomposition*, D. That is, the *derived alignment*, $\mathrm{algn}(A * D^{\mathrm{T}})$, may be separated into components of (i) the *skeletal contingent reduction alignment*, $\mathrm{algn}(A*D'^{\mathrm{T}})$, and (ii) sum *diagonal alignments*, $\sum(\mathrm{algn}(A*C*T) : (C,T) \in \mathrm{cont}(D))$, including the *root transform diagonalised alignment*, $\mathrm{algn}(A * T_{\mathrm{r}})$,

$$\mathrm{algn}(A * D^{\mathrm{T}}) = \mathrm{algn}(A * D'^{\mathrm{T}}) + \sum_{(C,T) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * T)$$

The *contingent slice diagonalisations*, $\{A*C*T : (C,T) \in \mathrm{cont}(D)\}$, are *axially independent* of eachother and *axially independent* of the *skeleton*, $A*D'^{\mathrm{T}}$.

Second, if it so happens to be the case that *contingently diagonalised decomposition*, D, is also such that *contingently* the *formal histogram* is equivalent to the *abstract histogram*

$$\forall (C,T) \in \mathrm{cont}(D)\ (A^{\mathrm{X}} * C * T \equiv (A * C * T)^{\mathrm{X}})$$

then the *skeletal contingent reduction* must also be such that *contingently* the *formal histogram* is equivalent to the *abstract histogram*

$$\forall (C,T') \in \mathrm{cont}(D')\ (A^{\mathrm{X}} * C * T' \equiv (A * C * T')^{\mathrm{X}})$$

but, because each of the *reduced transforms* is *mono-derived-variate*, $\forall T' \in \mathrm{transforms}(D')\ (|\mathrm{der}(T')| = 1)$ and so

$$\forall (C,T') \in \mathrm{cont}(D')\ (A * C * T' \equiv (A * C * T')^{\mathrm{X}})$$

the *reduced transforms* are *contingently formal*

$$\forall (C,T') \in \mathrm{cont}(D')\ (A^{\mathrm{X}} * C * T' \equiv A * C * T')$$

so the *skeletal derived* is *formal*, $A * D'^{\mathrm{T}} \equiv A^{\mathrm{X}} * D'^{\mathrm{T}}$, and the *content skeletal alignment* is zero,

$$\mathrm{algn}(A * D'^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D'^{\mathrm{T}}) = 0$$

Also, if *contingently* the *formal histogram* is equivalent to the *abstract histogram*, the *contingent content alignment* of each *transform*, T, with respect to the *slice*, C, equals the *contingent derived alignment*

$$\begin{aligned}\mathrm{algn}(A * C * T) - \mathrm{algn}(A^{\mathrm{X}} * C * T) &= \mathrm{algn}(A * C * T) - \mathrm{algn}((A * C * T)^{\mathrm{X}}) \\ &= \mathrm{algn}(A * C * T)\end{aligned}$$

Thus, given (i) *contingent diagonalisation* and (ii) *contingent formal-abstract equivalence*, the *content alignment* of the *nullable transform* of the *decomposition* equals the sum of the *contingent derived alignments* of the *contingently diagonalised transforms*,

$$\begin{aligned}&\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) \\ &\quad = \mathrm{algn}(A * D'^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D'^{\mathrm{T}}) + \\ &\qquad \sum_{(C,T) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * T) - \mathrm{algn}(A^{\mathrm{X}} * C * T) \\ &\quad = \sum_{(C,T) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * T)\end{aligned}$$

Note that a condition that the *transforms* are separately *non-overlapping*, $\forall T \in G\ (\neg\mathrm{overlap}(T))$, is insufficient to imply *contingent formal-abstract equivalence*. A *non-overlapping transform* implies that $(A * C)^{\mathrm{X}} * T \equiv ((A * C)^{\mathrm{X}} * T)^{\mathrm{X}}$ but does not constrain $A^{\mathrm{X}} * C * T$ to be *independent* nor imply that it is equivalent to $(A * C * T)^{\mathrm{X}}$.

Define the *summation alignment* as $\text{alignmentSum} \in \mathcal{A} \times \mathcal{D} \to \mathbf{R}$ as

$$\text{alignmentSum}(A, D) := \sum_{(C,T) \in \text{cont}(D)} \text{algn}(A * C * T)$$

Given (i) *contingent diagonalisation* and (ii) *contingent formal-abstract equivalence*, the *content alignment* of a *well behaved decomposition* having no *variable symmetries* equals the *summation alignment*

$$\text{algn}(A * D^{\text{T}}) - \text{algn}(A^{\text{X}} * D^{\text{T}}) \quad = \quad \text{alignmentSum}(A, D)$$

In order to calculate the *summation alignment*, $\text{alignmentSum}(A, D)$, only the *contingent alignments* of the recursive *contingents* tree need be computed. The *contingents* tree, $\text{contingents}(D)$, does not depend on the *nullable fud*, $\text{nullable}(U)(D)$, so there is no need to compute any of the *slice transforms* or their dependents. Thus the possibly impracticable *derived volume*, $|N^{\text{C}}|$, of the *nullable transform*, D^{T}, need not be realised.

Let the set of *summation aligned decompositions* $\mathcal{D}_{\Sigma}(A) \subset \mathcal{D}_{\text{w},U}$ be the subset of *well behaved distinct decompositions* having no *variable symmetries* that are subject to these two constraints with respect to *histogram* A,

$$\begin{aligned}
&\forall A \in \mathcal{A}_U \ (\mathcal{D}_{\Sigma}(A) = \\
&\quad \{D : D \in \mathcal{D}_{\text{w},U},\ \text{vars}(D) \subseteq \text{vars}(A), \\
&\qquad\qquad \text{isfunc}(\{(w, (C,T)) : (C,T) \in \text{cont}(D),\ w \in \text{der}(T)\}), \\
&\qquad\qquad \forall (C,T) \in \text{cont}(D)\ (\text{diagonal}(A * C * T)), \\
&\qquad\qquad \forall (C,T) \in \text{cont}(D)\ ((A * C * T)^{\text{X}} \equiv A^{\text{X}} * C * T)\})
\end{aligned}$$

Summation aligned decompositions are such that the *content alignment* equals the *summation alignment*,

$$\forall D \in \mathcal{D}_{\Sigma}(A)\ (\text{algn}(A * D^{\text{T}}) - \text{algn}(A^{\text{X}} * D^{\text{T}}) = \text{algnSum}(A, D))$$

where $\text{algnSum} = \text{alignmentSum}$.

Note that a *summation aligned decomposition* $D \in \mathcal{D}_{\Sigma}(A)$ could consist of *mono-derived-variate transforms*, $\forall T \in G\ (|\text{der}(T)| = 1)$. In this case, all of the *derived histograms* would be *diagonalised* regardless of the *histogram* A, and so the *decomposition* would already be *contingently reduced*, $D = D'$. The *contingent derived alignments*, however, would all be zero, and hence the *content alignment* would be zero, $\text{algn}(A * D^{\text{T}}) - \text{algn}(A^{\text{X}} * D^{\text{T}}) =$

$\text{alignmentSum}(A, D) = 0$. That is, the *summation aligned decomposition* would be *formal* with respect to A, $A * D^{\text{T}} \equiv A^{\text{X}} * D^{\text{T}}$. Therefore, as well as the (i) *contingent diagonalisation* and (ii) *contingent formal-abstract equivalence* constraints, *idealising summation aligned decompositions*, defined below, are also constrained (iii) such that the *contingent derived histograms* are not *independent*,

$$\forall (C, T) \in \text{cont}(D)\ (\neg(A * C * T \equiv (A * C * T)^{\text{X}}))$$

That is, the *contingent derived histogram* is not equivalent to the *contingent abstract histogram*. This constraint implies that an *idealising summation aligned decomposition* must be *pluri-derived-variate* everywhere, $\forall T \in G\ (|\text{der}(T)| > 1)$.

In the case where the *independent histogram* is *integral*, $A^{\text{X}} \in \mathcal{A}_{\text{i}}$, and given the *minimum alignment conjecture*, in section 'Minimum alignment' above, this implies that the *contingent derived alignment* of an *idealising summation aligned decomposition* is everywhere greater than zero,

$$\forall (C, T) \in \text{cont}(D)\ (\text{algn}(A * C * T) > 0)$$

because $(A^{\text{X}} \in \mathcal{A}_{\text{i}}) \wedge (A^{\text{X}} * C * T \equiv (A * C * T)^{\text{X}}) \implies (A * C * T)^{\text{X}} \in \mathcal{A}_{\text{i}}$.

Furthermore, because the *slices* are such that *contingently* the *formal histogram* is equivalent to the *abstract histogram*, $A^{\text{X}} * C * T \equiv (A * C * T)^{\text{X}}$, the *non-independent* constraint, $\neg(A * C * T \equiv (A * C * T)^{\text{X}})$, implies that the *slice* of the *histogram*, $A * C$, is not equivalent to the *slice* of the *independent*, $A^{\text{X}} * C$,

$$\neg(A * C * T \equiv A^{\text{X}} * C * T) \implies \neg(A * C \equiv A^{\text{X}} * C)$$

These three constraints allow an *independent slice*, $A * C \equiv (A * C)^{\text{X}}$, to have *derived alignment*, $\text{algn}((A * C)^{\text{X}} * T) > 0$. That is, where (i) $\text{diagonal}((A * C)^{\text{X}} * T)$, (ii) $((A * C)^{\text{X}} * T)^{\text{X}} \equiv A^{\text{X}} * C * T$ and (iii) $\neg((A * C)^{\text{X}} * T \equiv A^{\text{X}} * C * T)$. Therefore, in order to exclude this case, add a fourth constraint (iv) that the *formal slice* is *independent*

$$\forall (C, T) \in \text{cont}(D)\ ((A * C)^{\text{X}} * T \equiv ((A * C)^{\text{X}} * T)^{\text{X}})$$

Thus the *derived alignment* of an *independent slice* is zero, $\text{algn}((A * C)^{\text{X}} * T) = 0$. This constraint holds in the case where the *transform* is *non-overlapping*, $\neg\text{overlap}(T)$. If the *transform* is *non-overlapping* then the *formal slice* must be *independent*, $\neg\text{overlap}(T) \implies (A * C)^{\text{X}} * T \equiv$

$((A * C)^{\mathrm{X}} * T)^{\mathrm{X}}$. The *non-overlapping* constraint is stricter than necessary, but is such that any *formal slice*, $(A * C)^{\mathrm{X}} * T$, is *independent* regardless of the *histogram*, A, or *slice*, $A * C$.

Let the set of *idealising summation aligned decompositions* $\mathcal{D}_{\Sigma,\mathrm{k}}(A) \subset \mathcal{D}_{\Sigma}(A)$ be the subset of *summation aligned decompositions* that are subject to these two additional constraints with respect to *histogram* A which has *integral independent*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$,

$$\begin{aligned}
&\forall A \in \mathcal{A}_U \cap \mathcal{A}_{\mathrm{xi}}\ (\mathcal{D}_{\Sigma,\mathrm{k}}(A) = \\
&\quad \{D : D \in \mathcal{D}_{\Sigma}(A), \\
&\qquad\qquad \forall (C,T) \in \mathrm{cont}(D)\ (\neg(A * C * T \equiv (A * C * T)^{\mathrm{X}})), \\
&\qquad\qquad \forall (C,T) \in \mathrm{cont}(D)\ ((A * C)^{\mathrm{X}} * T \equiv ((A * C)^{\mathrm{X}} * T)^{\mathrm{X}})\})
\end{aligned}$$

where $\mathcal{A}_{\mathrm{xi}} = \{A : A \in \mathcal{A},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\}$.

There are no *idealising summation aligned decompositions* of an *independent substrate histogram*, $\mathcal{D}_{\Sigma,\mathrm{k}}(A^{\mathrm{X}}) = \emptyset$, because the *formal histogram* is *independent*, $A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * V^{\mathrm{C}})^{\mathrm{X}} * T \equiv ((A^{\mathrm{X}} * V^{\mathrm{C}})^{\mathrm{X}} * T)^{\mathrm{X}}$.

Idealising summation aligned decompositions are such that the *content alignment* is greater than zero,

$$\forall D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ (\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) = \mathrm{algnSum}(A, D) > 0)$$

The length of the *contingent diagonals* is at least two because of the *non-independent* constraint

$$\forall D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ \forall (C,T) \in \mathrm{cont}(D)\ (|(A * C * T)^{\mathrm{F}}| \geq 2)$$

Thus the *slice sizes* must also be at least two

$$\forall D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ \forall (C,T) \in \mathrm{cont}(D)\ (\mathrm{size}(A * C) \geq 2)$$

The length of the *contingent diagonals* is no greater than the *slice size*

$$\forall D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ \forall (C,T) \in \mathrm{cont}(D)\ (|(A * C * T)^{\mathrm{F}}| \leq \mathrm{size}(A * C))$$

The length of the *contingent diagonals* is no greater than the *underlying effective* and hence the *underlying volume*

$$\forall D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ \forall (C,T) \in \mathrm{cont}(D)\ (|(A * C * T)^{\mathrm{F}}| \leq |(A * C * T)^{\mathrm{C}}| \leq |A^{\mathrm{C}}|)$$

The *content alignment* of the *idealising summation aligned decomposition* $D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)$ equals the *summation* of the *derived alignments*, $\mathrm{algn}(A*D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) = \sum(\mathrm{algn}(A * C * T) : (C,T) \in \mathrm{cont}(D))$, so the *derived histograms* of the *slices* are *axially independent* of eachother. Therefore the *derived histogram* of a *slice*, $A*C*T$, is *axially independent* from the *derived histograms* its ancestor *slices*. For example, consider $((C_1,T_1),(C_2,T_2)) \in \mathrm{steps}(\mathrm{contingents}(D))$. The parent *transform* T_1 has non-zero *derived alignment* when constrained to its own *slice* $A*C_1$, $\mathrm{algn}(A*C_1*T_1) > 0$, but when constrained to its child *slice*, $A * C_2 \subseteq A * C_1$, its *derived alignment* is zero, $\mathrm{algn}(A * C_2 * T_1) = 0$, because the *derived histogram* is an *effective singleton*, $|(A * C_2 * T_1)^{\mathrm{F}}| = 1$. That is, a *slice* $A * C$, where $C \in \mathrm{dom}(\mathrm{cont}(D))$, may be said to contain a *derived alignment* $R \in \mathrm{transforms}(D)$ if the *derived histogram* is *aligned*, $\mathrm{algn}(A*C*R) > 0$. In this example, the parent *slice*, $A*C_1$, contains the parent *derived alignment*, T_1, because the *derived histogram* is a *non-singleton diagonal*, $\mathrm{diagonal}(A * C_1 * T_1)$ and $|(A * C_1 * T_1)^{\mathrm{F}}| > 1$, but the child *slice*, $A * C_2$, does not contain the parent *derived alignment*, T_1, because the *derived histogram* is *singleton*, $|(A * C_2 * T_1)^{\mathrm{F}}| = 1$. Conversely, the child *slice*, $A * C_2$, contains the child *derived alignment*, T_2, because the *derived histogram* is a *non-singleton diagonal*, $\mathrm{diagonal}(A * C_2 * T_2)$ and $|(A*C_2*T_2)^{\mathrm{F}}| > 1$, and the parent *slice*, $A*C_1$, also contains the child *derived alignment*, $\mathrm{algn}(A*C_1*T_2) > 0$, although not necessarily *diagonalised*. Thus a *slice* $A * C$ cannot contain any of the *derived alignments* of its ancestor *slices*, but contains all of the *derived alignments* of its descendant *slices*.

$$\begin{aligned}&\forall L \in \mathrm{paths}(\mathrm{contingents}(D))\ \forall (i,(C_i,T_i)),(j,(C_j,T_j)) \in L \\ &\quad (i < j \implies (\mathrm{algn}(A * C_i * T_j) > 0) \wedge (\mathrm{algn}(A * C_j * T_i) = 0))\end{aligned}$$

A path of an *idealising summation aligned decomposition* may then be viewed as the cumulative removal of *derived alignments* as the index increases, or the cumulative addition of *derived alignments* as the index decreases.

Consider a *slice* $A * C$ that contains the *derived alignment* T' of another *slice* $A * C'$, not necessarily a descendant *slice*, in an *idealising summation aligned decomposition* $D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)$. That is, $\mathrm{algn}(A * C * T') > 0$, where $(C,T),(C',T') \in \mathrm{cont}(D)$. If it is the case that the *derived alignment* is *diagonalised*, $\mathrm{diagonal}(A * C * T')$, then T' could be a *transform symmetry* of both *slices*, $A * C$ and $A * C'$, in another *idealising summation aligned decomposition* D'. That is, $(C,T'),(C',T') \in \mathrm{cont}(D')$. However, it may be the case that the *transform symmetry* does not have higher *alignment*, $\mathrm{algn}(A * C * T') < \mathrm{algn}(A * C * T)$. So it may be the case that the *decomposition* with the *transform symmetry*, D', has lower *alignment*, $\mathrm{algnSum}(A,D') < \mathrm{algnSum}(A,D)$.

Let $D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)$ be an *idealising summation aligned decomposition* with respect to *integral-independent histogram* $A \in \mathcal{A}_{\mathrm{xi}}$. The *content alignment* of each of the *idealising summation aligned super-decompositions* of D is greater than that of D,

$$\forall E \in \mathrm{desctrees}(A)(D)\ (\mathrm{algnSum}(A, E) > \mathrm{algnSum}(A, D))$$

where $\mathrm{desctrees}(A) \in \mathcal{D}_{\Sigma,\mathrm{k}}(A) \to \mathrm{P}(\mathcal{D}_{\Sigma,\mathrm{k}}(A))$ is defined as $\mathrm{desctrees}(A)(D) := \{E : E \in \mathcal{D}_{\Sigma,\mathrm{k}}(A),\ D \in \mathrm{subtrees}(E),\ E \neq D\}$. The *content alignments* of immediate children *idealising summation aligned super-decompositions* of D, $\{E : E \in \mathcal{D}_{\Sigma,\mathrm{k}}(A),\ D \in \mathrm{subtrees}(E),\ |\mathrm{nodes}(D)| = |\mathrm{nodes}(E)| - 1\}$, are greater than that of D, but less than the *content alignments* of their descendants,

$$\forall E \in \mathrm{childtrees}(A)(D)\ (\mathrm{algnSum}(A, E) > \mathrm{algnSum}(A, D))$$

and

$$\begin{aligned}&\forall E \in \mathrm{childtrees}(A)(D)\\ &\qquad \forall F \in \mathrm{desctrees}(A)(E)\ (\mathrm{algnSum}(A, F) > \mathrm{algnSum}(A, E))\end{aligned}$$

where $\mathrm{childtrees}(A) \in \mathcal{D}_{\Sigma,\mathrm{k}}(A) \to \mathrm{P}(\mathcal{D}_{\Sigma,\mathrm{k}}(A))$ is defined as $\mathrm{childtrees}(A)(D) := \{E : E \in \mathrm{desctrees}(A)(D),\ |\mathrm{nodes}(D)| = |\mathrm{nodes}(E)| - 1\}$.

The *non-idealisation alignment* of a *transform* T with respect to *histogram* A is defined as the difference between the *alignment* of *histogram* and the *alignment* of the *idealisation*, $\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A})$. The *non-idealisation alignment* is conjectured to be always positive in the case of *integral independent*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$. The *non-idealisation alignment* of an *idealising summation aligned decomposition* D with respect to A is $\mathrm{algn}(A) - \mathrm{algn}(A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A})$. The *non-idealisation alignment* of an *idealising summation aligned super-decomposition* E of D is $\mathrm{algn}(A) - \mathrm{algn}(A * E^{\mathrm{T}} * E^{\mathrm{T}\dagger A})$. Conjecture that the *non-idealisation alignment* of E is less than the *non-idealisation alignment* of D, $\mathrm{algn}(A) - \mathrm{algn}(A * E^{\mathrm{T}} * E^{\mathrm{T}\dagger A}) < \mathrm{algn}(A) - \mathrm{algn}(A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A})$. Hence the *idealisation alignment* of E is greater than the *idealisation alignment* of D,

$$\forall E \in \mathrm{desctrees}(A)(D)\ (\mathrm{algn}(A * E^{\mathrm{T}} * E^{\mathrm{T}\dagger A}) > \mathrm{algn}(A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}))$$

Conjecture that the *idealisation alignments* of the immediate children are greater than that for D, but less than their descendants,

$$\forall E \in \mathrm{childtrees}(A)(D)\ (\mathrm{algn}(A * E^{\mathrm{T}} * E^{\mathrm{T}\dagger A}) > \mathrm{algn}(A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}))$$

and

$$\forall E \in \text{childtrees}(A)(D)$$
$$\quad \forall F \in \text{desctrees}(A)(E)\ (\text{algn}(A * F^{\text{T}} * F^{\text{T}\dagger A}) > \text{algn}(A * E^{\text{T}} * E^{\text{T}\dagger A}))$$

If an *idealising summation aligned decomposition* D is *ideal* with respect to A, $\text{ideal}(A, D^{\text{T}})$, for example, if it is *effectively sliced*, $\forall C \in D^{\text{TP}}\ (|(A * C^{\text{U}})^{\text{F}}| \leq 1) \implies \text{ideal}(A, D^{\text{T}})$, then there are no *idealising summation aligned super-decompositions* of D,

$$\text{ideal}(A, D^{\text{T}}) \implies \text{desctrees}(A)(D) = \emptyset$$

Thus, conjecture that the *idealising summation aligned super-decompositions* of D are progressively more *ideal* as nodes are added.

Let $X \in \text{trees}(\mathcal{D}_{\Sigma,\text{k}}(A))$ be a tree of *idealising summation aligned decompositions* of *integral-independent histogram* A such that (i) the steps of the tree are immediate *super-decompositions*,

$$\forall (D, E) \in \text{steps}(X)\ (E \in \text{childtrees}(A)(D))$$

and (ii) the leaves of the tree are *ideal*,

$$\forall D \in \text{leaves}(X)\ (\text{ideal}(A, D^{\text{T}}))$$

The tree of *idealising summation aligned decompositions*, X, is said to be *fully searched*.

The *content alignment* increases along the paths of the tree, $\forall (D, E) \in \text{steps}(X)\ (\text{algnSum}(A, E) \geq \text{algnSum}(A, D))$. So the leaves of the tree, $\text{leaves}(X)$, have the maximum *content alignment* of their ancestors, $\forall L \in \text{paths}(X)\ \forall D \in \text{set}(L)\ (\text{algnSum}(A, L_{|L|}) \geq \text{algnSum}(A, D))$.

Conjecture that the *idealisation alignment* increases along the paths as the *idealising summation aligned super-decompositions* become progressively more *ideal*, $\forall (D, E) \in \text{steps}(X)\ (\text{algn}(A * E^{\text{T}} * E^{\text{T}\dagger A}) \geq \text{algn}(A * D^{\text{T}} * D^{\text{T}\dagger A}))$. The *idealisation alignment* of the leaves equals the *alignment* of the *histogram*, $\forall D \in \text{leaves}(X)\ (\text{algn}(A * D^{\text{T}} * D^{\text{T}\dagger A}) = \text{algn}(A))$.

The maximally *aligned* set of *idealising summation aligned decompositions*, $M = \text{maxd}(\{(D, \text{algnSum}(A, D)) : D \in \text{elements}(X)\})$, is a subset of the leaves, $M \subseteq \text{leaves}(X)$, and so consists only of *ideal decompositions*, $\forall D \in M\ (\text{ideal}(A, D^{\text{T}}))$.

In section 'Likely histograms', it is conjectured that there exists an intermediate *mid substrate transform* $T_{\mathrm{m}} \in \mathcal{T}_{U_A,V_A}$ which is neither *self* nor *unary*, $T_{\mathrm{m}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$, where the *formal* is *independent* and the *midisation entropy* is minimised,

$$T_{\mathrm{m}} \in \mathrm{mind}(\{(T, \mathrm{entropy}(A^{\mathrm{M}(T)})) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\})$$

At the *mid transform* the *formal* tends to the *abstract*, $A^{\mathrm{X}} * T_{\mathrm{m}} \approx (A * T_{\mathrm{m}})^{\mathrm{X}}$, and the *mid component size cardinality relative entropy* is small,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V_A^{\mathrm{C}} * T_{\mathrm{m}}) \quad \approx \quad 0$$

Subsequent minimisation of the *idealisation entropy*, where the *mid idealisation* is *integral*, $A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A} \in \mathcal{A}_{\mathrm{i}}$, tends to increase the *mid component size cardinality relative entropy*,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V_A^{\mathrm{C}} * T_{\mathrm{m}}) \quad \sim \quad -\ \mathrm{entropy}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A})$$

In section 'Transform alignment', it is conjectured that subsequent maximisation of the *idealisation alignment* also tends to increase the *relative entropy*,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V_A^{\mathrm{C}} * T_{\mathrm{m}}) \quad \sim \quad \mathrm{algn}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A})$$

The tree of *idealising summation aligned decompositions*, $X \in \mathrm{trees}(\mathcal{D}_{\Sigma,\mathrm{k}}(A))$, is *fully searched*, so, given any path $L \in \mathrm{paths}(X)$, the last *decomposition* is *ideal*, $A * L_l^{\mathrm{T}} * L_l^{\mathrm{T}\dagger A} = A$, where $l = |L|$. Consider the case where (i) the *root transform* is the *mid transform*, $L_1 = \{((\emptyset, T_{\mathrm{m}}), \emptyset)\}$, and (ii) the *idealisations* along the path are all *integral*, $\forall i \in \{1 \ldots l\}\ (A * L_i^{\mathrm{T}} * L_i^{\mathrm{T}\dagger A} \in \mathcal{A}_{\mathrm{i}})$. In this case the *idealisation alignment* increases along the path,

$$\forall i \in \{2 \ldots l\}\ (\mathrm{algn}(A * L_i^{\mathrm{T}} * L_i^{\mathrm{T}\dagger A}) > \mathrm{algn}(A * L_{i-1}^{\mathrm{T}} * L_{i-1}^{\mathrm{T}\dagger A}))$$

and so the *relative entropy* also increases along the path,

$$\begin{aligned}&\forall i \in \{2 \ldots l\}\\ &\quad (\mathrm{entropyRelative}(A * L_i^{\mathrm{T}}, V_A^{\mathrm{C}} * L_i^{\mathrm{T}}) > \mathrm{entropyRelative}(A * L_{i-1}^{\mathrm{T}}, V_A^{\mathrm{C}} * L_{i-1}^{\mathrm{T}}))\end{aligned}$$

The first *decomposition*, L_1, which is a *sub-decomposition* of all subsequent, has the least *relative entropy*, $\mathrm{entropyRelative}(A * L_1^{\mathrm{T}}, V_A^{\mathrm{C}} * L_1^{\mathrm{T}}) \approx 0$. The last *decomposition*, L_l, which is a *super-decomposition* of all previous, has the greatest *relative entropy*, $\mathrm{entropyRelative}(A * L_l^{\mathrm{T}}, V_A^{\mathrm{C}} * L_l^{\mathrm{T}}) > 0$.

That is, an *idealising summation aligned decomposition* D that (i) is *ideal*,

$A*D^{\mathrm{T}}*D^{\mathrm{T}\dagger A} = A$, and (ii) is *rooted* in the *mid transform*, $D = \{((\emptyset, T_{\mathrm{m}}), \cdot)\}$, tends to increase *relative entropy* as the cardinality of *decomposition* nodes increases,

$$\mathrm{entropyRelative}(A * D^{\mathrm{T}}, V_A^{\mathrm{C}} * D^{\mathrm{T}}) \quad \sim \quad |\mathrm{nodes}(D)|$$

In the case where each *transform* is the *mid transform* for the *component*,

$$\forall (C,T) \in \mathrm{cont}(D)\ (T \in \mathrm{mind}(\{(T', \mathrm{entropy}((A*C)^{\mathrm{M}(T')})) : \\ T' \in \mathcal{T}_{U_A,V_A},\ (A*C)^{\mathrm{X}} * T' = ((A*C)^{\mathrm{X}} * T')^{\mathrm{X}}\}))$$

then each *non-leaf decomposition node* $((\cdot,T),F) \in \mathrm{nodes}(D)$, where $F \neq \emptyset$, forms a *child decomposition* $E = \{((\emptyset,T),F)\}$ in *slice* $A*C$ which is *rooted* in the *slice mid transform*, T, so that the *slice formal* approximates to the *slice abstract*, $(A*C)^{\mathrm{X}}*T \approx (A*C*T)^{\mathrm{X}}$, but the *child decomposition relative entropy*, $\mathrm{entropyRelative}(A*C*E^{\mathrm{T}}, V_A^{\mathrm{C}} * E^{\mathrm{T}})$, is not necessarily small.

Conjecture that all *idealising summation aligned decompositions* can be *fully searched*,

$$\begin{aligned}&\forall A \in \mathcal{A}_U \cap \mathcal{A}_{\mathrm{xi}}\ \forall D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ \exists X \in \mathrm{trees}(\mathcal{D}_{\Sigma,\mathrm{k}}(A))\\ &\quad (\mathrm{roots}(X) = \{D\} \wedge \\ &\quad\ \forall (E,F) \in \mathrm{steps}(X)\ (F \in \mathrm{childtrees}(A)(E)) \wedge \\ &\quad\ \forall E \in \mathrm{leaves}(X)\ (\mathrm{ideal}(A, E^{\mathrm{T}})))\end{aligned}$$

That is, for any *idealising summation aligned decomposition* that is not *ideal* conjecture that there can always be found an *idealising summation aligned super-decomposition*,

$$\forall A \in \mathcal{A}_U \cap \mathcal{A}_{\mathrm{xi}}\ \forall D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ (\neg \mathrm{ideal}(A, D^{\mathrm{T}}) \implies \mathrm{desctrees}(A)(D) \neq \emptyset)$$

This requires that any *non-independent component* of the *partition*, $A*C^{\mathrm{U}} \neq (A*C^{\mathrm{U}})^{\mathrm{X}}$ where $C \in D^{\mathrm{P}}$, may be *diagonalised* such that the *formal histogram* is equivalent to *abstract histogram*, the *derived histogram* is not *independent* and the *formal histogram* is *independent*, $\exists T \in \mathcal{T}_{U,\mathrm{f},1}\ (\mathrm{diagonal}(A*C^{\mathrm{U}}*T) \wedge (A^{\mathrm{X}}*C^{\mathrm{U}}*T \equiv (A*C^{\mathrm{U}}*T)^{\mathrm{X}}) \wedge \neg(A*C^{\mathrm{U}}*T \equiv (A*C^{\mathrm{U}}*T)^{\mathrm{X}}) \wedge ((A*C^{\mathrm{U}})^{\mathrm{X}}*T \equiv ((A*C^{\mathrm{U}})^{\mathrm{X}}*T)^{\mathrm{X}}))$.

If it is the case that the *derived alignments* of a *decomposition* $D \in \mathcal{D}$ always decrease along the paths,

$$\begin{aligned}&\forall ((C_1,T_1),(C_2,T_2)) \in \mathrm{steps}(\mathrm{contingents}(D))\\ &\quad (\mathrm{algn}(A*C_1*T_1) > \mathrm{algn}(A*C_2*T_2))\end{aligned}$$

the *decomposition*, D, is said to be *slowing*. It is sometimes, but not necessarily, the case that an *idealising summation aligned decomposition* $D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)$ is *slowing*,

$$\begin{aligned}&\exists A \in \mathcal{A}_U \cap \mathcal{A}_{\mathrm{xi}}\ \exists D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\\&\quad\forall((C_1,T_1),(C_2,T_2)) \in \mathrm{steps}(\mathrm{contingents}(D))\\&\qquad(\mathrm{algn}(A * C_1 * T_1) > \mathrm{algn}(A * C_2 * T_2))\end{aligned}$$

There are several reasons why this is so. Firstly, the *slice sizes* must decrease,

$$\forall((C_1,\cdot),(C_2,\cdot)) \in \mathrm{steps}(\mathrm{contingents}(D))\ (\mathrm{size}(A * C_1) > \mathrm{size}(A * C_2))$$

The *contingent derived alignments* are each limited to the *maximum alignment*,

$$\mathrm{algn}(A * C * T) \leq \mathrm{alignmentMaximum}(U)(\mathrm{der}(T), \mathrm{size}(A * C))$$

The *maximum alignment* is limited by the *slice size*, $\mathrm{size}(A*C)$. For *regular derived histograms* of *slice size* $z = \mathrm{size}(A * C)$, *dimension* $n = |\mathrm{der}(T)|$ and *valency* $\{d\} = \{|U_w| : w \in \mathrm{der}(T)\}$, the *maximum alignment* approximates to $z(n-1)\ln d$.

Secondly, it is conjectured above that *idealising summation aligned super-decompositions* have higher *idealisation alignment* than their ancestor *idealising summation aligned decompositions*. That is, the *super-decompositions* are more *ideal*. Therefore the *slices* are progressively more *independent* along the paths of a *decomposition*,

$$\forall((C_1,\cdot),(C_2,\cdot)) \in \mathrm{steps}(\mathrm{contingents}(D))\ (\mathrm{algn}(A * C_1) > \mathrm{algn}(A * C_2))$$

If it is the case that the descendant *slice* is more *independent* because it is *partially independent*, that is $\exists Q \in \mathrm{B}(V)\ (A*C_2 \equiv Z_{A*C_2} * \prod(A/Z_{A*C_2}\ \%\ K : K \in Q))$ where $V = \mathrm{vars}(A)$ and $Z_X = \mathrm{scalar}(\mathrm{size}(X))$, but the ancestor *slice* is not *partially independent* according to this partition, Q, of the *variables*, $\neg(A * C_1 \equiv Z_{A*C_1} * \prod(A/Z_{A*C_1}\ \%\ K : K \in Q))$, then there must exist fewer *transforms* of the descendant *slice* that have non-zero *derived alignment*. This is because the *transform* T_Q constructed with *self partition transforms* of the components of the *variables* partition, Q, has zero *derived alignment*, $A * C_2 * T_Q \equiv (A * C_2 * T_Q)^{\mathrm{X}}$, where $T_Q = \{K^{\mathrm{CS\{\}T}} : K \in Q\}^{\mathrm{T}}$.

Thirdly, a *slice* $A * C$ cannot contain any of the *derived alignments* of its ancestor *slices*, because the *derived histogram* is an *effective singleton*,

$$\begin{aligned}&\forall L \in \mathrm{paths}(\mathrm{contingents}(D))\\&\quad\forall(i,(C_i,T_i)),(j,(C_j,T_j)) \in L\ (i < j \implies |(A * C_j * T_i)^{\mathrm{F}}| = 1))\end{aligned}$$

Exclusion of the *derived alignments* of the ancestor *slices* means that there must exist fewer *transforms* of the descendant *slice* that have non-zero *derived alignment*.

So, for the reasons that (i) *slice sizes* must decrease, (ii) *slice alignments* must decrease, and (iii) *slices* cannot contain ancestor *derived alignments*, it is sometimes the case that the *idealising summation aligned decomposition* is *slowing*.

It is therefore the case that the additional *content alignment* of an immediate child *slowing idealising summation aligned super-decomposition* E, of a *slowing idealising summation aligned decomposition* D, is less than that of D,

$$\text{algnSum}(A, E) - \text{algnSum}(A, D) < \text{algnSum}(A, D)$$

or

$$\text{algnSum}(A, D) < \text{algnSum}(A, E) < 2 \times \text{algnSum}(A, D)$$

So it is sometimes, but not necessarily, the case that the *content alignment* decreases in a *fully searched slowing idealising summation aligned decomposition* tree $X \in \text{trees}(\mathcal{D}_{\Sigma,\text{k}}(A))$. That is, sometimes, $\text{algnSum}(A, F) - \text{algnSum}(A, E) < \text{algnSum}(A, E) - \text{algnSum}(A, D)$, where $(D, E), (E, F) \in \text{steps}(X)$.

The set of *idealising summation aligned decompositions*, $\mathcal{D}_{\Sigma,\text{k}}(A)$, excludes *decompositions* containing *variable symmetries*. Consider a *well behaved decomposition* $D \in \mathcal{D}_{\text{w},U}$ that contains a *transform symmetry*, $|\text{nodes}(D)| > |\text{transforms}(D)|$, which is a special case of a *decomposition* containing *variable symmetries*. The *decomposition*, D, unions *slices* having the same *transform* T by means of *alternate slice transforms* in the *nullable fud*. For example, $\{C_1, C_2\} = \text{inverse}(\text{cont}(D))(T)$ implies a unioned *slice*, $A * (C_1 + C_2)$. The *alignment* of the unioned *slice* is greater than or equal to the sum of the *alignments* of the *slices* separately, $\text{algn}(A * (C_1 + C_2) * T) \geq \text{algn}(A * C_1 * T) + \text{algn}(A * C_2 * T)$. Therefore conjecture that *decompositions* that contain *variable symmetries*, but are otherwise subject to the same constraints as the *idealising summation aligned decompositions*, have *content alignment* greater than or equal to the *summation alignment*, $\text{algn}(A * D^{\text{T}}) - \text{algn}(A^{\text{X}} * D^{\text{T}}) \geq \text{algnSum}(A, D)$.

The *independent formal slice* constraint, $\forall (C, T) \in \text{cont}(D)\ ((A * C)^{\text{X}} * T \equiv ((A * C)^{\text{X}} * T)^{\text{X}})$, is not as strict as constraining each *transform* to be *non-overlapping*, $\forall T \in \text{ran}(\text{cont}(D))\ (\neg\text{overlap}(T))$, and therefore not as strict as constraining the entire *fud* of *transforms* to be *non-overlapping*, $\neg\text{overlap}(G)$ where $G = \text{transforms}(D) = \text{ran}(\text{cont}(D))$. In any case *non-overlapping transforms*, $\neg\text{overlap}(G)$, does not necessarily imply that the *nullable transform*, D^{T}, of an *idealising summation aligned decomposition* $D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)$, is *non-overlapping*. In fact, it is only in the case where the *decomposition*, D, contains only a *root transform*, $|G| = 1$, that it is *non-overlapping*

$$|\text{transforms}(D)| = 1 \iff \neg\text{overlap}(D^{\mathrm{T}})$$

A *well behaved decomposition* containing more than one *transform* is necessarily *overlapping* because the *nullable variables* depend on the *underlying variables* of their ancestor *slice variables*.

Similarly, the *contingent formal-abstract equivalence* constraint,

$$\forall (C,T) \in \text{cont}(D)\ (A^{\mathrm{X}} * C * T \equiv (A * C * T)^{\mathrm{X}})$$

does not necessarily imply that the *formal histogram* equals the *abstract histogram* for the *nullable transform*, D^{T}, of a *summation aligned decomposition* $D \in \mathcal{D}_{\Sigma}(A)$. In the case where the *decomposition*, D, contains only a *root transform*, $|G| = 1$, the *formal histogram* is necessarily equivalent to the *abstract histogram*

$$|\text{transforms}(D)| = 1 \implies A^{\mathrm{X}} * D^{\mathrm{T}} \equiv (A * D^{\mathrm{T}})^{\mathrm{X}}$$

4.18 Derived alignment and conditional probability

Consider the *complete integral congruent support sample histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ *drawn with replacement* from *distribution histogram* $E \in \mathcal{A}_{U,V,z_E}$, where the *distribution histogram* E is as *effective* as the *independent sample*, $E^{\mathrm{F}} \geq A^{\mathrm{XF}}$. Given some *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$, where $\text{und}(T) = V$, the set of *integral iso-transform-independents* is

$$Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$$

where the *integral iso-transform-independent* function, $Y_{U,\mathrm{i},T,z} \in \mathcal{A}_{U,\mathrm{i},V,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})$, is defined

$$Y_{U,\mathrm{i},T,z} = \{(B,((B^{\mathrm{X}} * T),(B * T)^{\mathrm{X}})) : B \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

where $W = \text{der}(T)$. For convenience, let the *integral iso-transform-independents* be abbreviated

$$\begin{aligned}\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) &= Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \\ &= \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}\end{aligned}$$

The *generalised multinomial probability* of the *sample histogram*, $\hat{Q}_{\mathrm{m},U}(E,z)(A)$, may be decomposed into (i) the *iso-transform-independents multinomial probability* and (ii) *iso-transform-independent conditional dependent multinomial probability*

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B) \times \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B)}$$

Compare to the *iso-independent conditional dependent* case, where the *generalised multinomial probability* of the *sample histogram*, $\hat{Q}_{\mathrm{m},U}(E,z)(A)$, is decomposed into (i) the *iso-independents multinomial probability* and (ii) the *iso-independent conditional dependent multinomial probability*

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B) \times \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B)}$$

where *integral iso-independents* is

$$Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} = A^{\mathrm{X}}\} \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$$

Compare to the *relative dependent* case, where the *generalised multinomial probability* is decomposed into (i) the *independent multinomial probability* and (ii) *relative dependent multinomial probability*

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = \hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}}) \times \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})}$$

Unlike in the *relative dependent* case, where the *independent histogram* must be *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, in the *iso-transform-independent conditional dependent* case there is no need for the *independent histogram* to be *integral* because the *integral iso-transform-independents* is non-empty, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) = Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \neq \emptyset$.

Defined in terms of the *generalised multinomial probability*, the *generalised iso-transform-independent conditional multinomial probability distribution* is

$$\hat{Q}_{\mathrm{m},\mathrm{y},T,U}(E,z) = \mathrm{normalise}(\{(A, \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\})$$

So

$$\hat{Q}_{\mathrm{m,y},T,U}(E,z)(A) = \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},T,z})|} \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B)}$$

and the *generalised multinomial probability* may be decomposed

$$\hat{Q}_{\mathrm{m},U}(E,z)(A) = \sum_{B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B) \times |\mathrm{ran}(Y_{U,\mathrm{i},T,z})| \hat{Q}_{\mathrm{m,y},T,U}(E,z)(A)$$

The cardinality of the components of the partition of $\mathcal{A}_{U,\mathrm{i},T,z}$ is the normalisation factor,

$$|\mathrm{ran}(Y_{U,\mathrm{i},T,z})| \leq \prod_{w \in V} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!} \times \prod_{w \in W} \frac{(z + |U_w| - 1)!}{z!\ (|U_w| - 1)!}$$

where $W = \mathrm{der}(T)$. The cardinality of the set of *integral iso-transform-independent* sets is also such that

$$|\mathrm{ran}(Y_{U,\mathrm{i},T,z})| \leq |\mathrm{dom}(Y_{U,\mathrm{i},T,z})| = |\mathcal{A}_{U,\mathrm{i},V,z}| = \frac{(z + v - 1)!}{z!\ (v - 1)!}$$

The *relative dependent multinomial probability* equals the *iso-transform-independent conditional dependent multinomial probability* if the *iso-transform-independents* set is a singleton containing the *independent*. In this case, however, the *sample* must be *independent*,

$$Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) = \{A^{\mathrm{X}}\} \implies A = A^{\mathrm{X}}$$

and therefore the *probability* is 1,

$$\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})}{\sum_{B \in \{A^{\mathrm{X}}\}} \hat{Q}_{\mathrm{m},U}(E,z)(B)} = \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})}{\hat{Q}_{\mathrm{m},U}(E,z)(A^{\mathrm{X}})} = 1$$

In this case, the *generalised iso-transform-independent conditional multinomial probability* does not depend on A^{X}

$$\hat{Q}_{\mathrm{m,y},T,U}(E,z)(A^{\mathrm{X}}) = \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},T,z})|}$$

The *iso-transform-independent conditional dependent multinomial probability* is greater than 0 and less than or equal to 1

$$0 < \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B)} \leq 1$$

because

$$0 < \hat{Q}_{\mathrm{m},U}(E,z)(A) \leq \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B) \leq 1$$

The *iso-transform-independent conditional dependent multinomial probability* is a probability proper because the *conditional probability* is always between zero and one, yielding a *probability function*,

$$\{(C, \frac{\hat{Q}_{\mathrm{m},U}(E,z)(C)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B)}) : C \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\} \in \mathcal{P}$$

The *iso-transform-independent conditional dependent multinomial probability* may be generalised to a *probability density*. Instead of *drawing* an *integral sample histogram* from the finite *integral congruent support*, $\mathcal{A}_{U,\mathrm{i},V,z}$, the *sample histogram* is *drawn* from the infinite *complete congruent histograms*, $A \in \mathcal{A}_{U,V,z}$. The *iso-transform-independent conditional dependent multinomial probability density* given the infinite *iso-transform-independents* is

$$\frac{\mathrm{mpdf}(U)(E,z)(A)}{\int_{B \in \mathcal{A}_{U,\mathrm{y},T,z}(A)} \mathrm{mpdf}(U)(E,z)(B)\ dB}$$

which is defined if the *distribution histogram* E is as *effective* as the *independent sample*, $E^{\mathrm{F}} \geq A^{\mathrm{XF}}$.

The *iso-transform-independent conditional dependent multinomial probability density* is greater than 0 and less than or equal to 1

$$0 < \frac{\mathrm{mpdf}(U)(E,z)(A)}{\int_{B \in \mathcal{A}_{U,\mathrm{y},T,z}(A)} \mathrm{mpdf}(U)(E,z)(B)\ dB} \leq 1$$

because

$$0 < \mathrm{mpdf}(U)(E,z)(A) \leq \int_{B \in \mathcal{A}_{U,\mathrm{y},T,z}(A)} \mathrm{mpdf}(U)(E,z)(B)\ dB \leq 1$$

The *iso-transform-independent conditional dependent multinomial probability* tends to the *iso-transform-independent conditional dependent multinomial probability density* as the *size* increases

$$\lim_{k \to \infty} \frac{\hat{Q}_{\mathrm{m},U}(E,kz)(Z_k * A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,kz}(Z_k * A)} \hat{Q}_{\mathrm{m},U}(E,kz)(B)} = \frac{\mathrm{mpdf}(U)(E,z)(A)}{\int_{B \in \mathcal{A}_{U,\mathrm{y},T,z}(A)} \mathrm{mpdf}(U)(E,z)(B)\ dB}$$

where $Z_k = \mathrm{scalar}(k)$. This is because the finite *integral iso-transform-independents* becomes a larger subset of the *iso-transform-independents* as the *size* increases,

$$Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \subset Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$$

If the *transform* is a *self partition transform*, for example $T_{\mathrm{s}} = V^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, or it is *value full functional*, for example $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, then the set of *integral iso-transform-independents* equals the set of *integral iso-independents*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$. In this case the *iso-transform-independent conditional dependent multinomial probability* equals the *iso-independent conditional dependent multinomial probability*

$$\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B)} = \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E,z)(B)}$$

If the *transform* is a *unary partition*, for example $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, then the set of *integral iso-transform-independents* equals the *integral congruent support*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A) = \mathcal{A}_{U,\mathrm{i},V,z}$. In this case the *iso-transform-independent conditional dependent multinomial probability* equals the *generalised multinomial probability*

$$\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B)} = \hat{Q}_{\mathrm{m},U}(E,z)(A)$$

Thus the *iso-transform-independent conditional dependent multinomial probability* for the *self partition transform* case, T_{s}, is greater than or equal to that for the *unary partition transform* case, T_{u},

$$\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B)} \geq \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B)}$$

That is, the *conditional probability* of the *sample* given a *self partition transform*, T_{s}, is greater than or equal to the *conditional probability* of the *sample* given a *unary partition transform*, T_{u}, regardless of the *distribution histogram*, E. The larger the intersection between the *integral iso-transform-independents* of the two *transforms*, $T_{\mathrm{s}}, T_{\mathrm{u}}$, which is the intersection between the *integral iso-independents* and the *integral congruent support*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_\mathrm{s},z}(A) \cap \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_\mathrm{u},z}(A) = Y^{-1}_{U,\mathrm{i},V,z}(A^\mathrm{X}) \cap \mathcal{A}_{U,\mathrm{i},V,z} = Y^{-1}_{U,\mathrm{i},V,z}(A^\mathrm{X})$, the smaller the difference in *conditional probabilities*. So the less *independent* the *sample*, $A \neq A^\mathrm{X}$, the greater the difference in *iso-transform-independent conditional dependent multinomial probability* between these extreme cases.

In the case where (i) the *sample* is *completely effective*, $A^\mathrm{F} = A^\mathrm{C}$, (ii) the *distribution histogram* equals the *sample*, $E = A$, and (iii) the *independent* is *integral*, $A^\mathrm{X} \in \mathcal{A}_\mathrm{i}$, which implies that the *idealisations* are *integral*, $A * T_\mathrm{s} * T_\mathrm{s}^{\dagger A} = A \in \mathcal{A}_\mathrm{i}$ and $A * T_\mathrm{u} * T_\mathrm{u}^{\dagger A} = A^\mathrm{X} \in \mathcal{A}_\mathrm{i}$, then the same inequality holds

$$\frac{\hat{Q}_{\mathrm{m},U}(A,z)(A * T_\mathrm{s} * T_\mathrm{s}^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_\mathrm{s},z}(A)} \hat{Q}_{\mathrm{m},U}(A,z)(B)} \geq \frac{\hat{Q}_{\mathrm{m},U}(A,z)(A * T_\mathrm{u} * T_\mathrm{u}^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_\mathrm{u},z}(A)} \hat{Q}_{\mathrm{m},U}(A,z)(B)}$$

because $\hat{Q}_{\mathrm{m},U}(A,z)(A) \geq \hat{Q}_{\mathrm{m},U}(A,z)(A^\mathrm{X})$. Note, however, that in the case where the *distribution histogram* equals the *independent sample*, $E = A^\mathrm{X}$, the inequality between the *idealisations* does not necessarily hold, because $\hat{Q}_{\mathrm{m},U}(A^\mathrm{X},z)(A) \leq \hat{Q}_{\mathrm{m},U}(A^\mathrm{X},z)(A^\mathrm{X})$, given the *integral mean multinomial probability distribution conjecture*. That is, in some cases

$$\frac{\hat{Q}_{\mathrm{m},U}(A^\mathrm{X},z)(A * T_\mathrm{s} * T_\mathrm{s}^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_\mathrm{s},z}(A)} \hat{Q}_{\mathrm{m},U}(A^\mathrm{X},z)(B)} < \frac{\hat{Q}_{\mathrm{m},U}(A^\mathrm{X},z)(A * T_\mathrm{u} * T_\mathrm{u}^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_\mathrm{u},z}(A)} \hat{Q}_{\mathrm{m},U}(A^\mathrm{X},z)(B)}$$

In the section 'Alignment and conditional probability', above, the negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability*, where the *distribution histogram* is (i) *independent*, $E = E^\mathrm{X}$, and (ii) sufficiently *effective*, $E^\mathrm{XF} \geq A^\mathrm{XF}$, was shown to be

$$\begin{aligned}
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^\mathrm{X},z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^\mathrm{X})} \hat{Q}_{\mathrm{m},U}(E^\mathrm{X},z)(B)} : E^\mathrm{XF} \geq A^\mathrm{XF} \right) \\
&\quad = \sum_{S \in A^\mathrm{S}} \ln A_S! \; + \; \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^\mathrm{X})} \frac{1}{\prod_{S \in B^\mathrm{S}} B_S!}
\end{aligned}$$

because

$$\begin{aligned}
&\forall B \in Y^{-1}_{U,\mathrm{i},V,z}(A^\mathrm{X}) \\
&\quad \left(\sum_{S \in B^\mathrm{S}} B_S \ln E^\mathrm{X}_S = \sum_{S \in B^\mathrm{XS}} B^\mathrm{X}_S \ln E^\mathrm{X}_S = \sum_{S \in A^\mathrm{XS}} A^\mathrm{X}_S \ln E^\mathrm{X}_S = \sum_{S \in A^\mathrm{S}} A_S \ln E^\mathrm{X}_S \right)
\end{aligned}$$

However, the same reasoning cannot be applied to the negative logarithm *independently-distributed iso transform independent conditional dependent multinomial probability* even given *independent distribution histogram.* This is because it is not necessarily the case that there is the common factor,

$$\prod_{S \in A^{\mathrm{XS}}} (E^{\mathrm{X}}_S)^{A^{\mathrm{X}}_S}$$

in the numerator and denominator of the *independently-distributed iso transform independent conditional dependent multinomial probability*,

$$\frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B)}$$

That is, in some cases $\exists B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\ (B^{\mathrm{X}} \neq A^{\mathrm{X}})$.

However, consider the case where (i) the *distribution histogram* is the *independent sample histogram*, $E = A^{\mathrm{X}}$, (ii) the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and (iii) *formal independent histogram* equals the *abstract histogram* which implies that the *independent* is in the *iso-transform-independents*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))$. So the *independent* is in the *integral iso-transform-independents*

$$(A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}) \wedge ((A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}) \implies A^{\mathrm{X}} \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))$$

Note that the third constraint is weaker than the case where the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$. The *mean histogram* of the *generalised multinomial probability distribution* is the *independent*, $\mathrm{mean}(\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)) = A^{\mathrm{X}}$. The *integral mean multinomial probability distribution conjecture*, defined above in 'Multinomial distributions', states that if the *mean* of the *multinomial probability distribution* is *integral* then it is also *modal*

$$\mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E,z)) \in \mathcal{A}_i \implies \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(E,z)) \in \mathrm{modes}(\hat{Q}_{\mathrm{m},U}(E,z))$$

If this conjecture is true then

$$\forall B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\ (\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B) \leq \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}}))$$

Hence the negative logarithm *independent-sample-distributed iso-transform-independent conditional dependent multinomial probability* is

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)} : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\right)$$

$$= \ -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} + \ln \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}$$

The *independent* is an *iso-transform-independent*, $A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, and hence the summation is such that

$$1 \leq \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} \leq |\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)|$$

Thus the negative logarithm *independent-sample-distributed iso-transform-independent conditional dependent multinomial probability*, where the *independent* is an *iso-transform-independent*, is bounded by the *alignment*

$$\begin{aligned}
& \mathrm{algn}(A) \\
\leq\ & \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)} : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\right) \\
\leq\ & \mathrm{algn}(A) + \ln |\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)|
\end{aligned}$$

The negative logarithm *independently-distributed relative dependent multinomial probability* of the *sample*, where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, which is the *alignment*,

$$\begin{aligned}
& \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A^{\mathrm{X}})} : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\right) \\
=\ & \sum_{S \in A^{\mathrm{S}}} \ln A_S! - \sum_{S \in A^{\mathrm{XS}}} \ln A^{\mathrm{X}}_S! \\
=\ & \mathrm{algn}(A)
\end{aligned}$$

does not depend on the *distribution histogram*. Nor does the negative logarithm *independently-distributed iso-independent conditional dependent multinomial probability*, where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, which is the *alignment-bounded iso-independent space*,

$$\begin{aligned}
& \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B)} : E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\right) \\
=\ & \sum_{S \in A^{\mathrm{S}}} \ln A_S!\ +\ \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} \\
=\ & \mathrm{algn}(A) + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}
\end{aligned}$$

Contrast the negative logarithm *independent-sample-distributed iso-transform-independent conditional dependent multinomial probability*, where the *independent* is an *integral iso-transform-independent*, $A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$,

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)} : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \right)$$

$$= \mathrm{algn}(A) + \ln \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}$$

which does depend on the *distribution histogram*, A^{X}.

Let *integral congruent delta* $(D,I) \in \mathcal{A}_{\mathrm{i}} \times \mathcal{A}_{\mathrm{i}}$ be such that its *perturbation*, $A - D + I$, is (i) *iso-transform-independence conserving*, $A - D + I \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, and (ii) *iso-independence conserving*, $A - D + I \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, so that $(A - D + I)^{\mathrm{X}} = A^{\mathrm{X}}$. That is, *delta*, (D,I), is *iso-independent* and *iso-abstract*, $A - D + I \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \cap Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$. Let (iii) the *formal independent* equal the *abstract*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$, so that the *integral independent*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, is an *integral iso-transform-independent*, $A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$. The change in negative logarithm *independent-sample-distributed iso-transform-independent conditional dependent multinomial probability*, given the *integral mean multinomial probability distribution conjecture*, because of the *application* of the *delta*, (D,I), is the difference in *alignments*,

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A-D+I)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)((A-D+I)^{\mathrm{X}})} + \ln \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A-D+I)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)((A-D+I)^{\mathrm{X}})} \right) -$$

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} + \ln \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} \right)$$

$$= \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A-D+I)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} + \ln \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A-D+I)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} \right) -$$

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} + \ln \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} \right)$$

$$= \mathrm{algn}(A - D + I) - \mathrm{algn}(A)$$

So the change in *conditional probability*, because of the *application* of the *delta*, (D, I), does not depend on the *transform*, T, under these constraints.

The *integral idealisation* of a *histogram* given an *effective transform*, $A * T * T^{\dagger A}$, is in both the *integral iso-transform-independents*, $A * T * T^{\dagger A} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, and the *integral iso-independents*, $A*T*T^{\dagger A} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$. The *integral idealisation* has a corresponding *iso-transform-independence* and *iso-independence conserving delta*, $A*T*T^{\dagger A} = A-D+I$. The change in negative logarithm *independent-sample-distributed iso-transform-independent conditional dependent multinomial probability*, given the *integral mean multinomial probability distribution conjecture*, where the *independent* is an *iso-transform-independent*, $A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, because of the *integral idealisation* of the *sample histogram* is the difference in *alignments*,

$$\begin{aligned}
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A*T*T^{\dagger A})}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} + \ln \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A*T*T^{\dagger A})} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}\right) - \\
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} + \ln \sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}\right) \\
= \quad & \mathrm{algn}(A*T*T^{\dagger A}) - \mathrm{algn}(A)
\end{aligned}$$

Consider the case where *distribution histogram* is not necessarily equal to the *independent sample histogram*. If (i) the *distribution histogram* is *independent*, $E = E^{\mathrm{X}}$, and (ii) sufficiently *effective*, $E^{\mathrm{XF}} \geq A^{\mathrm{XF}}$, (iii) the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and (iv) the *iso-transform-independents* equals the *iso-independents*, $Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}})) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, then the negative logarithm *independently-distributed iso-transform-independent conditional dependent multinomial probability* equals the negative logarithm *independently-distributed iso-independent conditional dependent multinomial* *probability*, which is the *alignment-bounded iso-independent space*

$$\begin{aligned}
&\Big(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)} : \\
&\qquad E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \Big) \\
=&\ \Big(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)} : E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \Big) \\
=&\ \sum_{S \in A^{\mathrm{S}}} \ln A_S! \ +\ \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{1}{\prod_{S \in B^{\mathrm{S}}} B_S!} \\
=&\ \mathrm{algn}(A) + \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}
\end{aligned}$$

This is the case, for example, if the *transform* is *value full functional*, $T = \{\{w\}^{\mathrm{CS\{\}T}} : w \in V\}^{\mathrm{T}}$.

In this case, where the *iso-transform-independents* equals the *iso-independents*, the negative logarithm *independently-distributed iso-transform-independent conditional dependent multinomial probability* is bounded by the *alignment*

$$\begin{aligned}
&\mathrm{algn}(A) \\
\leq&\ \Big(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)} : \\
&\qquad E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \Big) \\
\leq&\ \mathrm{algn}(A) + \ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|
\end{aligned}$$

Now consider the *lifted* case. For convenience, let the *lifted integral iso-transform-independents* be abbreviated

$$\begin{aligned}
\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A) &= \{B * T : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\} \\
&= \{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\} \\
&= \{B * T : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}
\end{aligned}$$

The *lifted iso-transform-independent quasi-conditional dependent multinomial probability* is

$$\frac{\hat{Q}_{\mathrm{m},U}(E * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E * T, z)(B')}$$

As noted above in section 'Iso-sets', it is only in the subset where the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, that the *lifted iso-transform-independent* relation is functional

$$\begin{aligned} &\{(A * T, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\} \\ \in\ \ & \mathcal{A}_{U,\mathrm{i},W,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z}) \end{aligned}$$

and hence the *lifted integral iso-transform-independent* sets do not partition the *integral congruent support* in the *derived variables*, $\mathcal{A}_{U,\mathrm{i},W,z}$, except where the *formal histogram* equals the *abstract histogram*

$$\begin{aligned} &\mathrm{ran}(\{(A * T, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : \\ &\quad (A, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \in Y_{U,\mathrm{i},T,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\}^{-1}) \\ \in\ \ & \mathrm{B}(\{B * T : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = (B * T)^{\mathrm{X}}\}) \end{aligned}$$

For this reason the *lifted iso-transform-independent quasi-conditional dependent multinomial probability* is only *quasi-conditional*. That is, *conditional* for the subset of the *derived integral congruent support* where the *formal histogram* equals the *abstract histogram*.

If the *transform* is a *self partition transform*, for example $T_{\mathrm{s}} = V^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, or it is *value full functional*, for example $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, then the set of *lifted integral iso-transform-independents* equals the set of *lifted integral iso-independents*, $\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A) = \{B * T_{\mathrm{s}} : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})\}$. The *lifted iso-transform-independent quasi-conditional dependent multinomial probability* equals the *iso-transform-independent conditional dependent multinomial probability*, which in turn equals the *iso-independent conditional dependent multinomial probability*

$$\begin{aligned} &\frac{\hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{s}}, z)(A * T_{\mathrm{s}})}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{s}}, z)(B')} \\ =\ \ & \frac{\hat{Q}_{\mathrm{m},U}(E, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E, z)(B)} \\ =\ \ & \frac{\hat{Q}_{\mathrm{m},U}(E, z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E, z)(B)} \end{aligned}$$

If the *transform* is a *unary partition*, for example $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, then the set of *lifted integral iso-transform-independents* equals the singleton *lifted integral congruent support*, $\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A) = \{B * T_{\mathrm{u}} : B \in \mathcal{A}_{U,\mathrm{i},V,z}\} = \{A * T_{\mathrm{u}}\}$. The *lifted generalised multinomial probability* is equal to one, $\hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{u}}, z)(A * T_{\mathrm{u}}) = 1$. So the *lifted iso-transform-independent quasi-conditional dependent multinomial probability* equals 1,

$$\frac{\hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{u}}, z)(A * T_{\mathrm{u}})}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{u}}, z)(B')} = 1$$

In the case of *unary partition transform*, the *lifted iso-transform-independent quasi-conditional dependent multinomial probability* is not necessarily equal to the corresponding *iso-transform-independent conditional dependent multinomial probability*. The *iso-transform-independent conditional dependent multinomial probability* equals the *generalised multinomial probability*, $\hat{Q}_{\mathrm{m},U}(E, z)(A)$, which may be less than one,

$$\frac{\hat{Q}_{\mathrm{m},U}(E, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(E, z)(B)} = \hat{Q}_{\mathrm{m},U}(E, z)(A) \leq 1$$

Whereas the *iso-transform-independent conditional dependent multinomial probability* for the *self partition transform* case is greater than or equal to that for the *unary partition transform* case,

$$\frac{\hat{Q}_{\mathrm{m},U}(E, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E, z)(B)} \geq \frac{\hat{Q}_{\mathrm{m},U}(E, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(E, z)(B)}$$

the *lifted iso-transform-independent quasi-conditional dependent multinomial probability* for the *self partition transform* case is less than or equal to that for the *unary partition transform* case

$$\frac{\hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{s}}, z)(A * T_{\mathrm{s}})}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{s}}, z)(B')} \leq \frac{\hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{u}}, z)(A * T_{\mathrm{u}})}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{u}}, z)(B')}$$

That is, the *lifted quasi-conditional probability* of the *sample* given a *self partition transform*, T_{s}, is less than or equal to the *lifted quasi-conditional probability* of the *sample* given a *unary partition transform*, T_{u}, regardless of the *distribution histogram*, E.

If the *distribution histogram* equals the *independent sample*, $E = A^{\mathrm{X}}$, however, it is sometimes the case that the direction of the *non-lifted conditional probability* inequality is the same as the *lifted* case for the corresponding *idealisations*, $A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger A} = A$ and $A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A} = A^{\mathrm{X}}$, because

$\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A) \leq \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A^{\mathrm{X}})$, given the *integral mean multinomial probability distribution conjecture*. So it is sometimes the case that

$$\frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)} < \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)}$$

In the case where (i) the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and (ii) the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, then the *independent* is an *integral iso-transform-independent*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$,

$$((A^{\mathrm{X}}*T)^{\mathrm{X}} = A^{\mathrm{X}}*T) \wedge (A^{\mathrm{X}}*T = (A*T)^{\mathrm{X}}) \implies A^{\mathrm{X}} \in Y_{U,\mathrm{i},T,z}^{-1}(((A^{\mathrm{X}}*T), (A*T)^{\mathrm{X}}))$$

and the *lifted integral iso-transform-independents* contains the *abstract histogram*

$$(A * T)^{\mathrm{X}} = A^{\mathrm{X}} * T \in \{B * T : B \in Y_{U,\mathrm{i},T,z}^{-1}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\}$$

That is, $(A * T)^{\mathrm{X}} \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)$.

The *lifted integral iso-abstracts* is a superset of the *lifted integral iso-transform-independents*

$$\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A) \subseteq \{B * T : B \in Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})\}$$

So the *lifted integral iso-transform-independents* is a subset of the *derived integral iso-independents*

$$\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A) \subseteq Y_{U,\mathrm{i},W,z}^{-1}((A * T)^{\mathrm{X}})$$

Thus the *independent derived*, $(A*T)^{\mathrm{X}}$, is the *independent* for all of the *lifted integral iso-transform-independents*, $\forall B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)\ (B'^{\mathrm{X}} = (A * T)^{\mathrm{X}})$.

The *abstract histogram*, $(A * T)^{\mathrm{X}}$, is *integral* because the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$,

$$(A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}) \wedge (A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}) \implies (A * T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$$

If, additionally, (iii) the *distribution histogram* is *independent*, $E = E^{\mathrm{X}}$, and (iv) *formal distribution histogram* is *independent*, $E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}}$, the *independent distribution histogram*, E^{X}, is *lifted* to an *independent derived distribution histogram*, $E^{\mathrm{X}} * T = (E * T)^{\mathrm{X}}$. That is, the *formal distribution histogram* equals the *abstract distribution histogram*.

Lastly, if (v) the *distribution histogram* is sufficiently *effective*, $E^{\mathrm{XF}} \geq A^{\mathrm{XF}}$, then the negative logarithm *lifted independently-distributed iso-transform-independent quasi-conditional dependent multinomial probability* can be rearranged in terms of *derived multinomial coefficients* and thence in terms of the *derived alignment*,

$$\begin{aligned}
&\Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(B')} : \\
&\qquad E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}},\ E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \Bigg) \\
&= \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A * T)_R! \ +\ \ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{1}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} \\
&= \mathrm{algn}(A * T) + \ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}
\end{aligned}$$

Now because $\forall B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)\ (B'^{\mathrm{X}} = (A * T)^{\mathrm{X}})$ and given the *minimum alignment conjecture*,

$$\forall B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A) \ \left(\frac{\prod_{S \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_S!}{\prod_{S \in B'^{\mathrm{S}}} B'_S!} \leq 1 \right)$$

and because $(A * T)^{\mathrm{X}} \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)$,

$$1 \leq \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} \leq |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)|$$

then

$$\begin{aligned}
& \mathrm{algn}(A * T) \\
\leq\ & \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A * T)_R! \ +\ \ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{1}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} \\
\leq\ & \mathrm{algn}(A * T) + \ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)|
\end{aligned}$$

In other words in this case, where (i) $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, (ii) $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, (iii) $E = E^{\mathrm{X}}$, (iv) $E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}}$, and (v) $E^{\mathrm{XF}} \geq A^{\mathrm{XF}}$, the negative logarithm *lifted independently-distributed iso-transform-independent quasi-conditional* *dependent multinomial probability* is such that

$$\begin{aligned}
&\mathrm{algn}(A * T) \\
\leq\ & \Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(B')} : \\
&\quad E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}},\ E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \Bigg) \\
\leq\ & \mathrm{algn}(A * T) + \ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)|
\end{aligned}$$

That is, given these conditions, the *derived alignment*, $\mathrm{algn}(A * T)$, is a bounded underestimate of the negative logarithm *lifted independently-distributed iso-transform-independent quasi-conditional dependent multinomial probability*.

The cardinality of the *lifted integral iso-transform-independents* must be less than or equal to the cardinality of the *derived integral congruent support*,

$$|\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)| \leq |\mathcal{A}_{U,\mathrm{i},W,z}| = \frac{(z + w - 1)!}{z!\ (w-1)!}$$

where $w = |W^{\mathrm{C}}|$. Thus $\ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)| < \overline{w} \ln z$ if $z > w$. So

$$\sum_{R \in (A*T)^{\mathrm{S}}} \ln (A * T)_R! \ +\ \ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{1}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} < \mathrm{algn}(A * T) + \overline{w} \ln z$$

Compare this to *derived maximum alignment*, $\mathrm{alignmentMaximum}(U)(W, z)$, which for large *size*, $z \gg w$, approximates to $z(n-1) \ln d$ for a *regular histogram* of *dimension* $n = |W|$ and *valency* $\{d\} = \{|U_u| : u \in V\}$. Therefore, in some cases the difference between the *derived alignment* and the negative logarithm *lifted independently-distributed iso-transform-independent quasi-conditional dependent multinomial probability* is less than the *derived alignment*, $\ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)| < \overline{w} \ln z < \mathrm{alignment}(A * T)$. That is, in some cases

$$\begin{aligned}
&\mathrm{algn}(A * T) \\
\leq\ & \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A * T)_R! \ +\ \ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{1}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} \\
\leq\ & 2 \times \mathrm{algn}(A * T)
\end{aligned}$$

In the case of the conditions given above, the negative logarithm *lifted independently-distributed iso-transform-independent quasi-conditional dependent multinomial probability*,

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(B')} : \right.$$
$$\left. E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}},\ E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \right)$$
$$= \left(\mathrm{algn}(A * T) + \ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} \right) \in \ln \mathbf{Q}_{>0}$$

may be abbreviated to the *alignment-bounded lifted iso-transform space*.

The corresponding negative logarithm *independently-distributed iso-transform-independent conditional dependent multinomial probability*,

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)} : \right.$$
$$\left. E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}},\ E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \right) \in \ln \mathbf{Q}_{>0}$$

may be abbreviated to the *alignment-bounded iso-transform space*. Strictly speaking, it is only the *lifted space* that is bounded by the *alignment*. However, the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, and so the *lifted iso-transform-independent* relation is functional,

$$\{(A * T, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\}$$
$$\in\ \mathcal{A}_{U,\mathrm{i},W,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})$$

which implies that each *derived histogram* maps to exactly one set of *iso-transform-independents*,

$$\{(A * T, Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})))) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\}$$
$$\in\ \mathcal{A}_{U,\mathrm{i},W,z} \to \mathrm{P}(\mathcal{A}_{U,\mathrm{i},V,z})$$

Thus the *alignment-bounded lifted iso-transform space* is correlated with the *alignment-bounded iso-transform space*.

The difference between the *alignment-bounded lifted iso-transform space* and the *derived alignment* is the *alignment-bounded lifted iso-transform error*

$$\ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A*T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}$$

The negative logarithm *independently-derived-distributed relative dependent derived multinomial probability* of the *sample*, where the *abstract histogram* is *integral*, $(A*T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, which is the *derived alignment*,

$$\begin{aligned}
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}((E*T)^{\mathrm{X}},z)(A*T)}{\hat{Q}_{\mathrm{m},U}((E*T)^{\mathrm{X}},z)((A*T)^{\mathrm{X}})} : (A*T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\right) \\
&= \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A*T)_R! - \sum_{R \in (A*T)^{\mathrm{XS}}} \ln (A*T)^{\mathrm{X}}_R! \\
&= \mathrm{algn}(A*T)
\end{aligned}$$

does not depend on the *derived distribution histogram*. Nor does the negative logarithm *independently-derived-distributed iso-independent conditional dependent derived multinomial probability*, where the *abstract histogram* is *integral*, $(A*T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, which is the *derived-alignment-bounded iso-independent space*,

$$\begin{aligned}
&\Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}((E*T)^{\mathrm{X}},z)(A*T)}{\sum_{B' \in Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}((E*T)^{\mathrm{X}},z)(B')} : \\
&\qquad (E*T)^{\mathrm{XF}} \geq (A*T)^{\mathrm{XF}},\ (A*T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\Bigg) \\
&= \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A*T)_R! \ + \ \ln \sum_{B' \in Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})} \frac{1}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} \\
&= \mathrm{algn}(A*T) + \ln \sum_{B' \in Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A*T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}
\end{aligned}$$

In the case of the conditions above, the negative logarithm *lifted independently-distributed iso-transform-independent quasi-conditional dependent multino-*

mial probability, which is the *alignment-bounded lifted iso-transform space*,

$$\Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(B')} :$$

$$E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}},\ E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \Bigg)$$

$$= \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A * T)_R! \ + \ \ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{1}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}$$

$$= \mathrm{algn}(A * T) + \ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}$$

does not depend on the *distribution histogram*.

In the case where the *distribution histogram* is *independent*, $E = E^{\mathrm{X}}$, the *alignment-bounded lifted iso-transform space* is less than or equal to the *derived-alignment-bounded iso-independent space*

$$\Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(B')} :$$

$$E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}},\ E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \Bigg)$$

$$\leq \Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}((E^{\mathrm{X}} * T)^{\mathrm{X}}, z)(A * T)}{\sum_{B' \in Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}((E^{\mathrm{X}} * T)^{\mathrm{X}}, z)(B')} :$$

$$(E^{\mathrm{X}} * T)^{\mathrm{XF}} \geq (A * T)^{\mathrm{XF}},\ (A * T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \Bigg)$$

because $\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A) \subseteq Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})$. Thus the *alignment-bounded lifted iso-transform error* is less than or equal to the *derived-alignment-bounded iso-independent error* where $E = E^{\mathrm{X}}$,

$$\ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} \leq \ln \sum_{B' \in Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}$$

The numerator in the *alignment-bounded lifted iso-transform error* expression is derived from the *independent* term of the *derived alignment*,

$\sum_{R \in (A*T)^{\rm XS}} \ln (A*T)^{\rm X}_R!$, which varies against the *entropy* of the *abstract histogram*, $\mathrm{entropy}((A*T)^{\rm X})$. In the case of *uniform abstract histogram* of *size* z and *derived volume* w where $z/w \in \mathbf{N}_{>0}$, the *independent* term is $w \ln(z/w)! \approx z \ln(z/w)$. So the *alignment-bounded lifted iso-transform error* with respect to the numerator varies with the *size*, z, and varies against the logarithm of the *derived volume*, $\ln w$. The *abstract histogram*, $(A*T)^{\rm X}$, tends to be more *uniform* at higher *derived alignments*.

The *alignment-bounded lifted iso-transform error* also varies with the cardinality of the *lifted integral iso-transform-independents*, $|\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)|$, which in turn varies with the cardinality of the *integral derived iso-independents*, $|Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\rm X})|$. As shown above, the average cardinality of the *integral derived iso-independents* is

$$\frac{|\mathcal{A}_{U,\mathrm{i},W,z}|}{|\mathrm{ran}(Y_{U,\mathrm{i},W,z})|} = \frac{(z+w-1)!}{z!\ (w-1)!} / \prod_{u \in W} \frac{(z+|U_u|-1)!}{z!\ (|U_u|-1)!}$$

The average cardinality of the *integral derived iso-independents* varies with both *size*, z, and *derived volume*, w.

In the case where the *size* is greater than the *derived volume*, $z > w$, the logarithm of the average cardinality is less than $\overline{w} \ln z$. In this case the negative contribution to the variation between the *error* and the *derived volume* from the numerator, $\ln w$, is outweighed by the positive contribution from the summation, $\overline{w}$. Hence, in the case where $z > w$, the *error* varies with both *size*, z, and *derived volume*, w.

For a given *derived volume*, w, the average cardinality of the *integral derived iso-independents*, $|Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\rm X})|$, varies with the entropy of the *valencies*, $\mathrm{entropy}(\{(u,|U_u|) : u \in W\})$. Hence the *error* also varies with *derived valency* entropy. The *error* tends to increase with *derived dimension*, $n = |W|$. *Regular derived histograms* tend to have higher *error* than *irregular*.

It is conjectured above that the cardinality of the set of *integral derived iso-independents*, $|Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\rm X})|$, corresponding to $(A*T)^{\rm X}$ varies with the *entropy* of the *abstract histogram*, $(A*T)^{\rm X}$. The cardinality of the subset *lifted integral iso-transform-independents*, $|\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)|$, also varies with the *entropy* of the *abstract histogram*. Therefore the *alignment-bounded lifted iso-transform error* varies with the *entropy* of the *abstract histogram*. The *formal histogram* equals the *abstract histogram*, $A^{\rm X} * T = (A*T)^{\rm X}$, so the *entropy* of the *abstract histogram* equals the *entropy* of the *formal indepen-*

dent histogram, $\mathrm{entropy}((A*T)^{\mathrm{X}}) = \mathrm{entropy}((A^{\mathrm{X}}*T)^{\mathrm{X}})$. The *entropy* of the *abstract histogram* equals the sum of the *entropies* of the *reductions*

$$\mathrm{entropy}((A*T)^{\mathrm{X}}) = \sum_{w \in W} \mathrm{entropy}(A*T \ \% \ \{w\}) = \sum_{w \in W} \mathrm{entropy}(A^{\mathrm{X}}*T \ \% \ \{w\})$$

That is, the more *uniform* the *perimeters*, the larger the cardinality of the set of *integral derived iso-independents*, and the higher the *error* with respect to the cardinality.

The ratio of the *alignment-bounded lifted iso-transform error* to the *derived alignment* is

$$\left(\ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A*T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} \right) / \mathrm{algn}(A*T)$$

where the *derived histogram* is not *independent*, $A * T \neq (A * T)^{\mathrm{X}} \implies \mathrm{algn}(A * T) > 0$.

As mentioned above, as the *derived alignment* increases to *maximum derived alignment*, $\mathrm{alignmentMaximum}(U)(W, z) \approx z \ln w$ where $z \gg w$, the ratio decreases, $\overline{w} \ln z / z \ln w$.

On the other hand, as noted above, the *derived alignment* approximates to the difference in *entropy* between the *abstract histogram* and the *derived histogram*, $\mathrm{algn}(A * T) \approx z \times \mathrm{entropy}((A * T)^{\mathrm{X}}) - z \times \mathrm{entropy}(A * T)$. Hence increases in *derived alignment* imply increases in the *entropy* of the *abstract histogram* to some degree. So there is a tendency to increase the ratio of the *alignment-bounded lifted iso-transform error* to the *alignment* at higher *derived alignments* due to the *abstract histogram entropy* which partly counteracts the tendency to decrease the ratio at higher *derived alignments* due to the *size*.

In the case where the *derived alignment* is approximately equal to the *expected derived alignment*, it is conjectured above ('Minimum alignment') that the *expected alignment* varies as the *volume*, w, for constant *size*, z, greater than the *volume*, $z > w$. So in the case of *expected derived alignment* the *alignment-bounded lifted iso-transform error* tends to be greater than the *derived alignment* and the ratio is greater than one, $(\overline{w} \ln z)/w > 1$.

If the *transform* is a *self partition transform*, for example $T_{\mathrm{s}} = V^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, or it is *value full functional*, for example $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in$

$V\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, then the set of *lifted integral iso-transform-independents* equals the set of *lifted integral iso-independents*, $\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A) = \{B * T_{\mathrm{s}} : B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})\}$, and the set of *integral iso-transform-independents* equals the set of *integral iso-independents*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$. So the *alignment-bounded iso-transform-independent space* equals the *alignment-bounded lifted iso-transform space*, which in turn equals the *alignment-bounded iso-independent space*

$$\begin{aligned}
& -\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B)} \\
= & -\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T_{\mathrm{s}},z)(A * T_{\mathrm{s}})}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T_{\mathrm{s}},z)(B')} \\
= & -\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B)}
\end{aligned}$$

Therefore, in this case, T_{s}, the *alignment-bounded iso-transform space* is bounded by the *derived alignment*,

$$\begin{aligned}
& \mathrm{algn}(A * T_{\mathrm{s}}) \\
= & \mathrm{algn}(A) \\
\leq & \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B)} : E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \right) \\
\leq & \mathrm{algn}(A) + \ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| \\
= & \mathrm{algn}(A * T_{\mathrm{s}}) + \ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)|
\end{aligned}$$

If the *transform* is a *unary partition*, for example $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, then the set of *integral iso-transform-independents* equals the singleton *lifted integral congruent support*, $\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A) = \{B * T_{\mathrm{u}} : B \in \mathcal{A}_{U,\mathrm{i},V,z}\} = \{A * T_{\mathrm{u}}\}$. The *lifted generalised multinomial probability* is equal to one, $\hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{u}},z)(A * T_{\mathrm{u}}) = 1$. So the *alignment-bounded lifted iso-transform space* is equal to zero,

$$-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T_{\mathrm{u}},z)(A * T_{\mathrm{u}})}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T_{\mathrm{u}},z)(B')} = 0$$

The lower bound is zero

$$\mathrm{algn}(A * T_{\mathrm{u}}) = 0$$

because $A * T_{\mathrm{u}} = (A * T_{\mathrm{u}})^{\mathrm{X}}$. The upper bound is also zero

$$\mathrm{algn}(A * T_{\mathrm{u}}) + \ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)| = \mathrm{algn}(A * T_{\mathrm{u}}) + \ln |\{A * T_{\mathrm{u}}\}| = 0$$

The set of *integral iso-transform-independents* equals the *integral congruent support*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A) = \mathcal{A}_{U,\mathrm{i},V,z}$, so the *alignment-bounded iso-transform space* equals the *generalised multinomial space*, which is greater than zero

$$\begin{aligned}
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B)} : E^{\mathrm{XF}} \geq A^{\mathrm{XF}}\right) \\
= \;& -\ln \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A) \\
> \;& 0
\end{aligned}$$

where the *distribution histogram* is *pluri-valent*, $|E^{\mathrm{XF}}| > 1$. In fact, it is greater than or equal to the *self partition* case

$$\begin{aligned}
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B)} : E^{\mathrm{XF}} \geq A^{\mathrm{XF}}\right) \\
\geq \;& \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}},z)(B)} : E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\right)
\end{aligned}$$

Therefore, in this case, T_{u}, the *alignment-bounded iso-transform space* is not bounded by the *derived alignment*.

In other words, although the *alignment-bounded iso-transform space* is functionally related to the *alignment-bounded lifted iso-transform space*, it is not always the case that the *alignment-bounded iso-transform space* is bounded by the *derived alignment*, $\mathrm{algn}(A * T)$, and so the prefix '*alignment-bounded*' of *alignment-bounded iso-transform space* is sometimes a misnomer with respect to *derived alignment* at least.

Let *integral congruent delta* $(D, I) \in \mathcal{A}_{\mathrm{i}} \times \mathcal{A}_{\mathrm{i}}$ be such that its *perturbation*, $A - D + I$, is *iso-transform-independence conserving*, $A - D + I \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$. So the *delta*, (D, I), is *iso-abstract*, $A - D + I \in Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$. The change in *alignment-bounded lifted iso-transform space* due to the *application* of the *iso-transform-independence conserving delta*, (D, I), is equal to the change in *derived alignment*

$$\begin{aligned}
&\Bigg(\text{algn}((A - D + I) * T) + \\
&\qquad \ln \sum_{B' \in \mathcal{A}'_{U,\text{i},\text{y},T,z}(A-D+I)} \frac{\prod_{R \in ((A-D+I)*T)^{\text{XS}}} ((A - D + I) * T)^{\text{X}}_R!}{\prod_{R \in B'^{\text{S}}} B'_R!}\Bigg) - \\
&\Bigg(\text{algn}(A * T) + \ln \sum_{B' \in \mathcal{A}'_{U,\text{i},\text{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\text{XS}}} (A * T)^{\text{X}}_R!}{\prod_{R \in B'^{\text{S}}} B'_R!}\Bigg) \\
= &\ \text{algn}((A - D + I) * T) - \text{algn}(A * T)
\end{aligned}$$

because $((A - D + I) * T)^{\text{X}} = (A * T)^{\text{X}}$ and therefore the *alignment-bounded lifted iso-transform error* is the same for both the *histogram*, A, and its *iso-transform-independent perturbation*, $A - D + I$.

A special case of an *iso-transform-independence conserving perturbation* is the *integral idealisation*, $A * T * T^{\dagger A} \in Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})$, where the change in *alignment-bounded lifted iso-transform space* because of the *integral idealisation* of the *sample histogram* is zero,

$$\begin{aligned}
&\Bigg(\text{algn}((A * T * T^{\dagger A}) * T) + \\
&\qquad \ln \sum_{B' \in \mathcal{A}'_{U,\text{i},\text{y},T,z}(A*T*T^{\dagger A})} \frac{\prod_{R \in ((A*T*T^{\dagger A})*T)^{\text{XS}}} ((A * T * T^{\dagger A}) * T)^{\text{X}}_R!}{\prod_{R \in B'^{\text{S}}} B'_R!}\Bigg) - \\
&\Bigg(\text{algn}(A * T) + \ln \sum_{B' \in \mathcal{A}'_{U,\text{i},\text{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\text{XS}}} (A * T)^{\text{X}}_R!}{\prod_{R \in B'^{\text{S}}} B'_R!}\Bigg) \\
= &\ \text{algn}(A * T * T^{\dagger A} * T) - \text{algn}(A * T) \\
= &\ 0
\end{aligned}$$

because $A * T * T^{\dagger A} * T = A * T$. That is, the *alignment-bounded lifted iso-transform space* is the same for a *sample histogram*, A, and its *integral idealisation*, $A * T * T^{\dagger A}$.

Consider the case where (i) the *independent distribution histogram* equals the *independent*, $E^{\text{X}} = A^{\text{X}}$, and (ii) *idealisation* is *integral*, $A * T * T^{\dagger A} \in \mathcal{A}_{\text{i}}$. The *integral idealisation* is in the *integral iso-idealisations* which is a subset of the *integral iso-transform-independents*,

$$A * T * T^{\dagger A} \in Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A}) \subseteq Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$$

where the *integral iso-idealisations* is defined $Y_{U,\mathrm{i},T,\dagger,z} = \{(A, A * T * T^{\dagger A}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$. The *integral iso-idealisations* is also the subset of the subset of the *integral iso-transform-independents* that have given *alignment-bounded lifted iso-transform space*, $(A - D + I) * T = A * T$, which is the intersection between the *integral iso-transform-independents* and the *integral iso-deriveds*, or the *integral iso-liftisations*,

$$\begin{aligned} &Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A}) \\ &\qquad \subseteq Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \cap D^{-1}_{U,\mathrm{i},T,z}(A * T) \\ &\qquad = Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap D^{-1}_{U\mathrm{i},T,z}(A * T) \end{aligned}$$

The *idealisation perturbation conjecture* states that of all the *integral iso-idealisations*, $A - D + I \in Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})$, that have given *alignment-bounded lifted iso-transform space*, $(A - D + I) * T = A * T$, the *integral sample idealisation*, $A * T * T^{\dagger A}$, has the greatest *multinomial probability*

$$\begin{aligned} A * T * T^{\dagger A} \quad \in \quad &\mathrm{maxd}(\{(A - D + I,\ \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A - D + I)) : \\ &\qquad A - D + I \in Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})\}) \end{aligned}$$

and hence the least *alignment-bounded iso-transform space*

$$\begin{aligned} A * T * T^{\dagger A} \quad \in \quad &\mathrm{mind}(\{(A - D + I,\ -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A - D + I)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)}) : \\ &\qquad A - D + I \in Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})\}) \end{aligned}$$

or

$$\begin{aligned} &\forall A - D + I \in Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A}) \\ &\quad \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A - D + I)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)} \geq -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A * T * T^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)} \right) \end{aligned}$$

In other words, conjecture that the *idealisation*, $A * T * T^{\dagger A}$, is the most conservative choice of *iso-idealisations* of the *sample histogram* given the *alignment-bounded lifted iso-transform space* of the *lifted sample*. Depending on the degree to which the *transform* is *formal*, $A * T \approx A^{\mathrm{X}} * T$, the *independent* approximates to the *neutralisation*, $A^{\mathrm{X}} = A^{\mathrm{X}} * T * T^{\odot A^{\mathrm{X}}} \approx A * T * T^{\odot A^{\mathrm{X}}}$, and thence to the *idealisation*, $A * T * T^{\odot A^{\mathrm{X}}} \approx A * T * T^{\dagger A}$. That is, the *idealisation* approximates most closely to the *independent* which is the *mean* of the *distribution*, $A * T * T^{\dagger A} \approx A^{\mathrm{X}} = \mathrm{mean}(\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z))$ and is therefore the most *probable* of the *integral iso-idealisations*.

The subset of the *integral iso-transform-independents* given $A * T$ is the intersection between the *integral iso-transform-independents* and the *integral iso-deriveds*, $Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \cap D^{-1}_{U,\mathrm{i},T,z}(A * T)$, which is the *integral iso-liftisations*, $Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap D^{-1}_{U,\mathrm{i},T,z}(A * T)$. Note that the *idealisation*, $A * T * T^{\dagger A}$, does not necessarily have the least *space* of all of the *iso-liftisations*. The *integral iso-liftisations* with the least *space* is in

$$\begin{aligned} \mathrm{mind}(\{(A - D + I,\ \ -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A - D + I)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)}) :\\ A - D + I \in Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap D^{-1}_{U\mathrm{i},T,z}(A * T)\}) \end{aligned}$$

For example, if the *liftisation-independent* is *integral* it is conjectured to be in the *integral iso-liftisations*, $A^{\mathrm{K}(T)} \in \mathcal{A}_{\mathrm{i}} \implies A^{\mathrm{K}(T)} \in Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap D^{-1}_{U\mathrm{i},T,z}(A * T)$, but in some cases it is not computable.

Given the conditions (i) $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, (ii) $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, and (iii) $A * T * T^{\dagger A} \in \mathcal{A}_{\mathrm{i}}$, let the negative logarithm *independent-sample-distributed iso-transform-independent conditional dependent multinomial probability* of the *idealisation*,

$$\begin{aligned} \Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A * T * T^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)} :\\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A * T * T^{\dagger A} \in \mathcal{A}_{\mathrm{i}} \Bigg) \in \ln \mathbf{Q}_{>0} \end{aligned}$$

be abbreviated to the *alignment-bounded iso-transform idealisation space*. The *alignment-bounded iso-transform idealisation space* is a special case of the *alignment-bounded iso-transform space* where $E^{\mathrm{X}} = A^{\mathrm{X}}$ and $A = A * T * T^{\dagger A}$. The *alignment-bounded iso-transform idealisation space* and the *alignment-bounded iso-transform space* both *lift* to the same *alignment-bounded lifted iso-transform space*.

Formal equals *abstract* implies *formal independent* equals *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$, so the *independent* is an *iso-transform-independent*, $A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, and therefore the *alignment-bounded iso-transform idealisation space* is bounded by the *idealisation alignment*, given the *integral mean multinomial probability distribution conjecture*,

$$\begin{aligned}
&\mathrm{algn}(A * T * T^{\dagger A}) \\
\leq\ & \Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A * T * T^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)} : \\
&\qquad A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A * T * T^{\dagger A} \in \mathcal{A}_{\mathrm{i}} \Bigg) \\
\leq\ & \mathrm{algn}(A * T * T^{\dagger A}) + \ln |\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)|
\end{aligned}$$

Note also that the set of *integral iso-transform-independents* given $A * T$, $Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \cap D^{-1}_{U,\mathrm{i},T,z}(A * T)$, is partitioned by sets of *integral iso-idealisations*. This is because the equivalence classes corresponding to the intersection between the *integral iso-transform-independents* and the *integral iso-deriveds* is a parent partition of the equivalence classes of the *integral iso-idealisation* function, $\mathrm{parent}(\{\{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ B * T = A * T\} : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}, \{\{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B * T * T^{\dagger B} = A * T * T^{\dagger A}\} : A \in \mathcal{A}_{U,\mathrm{i},V,z}\})$. That is, $Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \cap D^{-1}_{U,\mathrm{i},T,z}(A * T) \in \mathrm{P}(\mathrm{ran}(Y^{-1}_{U,\mathrm{i},T,\dagger,z}))$. However, only one of these sets of *integral iso-transform-independents* corresponds to the *sample histogram*, $Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A*T*T^{\dagger A})$. This is the only set of *integral iso-idealisations* which is such that the *independent* of the *idealisation* equals the *distribution histogram*, $(A*T*T^{\dagger A})^{\mathrm{X}} = A^{\mathrm{X}}$.

If the *transform* is a *self partition transform*, for example $T_{\mathrm{s}} = V^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, or it is *value full functional*, for example $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, then the *transform* is *ideal*, $A*T_{\mathrm{s}}*T_{\mathrm{s}}^{\dagger A} = A$, and the *alignment-bounded iso-transform idealisation space* equals the *alignment-bounded iso-transform-independent space*. Therefore, in this case, T_{s}, the *alignment-bounded iso-transform idealisation space* is bounded by the *derived alignment*.

If the *transform* is a *unary partition*, for example $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U,\mathrm{f},1}$, then the *idealisation* equals the *independent*, $A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A} = A^{\mathrm{X}}$. The *alignment-bounded lifted iso-transform space* is still equal to zero,

$$-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}} * T_{\mathrm{u}}, z)(A^{\mathrm{X}} * T_{\mathrm{u}})}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}} * T_{\mathrm{u}}, z)(B')} = 0$$

The lower bound is still zero

$$\mathrm{algn}(A^{\mathrm{X}} * T_{\mathrm{u}}) = 0$$

because $A^{\mathrm{X}} * T_{\mathrm{u}} = (A^{\mathrm{X}} * T_{\mathrm{u}})^{\mathrm{X}}$. The upper bound is also zero

$$\mathrm{algn}(A^{\mathrm{X}} * T_{\mathrm{u}}) + \ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)| = 0$$

The *alignment-bounded iso-transform idealised space* is greater than zero

$$\begin{aligned}
& -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)} \\
= & -\ln \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}}) \\
> & \ 0
\end{aligned}$$

However, given the *integral mean multinomial probability distribution conjecture*, the *alignment-bounded iso-transform idealised space* is less than or equal to the *alignment-bounded iso-transform space*

$$-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)} \leq -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}$$

So the *alignment-bounded iso-transform idealised space* is less out of bounds than the *alignment-bounded iso-transform space.*

Given the *complete integral congruent support sample histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$, consider the comparison of two *transforms* $T_1, T_2 \in \mathcal{T}_{U,\mathrm{f},1}$, where $\mathrm{und}(T_1) = \mathrm{und}(T_2) = V$. The sets of *iso-transform-independents* for each *transform*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)$ and $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$, may or may not be equal,

$$\begin{aligned}
& T_1 = T_2 \implies \\
& \quad Y^{-1}_{U,\mathrm{i},T_1,z}(((A^{\mathrm{X}} * T_1),(A * T_1)^{\mathrm{X}})) = Y^{-1}_{U,\mathrm{i},T_2,z}(((A^{\mathrm{X}} * T_2),(A * T_2)^{\mathrm{X}}))
\end{aligned}$$

Note that in the case that the *derived variables* are equal, $W_1 = W_2$ where $W_1 = \mathrm{der}(T_1)$ and $W_2 = \mathrm{der}(T_2)$, but the *transforms* are not equal, $T_1 \neq T_2$, it is not necessarily the case that the *deriveds* are not equal, $T_1 \neq T_2 \Longleftarrow A * T_1 \neq A * T_2$, the *formals* are not equal, $T_1 \neq T_2 \Longleftarrow A^{\mathrm{X}} * T_1 \neq A^{\mathrm{X}} * T_2$, or the *abstracts* are not equal, $(A * T_1)^{\mathrm{X}} \neq (A * T_2)^{\mathrm{X}}$. Conversely, even if the *formals* are equal and the *abstracts* are equal, $(A^{\mathrm{X}} * T_1 = A^{\mathrm{X}} * T_2) \wedge ((A * T_1)^{\mathrm{X}} = (A * T_2)^{\mathrm{X}})$, it is not necessarily the case that the *iso-transform-independents* are equal, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) = \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$.

In the case where the *iso-transform-independents* are not equal, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \neq \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$, the difference in negative logarithm *iso-transform-independent* *conditional dependent multinomial probability* is non-zero

$$\begin{aligned}&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B_2 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B_2)}\right) - \\ &\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\sum_{B_1 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B_1)}\right) \\ \neq\ & 0\end{aligned}$$

unless it so happens that denominators are equal

$$\sum_{B_1 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B_1) = \sum_{B_2 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B_2)$$

In the case that the denominators are not equal, conjecture that in general the larger the intersection, $|\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \cap \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)|$, the less the difference in the *iso-transform-independent conditional dependent multinomial probability* because the denominators are more nearly equal. For example, consider the case where the *derived variables* are equal, $W_1 = W_2$. Even if the *transforms* are not equal, $T_1 \neq T_2$, the *formals* may sometimes be equal, $A^{\mathrm{X}} * T_1 = A^{\mathrm{X}} * T_2$. If that is the case, then the intersection of the *iso-formals*, $|Y^{-1}_{U,\mathrm{i},T_1,\mathrm{V},z}(A^{\mathrm{X}} * T_1) \cap Y^{-1}_{U,\mathrm{i},T_2,\mathrm{V},z}(A^{\mathrm{X}} * T_2)|$, tends to be larger because it is more often the case that *iso-formal histograms* $B_1 \in Y^{-1}_{U,\mathrm{i},T_1,\mathrm{V},z}(A^{\mathrm{X}} * T_1)$ and $B_2 \in Y^{-1}_{U,\mathrm{i},T_2,\mathrm{V},z}(A^{\mathrm{X}} * T_2)$ are equal, $B_1 = B_2$, because $B_1^{\mathrm{X}} * T_1 = A^{\mathrm{X}} * T_1 = A^{\mathrm{X}} * T_2 = B_2^{\mathrm{X}} * T_2$. Similarly, the *abstracts* may sometimes be equal, $(A*T_1)^{\mathrm{X}} = (A*T_2)^{\mathrm{X}}$. If that is the case, then the intersection of the *iso-abstracts*, $|Y^{-1}_{U,\mathrm{i},T_1,\mathrm{W},z}((A * T_1)^{\mathrm{X}}) \cap Y^{-1}_{U,\mathrm{i},T_2,\mathrm{W},z}((A*T_2)^{\mathrm{X}})|$, tends to be larger because it is more often the case that *iso-abstract histograms* $B_1 \in Y^{-1}_{U,\mathrm{i},T_1,\mathrm{W},z}((A * T_1)^{\mathrm{X}})$ and $B_2 \in Y^{-1}_{U,\mathrm{i},T_2,\mathrm{W},z}((A * T_2)^{\mathrm{X}})$ are equal, $B_1 = B_2$, because $(B_1 * T_1)^{\mathrm{X}} = (A * T_1)^{\mathrm{X}} = (A * T_2)^{\mathrm{X}} = (B_2 * T_2)^{\mathrm{X}}$. If either the *iso-formals* intersect or the *iso-abstracts* intersect, then it is sometimes the case that the *iso-transform-independents* intersect, because the *iso-transform-independents* is the intersection between the *iso-formals* and *iso-abstracts*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) = Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) = Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T) \cap Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$.

Now consider the case where (i) the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, (ii) the *idealisations* are *integral*, $A * T_1 * T_1^{\dagger A}, A * T_2 * T_2^{\dagger A} \in \mathcal{A}_{\mathrm{i}}$, and so are in the same set of *integral iso-independents*, $A * T_1 * T_1^{\dagger A}, A * T_2 * T_2^{\dagger A} \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$. The difference in negative logarithm *iso-transform-independent conditional* *dependent multinomial idealisation probability* is

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A * T_2 * T_2^{\dagger A})}{\sum_{B_2 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B_2)}\right) - \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E,z)(A * T_1 * T_1^{\dagger A})}{\sum_{B_1 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)} \hat{Q}_{\mathrm{m},U}(E,z)(B_1)}\right)$$

which is not necessarily zero even if the *iso-transform-independents* are equal, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) = \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$, unless it so happens that the numerators are also equal, $\hat{Q}_{\mathrm{m},U}(E,z)(A * T_1 * T_1^{\dagger A}) = \hat{Q}_{\mathrm{m},U}(E,z)(A * T_2 * T_2^{\dagger A})$.

If it is the case that (iii) the *iso-transform-independents* are not equal, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \neq \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$, but (iv) the *derived variables* are equal, $W_1 = W_2$, (v) the *formals* are equal, $A^{\mathrm{X}} * T_1 = A^{\mathrm{X}} * T_2$, and (vi) the *abstracts* are equal, $(A * T_1)^{\mathrm{X}} = (A * T_2)^{\mathrm{X}}$, then the intersection of the *iso-transform-independents* includes both *idealisations*, $\{A * T_1 * T_1^{\dagger A}, A * T_2 * T_2^{\dagger A}\} \subset \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \cap \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$. That is, the *iso-transform-independents* of the first *transform* includes the *idealisation* of the second *transform*, $A*T_2*T_2^{\dagger A} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)$, and vice-versa, $A * T_1 * T_1^{\dagger A} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$.

If it is the case instead that the *derived variables* are not necessarily equal, (iii) the *distribution histogram* is the *independent sample*, $E = A^{\mathrm{X}}$, and (iv) the *formal independent* equals the *abstract* of both *transforms*, $(A^{\mathrm{X}} * T_1)^{\mathrm{X}} = (A * T_1)^{\mathrm{X}}$ and $(A^{\mathrm{X}} * T_2)^{\mathrm{X}} = (A * T_2)^{\mathrm{X}}$, so that the *integral independent*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, is an *integral iso-transform-independent* for both *transforms*, $A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)$ and $A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$, then the change in negative logarithm *independent-sample-distributed iso-transform-independent conditional dependent multinomial idealisation probability*, given the *integral mean multinomial probability distribution conjecture*, is the difference in the *alignments* of the *idealisations* plus a difference in terms that do not depend on the *idealisations* but only on the *iso-transform-independents*

$$\begin{aligned}
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A * T_2 * T_2^{\dagger A})}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} + \ln \sum_{B_2 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B_2)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}\right) - \\
&\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A * T_1 * T_1^{\dagger A})}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} + \ln \sum_{B_1 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B_1)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}\right) \\
= \ & \mathrm{algn}(A * T_2 * T_2^{\dagger A}) - \mathrm{algn}(A * T_1 * T_1^{\dagger A}) + \\
&\left(\ln \sum_{B_2 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B_2)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})} - \ln \sum_{B_1 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)} \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B_1)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}\right)
\end{aligned}$$

In the case where the *iso-transform-independents* are equal, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) = \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$, then the difference is just the difference in *alignments*, $\mathrm{algn}(A * T_2 * T_2^{\dagger A}) - \mathrm{algn}(A * T_1 * T_1^{\dagger A})$.

If it is the case instead that (iii) the *distribution histogram* is the *independent sample*, $E = A^{\mathrm{X}}$, and more strictly (iv) the *formal* equals the *abstract* of both *transforms*, $A^{\mathrm{X}} * T_1 = (A * T_1)^{\mathrm{X}}$ and $A^{\mathrm{X}} * T_2 = (A * T_2)^{\mathrm{X}}$, then the change in negative logarithm *independent-sample-distributed iso-transform-independent conditional dependent multinomial idealisation probability* is the change in *alignment-bounded iso-transform idealisation space*

$$\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A * T_2 * T_2^{\dagger A})}{\sum_{B_2 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B_2)}\right) - \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A * T_1 * T_1^{\dagger A})}{\sum_{B_1 \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B_1)}\right)$$

and the difference in *alignment-bounded lifted iso-transform space* is difference in *derived alignments* plus the difference in *alignment-bounded lifted iso-transform errors*

$$\begin{aligned}
& \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}} * T_2,z)(A * T_2)}{\sum_{B'_2 \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}} * T_2,z)(B'_2)}\right) - \\
& \left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}} * T_1,z)(A * T_1)}{\sum_{B'_1 \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}} * T_1,z)(B'_1)}\right) \\
= \quad & \mathrm{algn}(A * T_2) - \mathrm{algn}(A * T_1) + \\
& \left(\ln \sum_{B'_2 \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A)} \frac{\prod_{R \in (A*T_2)^{\mathrm{XS}}} (A * T_2)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}_2} (B'_2)_R!} - \right. \\
& \left. \ln \sum_{B'_1 \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)} \frac{\prod_{R \in (A*T_1)^{\mathrm{XS}}} (A * T_1)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}_1} (B'_1)_R!}\right)
\end{aligned}$$

The difference in *alignment-bounded lifted iso-transform errors* is bounded

$$\begin{aligned}
& -\ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)| \\
\leq \quad & \left(\ln \sum_{B'_2 \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A)} \frac{\prod_{R \in (A*T_2)^{\mathrm{XS}}} (A * T_2)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}_2} (B'_2)_R!} - \right. \\
& \left. \ln \sum_{B'_1 \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)} \frac{\prod_{R \in (A*T_1)^{\mathrm{XS}}} (A * T_1)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}_1} (B'_1)_R!}\right) \\
\leq \quad & \ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A)|
\end{aligned}$$

In the case where the *derived variables* are equal, $W_1 = W_2$, and the *abstracts* are equal, $(A * T_1)^{\mathrm{X}} = (A * T_2)^{\mathrm{X}}$, then the numerators of the *alignment-bounded lifted iso-transform errors* are equal and so the difference in *errors* tends to be smaller. Conjecture that in the case where the *derived variables* are equal, $W_1 = W_2$, in general the larger the intersection between the *lifted iso-transform-independents*, $|\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \cap \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A)|$, the less the difference in *alignment-bounded lifted iso-transform errors* and the more nearly the difference in *alignment-bounded lifted iso-transform space* equals the difference in *derived alignments*, $\mathrm{algn}(A * T_2) - \mathrm{algn}(A * T_1)$. Conjecture also that in the case where the *derived variables* are not equal, $W_1 \neq W_2$, in general the larger the intersection between the *iso-transform-independents*, $|\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \cap \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)|$, the smaller the difference in *alignment-bounded lifted iso-transform errors*. As conjectured above, the *alignment-bounded iso-transform idealisation space* is functionally related to the *alignment-bounded lifted iso-transform space*, so conjecture that in general the smaller the difference in *alignment-bounded lifted iso-transform errors* the more nearly the difference in *alignment-bounded iso-transform idealisation space* equals the difference in *derived alignments*, $\mathrm{algn}(A * T_2 * T_2^{\dagger A}) - \mathrm{algn}(A * T_1 * T_1^{\dagger A})$.

For convenience, define $\mathrm{ln!} \in \mathcal{A}_{\mathrm{i}} \to \ln \mathbf{Q}_{>0}$ as

$$\mathrm{ln!}(A) := \sum_{S \in A^{\mathrm{S}}} \ln A_S! = \ln \prod_{S \in A^{\mathrm{S}}} A_S!$$

ln! is undefined where $A = \emptyset$. The *alignment* of an *integral-independent histogram*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, may be expressed in terms of the *non-independent* term and the *independent* term as

$$\mathrm{algn}(A) = \mathrm{ln!}(A) - \mathrm{ln!}(A^{\mathrm{X}})$$

Define $\mathrm{lnhar!} \in \mathrm{P}(\mathcal{A}_{\mathrm{i}}) \to \ln \mathbf{Q}_{>0}$ as

$$\mathrm{lnhar!}(X) := \ln \sum_{B \in X} \frac{1}{\prod_{R \in B^{\mathrm{S}}} B_R!}$$

lnhar! is undefined where $X = \emptyset$ or $\emptyset \in X$. The *alignment-bounded lifted iso-transform space* may be expressed

$$\begin{aligned}
&\Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(B')} : \\
&\qquad E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}},\ E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \Bigg) \\
&= \mathrm{algn}(A * T) + \ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!} \\
&= \mathrm{algn}(A * T) + \ln!((A * T)^{\mathrm{X}}) + \mathrm{lnhar!}(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)) \\
&= \ln!(A * T) + \mathrm{lnhar!}(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A))
\end{aligned}$$

The *alignment-bounded lifted iso-transform space error* may be expressed as $\ln!((A * T)^{\mathrm{X}}) + \mathrm{lnhar!}(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A))$ which is such that

$$0 \leq \ln!((A * T)^{\mathrm{X}}) + \mathrm{lnhar!}(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)) \leq \ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)|$$

A *value roll* $(V, v, s, t) \in \mathrm{rollValues}(U)$ is such that the *independent* of the application of the *value roll*, (V, v, s, t), to a *histogram* A is equal to the application of the *value roll* to the *independent histogram*, $(A*(V, v, s, t)^{\mathrm{R}})^{\mathrm{X}} = A^{\mathrm{X}} * (V, v, s, t)^{\mathrm{R}}$. The *transform* of the *value roll*, $T = (V, v, s, t)^{\mathrm{T}}$, is therefore such that the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * (V, v, s, t)^{\mathrm{T}} = (A * (V, v, s, t)^{\mathrm{T}})^{\mathrm{X}}$. The *iso-transform-independents*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, are the set of *complete congruent histograms* having the same *perimeters* as the *value rolled histogram*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ (B * (V, v, s, t)^{\mathrm{R}})^{\mathrm{X}} = (A * (V, v, s, t)^{\mathrm{R}})^{\mathrm{X}}\}$.

The *derived alignment* in *variables* $\mathrm{der}(T)$ is equal to the *alignment* in *variables* V of the *rolled histogram*, $\mathrm{algn}(A*T) = \mathrm{algn}(A*(V, v, s, t)^{\mathrm{R}})$. If the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, then the *value roll transform*, T, satisfies the constraints required so that the negative logarithm *lifted independent-sample-distributed iso-transform-independent quasi-conditional dependent multinomial probability* is bounded by the *derived alignment*, $\mathrm{algn}(A * (V, v, s, t)^{\mathrm{T}})$. That is, the *alignment-bounded lifted iso-transform space* of the *value roll*,

(V,v,s,t), is bounded by the *value rolled histogram alignment*

$$\begin{aligned}
&\mathrm{algn}(A * (V,v,s,t)^{\mathrm{R}}) \\
\leq\ &\left(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}} * T, z)(B')} : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \right) \\
\leq\ &\mathrm{algn}(A * (V,v,s,t)^{\mathrm{R}}) + \ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)|
\end{aligned}$$

The *value roll transform*, $T = (V,v,s,t)^{\mathrm{T}}$, is *ideal*, $A = A * T * T^{\dagger A}$, if the source and target *values*, s and t, are *independent*, $A * (\{v\} \times \{s,t\})^{\mathrm{U}} = A^{\mathrm{X}} * (\{v\} \times \{s,t\})^{\mathrm{U}}$.

As shown above in section 'Deltas and Perturbations', the application of a *value roll* $(V,v,s,t) \in \mathrm{rollValues}(U)$ cannot be *iso-independence conserving*, $A * (V,v,s,t)^{\mathrm{R}} \notin Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, but it is *iso-transform-independence conserving*, $A * (V,v,s,t)^{\mathrm{R}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, where $T = (V,v,s,t)^{\mathrm{T}}$.

A pair of *non-circular value rolls* $(V,v_1,s_1,t_1),(V,v_2,s_2,t_2) \in \mathrm{rollValues}(U)$ cannot be in the same set of *iso-transform-independents*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \neq \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$ where $T_1 = (V,v_1,s_1,t_1)^{\mathrm{T}}$ and $T_2 = (V,v_2,s_2,t_2)^{\mathrm{T}}$, unless it so happens that $(A * (V,v_1,s_1,t_1)^{\mathrm{R}})^{\mathrm{X}} = (A * (V,v_2,s_2,t_2)^{\mathrm{R}})^{\mathrm{X}}$, for example where $v_1 = v_2$, $t_1 = t_2$ and $A\%\{v_1\}(\{(v_1,s_1)\}) = A\%\{v_1\}(\{(v_1,s_2)\})$. The *alignment-bounded lifted iso-transform error* varies with the *derived volume*. So the *errors* are more comparable if the *derived volumes* are equal, $|W_1^{\mathrm{C}}| = |W_2^{\mathrm{C}}|$ where $W_1 = \mathrm{der}(T_1)$ and $W_2 = \mathrm{der}(T_2)$. That is, if the *histogram* is *effectively regular* in *variables* v_1 and v_2, $|(A\%\{v_1\})^{\mathrm{F}}| = |(A\%\{v_2\})^{\mathrm{F}}|$.

Consider a pair of *value rolls* applied in sequence, $A*(V,v_1,s_1,t_1)^{\mathrm{R}}*(V,v_2,s_2,t_2)^{\mathrm{R}}$ where $T_1 = (V,v_1,s_1,t_1)^{\mathrm{T}}$ and $T_2 = (V,v_2,s_2,t_2)^{\mathrm{T}}$. The first *value rolled histogram*, $A * (V,v_1,s_1,t_1)^{\mathrm{R}}$, is a member of both sets of *iso-transform-independents*, $A*(V,v_1,s_1,t_1)^{\mathrm{R}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \cap \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A*(V,v_1,s_1,t_1)^{\mathrm{R}})$.

The *iso-transform-independents* of the pair of *value rolls* applied in sequence, $A * (V,v_1,s_1,t_1)^{\mathrm{R}} * (V,v_2,s_2,t_2)^{\mathrm{R}}$ is $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$ where $T = ((V,v_2,s_2,t_2)^{\mathrm{R}} \circ (V,v_1,s_1,t_1)^{\mathrm{R}})^{\mathrm{T}}$. This set is the union

$$\begin{aligned}
&\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) = \\
&\quad \{B : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A),\ B * (V,v_1,s_1,t_1)^{\mathrm{R}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A * (V,v_1,s_1,t_1)^{\mathrm{R}})\} \cup \\
&\quad \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A * (V,v_1,s_1,t_1)^{\mathrm{R}})
\end{aligned}$$

The sets of *iso-transform-independents* intersect, $A*(V,v_1,s_1,t_1)^{\mathrm{R}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \cap \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A * (V,v_1,s_1,t_1)^{\mathrm{R}})$. Therefore the cardinality of the sequence *iso-*

transform-independents is less than or equal to the sum of the cardinalities of the *value rolls* cumulatively or separately applied,

$$\begin{aligned} |\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)| &\leq |\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)\}| + |\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A * (V, v_1, s_1, t_1)^{\mathrm{R}})| \\ &\leq |\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A)| + |\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)| \end{aligned}$$

The cardinality of the sequence *lifted iso-transform-independents* is also less than or equal to the sum of the cardinalities of the *lifted iso-transform-independents* of the *value rolls* cumulatively or separately applied,

$$\begin{aligned} |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)| &\leq |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)\}| + |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A * (V, v_1, s_1, t_1)^{\mathrm{R}})| \\ &\leq |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)| + |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A)| \end{aligned}$$

The *alignment-bounded lifted iso-transform error* depends on the set of *lifted iso-transform-independents*, $\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)$, and so the *alignment-bounded lifted iso-transform error* of a pair of *value rolls* in sequence is conjectured to be less than or equal to the sum of the *alignment-bounded lifted iso-transform errors* of the *value rolls* cumulatively applied

$$\begin{aligned} &\mathrm{ln}!((A * T)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)) \\ \leq\ &\mathrm{ln}!((A * T_1)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)) + \\ &\mathrm{ln}!((A * (V, v_1, s_1, t_1)^{\mathrm{R}} * T_2)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A * (V, v_1, s_1, t_1)^{\mathrm{R}})) \end{aligned}$$

or separately applied

$$\begin{aligned} &\mathrm{ln}!((A * T)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)) \\ \leq\ &\mathrm{ln}!((A * T_1)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)) + \\ &\mathrm{ln}!((A * T_2)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A)) \end{aligned}$$

Let $\mathcal{J}_{U,V}$ be the set of lists of *value rolls* in *variables* V and *system* U, $\mathcal{J}_{U,V} = \{L : L \in \mathcal{L}(\mathrm{rollValues}(U)),\ (\forall (W, \cdot, \cdot, \cdot) \in \mathrm{set}(L)\ (W = V))\}$. The *transform* of a *non-circular value roll list* $J \in \mathcal{J}_{U,V}$ is also *formal abstract equivalent*, $A^{\mathrm{X}} * J^{\mathrm{T}} = (A * J^{\mathrm{T}})^{\mathrm{X}}$, so the *alignment-bounded lifted iso-transform space* of the *value roll list*, J, is also bounded by the *derived alignment*, $\mathrm{algn}(A * J^{\mathrm{T}})$. That is, because the *alignment-bounded lifted iso-transform space* of each of the successive *formal abstract equivalent* applications of the *value rolls* in the *value roll list*, $A * J^{\mathrm{R}}_{\{1 \ldots i-1\}} * J^{\mathrm{R}}_i$, is bounded, the application of the entire *value roll list* at once, $A * J^{\mathrm{R}}$, must also be bounded. The change in *derived alignment* of the application *value roll list* equals the sum of the changes in *derived alignment*, $\mathrm{algn}(A * J^{\mathrm{R}}) - \mathrm{algn}(A) = (\mathrm{algn}(A * J^{\mathrm{R}}_1) - \mathrm{algn}(A)) + \sum_{i \in 2 \ldots |J|} (\mathrm{algn}(A * J^{\mathrm{R}}_{\{1 \ldots i-1\}} * J^{\mathrm{R}}_i) - \mathrm{algn}(A * J^{\mathrm{R}}_{\{1 \ldots i-2\}} * J^{\mathrm{R}}_{i-1}))$. The *alignment-bounded lifted iso-transform error* of the *value roll list* is conjectured to be less than or equal to the sum of the *alignment-bounded lifted iso-transform errors* of the *value rolls* cumulatively applied,

$$\begin{aligned} & \ln!((A * J^{\mathrm{R}})^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},J^{\mathrm{T}},z}(A)) \\ \leq & \sum_{i \in 1 \ldots |J|} (\ln!((A * J^{\mathrm{R}}_{\{1 \ldots i\}})^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},J^{\mathrm{T}}_{\{1 \ldots i\}},z}(A)) \end{aligned}$$

or individually applied

$$\begin{aligned} & \ln!((A * J^{\mathrm{R}})^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},J^{\mathrm{T}},z}(A)) \\ \leq & \sum_{i \in 1 \ldots |J|} (\ln!((A * J^{\mathrm{R}}_i)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},J^{\mathrm{T}}_i,z}(A)) \end{aligned}$$

This is the case regardless of the order of the *value roll list*.

A *reduction* $A\%K$ of the *histogram*, A, to *variables* $K \subset V$ can be viewed as a set of *value roll lists* for each of the *reduced variables*, $V \setminus K$, such that the *values* are *rolled* to a single *value*. For example, in the case of a *reduction* by a single *variable*, let $\{v\} = V \setminus K$, $M \in \mathrm{enums}(U_v)$, $L = \mathrm{flip}(M)$, $d = |U_v|$ and $J = \{(i, (V, v, L_i, L_d)) : i \in \{1 \ldots d-1\}\} \in \mathcal{J}_{U,V}$. Then $\mathrm{algn}(A * J^{\mathrm{R}}) = \mathrm{algn}(A \ \% \ (V \setminus \{v\})) = \mathrm{algn}(A\%K)$. Therefore the *alignment-bounded lifted iso-transform space* of a *reduction transform*, $T = \{w^{\mathrm{CSVT}} : w \in K\}^{\mathrm{T}}$, is also bounded by the *derived alignment*, $\mathrm{algn}(A * T) = \mathrm{algn}(A\%K)$.

As noted above, two *value rolls* $(V, v_1, s_1, t_1), (V, v_2, s_2, t_2) \in \mathrm{rollValues}(U)$ cannot be in the same set of *iso-transform-independents*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \neq \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$ where $T_1 = (V, v_1, s_1, t_1)^{\mathrm{T}}$ and $T_2 = (V, v_2, s_2, t_2)^{\mathrm{T}}$, if the *abstract histograms* are not equal, $(A * (V, v_1, s_1, t_1)^{\mathrm{R}})^{\mathrm{X}} \neq (A * (V, v_2, s_2, t_2)^{\mathrm{R}})^{\mathrm{X}}$. Thus the difference in *alignment-bounded lifted iso-transform errors*, $(\ln!((A * T_2)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A))) - (\ln!((A * T_1)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)))$, is sometimes non-zero. However, the intersection of the *iso-transform independents* includes both the *sample*, A, and the *integral independent sample*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$. That is, $A, A^{\mathrm{X}} \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_1,z}(A) \cap \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$. In fact, although the *abstract histograms* are different, parts of the *rolled histograms* are common in each case, $A * (V, v_1, s_1, t_1)^{\mathrm{R}} * X = A * (V, v_2, s_2, t_2)^{\mathrm{R}} * X$ where $X = (\{v_1\}^{\mathrm{CS}} \setminus (\{v_1\} \times \{s_1, t_1\}))^{\mathrm{U}} * (\{v_2\}^{\mathrm{CS}} \setminus (\{v_2\} \times \{s_2, t_2\}))^{\mathrm{U}}$. That is, the *perimeters* of A are unchanged except at four *values*, $\forall w \in V \ (A * (V, v_1, s_1, t_1)^{\mathrm{R}} \ \% \ \{w\} * X = A * (V, v_2, s_2, t_2)^{\mathrm{R}} \ \% \ \{w\} * X)$. The differences are fewer if $v_1 = v_2$ and $|\{s_1, s_2, t_1, t_2\}| < 4$. The difference in *alignment-bounded lifted iso-transform errors* is therefore sometimes smaller than would be the case if the *value rolls* were applied to different *sample histograms*,

$A * (V, v_1, s_1, t_1)^{\mathrm{R}}$ and $B * (V, v_2, s_2, t_2)^{\mathrm{R}}$ where $B \in \mathcal{A}_{U,\mathrm{i},V,z} \setminus \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_2,z}(A)$. That is, in some cases $|(\ln!((A * T_2)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(A))) - (\ln!((A * T_1)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)))| \leq |(\ln!((B * T_2)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_2,z}(B))) - (\ln!((A * T_1)^{\mathrm{X}}) + \mathrm{lnhar}!(\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T_1,z}(A)))|$. Similarly, two *non-circular value roll lists* $J_x, J_y \in \mathcal{J}_{U,V}$ applied to the same *sample*, $A * J_x^{\mathrm{R}}$ and $A * J_y^{\mathrm{R}}$, will sometimes have a smaller difference in the sum of *alignment-bounded lifted iso-transform errors* if they intersect, $\mathrm{set}(J_x) \cap \mathrm{set}(J_y) \neq \emptyset$.

In section 'Substrate structures', above, it is shown that the *non-overlapping substrate transforms set*, $\mathcal{T}_{U,V,\mathrm{n}}$, can be constructed from *linear fuds* where the first *transform* is a *non-overlapping substrate self-cartesian transform*, $\mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n}}$, and the subsequent *transforms* are *self substrate decremented transforms*, $\mathcal{T}_{U,W,-} \cap \mathcal{T}_{U,W,\mathrm{n,s}}$. In turn, the *linear fuds* of *self substrate decremented transforms* correspond bijectively to the *non-circular unique-source value roll lists*, $\mathcal{J}_{U,V,-} \subset \mathcal{J}_{U,V}$.

A *non-overlapping transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$, where $\mathrm{und}(T) = V$ and $\neg\mathrm{overlap}(T)$, is such that the *formal histogram* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, but does not necessarily imply that the *formal* is *abstract*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}} \Longleftarrow A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$. Therefore, even if the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, the *transform*, T, does not necessarily satisfy the constraint required so that the negative logarithm *lifted independent-sample-distributed iso-transform-independent quasi-conditional dependent multinomial probability* is bounded by the *derived alignment*, $\mathrm{algn}(A * T)$.

Even where the *non-overlapping transform* is a *non-overlapping substrate self-cartesian transform*, $T \in \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n}}$, it is not necessarily the case that the *formal* is *abstract*. However, in the special case where the *transform* is the singleton *self substrate self-cartesian transform*, $\{T\} = \mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n,s}}$, it is then a *value full functional transform*, $T = \{\{v\}^{\mathrm{CS}\{\}V\mathrm{T}} : v \in V\}^{\mathrm{T}}$, and hence the *formal* equals the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$.

In the case where a *non-overlapping transform*, $\neg\mathrm{overlap}(T)$, is *mono-derived-variate*, $|\mathrm{der}(T)| = 1$, then the *derived histogram* is necessarily *independent*, $A * T = (A * T)^{\mathrm{X}}$. In the case where the *formal* also equals the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, the *derived histogram* must be purely *formal*, $A * T = (A * T)^{\mathrm{X}} = A^{\mathrm{X}} * T$. In any case both the *derived* and *formal* have zero *alignment*, $\mathrm{algn}(A * T) = \mathrm{algn}(A^{\mathrm{X}} * T) = 0$.

As shown above, *value roll transforms*, which correspond to *self substrate* *decremented transforms*, are such that the *formal* equals the *abstract*. So a *non-overlapping transform* that can be constructed from a *linear fud* of the *value full functional transform*, $\mathcal{T}_{U,V,\mathrm{c}} \cap \mathcal{T}_{U,V,\mathrm{n,s}}$, followed by sequence of *self substrate decremented transforms*, $\mathcal{T}_{U,W,-} \cap \mathcal{T}_{U,W,\mathrm{n,s}}$, must also be such that the *formal* equals the *abstract*. This is because in this case the *non-overlapping transform* is a *self non-overlapping substrate transform*, $\forall T \in \mathcal{T}_{U,V,\mathrm{n,s}}\ (A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}})$.

4.19 Substrate structures alignment

Some of the conjectures of approximations and relations between variables stated in the previous discussion of *alignment* may be formalised in terms of the statistics of real-valued functions on a support of *distinct geometry sized cardinal substrate histograms*. The set of *sized cardinal substrate histograms* $\mathcal{A}_z$, defined above in section 'Distinct geometry sized cardinal substrate histograms', is the set of *complete integral cardinal substrate histograms* of *size* z and *dimension* less than or equal to the *size* such that the *independent* is *completely effective*

$$\mathcal{A}_z = \{A : A \in \mathcal{A}_\mathrm{c} \cap \mathcal{A}_\mathrm{i},\ \mathrm{size}(A) = z,\ |V_A| \leq z,\ A^{\mathrm{U}} = A^{\mathrm{XF}} = A^{\mathrm{C}}\}$$

Each *substrate histogram* $A \in \mathcal{A}_z$ has $|V_A|! \prod_{w \in V_A} |U_A(w)|!$ *cardinal substrate permutations*. These *frame* mappings partition the *substrate histograms* into equivalence classes having the same *geometry*. Let P_z be the partition, $P_z \in \mathrm{B}(\mathcal{A}_z)$, such that the components of P_z are the equivalence classes by *cardinal substrate permutation*, $\forall C \in P_z\ \forall A \in C\ (|C| = |V_A|! \prod_{w \in V_A} |U_A(w)|)$.

Each of the *substrate histograms* in a component of P_z, that are equivalent by *cardinal substrate permutation*, have the same *alignment*, $\forall C \in P_z\ \forall A, B \in C\ (\mathrm{algn}(A) = \mathrm{algn}(B))$ where algn = alignment.

If the *substrate histograms* are partitioned, for example to analyse correlations grouped by low or high *alignment*, then the partition should be a parent partition of P_z. That is, the *substrate histograms* partition should be independent of *cardinal substrate permutation*.

The central moment functions of the *renormalised geometry-weighted probability function*, $\hat{R}_z$, that operate on real-valued functions of the *sized cardinal* *substrate histograms*, $\mathcal{A}_z \to \mathbf{R}$, are defined

$$\begin{aligned}
\mathrm{ex}(z)(F) &:= \mathrm{expected}(\hat{R}_z)(F) \\
\mathrm{var}(z)(F) &:= \mathrm{variance}(\hat{R}_z)(F) \\
\mathrm{cov}(z)(F,G) &:= \mathrm{covariance}(\hat{R}_z)(F,G) \\
\mathrm{corr}(z)(F,G) &:= \mathrm{correlation}(\hat{R}_z)(F,G)
\end{aligned}$$

where

$$\hat{R}_z = \mathrm{normalise}(\{(A, \frac{1}{|V_A|! \prod_{w \in V_A} |U_A(w)|!}) : A \in \mathrm{dom}(F)\}) \in \mathcal{P}$$

4.19.1 Iso-independent conditional

Define the subset of the *sized cardinal substrate histograms*, $\mathcal{A}_z$, for which the *independent*, A^{X}, is *integral*, and therefore also a *substrate histogram*, as the *integral-independent substrate histograms*,

$$\mathcal{A}_{z,\mathrm{xi}} = \{A : A \in \mathcal{A}_z,\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\} = \{A : A, A^{\mathrm{X}} \in \mathcal{A}_z\} \subset \mathcal{A}_z$$

Define the *alignment substrate function* for some *size* z, $X_{z,\mathrm{a}} \in \mathcal{A}_z \to \mathbf{R}$, as

$$X_{z,\mathrm{a}} = \{(A, \mathrm{algn}(A)) : A \in \mathcal{A}_z\}$$

This may be equally be expressed as the *negative logarithm independent-sample-distributed relative dependent multinomial probability density substrate function*

$$X_{z,\mathrm{a}} = \{(A, -\ln \frac{\mathrm{mpdf}(U_A)(A^{\mathrm{X}}, z)(A)}{\mathrm{mpdf}(U_A)(A^{\mathrm{X}}, z)(A^{\mathrm{X}})}) : A \in \mathcal{A}_z\}$$

where for some $(E, z) \in \mathcal{A}_U \times \mathbf{Q}_{\geq 0}$ the *multinomial probability density function*, $\mathrm{mpdf}(U)(E, z) \in \mathcal{A}_{U,V,z} \to \mathbf{R}_{\geq 0}$, is defined

$$\mathrm{mpdf}(U)(E,z)(A) := \frac{\Gamma_! z}{\prod_{S \in A^{\mathrm{S}}} \Gamma_! A_S} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S}$$

Let the subset of the *alignment substrate function* for which the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, be defined the *alignment integral-independent substrate function*, $X_{z,\mathrm{xi,a}} \in \mathcal{A}_{z,\mathrm{xi}} \to \ln \mathbf{Q}_{>0}$, as

$$X_{z,\mathrm{xi,a}} = \mathrm{filter}(\mathcal{A}_{z,\mathrm{xi}}, X_{z,\mathrm{a}}) = \{(A, \mathrm{algn}(A)) : A \in \mathcal{A}_z,\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\} \subset X_{z,\mathrm{a}}$$

which may be be expressed in terms of the *generalised multinomial probability distribution* as the *independent-sample-distributed relative dependent multinomial space substrate function*

$$X_{z,\mathrm{xi,a}} = \{(A, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A^{\mathrm{X}})}) : A \in \mathcal{A}_{z,\mathrm{xi}}\}$$

where for some $(E, z) \in \mathcal{A}_U \times \mathbf{N}$ the *generalised multinomial probability distribution* $\hat{Q}_{\mathrm{m},U}(E, z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is defined

$$\hat{Q}_{\mathrm{m},U}(E, z)(A) := \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S}$$

Define the *independent-sample-distributed iso-independent conditional dependent multinomial space substrate function* for some *size* z, $X_{z,\mathrm{y}} \in \mathcal{A}_z \to \ln \mathbf{Q}_{>0}$, as

$$X_{z,\mathrm{y}} = \{(A, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\sum_{B \in Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)}) : A \in \mathcal{A}_z\}$$

where the *integral iso-independent* function, $Y_{U,\mathrm{i},V,z} \in \mathcal{A}_{U,\mathrm{i},V,z} \to \mathcal{A}_{U,V,z}$, is defined

$$Y_{U,\mathrm{i},V,z} = \{(A, A^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\} \subset Y_{U,V,z} \subset \text{independent}$$

In the case where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, the *independent-sample-distributed iso-independent conditional dependent multinomial space*, $X_{z,\mathrm{y}}(A)$, is the *alignment-bounded iso-independent space*,

$$\begin{aligned}
&\left(X_{z,\mathrm{y}}(A) : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\right) \\
= \quad &\left(-\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\sum_{B \in Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)} : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\right) \\
= \quad &\mathrm{algn}(A) + \ln \sum_{B \in Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}
\end{aligned}$$

As shown above, in 'Alignment and conditional probability', given the *minimum alignment conjecture*, the *alignment-bounded iso-independent space* is bounded

$$\begin{aligned}
& \text{algn}(A) \\
\leq\ & \text{algn}(A) + \ln \sum_{B \in Y^{-1}_{U_A,\text{i},V_A,z}(A^{\text{X}})} \frac{\prod_{S \in A^{\text{XS}}} A^{\text{X}}_S!}{\prod_{S \in B^{\text{S}}} B_S!} \\
\leq\ & \text{algn}(A) + \ln |Y^{-1}_{U_A,\text{i},V_A,z}(A^{\text{X}})|
\end{aligned}$$

Therefore the *independent-sample-distributed iso-independent conditional dependent multinomial space integral-independent substrate function*, defined $X_{z,\text{xi},\text{y}} = \text{filter}(\mathcal{A}_{z,\text{xi}}, X_{z,\text{y}}) \subset X_{z,\text{y}}$, also known as the *alignment-bounded iso-independent space substrate function*, is correlated with the *alignment integral-independent substrate function*, $X_{z,\text{xi},\text{a}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(X_{z,\text{xi},\text{y}}, X_{z,\text{xi},\text{a}}) \geq 0)$$

The *alignment* is an underestimate of the *alignment-bounded iso-independent space* and hence the expected *alignment* must be less than or equal to the expected *alignment-bounded iso-independent space* for all *sizes* of *substrate histograms*

$$\forall z \in \mathbf{N}_{>0}\ (\text{ex}(z)(X_{z,\text{xi},\text{y}}) \geq \text{ex}(z)(X_{z,\text{xi},\text{a}}))$$

This is derived from the expected *alignment-bounded iso-independent error*

$$\forall z \in \mathbf{N}_{>0}\ (\text{ex}(z)(X_{z,\text{xi},\text{y}} - X_{z,\text{xi},\text{a}}) \geq 0)$$

where the *alignment-bounded iso-independent error* is

$$X_{z,\text{xi},\text{y}}(A) - X_{z,\text{xi},\text{a}}(A) = \ln \sum_{B \in Y^{-1}_{U_A,\text{i},V_A,z}(A^{\text{X}})} \frac{\prod_{S \in A^{\text{XS}}} A^{\text{X}}_S!}{\prod_{S \in B^{\text{S}}} B_S!}$$

The *alignment-bounded iso-independent error* increases with *size*

$$\forall z_1, z_2 \in \mathbf{N}_{>0}\ (z_2 > z_1 \implies \text{ex}(z_2)(X_{z_2,\text{xi},\text{y}} - X_{z_2,\text{xi},\text{a}}) \geq \text{ex}(z_1)(X_{z_1,\text{xi},\text{y}} - X_{z_1,\text{xi},\text{a}}))$$

The *alignment-bounded iso-independent error* varies with *volume*

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(X_{z,\text{xi},\text{y}} - X_{z,\text{xi},\text{a}}, \{(A, |A^{\text{C}}|) : A \in \mathcal{A}_z\}) \geq 0)$$

The *alignment-bounded iso-independent error* varies with *valency* entropy for given *volume*

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(X_{z,\text{xi},\text{y}} - X_{z,\text{xi},\text{a}}, X_{z,\text{h},\text{U}}) \geq 0)$$

where $X_{z,\mathrm{h,U}} = \{(A, \mathrm{entropy}(\{(w, |U_A(w)|) : w \in V_A\})/|A^{\mathrm{C}}|) : A \in \mathcal{A}_z\}$.

Conjecture that the *alignment-bounded iso-independent error* varies with the entropy of the *independent*

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(X_{z,\mathrm{xi,y}} - X_{z,\mathrm{xi,a}}, X_{z,\mathrm{h,x}}) \geq 0)$$

where $X_{z,\mathrm{h,x}} = \{(A, \mathrm{entropy}(A^{\mathrm{X}})) : A \in \mathcal{A}_z\}$.

Therefore conjecture that the *alignment-bounded iso-independent error* varies with the *alignment*

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(X_{z,\mathrm{xi,y}} - X_{z,\mathrm{xi,a}}, X_{z,\mathrm{xi,a}}) \geq 0)$$

but that overall the *alignment-bounded iso-independent error* ratio varies against the *alignment*

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)((X_{z,\mathrm{xi,y}} - X_{z,\mathrm{xi,a}})/X_{z,\mathrm{xi,a}}, X_{z,\mathrm{xi,a}}) \leq 0)$$

where the *alignment-bounded iso-independent error* ratio is

$$\frac{X_{z,\mathrm{xi,y}}(A) - X_{z,\mathrm{xi,a}}(A)}{X_{z,\mathrm{xi,a}}(A)} = \left(\ln \sum_{B \in Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} A^{\mathrm{X}}_S!}{\prod_{S \in B^{\mathrm{S}}} B_S!}\right) / \mathrm{algn}(A)$$

which is defined for *non-independent sample*, $A \neq A^{\mathrm{X}}$.

Conjecture that the correlation between the *alignment-bounded iso independent space substrate function*, $X_{z,\mathrm{xi,y}}$, and the *alignment integral independent substrate function*, $X_{z,\mathrm{xi,a}}$, can be extended from the *integral-independent substrate histograms*, $\mathcal{A}_{z,\mathrm{xi}}$, to all *substrate histograms*, $\mathcal{A}_z$. That is, conjecture that the *independent-sample-distributed iso-independent conditional dependent multinomial space substrate function*, $X_{z,\mathrm{y}}$, and the *alignment substrate function*, $X_{z,\mathrm{a}}$, are also correlated,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(X_{z,\mathrm{y}}, X_{z,\mathrm{a}}) \geq 0)$$

Further, conjecture that the *independent-sample-distributed iso-independent conditional dependent multinomial space error*, $X_{z,\mathrm{y}}(A) - X_{z,\mathrm{a}}(A)$, increases with *size*,

$$\forall z_1, z_2 \in \mathbf{N}_{>0}\ (z_2 > z_1 \implies \mathrm{ex}(z_2)(X_{z_2,\mathrm{y}} - X_{z_2,\mathrm{a}}) \geq \mathrm{ex}(z_1)(X_{z_1,\mathrm{y}} - X_{z_1,\mathrm{a}}))$$

varies with *volume*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(X_{z,\mathrm{y}} - X_{z,\mathrm{a}}, \{(A, |A^{\mathrm{C}}|) : A \in \mathcal{A}_z\}) \geq 0)$$

varies with *valency* entropy for given *volume*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(X_{z,\mathrm{y}} - X_{z,\mathrm{a}}, X_{z,\mathrm{h,U}}) \geq 0)$$

varies with the entropy of the *independent*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(X_{z,\mathrm{y}} - X_{z,\mathrm{a}}, X_{z,\mathrm{h,x}}) \geq 0)$$

and varies with the *alignment*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(X_{z,\mathrm{y}} - X_{z,\mathrm{a}}, X_{z,\mathrm{a}}) \geq 0)$$

Conjecture that the *independent-sample-distributed iso-independent conditional dependent multinomial space error* ratio, $(X_{z,\mathrm{y}}(A) - X_{z,\mathrm{a}}(A))/X_{z,\mathrm{a}}(A)$, varies against the *alignment*

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)((X_{z,\mathrm{y}} - X_{z,\mathrm{a}})/X_{z,\mathrm{a}}, X_{z,\mathrm{a}}) \leq 0)$$

However, although there is a correlation between the *independent-sample-distributed iso-independent conditional dependent multinomial space substrate function*, $X_{z,\mathrm{y}}$, and the *alignment substrate function*, $X_{z,\mathrm{a}}$, the correlation is less than for the subset where the *independent* is *integral*,

$$\forall z \in \mathbf{N}_{\geq t}\ (\mathrm{corr}(z)(X_{z,\mathrm{xi,y}}, X_{z,\mathrm{xi,a}}) \geq \mathrm{corr}(z)(X_{z,\mathrm{y}}, X_{z,\mathrm{a}}))$$

where threshold $t \in \mathbf{N}_{>0}$ is the minimum *size* such that the variances are non-zero, $\forall z \in \mathbf{N}_{\geq t}\ (\mathrm{var}(z)(X_{z,\mathrm{xi,y}}) > 0 \wedge \mathrm{var}(z)(X_{z,\mathrm{xi,a}}) > 0)$. The correlation is lower because the *independent-sample-distributed iso-independent conditional dependent multinomial space*,

$$X_{z,\mathrm{y}}(A) = -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\sum_{B \in Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)}$$

is not bounded by the *alignment*, $X_{z,\mathrm{a}}(A) = \mathrm{algn}(A)$. First, the upper bound, $\mathrm{algn}(A) + \ln |Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})|$, of the *alignment-bounded iso-independent space*, $X_{z,\mathrm{xi,y}}(A) = (X_{z,\mathrm{y}}(A) : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}})$, may be exceeded because *non-integral independents* are excluded from the *minimum alignment conjecture*. So, in some cases the *alignment* of an *iso-independent histogram* $B \in Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})$ may be negative

$$\mathrm{algn}(B) < 0 \implies \frac{\prod_{S \in A^{\mathrm{XS}}} \Gamma_! A^{\mathrm{X}}_S}{\prod_{S \in B^{\mathrm{S}}} B_S!} > 1$$

and thus in some cases

$$\ln \sum_{B \in Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} \Gamma_! A^{\mathrm{X}}_S}{\prod_{S \in B^{\mathrm{S}}} B_S!} > \ln |Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})|$$

Second, the lower bound, $\mathrm{algn}(A)$, of the *alignment-bounded iso-independent space*, $X_{z,\mathrm{xi,y}}(A)$, may not be reached in some other cases because the *non-integral independent* is not an *iso-independent*, $A^{\mathrm{X}} \notin Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})$, and hence there does not necessarily exist a term in

$$\sum_{B \in Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} \Gamma_! A^{\mathrm{X}}_S}{\prod_{S \in B^{\mathrm{S}}} B_S!}$$

which is equal to 1. Thus in some cases

$$\ln \sum_{B \in Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})} \frac{\prod_{S \in A^{\mathrm{XS}}} \Gamma_! A^{\mathrm{X}}_S}{\prod_{S \in B^{\mathrm{S}}} B_S!} < 0$$

4.19.2 Iso-transform-independent conditional

Define the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space substrate transform search set*, for some *size* z, $X_{z,\mathrm{T,y}} \in \mathcal{A}_z \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_z$ define $X_{z,\mathrm{T,y}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$X_{z,\mathrm{T,y}}(A) = \{(T, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\sum_{B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)}) : T \in \mathcal{T}_{U_A,V_A}\}$$

where the *integral iso-transform-independents* is abbreviated

$$\begin{aligned} \mathcal{A}_{U,\mathrm{i,y},T,z}(A) &= Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \\ &= \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\} \end{aligned}$$

and the *substrate transforms set* is defined

$$\mathcal{T}_{U,V} = \{F^{\mathrm{T}} : F \subseteq \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\}\}$$

In the case where (i) the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and (ii) the *formal independent histogram* equals the *abstract histogram*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} =$

$(A * T)^{\mathrm{X}}$, which together imply that the *independent* is in the *integral iso-transform-independents*,

$$(A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}) \wedge ((A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}) \implies A^{\mathrm{X}} \in \mathcal{A}_{U_A,\mathrm{i},\mathrm{y},T,z}(A)$$

the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space* can be expressed in terms of the *alignment*,

$$\begin{aligned}
&(X_{z,\mathrm{T},\mathrm{y}}(A)(T) : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}) \\
= &\left(-\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\sum_{B \in \mathcal{A}_{U_A,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)} : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\right) \\
= &\ \mathrm{algn}(A) + \ln \sum_{B \in \mathcal{A}_{U_A,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)}{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A^{\mathrm{X}})}
\end{aligned}$$

As conjectured above, in 'Derived alignment and conditional probability', given the *integral mean multinomial probability distribution conjecture*, the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space* is bounded,

$$\begin{aligned}
&\mathrm{algn}(A) \\
\leq\ &\mathrm{algn}(A) + \ln \sum_{B \in \mathcal{A}_{U_A,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)}{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A^{\mathrm{X}})} \\
\leq\ &\mathrm{algn}(A) + \ln |\mathcal{A}_{U_A,\mathrm{i},\mathrm{y},T,z}(A)|
\end{aligned}$$

Let the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate independent-formal-abstract transform search set*, which is constrained such that (i) $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$ and (ii) $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$, be defined $X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_{z,\mathrm{xi}}$ define $X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}(A) = \{(T, y) : (T, y) \in X_{z,\mathrm{T},\mathrm{y}}(A),\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}$$

Conjecture that, given these constraints, the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate independent-formal-abstract transform minimum function*, $\mathrm{minr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}$, is correlated with the *alignment integral-independent substrate function*, $X_{z,\mathrm{xi},\mathrm{a}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{minr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}, X_{z,\mathrm{xi},\mathrm{a}}) \geq 0)$$

As shown in section 'Derived alignment and conditional probability', above, the *iso-transform-independent conditional dependent multinomial probability* for the *self partition transform* case, for example $T_{\mathrm{s}} = V_A^{\mathrm{CS}\{\}\mathrm{T}} \in \mathcal{T}_{U_A,V_A}$, or the *value full functional transform* case, for example $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V_A\}^{\mathrm{T}} \in \mathcal{T}_{U_A,V_A}$, is greater than or equal to that for the *unary partition transform* case, $T_{\mathrm{u}} = \{V_A^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U_A,V_A}$,

$$\frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)} \geq \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}$$

and hence the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space* of the *self partition* case, T_{s}, is less than or equal to that of the *unary partition transform* case, T_{u},

$$-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)} \leq -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(B)}$$

So, to the degree that the *minimum transforms*, $\mathrm{mind}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}(A)) \subset \mathcal{T}_{U_A,V_A}$, tend to be closer to *self partition transforms*, T_{s}, rather than to the *unary partition transforms*, T_{u}, the correlation between $\mathrm{minr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}$ and $X_{z,\mathrm{xi},\mathrm{a}}$ is similar to that between $X_{z,\mathrm{xi},\mathrm{y}}$ and $X_{z,\mathrm{xi},\mathrm{a}}$. In the special case that the *self partition* is a *minimum transform*, $T_{\mathrm{s}} \in \mathrm{mind}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}(A))$, the set of *integral iso-transform-independents* equals the set of *integral iso-independents*, $\mathcal{A}_{U_A,\mathrm{i},\mathrm{y},T_{\mathrm{s}},z}(A) = Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})$ and hence

$$\begin{aligned} &T_{\mathrm{s}} \in \mathrm{mind}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}(A)) \implies \\ &\quad \mathrm{minr}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}(A)) = X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}(A)(T_{\mathrm{s}}) = X_{z,\mathrm{xi},\mathrm{y}}(A) \end{aligned}$$

Consider, in contrast, the *transform maximum function*. The *independent-sample-distributed iso-transform-independent conditional dependent multinomial probability*

$$\frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U_A,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}},z)(B)}$$

is always least where the *transform* is a *unary partition*, $T_{\mathrm{u}} = \{V_A^{\mathrm{CS}}\}^{\mathrm{T}} \in \mathcal{T}_{U_A,V_A}$, because the set of *integral iso-transform-independents* equals the *integral congruent support*, $\mathcal{A}_{U_A,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A) = \mathcal{A}_{U_A,\mathrm{i},V_A,z}$. The maximum counterpart, the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate independent-formal-abstract transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{xfa}}$, is such that

$T_{\rm u} \in \mathrm{maxd}(X_{z,\mathrm{xi,T,y,xfa}}(A))$. The *maximum function* simply equals the *generalised multinomial space*,

$$\begin{aligned}
\mathrm{maxr}(X_{z,\mathrm{xi,T,y,xfa}}(A)) &= -\ln \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}},z)(A) \\
&= -\ln \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{A^{\mathrm{X}}_S}{z}\right)^{A_S} \\
&= \sum_{S \in A^{\mathrm{S}}} \ln A_S! - \sum_{S \in A^{\mathrm{S}}} A_S \ln A^{\mathrm{X}}_S + z \ln z - \ln z! \\
&= \mathrm{algn}(A) - \sum_{S \in A^{\mathrm{S}}} A_S \ln A^{\mathrm{X}}_S + \sum_{S \in A^{\mathrm{XS}}} \ln A^{\mathrm{X}}_S! + z \ln z - \ln z!
\end{aligned}$$

Therefore it is correlated with the *alignment integral-independent substrate function*

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,xfa}}, X_{z,\mathrm{xi,a}}) \geq 0)$$

but the correlation is lower

$$\forall z \in \mathbf{N}_{\geq t}\ (\mathrm{corr}(z)(\mathrm{minr} \circ X_{z,\mathrm{xi,T,y,xfa}}, X_{z,\mathrm{xi,a}}) \geq \mathrm{corr}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,xfa}}, X_{z,\mathrm{xi,a}}))$$

where the threshold $t \in \mathbf{N}_{>0}$ is the minimum *size* such that the variances are non-zero, $\forall z \in \mathbf{N}_{\geq t}\ \forall F \in \{\mathrm{minr} \circ X_{z,\mathrm{xi,T,y,xfa}}, \mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,xfa}}, X_{z,\mathrm{xi,a}}\}\ (\mathrm{var}(z)(F) > 0)$.

Conjecture that the *transform minimum function* correlation between the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate independent-formal-abstract transform minimum function*, $\mathrm{minr} \circ X_{z,\mathrm{xi,T,y,xfa}}$, and the *alignment integral-independent substrate function*, $X_{z,\mathrm{xi,a}}$, is less than the corresponding correlation between the *alignment-bounded iso-independent space substrate function*, $X_{z,\mathrm{xi,y}}$, and the *alignment integral-independent substrate function*, $X_{z,\mathrm{xi,a}}$

$$\forall z \in \mathbf{N}_{\geq t}\ (\mathrm{corr}(z)(X_{z,\mathrm{xi,y}}, X_{z,\mathrm{xi,a}}) \geq \mathrm{corr}(z)(\mathrm{minr} \circ X_{z,\mathrm{xi,T,y,xfa}}, X_{z,\mathrm{xi,a}}))$$

where the threshold $t \in \mathbf{N}_{>0}$ is the minimum *size* such that the variances are non-zero, $\forall z \in \mathbf{N}_{\geq t}\ \forall F \in \{\mathrm{minr} \circ X_{z,\mathrm{xi,T,y,xfa}}, X_{z,\mathrm{xi,y}}, X_{z,\mathrm{xi,a}}\}\ (\mathrm{var}(z)(F) > 0)$.

Extending the *transform minimum function* correlation to the cases where the *independent* is not necessarily *integral* and hence not in the *integral iso-transform-independents*, conjecture that the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space substrate* *transform minimum function*, $\text{minr} \circ X_{z,\text{T},\text{y}}$, is correlated with the *alignment substrate function*,

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(\text{minr} \circ X_{z,\text{T},\text{y}}, X_{z,\text{a}}) \geq 0)$$

but conjecture that this correlation is less than the corresponding constrained correlation

$$\forall z \in \mathbf{N}_{\geq t}\ (\text{corr}(z)(\text{minr} \circ X_{z,\text{xi},\text{T},\text{y},\text{xfa}}, X_{z,\text{xi},\text{a}}) \geq \text{corr}(z)(\text{minr} \circ X_{z,\text{T},\text{y}}, X_{z,\text{a}}))$$

above the variance threshold *size* t.

Also conjecture that this correlation is less than the corresponding correlation between the *independent-sample-distributed iso-independent conditional dependent multinomial space substrate function*, $X_{z,\text{y}}$, and the *alignment substrate function*, $X_{z,\text{a}}$

$$\forall z \in \mathbf{N}_{\geq t}\ (\text{corr}(z)(X_{z,\text{y}}, X_{z,\text{a}}) \geq \text{corr}(z)(\text{minr} \circ X_{z,\text{T},\text{y}}, X_{z,\text{a}}))$$

above the variance threshold *size* t.

Consider a stricter case of the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate independent-formal-abstract transform search set*, $X_{z,\text{xi},\text{T},\text{y},\text{xfa}}$, which is also constrained such that the *transforms* are *ideal*. That is, given constraints (i) $A^{\text{X}} \in \mathcal{A}_{\text{i}}$, (ii) $(A^{\text{X}} * T)^{\text{X}} = (A * T)^{\text{X}}$, and (iii) $A = A * T * T^{\dagger A}$, define the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal independent formal-abstract transform search set* $X_{z,\text{xi},\text{T},\text{y},\text{xfa},\text{j}} \in \mathcal{A}_{z,\text{xi}} \to (\mathcal{T}_{\text{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_{z,\text{xi}}$ define $X_{z,\text{xi},\text{T},\text{y},\text{xfa},\text{j}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$X_{z,\text{xi},\text{T},\text{y},\text{xfa},\text{j}}(A) = \{(T, y) : (T, y) \in X_{z,\text{xi},\text{T},\text{y},\text{xfa}}(A),\ A = A * T * T^{\dagger A}\}$$

Conjecture that the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal independent-formal-abstract transform maximum function*, $\text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{xfa},\text{j}}$, is better correlated with the *alignment integral-independent substrate function*, $X_{z,\text{xi},\text{a}}$, than the case where *non-ideal transforms* are allowed

$$\forall z \in \mathbf{N}_{\geq t}\ (\text{corr}(z)(\text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{xfa},\text{j}}, X_{z,\text{xi},\text{a}}) \geq \text{corr}(z)(\text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{xfa}}, X_{z,\text{xi},\text{a}}))$$

above the variance threshold *size* t. Here it is no longer the case that the *unary partition transform*, $T_{\text{u}} = \{V_A^{\text{CS}}\}^{\text{T}}$, necessarily has the minimum *probability* and hence the maximum *space*. It is only for the *independent sample*,

$A = A^{\mathrm{X}}$, that the *unary partition transform* is *ideal*, $\mathrm{ideal}(A^{\mathrm{X}}, T_{\mathrm{u}})$.

Extending the *ideal transform search set* to the cases where the *independent* is not necessarily *integral* and hence not in the *integral iso-transform-independents*, define the *independent-sample-distributed iso-transform independent conditional dependent multinomial space substrate ideal transform search set* $X_{z,\mathrm{T,y,j}} \in \mathcal{A}_z \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_z$ define $X_{z,\mathrm{T,y,j}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$\begin{aligned} X_{z,\mathrm{T,y,j}}(A) &= \{(T,y) : (T,y) \in X_{z,\mathrm{T,y}}(A),\ A = A * T * T^{\dagger A}\} \\ &= \{(T, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}},z)(B)}) : \\ &\qquad T \in \mathcal{T}_{U_A,V_A},\ A = A * T * T^{\dagger A}\} \end{aligned}$$

Conjecture that the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space substrate ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{T,y,j}}$, is correlated with the *alignment substrate function*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{T,y,j}}, X_{z,\mathrm{a}}) \geq 0)$$

Certainly it is still the case that where the *maximum transform* is a *self partition transform* or a *value full functional*, the set of *integral iso-transform-independents* equals the set of *integral iso-independents*, $\mathcal{A}_{U_A,\mathrm{i,y},T_{\mathrm{s}},z}(A) = Y^{-1}_{U_A,\mathrm{i},V_A,z}(A^{\mathrm{X}})$, and so $X_{z,\mathrm{T,y,j}}(A)(T_{\mathrm{s}}) = X_{z,\mathrm{y}}(A)$. Even so, conjecture that this correlation is less than the corresponding correlation between the *independent-sample-distributed iso-independent conditional dependent multinomial space substrate function*, $X_{z,\mathrm{y}}$, and the *alignment substrate function*, $X_{z,\mathrm{a}}$

$$\forall z \in \mathbf{N}_{\geq t}\ (\mathrm{corr}(z)(X_{z,\mathrm{y}}, X_{z,\mathrm{a}}) \geq \mathrm{corr}(z)(\mathrm{maxr} \circ X_{z,\mathrm{T,y,j}}, X_{z,\mathrm{a}}))$$

above the variance threshold *size* t.

Alignment, $X_{z,\mathrm{a}}$, by itself is a weaker proxy for the *iso-transform independent* case, $\mathrm{maxr} \circ X_{z,\mathrm{T,y,j}}$, than for the *iso-independent* case, $X_{z,\mathrm{y}}$, because the *alignment* expression does not depend on *transform*. To obtain a better correlated expression in terms of *derived alignment*, *idealised alignment* and *actualised alignment*, consider the *lifted* case.

Define the *derived alignment substrate transform search set*, for some *size* z, $X'_{z,\mathrm{T,a}} \in \mathcal{A}_z \to (\mathcal{T}_{\mathrm{f}} \to \mathbf{R})$, and for some $A \in \mathcal{A}_z$ define $X'_{z,\mathrm{T,a}}(A) \in \mathcal{T}_{U_A,V_A} \to \mathbf{R}$ as

$$X'_{z,\mathrm{T,a}}(A) = \{(T, \mathrm{algn}(A * T)) : T \in \mathcal{T}_{U_A,V_A}\}$$

In terms of the *multinomial probability density function*, $\mathrm{mpdf}(U)(E,z) \in \mathcal{A}_{U,V,z} \to \mathbf{R}_{\geq 0}$, the *derived alignment substrate transform search set* is

$$X'_{z,\mathrm{T,a}}(A) = \{(T, -\ln \frac{\mathrm{mpdf}(U_A)((A*T)^{\mathrm{X}},z)(A*T)}{\mathrm{mpdf}(U_A)((A*T)^{\mathrm{X}},z)((A*T)^{\mathrm{X}})}) : T \in \mathcal{T}_{U_A,V_A}\}$$

Define the *derived alignment integral-independent substrate transform search set* as $X'_{z,\mathrm{xi,T,a}} = \mathrm{filter}(\mathcal{A}_{z,\mathrm{xi}}, X'_{z,\mathrm{T,a}}) \subset X'_{z,\mathrm{T,a}}$, which is such that $X'_{z,\mathrm{xi,T,a}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \mathbf{R})$.

Define the *derived alignment integral-independent substrate formal-abstract transform search set*, which is constrained such that (i) the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$ and (ii) the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}}$, for some *size* z, $X'_{z,\mathrm{xi,T,a,fa}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_{z,\mathrm{xi}}$ define $X'_{z,\mathrm{xi,T,a,fa}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$\begin{aligned} X'_{z,\mathrm{xi,T,a,fa}}(A) &= \{(T, \mathrm{algn}(A*T)) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}}\} \\ &= \{(T,a) : (T,a) \in X'_{z,\mathrm{T,a}}(A),\ A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}}\} \end{aligned}$$

The *independent* is *integral* and the *formal histogram* equals the *abstract histogram*, so the *independent derived histogram*, or *abstract histogram*, must be *integral*, $(A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}) \wedge (A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}}) \implies (A*T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$. Thus the *derived alignment integral-independent substrate formal-abstract transform search set* can be defined in terms of the rational *generalised multinomial probability distribution* $\hat{Q}_{\mathrm{m},U}(E,z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$

$$X'_{z,\mathrm{xi,T,a,fa}}(A) =$$
$$\{(T, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}((A*T)^{\mathrm{X}},z)(A*T)}{\hat{Q}_{\mathrm{m},U_A}((A*T)^{\mathrm{X}},z)((A*T)^{\mathrm{X}})}) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}}\}$$

The *derived alignment integral-independent substrate formal-abstract transform search set* can also be defined as the *lifted alignment integral-independent substrate formal-abstract transform search set*

$$X'_{z,\mathrm{xi,T,a,fa}}(A) =$$
$$\{(T, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T,z)(A*T)}{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T,z)(A^{\mathrm{X}} * T)}) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}}\}$$

because *lifted* equals *derived* when the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}} \implies (A*T = (A*T)) \wedge (A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}})$.

Define the *derived alignment integral-independent substrate ideal formal-abstract transform search set*, which is additionally constrained such that (iii) the *transform* is *ideal*, $A = A * T * T^{\dagger A}$, for some *size* z, $X'_{z,\mathrm{xi,T,a,fa,j}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_{z,\mathrm{xi}}$ define $X'_{z,\mathrm{xi,T,a,fa,j}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$X'_{z,\mathrm{xi,T,a,fa,j}}(A) = \{(T,a) : (T,a) \in X'_{z,\mathrm{xi,T,a,fa}}(A),\ A = A * T * T^{\dagger A}\}$$

Define the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate formal-abstract transform search set*, which is constrained such that (i) the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, and (ii) the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, for some *size* z, $X_{z,\mathrm{xi,T,y,fa}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_{z,\mathrm{xi}}$ define $X_{z,\mathrm{xi,T,y,fa}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$\begin{aligned} X_{z,\mathrm{xi,T,y,fa}}(A) &= \{(T, X_{z,\mathrm{T,y}}(A)(T)) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\} \\ &= \{(T,y) : (T,y) \in X_{z,\mathrm{T,y}}(A),\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\} \end{aligned}$$

The *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate formal-abstract transform search set*, $X_{z,\mathrm{xi,T,y,fa}}(A)$, is a subset of the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate independent-formal-abstract transform search set*,

$$X_{z,\mathrm{xi,T,y,fa}}(A) \subseteq X_{z,\mathrm{xi,T,y,xfa}}(A)$$

because the constraints of (i) *integral independent*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$ and (ii) *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, imply that the *formal independent histogram* equals the *abstract histogram*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$, and the *independent* is an *iso-transform-independent*, $A^{\mathrm{X}} \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)$.

The *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate formal-abstract transform search set*, $X_{z,\mathrm{xi,T,y,fa}}$, can be abbreviated to the *alignment-bounded iso-transform space transform search set*,

$$\begin{aligned} & X_{z,\mathrm{xi,T,y,fa}}(A)(T) \\ = \ & (X_{z,\mathrm{T,y}}(A)(T) : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}) \\ = \ & \left(-\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}},z)(A)}{\sum_{B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}},z)(B)} : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \right) \end{aligned}$$

Note that '*alignment-bounded*' is a misnomer. That is, the *alignment-bounded iso-transform space*, $X_{z,\mathrm{xi,T,y,fa}}(A)(T)$, is not strictly bounded by *alignment* or *derived alignment*. However, its *lifted* counterpart, $X'_{z,\mathrm{xi,T,y,fa}}(A)(T)$, below, is bounded by *derived alignment* under the same constraints.

Define the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*, which is additionally constrained such that (iii) the *transform* is *ideal*, $A = A*T*T^{\dagger A}$, for some *size* z, $X_{z,\mathrm{xi,T,y,fa,j}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_{z,\mathrm{xi}}$ define $X_{z,\mathrm{xi,T,y,fa,j}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$X_{z,\mathrm{xi,T,y,fa,j}}(A) = \{(T,y) : (T,y) \in X_{z,\mathrm{xi,T,y,fa}}(A),\ A = A*T*T^{\dagger A}\}$$

As for the *ideal-agnostic* case, $X_{z,\mathrm{xi,T,y,fa}}(A)$, the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*, $X_{z,\mathrm{xi,T,y,fa,j}}(A)$, is a subset of the corresponding *ideal independent-formal-abstract transform search set*, $X_{z,\mathrm{xi,T,y,fa,j}}(A) \subseteq X_{z,\mathrm{xi,T,y,xfa,j}}(A)$.

The *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*, $X_{z,\mathrm{xi,T,y,fa,j}}$, can be abbreviated to the *alignment-bounded iso-transform space ideal transform search set*,

$$\begin{aligned}
& X_{z,\mathrm{xi,T,y,fa,j}}(A)(T) \\
= \ & (X_{z,\mathrm{T,y}}(A)(T) : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}},\ A = A*T*T^{\dagger A} \in \mathcal{A}_{\mathrm{i}}) \\
= \ & \Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}},z)(A*T*T^{\dagger A})}{\sum_{B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}},z)(B)} : \\
& \quad A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A*T)^{\mathrm{X}},\ A*T*T^{\dagger A} \in \mathcal{A}_{\mathrm{i}} \Bigg)
\end{aligned}$$

Again, this is a misnomer. That is, the *alignment-bounded iso-transform idealisation space*, $X_{z,\mathrm{xi,T,y,fa,j}}(A)(T)$, is not strictly bounded by *alignment* or *derived alignment*. However, its *lifted* counterpart, $X'_{z,\mathrm{xi,T,y,fa,j}}(A)(T)$, below, is bounded by *derived alignment* under the same constraints.

Define the *lifted independent-sample-distributed iso-transform-independent quasi-conditional dependent multinomial space substrate transform search set*, for some *size* z, $X'_{z,\mathrm{T,y}} \in \mathcal{A}_z \to (\mathcal{T}_\mathrm{f} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_z$ define $X'_{z,\mathrm{T,y}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$X'_{z,\mathrm{T,y}}(A) = \{(T, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U_A,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(B')}) : T \in \mathcal{T}_{U_A,V_A}\}$$

where the *lifted integral iso-transform-independents* is abbreviated

$$\begin{aligned}
\mathcal{A}'_{U,\mathrm{i,y},T,z}(A) &= \{B * T : B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A)\} \\
&= \{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\} \\
&= \{B * T : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}
\end{aligned}$$

In the case where (i) the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_\mathrm{i}$, and (ii) the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, the *lifted independent-sample-distributed iso-transform-independent quasi-conditional dependent multinomial space*, $X'_{z,\mathrm{T,y}}(A)(T)$, is the *alignment-bounded lifted iso-transform space*

$$\begin{aligned}
&(X'_{z,\mathrm{T,y}}(A)(T) : A^{\mathrm{X}} \in \mathcal{A}_\mathrm{i},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}) \\
&= \Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U_A,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(B')} : \\
&\qquad A^{\mathrm{X}} \in \mathcal{A}_\mathrm{i},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\Bigg) \\
&= \left(\mathrm{algn}(A * T) + \ln \sum_{B' \in \mathcal{A}'_{U_A,\mathrm{i,y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}\right)
\end{aligned}$$

As shown above, in 'Derived alignment and conditional probability', given the *minimum alignment conjecture*, the *alignment-bounded lifted iso-transform space* is bounded

$$\begin{aligned}
&\mathrm{algn}(A * T) \\
\leq\ &\Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U_A,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(B')} : \\
&\qquad A^{\mathrm{X}} \in \mathcal{A}_\mathrm{i},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\Bigg) \\
\leq\ &\mathrm{algn}(A * T) + \ln |\mathcal{A}'_{U_A,\mathrm{i,y},T,z}(A)|
\end{aligned}$$

So conjecture that the *lifted independent-sample-distributed iso-transform-independent quasi-conditional dependent multinomial space substrate transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{T},\mathrm{y}}$, is correlated with the *derived alignment substrate transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{T},\mathrm{a}}$, when constrained such that (i) $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, (ii) $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$.

Define the *lifted independent-sample-distributed iso-transform-independent quasi-conditional dependent multinomial space integral-independent substrate formal-abstract transform search set*, also known as the *alignment-bounded lifted iso-transform space transform search set*, for some *size* z, $X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_{z,\mathrm{xi}}$ define $X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}(A) = \{(T, X'_{z,\mathrm{T},\mathrm{y}}(A)(T)) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\}$$

That is,

$$X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}(A)(T) = (X'_{z,\mathrm{T},\mathrm{y}}(A)(T) : A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}})$$

Then the correlation between the *alignment-bounded lifted iso-transform space transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}$, and the *derived alignment integral-independent substrate formal-abstract transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa}}$, is such that

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa}}) \geq 0)$$

The *derived alignment* is an underestimate of the *alignment-bounded lifted iso-transform space* and hence the expected maximum *derived alignment* must be less than or equal to the expected maximum *alignment-bounded lifted iso-transform space* for all *sizes* of *substrate histograms*

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{ex}(z)(\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}) \geq \mathrm{ex}(z)(\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa}}))$$

This is derived from the expected maximum *alignment-bounded lifted iso-transform error*

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{ex}(z)(\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}} - \mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa}}) \geq 0)$$

which is related to the expected average *alignment-bounded lifted iso-transform error*

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{ex}(z)(\mathrm{average} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}} - \mathrm{average} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa}}) \geq 0)$$

where the *alignment-bounded lifted iso-transform error* for *substrate transform* $T \in \mathcal{T}_{U_A,V_A}$ is

$$X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}(A)(T) - X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa}}(A)(T) = \ln \sum_{B' \in \mathcal{A}'_{U_A,\mathrm{i},\mathrm{y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}$$

Now consider the correlation relationship between the *alignment-bounded iso-transform space* and its *lifted* counterpart, the *alignment-bounded lifted iso-transform space*. Conjecture that the *alignment-bounded lifted iso-transform space transform maximum function*, $\text{maxr} \circ X'_{z,\text{xi},\text{T},\text{y},\text{fa}}$, is correlated with the corresponding *alignment-bounded iso-transform space transform maximum function*, $\text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa}}$

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(\text{maxr} \circ X'_{z,\text{xi},\text{T},\text{y},\text{fa}}, \text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa}}) \geq 0)$$

Further, conjecture that the *alignment-bounded lifted iso-transform space transform average function*, $\text{average} \circ X'_{z,\text{xi},\text{T},\text{y},\text{fa}}$, is correlated with the corresponding *alignment-bounded iso-transform space transform average function*, $\text{average} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa}}$

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(\text{average} \circ X'_{z,\text{xi},\text{T},\text{y},\text{fa}}, \text{average} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa}}) \geq 0)$$

and that the *average function* correlation is greater than the *maximum function* correlation

$$\begin{aligned}
&\forall z \in \mathbf{N}_{\geq t} \\
&\quad (\text{corr}(z)(\text{average} \circ X'_{z,\text{xi},\text{T},\text{y},\text{fa}}, \text{average} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa}}) \\
&\quad \geq \ \text{corr}(z)(\text{maxr} \circ X'_{z,\text{xi},\text{T},\text{y},\text{fa}}, \text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa}}))
\end{aligned}$$

above the variance threshold *size* t.

Noting that the domains of the *transform functions* are equal,

$$\text{dom}(X'_{z,\text{xi},\text{T},\text{y},\text{fa}}(A)) = \text{dom}(X_{z,\text{xi},\text{T},\text{y},\text{fa}}(A))$$

consider a *transform* $T \in \text{dom}(X_{z,\text{xi},\text{T},\text{y},\text{fa}}(A))$.

If the *transform* T approximates more closely to the *self partition transform*, $T_\text{s} = V_A^{\text{CS}\{\}\text{T}} \in \mathcal{T}_{U_A,V_A}$, or the *value full functional transform*, $T_\text{s} = \{\{w\}^{\text{CS}\{\}\text{T}} : w \in V_A\}^\text{T} \in \mathcal{T}_{U_A,V_A}$, than it does to the *unary partition transform*, $T_\text{u} = \{V_A^{\text{CS}}\}^\text{T} \in \mathcal{T}_{U_A,V_A}$, then the *alignment-bounded iso-transform space* is approximately equal to the *alignment-bounded lifted iso-transform space*, $X_{z,\text{xi},\text{T},\text{y},\text{fa}}(A)(T) \approx X'_{z,\text{xi},\text{T},\text{y},\text{fa}}(A)(T)$. In the *self partition* case, T_s, the set of *integral iso-transform-independents* is bijective to the set of *lifted integral iso-transform-independents*, and so the *alignment-bounded iso-transform space* equals the *alignment-bounded lifted iso-transform space*,

$$\begin{aligned}
&-\ln \frac{\hat{Q}_{\text{m},U_A}(A^\text{X}, z)(A)}{\sum_{B \in \mathcal{A}_{U_A,\text{i},\text{y},T_\text{s},z}(A)} \hat{Q}_{\text{m},U_A}(A^\text{X}, z)(B)} \\
= \ &-\ln \frac{\hat{Q}_{\text{m},U_A}(A^\text{X} * T_\text{s}, z)(A * T_\text{s})}{\sum_{B' \in \mathcal{A}'_{U_A,\text{i},\text{y},T_\text{s},z}(A)} \hat{Q}_{\text{m},U}(A^\text{X} * T_\text{s}, z)(B')}
\end{aligned}$$

That is, $X_{z,\mathrm{xi,T,y,fa}}(A)(T_{\mathrm{s}}) = X'_{z,\mathrm{xi,T,y,fa}}(A)(T_{\mathrm{s}})$.

If the *transform* T approximates more closely to the *unary partition transform*, T_{u}, then the relationship between the *alignment-bounded lifted iso-transform space*, $X'_{z,\mathrm{xi,T,y,fa}}(A)(T)$, and the *alignment-bounded iso-transform space*, $X_{z,\mathrm{xi,T,y,fa}}(A)(T)$, is weaker. The *unary partition transform*, $T_{\mathrm{u}} \in \mathrm{maxd}(X_{z,\mathrm{xi,T,y,fa}}(A))$, has the largest *alignment-bounded iso-transform space*, $X_{z,\mathrm{xi,T,y,fa}}(A)(T_{\mathrm{u}}) = -\ln \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)$, but the *alignment-bounded lifted iso-transform space* is zero, $X'_{z,\mathrm{xi,T,y,fa}}(A)(T_{\mathrm{u}}) = 0$. Thus the *maximum function* correlation is lower than the *average function* correlation. In *pluri-valent* cases, the *maximum transforms* do not intersect,

$$\mathrm{maxd}(X_{z,\mathrm{xi,T,y,fa}}(A)) \cap \mathrm{maxd}(X'_{z,\mathrm{xi,T,y,fa}}(A)) = \emptyset$$

because $T_{\mathrm{u}} \notin \mathrm{maxd}(X'_{z,\mathrm{xi,T,y,fa}}(A))$.

In spite of the relatively weak correlation between the *alignment-bounded iso-transform space transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa}}$, and its *lifted* counterpart, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,y,fa}}$, conjecture that the *alignment-bounded iso-transform space transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa}}$, is transitively correlated with the *derived alignment integral-independent substrate formal-abstract transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa}}) \geq 0)$$

Define the *lifted independent-sample-distributed iso-transform-independent quasi-conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*, also known as the *alignment-bounded lifted iso-transform space ideal transform search set*, which is additionally constrained such that (iii) the *transform* is *ideal*, $A = A * T * T^{\dagger A}$, for some *size* z, $X'_{z,\mathrm{xi,T,y,fa,j}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, and for some $A \in \mathcal{A}_{z,\mathrm{xi}}$ as $X'_{z,\mathrm{xi,T,y,fa,j}}(A) \in \mathcal{T}_{U_A,V_A} \to \ln \mathbf{Q}_{>0}$ as

$$X'_{z,\mathrm{xi,T,y,fa,j}}(A) = \{(T, y) : (T, y) \in X'_{z,\mathrm{xi,T,y,fa}}(A),\ A = A * T * T^{\dagger A}\}$$

Just as for the *ideal-agnostic* case, $X'_{z,\mathrm{xi,T,y,fa}}$, above, there is a correlation between the *alignment-bounded lifted iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,y,fa,j}}$, and the *derived alignment integral-independent substrate ideal formal-abstract transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}) \geq 0)$$

But in this *ideal* case the correlation between the *alignment-bounded lifted iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}$, and the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}, \mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}) \geq 0)$$

is stronger than for the *ideal-agnostic* case

$$\begin{aligned} &\forall z \in \mathbf{N}_{\geq t} \\ &\qquad (\mathrm{corr}(z)(\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}, \mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}) \\ &\quad \geq \ \mathrm{corr}(z)(\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}, \mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}})) \end{aligned}$$

The reason is that now a *transform* $T \in \mathrm{dom}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}(A))$ that approximates more closely to the *unary partition transform*, T_{u}, must be *ideal* which is the case only for the subset of *substrate histograms* that are nearly *independent*, $A \approx A^{\mathrm{X}}$. So the variance of $\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}$ is less than that for $\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}$

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{var}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}) \geq \mathrm{var}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}))$$

The variance is also conjectured to be lower as a consequence of the *idealisation perturbation conjecture*. Let *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ have *alignment-bounded iso-transform space ideal maximum transform*, $T_{\mathrm{y}} \in \mathrm{maxd}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}(A)) \subset \mathcal{T}_{U_A,V_A}$, which is such that

$$\mathrm{maxr}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}(A)) = X_{z,\mathrm{T},\mathrm{y}}(A)(T_{\mathrm{y}}) = X_{z,\mathrm{T},\mathrm{y}}(A * T_{\mathrm{y}} * T_{\mathrm{y}}^{\dagger A})(T_{\mathrm{y}})$$

The *idealisation perturbation conjecture* states that of all the *integral iso-idealisations*, $B \in Y^{-1}_{U_A,\mathrm{i},T_{\mathrm{y}},\dagger,z}(A * T_{\mathrm{y}} * T_{\mathrm{y}}^{\dagger A})$, which have the given *alignment-bounded lifted iso-transform space*,

$$B * T_{\mathrm{y}} = A * T_{\mathrm{y}} \implies X'_{z,\mathrm{T},\mathrm{y}}(B)(T_{\mathrm{y}}) = X'_{z,\mathrm{T},\mathrm{y}}(A)(T_{\mathrm{y}})$$

the *integral sample idealisation*, $B = A * T_{\mathrm{y}} * T_{\mathrm{y}}^{\dagger A} = A$, has the least *alignment-bounded iso-transform space*. The *integral iso-idealisations* are a subset of the *integral-independent substrate histograms*, $Y^{-1}_{U_A,\mathrm{i},T_{\mathrm{y}},\dagger,z}(A * T_{\mathrm{y}} * T_{\mathrm{y}}^{\dagger A}) \subset \mathcal{A}_{z,\mathrm{xi}}$. According to the *idealisation perturbation conjecture*, this subset is such that

$$\forall B \in Y^{-1}_{U_A,\mathrm{i},T_{\mathrm{y}},\dagger,z}(A * T_{\mathrm{y}} * T_{\mathrm{y}}^{\dagger A})\ (X_{z,\mathrm{T},\mathrm{y}}(B)(T_{\mathrm{y}}) \geq X_{z,\mathrm{T},\mathrm{y}}(A)(T_{\mathrm{y}}))$$

But the *iso-independents* cannot have the same *maximum ideal transform*, $B \neq A \implies T_{\mathrm{y}} \notin \mathrm{maxd}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}(B))$. Let the *maximum ideal transform* of the *iso-independent* B be $T_{\mathrm{n}} \in \mathrm{maxd}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}(B)) \subset \mathcal{T}_{U_B,V_B} = \mathcal{T}_{U_A,V_A}$. The set of *transforms* of B which are such that the *alignment-bounded iso-transform space* is greater than the *alignment-bounded iso-transform maximum space* of A is

$$\{T : T \in \mathrm{dom}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}(B)),\ X_{z,\mathrm{T},\mathrm{y}}(B)(T) \geq X_{z,\mathrm{T},\mathrm{y}}(A)(T_{\mathrm{y}})\}$$

But in order for the *maximum transform* of B to have greater *alignment-bounded iso-transform space* than the *alignment-bounded iso-transform maximum space* of A, $X_{z,\mathrm{T},\mathrm{y}}(B)(T_{\mathrm{n}}) \geq X_{z,\mathrm{T},\mathrm{y}}(A)(T_{\mathrm{y}})$, it must be a member of a subset of these which has cardinality of one less,

$$\begin{aligned}&X_{z,\mathrm{T},\mathrm{y}}(B)(T_{\mathrm{n}}) \geq X_{z,\mathrm{T},\mathrm{y}}(A)(T_{\mathrm{y}}) \implies \\ &\quad T_{\mathrm{n}} \in \{T : T \in \mathrm{dom}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}(B)),\ X_{z,\mathrm{T},\mathrm{y}}(B)(T) \geq X_{z,\mathrm{T},\mathrm{y}}(A)(T_{\mathrm{y}})\} \setminus \{T_{\mathrm{y}}\}\end{aligned}$$

Thus the *maximum transform*, T_{n}, of the *iso-independent*, B, is weakly constrained by the *idealisation perturbation conjecture*. In the *ideal-agnostic* case, by contrast, the *maximum transform* for *iso-independents* subset of the *integral-independent substrate histograms* is always the *unary partition transform*, $T_{\mathrm{u}} \in \mathrm{maxd}(X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}(B))$, and so the *ideal-agnostic transform maximum space* correlation is sometimes lower.

Therefore, conjecture that the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform maximum function*, also known as the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}$, is transitively correlated with the *derived alignment integral-independent substrate ideal formal-abstract transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa},\mathrm{j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa},\mathrm{j}}) \geq 0)$$

Conjecture, further, that this correlation is greater than the *ideal-agnostic* correlation

$$\begin{aligned}&\forall z \in \mathbf{N}_{\geq t} \\ &\qquad (\mathrm{corr}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa},\mathrm{j}}) \\ &\quad \geq\ \mathrm{corr}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi},\mathrm{T},\mathrm{a},\mathrm{fa}}))\end{aligned}$$

above the variance threshold *size* t.

4.20 Computation of alignment

4.20.1 Alignmenter

Given a *histogram* $A \in \mathcal{A}$ consider the computation *time* to calculate its *alignment* $\mathrm{alignment}(A)$. Let $I_{\mathrm{a}} = \mathrm{alignmenter} \in \mathrm{computers}$, $\mathrm{domain}(I_{\mathrm{a}}) = \mathcal{A}$, $\mathrm{range}(I_{\mathrm{a}}) = \mathbf{Q}$, and $\mathrm{apply}(I_{\mathrm{a}})(A) \approx \mathrm{alignment}(A)$. The *alignmenter* relies on *independenter* I_{X} to calculate exactly the *independent histogram* A^{X}. The *alignmenter* also delegates the calculation of the logarithm of the gamma function. In order for the *alignmenter* to have finite *time*, $I_{\mathrm{a}}^{\mathrm{t}}(A) < \infty$, the real log gamma function must be approximated to some rational, for example by means of the Stirling's approximation or the Lanczos approximation. Let $I_{\approx\ln!} = \mathrm{logfactorialer} \in \mathrm{computers}$ be some implementation such that $\mathrm{domain}(I_{\approx\ln!}) = \mathbf{Q}_{\geq 0}$, $\mathrm{range}(I_{\approx\ln!}) = \mathbf{Q}$, and $\mathrm{apply}(I_{\approx\ln!})(x) \approx \ln \Gamma_! x$ and such that $\mathrm{apply}(I_{\approx\ln!})(0) = \mathrm{apply}(I_{\approx\ln!})(1) = 0$. Then

$$\begin{aligned} &I_{\mathrm{a}}^{\mathrm{t}}(A) > \\ &\quad I_{\mathrm{X}}^{\mathrm{t}}(A) + \sum_{S \in A^{\mathrm{FS}}} I_{\approx\ln!}^{\mathrm{t}}(A_S) + \sum_{S \in A^{\mathrm{XFS}}} I_{\approx\ln!}^{\mathrm{t}}(A_S^{\mathrm{X}}) + (|A^{\mathrm{F}}| + |A^{\mathrm{XF}}| + 1) I_{+}^{\mathrm{t}}(0,0) \end{aligned}$$

where $I_{+} = \mathrm{adder}$. Reducing the *independenter* to its underlying *adder* and *multiplier*

$$\begin{aligned} &I_{\mathrm{a}}^{\mathrm{t}}(A) > \\ &\quad (|A^{\mathrm{F}}|n + |A^{\mathrm{XF}}| + 1) I_{+}^{\mathrm{t}}(0,0) + |A^{\mathrm{XF}}| n I_{\times}^{\mathrm{t}}(1,1) + (|A^{\mathrm{F}}| + |A^{\mathrm{XF}}|) I_{\approx\ln!}^{\mathrm{t}}(1) \end{aligned}$$

where $V = \mathrm{vars}(A)$, $n = |V|$ and $I_{\times} = \mathrm{multiplier}$. If it is the case that the *log factorialer* has constant *time*, $\exists m \in \mathbf{N}_{>0}\ (I_{\approx\ln!}^{\mathrm{t}} \in \mathrm{O}(\mathbf{Q} \times \{1\}, m))$, and if the *histograms* are implemented in an *array histogram representation* on *ordered list state representations*

$$\exists m \in \mathbf{N}_{>0}\ (I_{\mathrm{a}}^{\mathrm{t}} \in \mathrm{O}(\{(A, ny) : A \in \mathcal{A},\ y = |A^{\mathrm{XF}}|,\ n = |\mathrm{vars}(A)\}|, m))$$

If A is a *regular histogram* in a *system* U of *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$ for which the *independent* is *completely effective*, $A^{\mathrm{XF}} = A^{\mathrm{C}}$, then the *alignmenter time* is of the same complexity as the *independenter time*, nd^n.

If the *independent* is *completely effective*, $A^{\mathrm{XF}} = A^{\mathrm{C}}$, then the *space* complexity of an *array histogram representation*, v, is less than the *space* complexity of a *binary map histogram representation*, $v \ln v$, where *volume* $v = |A^{\mathrm{C}}|$.

4.20.2 Single state roll computers

Define the set of *single state rolls* $\mathrm{rollStateSingles} \subset \mathrm{rolls}$ as the subset of *rolls* that are singletons, $\mathrm{rollStateSingles} = \{R : R \in \mathrm{rolls},\ |R| = 1\}$.

Consider the application of a *single state roll* $R \in \mathrm{rollStateSingles}$ to a *histogram* A and the computation *time* of the pair of the *rolled histogram* and *independent rolled histogram* $(A * R, (A * R)^{\mathrm{X}})$ given the pair of the *histogram* and *independent histogram* (A, A^{X}) prior to *rolling*. All four *histograms* are implemented in *array histogram representations* on *ordered list state representations*. Let $I = \mathrm{rollStateSingler}(U) \in \mathrm{computers}$ in *system* U such that $\mathrm{range}(I) = \{(A, A^{\mathrm{X}}) : A \in \mathcal{A}_U\} \subset \mathcal{A}_U \times \mathcal{A}_U$, and $\mathrm{domain}(I) = \{(R, (A, A^{\mathrm{X}})) : A \in \mathcal{A}_U,\ R \in A^{\mathrm{CS}} \to A^{\mathrm{CS}},\ |R| = 1\} \subset \mathrm{rollStateSingles} \times \mathrm{range}(I)$ such that $\mathrm{apply}(I)((R, (A, A^{\mathrm{X}}))) = (A*R, (A*R)^{\mathrm{X}})$. Let $\{(S, T)\} = R$. The *rolled* pair can be separated into mutable and immutable parts. Thus for the first of the pair

$$A * R = \{(S, 0), (T, A_S + A_T)\} + (A \setminus \{(S, A_S), (T, A_T)\})$$

To calculate the mutable parts of the *independent* only those *states* that are *incident* on either of the source or target *states* need be considered. Let $B = \bigcup\{\mathrm{incidence}(A, S, i) \cup \mathrm{incidence}(A, T, i) : i \in \{1 \ldots |V|\}\}$ and $B_{\mathrm{X}} = \bigcup\{\mathrm{incidence}(A^{\mathrm{X}}, S, i) \cup \mathrm{incidence}(A^{\mathrm{X}}, T, i) : i \in \{1 \ldots |V|\}\}$. Then separating into mutable and immutable

$$(A * R)^{\mathrm{X}} = (B_{\mathrm{X}}^{\mathrm{F}} \setminus (B * R)^{\mathrm{XF}})^{\mathrm{Z}} + (B * R)^{\mathrm{X}} + (A^{\mathrm{X}} \setminus B_{\mathrm{X}})$$

In the special case where A is a *regular histogram* of *dimension* $n = |V|$ and *valency* d, where $\{d\} = \{|U_v| : v \in V\}$, and where A^{X} is *completely effective*, $A^{\mathrm{XF}} = A^{\mathrm{C}}$, and where the source *state* S and target *state* T have no *degree of incidence*, $\{T\}^{\mathrm{U}} \in \mathrm{incidence}(V^{\mathrm{C}}, S, 0)$, then the subset B_{X} forms a *cartesian sub-volume* of cardinality $d^n - (d-2)^n$ where $d \geq 2$. Compare this to the *independenter* $I_{\mathrm{X}} \in \mathrm{computers}$ that calculates the *independent histogram*, $\mathrm{domain}(I_{\mathrm{X}}) = \mathrm{range}(I_{\mathrm{X}}) = \mathcal{A}$ and $\mathrm{apply}(I_{\mathrm{X}})(A) = A^{\mathrm{X}}$, also implemented in the *array histogram representation*. As conjectured above, I_{X} has *time* complexity of ny where $n = |\mathrm{vars}(A)|$ and y is the *effective independent cartesian sub-volume* $y = |A^{\mathrm{XF}}|$. If $A \in \mathrm{dom}(\mathcal{O}_{U,z})$ then A^{X} is *completely effective* and hence $y = |V^{\mathrm{C}}| = d^n$ in the case of the *regular histogram*. In this case the part of the computation *time* $I^{\mathrm{t}}((R, (A, A^{\mathrm{X}})))$ for the *independent rolled histogram* is a fraction $(d^n - (d-2)^n)/d^n = 1 - (1 - 2/d)^n$ of the *time* of the *independenter* $I_{\mathrm{X}}^{\mathrm{t}}(A * R)$. In other words, the *roll* only needs *time* $I_{\mathrm{X}}^{\mathrm{t}}(B * R)$ to calculate the *independent rolled histogram* as though A

was equal to its subset B which corresponds to the *cartesian sub-volume* B_{X} of the *independent* A^{X} *incident* on the source or target *states*. This subset requires *time* complexity $nd^n - n(d-2)^n$. Given that the calculation of the *rolled histogram* requires only one addition and one reset then

$$I^{\mathrm{t}}((R,(A,A^{\mathrm{X}}))) > I_0^{\mathrm{t}}(1) + I_+^{\mathrm{t}}(0,0) + |B_{\mathrm{X}}^{\mathrm{F}} \setminus (B*R)^{\mathrm{XF}}|I_0^{\mathrm{t}}(1) + I_{\mathrm{X}}^{\mathrm{t}}(B*R)$$

where $I_0 = \text{resetter} \in \text{computers}$ and $I_+ = \text{adder} \in \text{computers}$. Conjecture that the overall *time* complexity is $nd^n - n(d-2)^n$.

Note that the *roll computers* are here defined such that the operations are in-place mutations to the *array histogram representation*. That is, the *roll computers* conclude the computations with *list setter* $I_{\mathrm{L,s}}$ operations on the *array representation*. The implicit update to array has *time* complexity of the *ordered list indexer*, n, so the overall complexity is unchanged. Thus the domain of the *roll computers* is the *roll* crossed with the range. Contrast this to other *computers*, such as the *independenter*, which are defined here such that *time* of the implementation implies an instantiation of the array. *Roll computers* compute the *roll* by resetting and adding regardless of the *effectiveness* of the argument, whereas the *independenter* need only compute where *effective*. On the other hand, the *independenter* requires *time* for the cost of instantiation.

4.20.3 Value roll computers

Consider the application of a *value roll* $(V,v,s,t) \in \text{rollValues}(U)$ in *system* U to a *histogram* $A \in \mathcal{A}_U$ having *variables* $V = \text{vars}(A)$, and the computation *time* of the pair of (i) the *rolled histogram* $A * (V,v,s,t)^{\mathrm{R}}$, and the (ii) the *independent rolled histogram* $(A * (V,v,s,t)^{\mathrm{R}})^{\mathrm{X}}$, given the corresponding pair prior to *rolling*, (A,A^{X}), where all *histograms* are implemented in *array histogram representations* on *ordered list state representations*.

Let $I_{\mathrm{R}} = \text{rollValuer} \in \text{computers}$ such that $\text{range}(I_{\mathrm{R}}) \subset \mathcal{A} \times \mathcal{A}$ is defined $\text{range}(I_{\mathrm{R}}) = \{(A,A^{\mathrm{X}}) : A \in \mathcal{A}\}$ and $\text{domain}(I_{\mathrm{R}}) \subset \bigcup\{\text{rollValues}(U) : U \in \mathcal{U}\} \times \text{range}(I_{\mathrm{R}})$ is defined $\text{domain}(I_{\mathrm{R}}) = \{((V,v,s,t),(A,A^{\mathrm{X}}) : U \in \mathcal{U},\ A \in \mathcal{A}_U,\ (V,v,s,t) \in \text{rollValues}(U),\ V = \text{vars}(A)\}$, such that

$$\text{apply}(I_{\mathrm{R}})(((V,v,s,t),(A,A^{\mathrm{X}}))) = (A * (V,v,s,t)^{\mathrm{R}}, (A * (V,v,s,t)^{\mathrm{R}})^{\mathrm{X}})$$

The *independent rolled histogram* equals the *rolled independent histogram* $(A*(V,v,s,t)^{\mathrm{R}})^{\mathrm{X}} = A^{\mathrm{X}}*(V,v,s,t)^{\mathrm{R}}$, so only those *states* that *reduce* to either of the source or target *states*, $\{(v,s)\}$ or $\{(v,t)\}$, are changed by the *value* *roll*, $\text{states}(A * \{\{(v,s)\},\{(v,t)\}\}^{\text{U}})$ and $\text{states}(A^{\text{X}} * \{\{(v,s)\},\{(v,t)\}\}^{\text{U}})$. In the case where the source and target *values* differ, $s \neq t$, the mutable and immutable parts of the application of the *value roll* (V,v,s,t) to the *histogram* A can be separated out

$$\begin{aligned} A * (V,v,s,t)^{\text{R}} \quad = \quad & (A * \{\{(v,s)\}\}^{\text{U}})^{\text{Z}} + \\ & A * \{\{(v,t)\}\}^{\text{U}} + \\ & \{(S \setminus \{(v,s)\} \cup \{(v,t)\}, c) : (S,c) \in A * \{\{(v,s)\}\}^{\text{U}}\} + \\ & (A \setminus (A * \{\{(v,s)\},\{(v,t)\}\}^{\text{U}})) \end{aligned}$$

Similarly the mutable and immutable parts of the application of the *value roll* (V,v,s,t) to the *independent* A^{X} can be separated out where $s \neq t$

$$\begin{aligned} A^{\text{X}} * (V,v,s,t)^{\text{R}} \quad = \quad & (A^{\text{X}} * \{\{(v,s)\}\}^{\text{U}})^{\text{Z}} + \\ & A^{\text{X}} * \{\{(v,t)\}\}^{\text{U}} + \\ & \{(S \setminus \{(v,s)\} \cup \{(v,t)\}, c) : (S,c) \in A^{\text{X}} * \{\{(v,s)\}\}^{\text{U}}\} + \\ & (A^{\text{X}} \setminus (A^{\text{X}} * \{\{(v,s)\},\{(v,t)\}\}^{\text{U}})) \end{aligned}$$

The *time* is thus constrained to have a lower bound

$$\begin{aligned} I_{\text{R}}^{\text{t}}((((V,v,s,t),(A,A^{\text{X}}))) \quad > \quad & 4x(I_{\text{S,o}}^{\text{t}}(\cdot) + I_{\text{L,g}}^{\text{t}}(\cdot)) + 2xI_{+}^{\text{t}}(\cdot) + 4x(J_{\text{S,o}}^{\text{t}}(\cdot) + I_{\text{L,s}}^{\text{t}}(\cdot)) \\ > \quad & 4xI_{\text{L,g}}^{\text{t}}(\cdot) + 8nxI_{\times}^{\text{t}}(\cdot) + 2x(4n+1)I_{+}^{\text{t}}(\cdot) + 4xI_{\text{L,s}}^{\text{t}}(\cdot) \end{aligned}$$

where $x = |V^{\text{C}}|/|U_v| = |V^{\text{C}} * \{\{(v,s)\}\}^{\text{U}}| = |V^{\text{C}} * \{\{(v,t)\}\}^{\text{U}}|$ is the cardinality of the *volumes incident* on $\{(v,s)\}$ and on $\{(v,t)\}$, and $I_{\text{L,g}}$ and $I_{\text{L,s}}$ are the *list getter* and *setter*, and $I_{\text{S,o}}$ and $J_{\text{S,o}}$ are the *state ordered indexer* and its *inverse*.

In the special case where A is a *regular histogram* of *dimension* $n = |V|$ and *valency* d, where $\{d\} = \{|U_w| : w \in V\}$, then the *incident sub-volume* has cardinality $x = |\text{incidence}(V^{\text{C}}, \{(v,t)\}, 1)| = d^{n-1}$, so the complexity is nd^{n-1}. This complexity of *time* is less than or equal to that for the *single state roller* and for the *independenter*, $nd^{n-1} \leq n(d^n - (d-2)^n) \leq nd^n$.

Consider the application of a *value roll* $(V,v,s,t) \in \text{rollValues}(U)$ in *system* U to a *histogram* $A \in \mathcal{A}_U$, having *variables* $V = \text{vars}(A)$, and the computation *time* of a triple of (i) an approximation to the *rolled alignment* $\text{algn}(A * (V,v,s,t)^{\text{R}})$, (ii) the *rolled histogram* $A * (V,v,s,t)^{\text{R}}$, and (iii) the *independent rolled histogram* $(A * (V,v,s,t)^{\text{R}})^{\text{X}}$, given (a) the reductions by *variable* of the difference in log factorial approximations between A and A^{X}, and (b) the triple prior to *rolling*, $(\text{algn}(A), A, A^{\text{X}})$.

Let $\text{rals} \in \mathcal{A} \to (\mathcal{V} \to (\mathcal{S} \to \mathrm{Q}))$ be defined as

$$\begin{aligned}&\text{rals}(A) := \\ &\quad \{(w, \{(S, \sum(I^*_{\approx\text{ln!}}(A_T) : T \in A^{\mathrm{S}},\ T \supseteq S) - \sum(I^*_{\approx\text{ln!}}(A^{\mathrm{X}}_T) : T \in A^{\mathrm{XS}},\ T \supseteq S)) : \\ &\quad S \in (A\%\{w\})^{\mathrm{S}}\}) : w \in V\}\end{aligned}$$

where $I_{\approx\text{ln!}} = \text{logfactorialer} \in \text{computers}$ and $I^* := \text{apply}(I)$.

Let $I_{\mathrm{R,a}} = \text{rollValueAlignmenter} \in \text{computers}$ be such that $\text{range}(I_{\mathrm{R,a}}) \subset \mathbf{Q} \times \mathcal{A} \times \mathcal{A}$ defined as

$$\text{range}(I_{\mathrm{R,a}}) = \{(I^*_{\mathrm{a}}(A), A, A^{\mathrm{X}}) : A \in \mathcal{A}\}$$

and such that $\text{domain}(I_{\mathrm{R,a}}) \subset \bigcup\{\text{rollValues}(U) : U \in \mathcal{U}\} \times (\mathcal{V} \to (\mathcal{S} \to \mathrm{Q})) \times \text{range}(I_{\mathrm{R,a}})$ is defined as

$$\begin{aligned}\text{domain}(I_{\mathrm{R,a}}) &= \{((V,v,s,t), \text{rals}(A), (a, A, A^{\mathrm{X}})) : (a, A, A^{\mathrm{X}}) \in \text{range}(I_{\mathrm{R,a}}), \\ &\qquad U \in \mathcal{U},\ (V,v,s,t) \in \text{rollValues}(U),\ V = \text{vars}(A)\}\end{aligned}$$

and such that

$$\begin{aligned}&\text{apply}(I_{\mathrm{R,a}})(((V,v,s,t), Y, (a, A, A^{\mathrm{X}}))) = \\ &\quad (I^*_{\mathrm{a}}(A * (V,v,s,t)^{\mathrm{R}}), A * (V,v,s,t)^{\mathrm{R}}, (A * (V,v,s,t)^{\mathrm{R}})^{\mathrm{X}})\end{aligned}$$

where $Y = \text{rals}(A)$.

The computation of both the *rolled histogram* $A * (V,v,s,t)^{\mathrm{R}}$ and the *independet rolled histogram* $(A * (V,v,s,t)^{\mathrm{R}})^{\mathrm{X}}$ is the same in the *value rolled alignmenter* $I_{\mathrm{R,a}}$ as it is in the *value roller* I_{R}. That is,

$$\begin{aligned}&A * (V,v,s,t)^{\mathrm{R}} = \\ &\quad (A * \{Q_s\}^{\mathrm{U}})^{\mathrm{Z}} + A * \{Q_t\}^{\mathrm{U}} + \{(S \setminus Q_s \cup Q_t, c) : (S,c) \in A * \{Q_s\}^{\mathrm{U}}\} + \\ &\quad (A \setminus (A * \{Q_s, Q_t\}^{\mathrm{U}}))\end{aligned}$$

where $s \neq t$, and $Q = \{(u, \{(v,u)\}) : u \in U_v\} \in \mathcal{W}_U \to \mathcal{S}_U$. Similarly

$$\begin{aligned}&(A * (V,v,s,t)^{\mathrm{R}})^{\mathrm{X}} = \\ &\quad (A^{\mathrm{X}} * \{Q_s\}^{\mathrm{U}})^{\mathrm{Z}} + A^{\mathrm{X}} * \{Q_t\}^{\mathrm{U}} + \{(S \setminus Q_s \cup Q_t, c) : (S,c) \in A^{\mathrm{X}} * \{Q_s\}^{\mathrm{U}}\} + \\ &\quad (A^{\mathrm{X}} \setminus (A^{\mathrm{X}} * \{Q_s, Q_t\}^{\mathrm{U}}))\end{aligned}$$

Then, given the approximate *alignment* prior to *rolling*, $a = I^*_{\mathrm{a}}(A) \in \mathbf{Q}$, the approximate *rolled alignment* can be calculated

$$\begin{aligned} I^*_{\mathrm{a}}(A * (V,v,s,t)^{\mathrm{R}}) &= a - Y_v(Q_s) - Y_v(Q_t) + \\ &\quad \sum (I^*_{\approx\ln!}(c) : (S,c) \in A * (V,v,s,t)^{\mathrm{R}} * \{Q_t\}^{\mathrm{U}}) - \\ &\quad \sum (I^*_{\approx\ln!}(c) : (S,c) \in (A * (V,v,s,t)^{\mathrm{R}})^{\mathrm{X}} * \{Q_t\}^{\mathrm{U}}) \end{aligned}$$

The *time* is thus constrained to have a lower bound

$$\begin{aligned} I^{\mathrm{t}}_{\mathrm{R,a}}(((V,v,s,t),Y,(a,A,A^{\mathrm{X}}))) &> I^{\mathrm{t}}_{\mathrm{R}}(((V,v,s,t),(A,A^{\mathrm{X}}))) + \\ &\quad (2x+1)I^{\mathrm{t}}_{+}(\cdot) + 2xI^{\mathrm{t}}_{\approx\ln!}(\cdot) \end{aligned}$$

where $x = |V^{\mathrm{C}}|/|U_v| = |V^{\mathrm{C}} * \{Q_t\}^{\mathrm{U}}|$ is the cardinality of the *volumes incident* on Q_s and on Q_t.

Therefore the *value rolled alignmenter* has *time* of the same complexity as the *value roller*, which is nd^{n-1} in the case of the *regular histogram* of *dimension* n and *valency* d.

4.21 Tractable alignment-bounding

Let the set of *inducers* be the subset of *computers* parameterised by *integral size*, $\mathrm{inducers}(z) \subset \mathrm{computers}$, such that (i) the domain is a subset of the *substrate histograms* and a superset of the *integral-independent substrate histograms*, $\forall I_z \in \mathrm{inducers}(z)\ (\mathcal{A}_{z,\mathrm{xi}} \subseteq \mathrm{domain}(I_z) \subseteq \mathcal{A}_z)$, (ii) the range is a non-empty rational-valued function, $\forall I_z \in \mathrm{inducers}(z)\ (I^*_z \in \mathrm{domain}(I_z) \to ((\mathcal{X} \to \mathbf{Q}) \setminus \{\emptyset\}))$, such that application to a domain *substrate histogram*, $A \in \mathrm{domain}(I_z)$, returns a rational-valued function of the *substrate models set*, $I^*_z(A) \in \mathcal{M}_{U_A,V_A} \to \mathbf{Q}$, (iii) both the *inducer time* and *space* are finite, $I^{\mathrm{t}}_z(A) < \infty$ and $I^{\mathrm{s}}_z(A) < \infty$, and (iv) the maximum of the *inducer* application, $\mathrm{maxr} \circ I^*_z$, is positively correlated with the finite *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ \forall I_z \in \mathrm{inducers}(z)\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I^*_z) \geq 0)$$

The correlation, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I^*_z)$, is called the *inducer correlation* for *size* z. The *substrate models* of the *substrate histogram*, $\mathcal{M}_{U_A,V_A}$, each map to a *substrate transform*, $\mathrm{transform}(U,V) \in \mathcal{M}_{U,V} \to \mathcal{T}_{U,V}$, so that the maximum *induced transforms* of *inducer* I_z applied to *substrate histogram* A are $\{\mathrm{transform}(U_A,V_A)(M) : M \in \mathrm{maxd}(I^*_z(A))\} \subset \mathcal{T}_{U_A,V_A}$.

Given an *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$, let the domain of the *alignment-bounded iso-transform space ideal transform search set*, $\mathrm{dom}(X_{z,\mathrm{xi,T,y,fa,j}}(A)) \subset \mathcal{T}_{U_A,V_A}$, which consists of *substrate transforms* subject to the constraints of (i) *formal-abstract equality*, and (ii) *ideality*, be abbreviated to the *literal substrate transforms*

$$\begin{aligned}\mathcal{T}_{\mathrm{fa,j}}(A) &= \mathrm{dom}(X_{z,\mathrm{xi,T,y,fa,j}}(A)) \\ &= \{T : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}\end{aligned}$$

The domain of the *derived alignment integral-independent substrate ideal formal-abstract transform search set*, $\mathrm{dom}(X'_{z,\mathrm{xi,T,a,fa,j}}(A))$, equals the domain of the *alignment-bounded iso-transform space ideal transform search set*, $\mathrm{dom}(X'_{z,\mathrm{xi,T,a,fa,j}}(A)) = \mathrm{dom}(X_{z,\mathrm{xi,T,y,fa,j}}(A))$, and may also be abbreviated, $\mathcal{T}_{\mathrm{fa,j}}(A) = \mathrm{dom}(X'_{z,\mathrm{xi,T,a,fa,j}}(A))$.

Let the *literal substrate decompositions* $\mathcal{D}_{\mathrm{fa,j}}(A) \subset \mathcal{D}_{U_A,V_A}$, which are *substrate decompositions* subject to the constraints of (i) *formal-abstract equality*, and (ii) *ideality*, be defined

$$\mathcal{D}_{\mathrm{fa,j}}(A) = \{D : D \in \mathcal{D}_{U_A,V_A},\ A^{\mathrm{X}} * D^{\mathrm{T}} = (A * D^{\mathrm{T}})^{\mathrm{X}},\ A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}\}$$

The *literal substrate decompositions* map to the *literal substrate transforms*, $\{D^{\mathrm{T}V_A} : D \in \mathcal{D}_{\mathrm{fa,j}}(A)\} = \mathcal{T}_{\mathrm{fa,j}}(A)$. The cardinality of the *literal substrate decompositions* is greater than or equal to the *literal substrate transforms*, $|\mathcal{D}_{\mathrm{fa,j}}(A)| \geq |\mathcal{T}_{\mathrm{fa,j}}(A)|$, because the *literal substrate decompositions* includes those *decompositions* which consist solely of a *literal substrate transform*, $\{\{((\emptyset, T), \emptyset)\} : T \in \mathcal{T}_{\mathrm{fa,j}}(A)\} \subset \mathcal{D}_{\mathrm{fa,j}}(A)$.

Note that if a *substrate decomposition* $D \in \mathcal{D}_{U_A,V_A}$ contains more than a *root transform*, $|\mathrm{transforms}(D)| > 1$, then the *expanded nullable transform*, $D^{\mathrm{T}V_A}$, is not necessarily in the *literal substrate transforms*, $\mathcal{T}_{\mathrm{fa,j}}(A)$. This is because the *nullable transform* of a multiple *decomposition* is *overlapping*, $|\mathrm{transforms}(D)| > 1 \implies \mathrm{overlap}(D^{\mathrm{T}})$, even if it so happens that the *transforms* of the *decomposition* are each individually *non-overlapping*, $\forall T \in \mathrm{transforms}(D)\ (\neg\mathrm{overlap}(T))$. A *non-overlapping transform* implies that the *formal histogram* is *independent*, $\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, so it is sometimes the case that an *overlapping transform*, $\mathrm{overlap}(T)$, has *non-independent formal histogram*, $A^{\mathrm{X}} * T \neq (A^{\mathrm{X}} * T)^{\mathrm{X}}$. If it is indeed the case that an *overlapping nullable transform*, $\mathrm{overlap}(D^{\mathrm{T}})$, has *non-independent formal histogram*, $A^{\mathrm{X}} * D^{\mathrm{T}} \neq (A^{\mathrm{X}} * D^{\mathrm{T}})^{\mathrm{X}}$, then the *formal histogram* cannot be equal to the *abstract histogram*, $A^{\mathrm{X}} * D^{\mathrm{T}} \neq (A * D^{\mathrm{T}})^{\mathrm{X}}$. So, in this case of *non-independent formal histogram*, the *substrate decomposition*, D, is not a *literal substrate decomposition*, $D \notin \mathcal{D}_{\mathrm{fa,j}}(A)$. The multiple *literal substrate decompositions*, $\{D : D \in \mathcal{D}_{\mathrm{fa,j}}(A),\ \mathrm{transforms}(D) > 1\} \subset \mathcal{D}_{\mathrm{fa,j}}(A)$, all have *overlapping nullable transform*, but nonetheless all have *independent formal histogram*.

Similarly, let the *literal substrate fuds* $\mathcal{F}_{\mathrm{fa,j}}(A) \subset \mathcal{F}_{U_A,V_A}$, which are *substrate fuds* subject to the constraints of (i) *formal-abstract equality*, and (ii) *ideality*, be defined

$$\mathcal{F}_{\mathrm{fa,j}}(A) = \{F : F \in \mathcal{F}_{U_A,V_A},\ A^{\mathrm{X}} * F^{\mathrm{T}} = (A * F^{\mathrm{T}})^{\mathrm{X}},\ A = A * F^{\mathrm{T}} * F^{\mathrm{T}\dagger A}\}$$

The *literal substrate fuds* map to the *literal substrate transforms*, $\{F^{\mathrm{T}V_A} : F \in \mathcal{F}_{\mathrm{fa,j}}(A)\} = \mathcal{T}_{\mathrm{fa,j}}(A)$. The cardinality of the *literal substrate fuds* is greater than or equal to the *literal substrate transforms*, $|\mathcal{F}_{\mathrm{fa,j}}(A)| \geq |\mathcal{T}_{\mathrm{fa,j}}(A)|$, because the *literal substrate fuds* includes those *fuds* which consist solely of a *literal substrate transform*, $\{\{T\} : T \in \mathcal{T}_{\mathrm{fa,j}}(A)\} \subset \mathcal{F}_{\mathrm{fa,j}}(A)$.

Finally, let the *literal substrate fud decompositions* $\mathcal{D}_{\mathrm{F,fa,j}}(A) \subset \mathcal{D}_{\mathrm{F},U_A,V_A}$, which are *substrate fud decompositions* subject to the constraints of (i) *formal-abstract equality*, and (ii) *ideality*, be defined

$$\mathcal{D}_{\mathrm{F,fa,j}}(A) = \{D : D \in \mathcal{D}_{\mathrm{F},U_A,V_A},\ A^{\mathrm{X}} * D^{\mathrm{T}} = (A * D^{\mathrm{T}})^{\mathrm{X}},\ A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}\}$$

The *substrate fud decompositions*, $\mathcal{D}_{\mathrm{F},U_A,V_A}$, is finite because the *substrate fuds* are constrained to appear no more than once in any path, $\forall L \in \mathrm{paths}(D)\ (\{(i,F) : (i,(\cdot,F)) \in L\} \in \mathbf{N} \leftrightarrow \mathcal{F}_{U_A,V_A})$. The *literal substrate fud decompositions* map to the *literal substrate transforms*, $\{D^{\mathrm{T}V_A} : D \in \mathcal{D}_{\mathrm{F,fa,j}}(A)\} = \mathcal{T}_{\mathrm{fa,j}}(A)$. The cardinality of the *literal substrate fud decompositions* is greater than or equal to the *literal substrate fuds*, $|\mathcal{D}_{\mathrm{F,fa,j}}(A)| \geq |\mathcal{F}_{\mathrm{fa,j}}(A)|$, because the *literal substrate fud decompositions* includes those *decompositions* which consist solely of a *literal substrate fud*, $\{\{((\emptyset,F),\emptyset)\} : F \in \mathcal{F}_{\mathrm{fa,j}}(A)\} \subset \mathcal{D}_{\mathrm{F,fa,j}}(A)$. The cardinality of the *literal substrate fud decompositions* is therefore also greater than or equal to the *literal substrate transforms*, $|\mathcal{D}_{\mathrm{F,fa,j}}(A)| \geq |\mathcal{T}_{\mathrm{fa,j}}(A)|$.

The set of *transforms* of the *substrate models* searched by the *inducer* for a given *substrate histogram* A, $\{\mathrm{transform}(U_A,V_A)(M) : M \in \mathrm{dom}(I_z^*(A))\} \subset \mathcal{T}_{U_A,V_A}$, need not intersect with the *literal substrate transforms*, $\mathcal{T}_{\mathrm{fa,j}}(A)$. So in some cases $\{\mathrm{transform}(U_A,V_A)(M) : M \in \mathrm{dom}(I_z^*(A))\} \cap \mathcal{T}_{\mathrm{fa,j}}(A) = \emptyset$. All that is required of *inducers* is that there is a positive correlation between the *maximum functions*, not that the domains of the *search sets* intersect. This allows definitions of *inducers* that search *substrate models* which are *overlapping*, *formal-abstract-inequivalent* or *non-ideal*.

In addition, the definition of an *inducer* may be such that its domain is a proper superset of the *integral-independent substrate histograms*, $\mathcal{A}_{z,\mathrm{xi}} \subset \mathrm{domain}(I_z)$, and thus is a proper superset of the domain of the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{dom}(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}) = \mathrm{dom}(X_{z,\mathrm{xi,T,y,fa,j}}) = \mathcal{A}_{z,\mathrm{xi}} \subset \mathrm{domain}(I_z)$. The correlation, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I_z^*)$, is restricted to the intersection of the domains of the argument functions, which is $\mathcal{A}_{z,\mathrm{xi}}$. Thus no constraint is made by the correlation on the definition of the *inducer* for the part of its domain which is the disjoint set of *non-integral-independent substrate histograms*, $\mathcal{A}_z \setminus \mathcal{A}_{z,\mathrm{xi}}$. For example, the definition of an *inducer* could be extended into *non-integral-independent substrate histograms* by the use of approximations to the *multinomial probability density function*, $\mathrm{mpdf}(U)(E, z) \in \mathcal{A}_{U,V,z} \to \mathbf{R}_{\geq 0}$, to interpolate, via the unit-translated gamma function, $\Gamma_!$, from the *generalised multinomial probability distribution* $\hat{Q}_{\mathrm{m},U}(E, z) \in \mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}$, but the correlation would not imply this interpolation.

The *tractable inducers* subset of *inducers* is such that neither the computational *time* complexity nor the representational *encoding space* complexity of the application, I_z^*, is greater than polynomial. The complexities are always with respect to some underlying variable, for example *valency* or *dimension*. So the application of an *intractable inducer* may still be practicable. That is, its parameters may be such that the values of the underlying variables are small enough that the computation *time* and representation *space* are practicable. Conversely, a *tractable inducer* is not necessarily practicable. Practicability implies absolute limits on the *time* and *space* of the computation of the *inducer*, whereas *tractable inducers* may exceed these limits so long as the complexities are not exponential or higher.

Let the *log-rational approxer* $I_{\approx \ln \mathbf{Q}} \in \mathrm{computers}$ be a *computer* that finitely approximates, to some degree of accuracy, between the log-positive-rational numbers, $\ln \mathbf{Q}_{>0}$, and the rational numbers, $\mathbf{Q}$. The domain is $\mathrm{domain}(I_{\approx \ln \mathbf{Q}}) = \ln \mathbf{Q}_{>0}$. The range is $\mathrm{range}(I_{\approx \ln \mathbf{Q}}) = \mathbf{Q}$. The application is such that, $\forall x \in \ln \mathbf{Q}_{>0}\ (I_{\approx \ln \mathbf{Q}}^*(x) \approx x)$.

Let the *real approxer* $I_{\approx \mathbf{R}} \in \mathrm{computers}$ be a *computer* that finitely approximates, to some degree of accuracy, between the real numbers, $\mathbf{R}$, and the rational numbers, $\mathbf{Q}$. The domain is $\mathrm{domain}(I_{\approx \mathbf{R}}) = \mathbf{R}$. The range is $\mathrm{range}(I_{\approx \mathbf{R}}) = \mathbf{Q}$. The application is such that, $\forall x \in \mathbf{R}\ (I_{\approx \mathbf{R}}^*(x) \approx x)$. The application of the *real approxer*, $I_{\approx \mathbf{R}}$, is a superset of the application of the *log-rational approxer*, $I_{\approx \ln \mathbf{Q}}$. That is, $I_{\approx \ln \mathbf{Q}}^* \subset I_{\approx \mathbf{R}}^*$.

4.21.1 Literal inducers

Let the *literal alignment-bounded iso-transform space ideal transform inducer* $I_{z,\mathrm{y,l}} \in \mathrm{inducers}(z)$ be a literal implementation of the *alignment-bounded iso-transform space ideal transform search set*, $X_{z,\mathrm{xi,T,y,fa,j}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, except that the finite approximation of the *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$, is made between the log-positive-rational, $\ln \mathbf{Q}_{>0}$, valued range of the *search set* to the rational, $\mathbf{Q}$, valued range of the *inducer* application, $I^{*}_{z,\mathrm{y,l}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \mathbf{Q})$. That is,

$$\mathrm{domain}(I_{z,\mathrm{y,l}}) = \mathrm{dom}(X_{z,\mathrm{xi,T,y,fa,j}}) = \mathcal{A}_{z,\mathrm{xi}}$$

and

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I^{*}_{z,\mathrm{y,l}}(A) = \{(T, I^{*}_{\approx\ln\mathbf{Q}}(y)) : (T,y) \in X_{z,\mathrm{xi,T,y,fa,j}}(A)\})$$

The domain of the application of the *literal alignment-bounded inducer*, $I_{z,\mathrm{y,l}}$, to *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ is the *literal substrate transforms*, $\mathrm{dom}(I^{*}_{z,\mathrm{y,l}}(A)) = \mathrm{dom}(X_{z,\mathrm{xi,T,y,fa,j}}(A)) = \mathcal{T}_{\mathrm{fa,j}}(A)$.

The correlation between its *maximum function* and the *alignment-bounded iso-transform space ideal transform maximum function* is almost, but not exactly, one, because of the approximation made by the *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\ &\quad(\mathrm{var}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}) > 0 \wedge \mathrm{var}(z)(\mathrm{maxr} \circ I^{*}_{z,\mathrm{y,l}}) > 0 \implies\\ &\qquad\qquad \mathrm{corr}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I^{*}_{z,\mathrm{y,l}}) \approx 1)\end{aligned}$$

Similarly, let the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer* $I'_{z,\mathrm{a,l}} \in \mathrm{inducers}(z)$ be a literal implementation of the *derived alignment integral-independent substrate ideal formal-abstract transform search set*, $X'_{z,\mathrm{xi,T,a,fa,j}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$, except that the finite approximation of the *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$, is made between the log-positive-rational, $\ln \mathbf{Q}_{>0}$, valued range of the *search set* to the rational, $\mathbf{Q}$, valued range of the *inducer* application, $I'^{*}_{z,\mathrm{a,l}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \mathbf{Q})$. That is,

$$\mathrm{domain}(I'_{z,\mathrm{a,l}}) = \mathrm{dom}(X'_{z,\mathrm{xi,T,a,fa,j}}) = \mathcal{A}_{z,\mathrm{xi}}$$

and

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{*}_{z,\mathrm{a,l}}(A) = \{(T, I^{*}_{\approx\ln\mathbf{Q}}(a)) : (T,a) \in X'_{z,\mathrm{xi,T,a,fa,j}}(A)\})$$

The domain of the application of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, to *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ is the *literal substrate transforms*, $\mathrm{dom}(I'^{*}_{z,\mathrm{a,l}}(A)) = \mathrm{dom}(X'_{z,\mathrm{xi,T,a,fa,j}}(A)) = \mathcal{T}_{\mathrm{fa,j}}(A)$. The application can therefore be expressed

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{*}_{z,\mathrm{a,l}}(A) = \{(T, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algn}(A * T))) : T \in \mathcal{T}_{\mathrm{fa,j}}(A)\})$$

In order to allow the comparison of *space* and *time* complexities between them, the two *literal inducers*, $I_{z,\mathrm{y,l}}$ and $I'_{z,\mathrm{a,l}}$, are defined with (i) the same degree of accuracy of the log-positive-rational approximation of the *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$, and (ii) the same representations of *histograms* and *transforms.*

It is conjectured, in section 'Substrate structures alignment' above, that the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$, is correlated with the *derived alignment integral-independent substrate ideal formal-abstract transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}) \geq 0)$$

Hence conjecture that the *maximum transform* function of the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I'_{z,\mathrm{a,l}}$, is positively correlated with that of the *alignment-bounded iso-transform space ideal transform search set*, $X_{z,\mathrm{xi,T,y,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}) \geq 0)$$

but because of the introduction of log-positive-rational approximations, the correlation is lower

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\ &\quad (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}) \geq\\ &\qquad\qquad \mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}))\end{aligned}$$

Although it is conjectured that the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$, is correlated with the *derived alignment integral-independent substrate ideal formal-abstract transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}$, it is not necessarily the case that the sets of *maximum transforms* intersect for any given *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$. That is, it is sometimes the case that

$$\mathrm{maxd}(X_{z,\mathrm{xi,T,y,fa,j}}(A)) \cap \mathrm{maxd}(X'_{z,\mathrm{xi,T,a,fa,j}}(A)) = \emptyset$$

and so the correlation is not perfect,

$$\exists z \in \mathbf{N}_{>0}\ (\mathrm{corr}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}) < 1)$$

This is also true for the correlation between the *literal inducers*,

$$\exists z \in \mathbf{N}_{>0}\ (\mathrm{corr}(z)(\mathrm{maxr} \circ I^*_{z,\mathrm{y,l}}, \mathrm{maxr} \circ I'^*_{z,\mathrm{a,l}}) < 1)$$

The compromise of the less than perfect correlation may be required by a practicable computation, however, because the computation *time* of the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A)$, is less than that of the *literal alignment-bounded iso-transform space ideal transform inducer*, $I^{\mathrm{t}}_{z,\mathrm{y,l}}(A)$.

For given *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ the domains of the *space* functions are the *literal substrate transforms*, $\mathrm{dom}(X'_{z,\mathrm{xi,T,a,fa,j}}(A)) = \mathrm{dom}(X_{z,\mathrm{xi,T,y,fa,j}}(A)) = \mathcal{T}_{\mathrm{fa,j}}(A)$. Consider the difference in computation of *literal substrate transform* $T \in \mathcal{T}_{\mathrm{fa,j}}(A)$ in a literal implementation of expression

$$\begin{aligned}
X_{z,\mathrm{xi,T,y,fa,j}}(A)(T) &= X_{z,\mathrm{T,y}}(A)(T) \\
&= -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\sum_{B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)}
\end{aligned}$$

and of expression

$$\begin{aligned}
X'_{z,\mathrm{xi,T,a,fa,j}}(A)(T) &= X'_{z,\mathrm{T,a}}(A)(T) \\
&= \mathrm{algn}(A * T) \\
&= -\ln \frac{\hat{Q}_{\mathrm{m},U_A}((A * T)^{\mathrm{X}}, z)(A * T)}{\hat{Q}_{\mathrm{m},U_A}((A * T)^{\mathrm{X}}, z)((A * T)^{\mathrm{X}})} \\
&= -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(A * T)}{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(A^{\mathrm{X}} * T)}
\end{aligned}$$

Define the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space computer* $I_{U,V,z,\mathrm{T,y}} \in \mathrm{computers}$ such that $\mathrm{domain}(I_{U,V,z,\mathrm{T,y}}) = \mathcal{A}_{U,\mathrm{i},V,z} \times \mathcal{T}_{U,V}$ and

$$I^*_{U,V,z,\mathrm{T,y}}((A,T)) \approx -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},y,T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)}$$

The *independent-sample-distributed iso-transform-independent conditional dependent multinomial space computer*, $I_{U,V,z,\mathrm{T,y}}$, is, in turn, defined in terms of the *multinomial space computer* $I_{\mathrm{m}} \in$ computers. The *multinomial space computer* is such that $I_{\mathrm{m}}^* \in \{(E,A) : E \in \mathcal{A},\ A \in \mathcal{A}_{\mathrm{i}},\ E^{\mathrm{F}} \geq A^{\mathrm{F}},\ z_E > 0\} \to \mathbf{Q}$,

$$I_{\mathrm{m}}^*((E,A)) \approx -\ln \hat{Q}_{\mathrm{m},U}(E,z_A)(A) = -\ln \frac{z_A!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S}$$

It is in the *multinomial space computer*, I_{m}, that the approximation between log-positive-rational, $\ln \mathbf{Q}_{>0}$, and rational, $\mathbf{Q}$, is made by means of the *log-rational approxer*, $I_{\approx \ln \mathbf{Q}}$. Note that there is no need yet for an implementation of the unit-translated gamma function, $\Gamma_!$, such as in the *log factorialer*, $I_{\approx \ln !}$, because the factorial computations are integral.

The *independent-sample-distributed iso-transform-independent conditional dependent multinomial space computer*, $I_{U,V,z,\mathrm{T},\mathrm{y}}$, also requires an *independenter*, $I_{\mathrm{X}}^*(A) = A^{\mathrm{X}}$, where $I_{\mathrm{X}} =$ independenter. Thus the computation *time* of the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space computer*, $I_{U,V,z,\mathrm{T},\mathrm{y}}^{\mathrm{t}}((A,T))$, is such that

$$I_{U,V,z,\mathrm{T},\mathrm{y}}^{\mathrm{t}}((A,T)) > I_{\mathrm{X}}^{\mathrm{t}}(A) + I_{\mathrm{m}}^{\mathrm{t}}((A^{\mathrm{X}},A)) + \sum I_{\mathrm{m}}^{\mathrm{t}}((A^{\mathrm{X}},B)) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$$

Let the *literal alignment-bounded iso-transform space ideal transform inducer*, $I_{z,\mathrm{y},\mathrm{l}}$, be implemented in terms of an *independent-sample-distributed iso-transform-independent conditional dependent multinomial space computer*, $I_{U,V,z,\mathrm{T},\mathrm{y}}$, so that for *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ and *literal substrate transform* $T \in \mathcal{T}_{\mathrm{fa},\mathrm{j}}(A)$

$$I_{z,\mathrm{y},\mathrm{l}}^*(A)(T) = I_{U_A,V_A,z,\mathrm{T},\mathrm{y}}^*((A,T)) \approx X_{z,\mathrm{xi},\mathrm{T},\mathrm{y},\mathrm{fa},\mathrm{j}}(A)(T)$$

The computation *time* of *histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ of the *alignment-bounded inducer* is therefore

$$I_{z,\mathrm{y},\mathrm{l}}^{\mathrm{t}}(A) > \sum I_{U_A,V_A,z,\mathrm{T},\mathrm{y}}^{\mathrm{t}}((A,T)) : T \in \mathcal{T}_{\mathrm{fa},\mathrm{j}}(A)$$

Define the *independent-sample-distributed relative dependent multinomial space computer* $I_{U,V,z,\mathrm{a}} \in$ computers such that $\mathrm{domain}(I_{U,V,z,\mathrm{a}}) = \{A : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\}$ and

$$I_{U,V,z,\mathrm{a}}^*(A) \approx -\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A)}{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}},z)(A^{\mathrm{X}})}$$

The *independent-sample-distributed relative dependent multinomial space computer*, $I_{U,V,z,\mathrm{a}}$, is also defined in terms of an *independenter*, I_{X}, and a *multinomial space computer*, I_{m}. Thus the computation *time* of the *independent-sample-distributed relative dependent multinomial space computer*, $I_{U,V,z,\mathrm{a}}^{\mathrm{t}}(A)$, is such that

$$I^{\mathrm{t}}_{U,V,z,\mathrm{a}}(A) > I^{\mathrm{t}}_{\mathrm{X}}(A) + I^{\mathrm{t}}_{\mathrm{m}}((A^{\mathrm{X}}, A)) + I^{\mathrm{t}}_{\mathrm{m}}((A^{\mathrm{X}}, A^{\mathrm{X}}))$$

Consider a variation of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, which is the *relative literal derived alignment integral-independent substrate ideal formal-abstract transform inducer* $I'_{z,\mathrm{a,l,r}} \in \mathrm{inducers}(z)$. The *relative literal derived alignment inducer*, $I'_{z,\mathrm{a,l,r}}$, is implemented with a *transformer*, $I_{*\mathrm{T}} = \mathrm{transformer}$, followed by an *independent-sample-distributed relative dependent multinomial space computer*, $I_{U,V,z,\mathrm{a}}$, so that for *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ and *literal substrate transform* $T \in \mathcal{T}_{\mathrm{fa,j}}(A)$

$$I'^{*}_{z,\mathrm{a,l,r}}(A)(T) = I^{*}_{U_A,W,z,\mathrm{a}}(I^{*}_{*\mathrm{T}}((T, A))) \approx X'_{z,\mathrm{xi,T,a,fa,j}}(A)(T)$$

where $W = \mathrm{der}(T)$. In this case the computation *time* of *histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ of the *relative literal derived alignment inducer* is therefore

$$I'^{\mathrm{t}}_{z,\mathrm{a,l,r}}(A) > \sum I^{\mathrm{t}}_{*\mathrm{T}}((T, A)) + I^{\mathrm{t}}_{U_A,W,z,\mathrm{a}}(A * T) : T \in \mathcal{T}_{\mathrm{fa,j}}(A),\ W = \mathrm{der}(T)$$

In this case, for the sake of comparison, the *derived alignment*, $\mathrm{algn}(A*T)$, is computed in the *independent-sample-distributed relative dependent multinomial space computer*, $I^{*}_{U_A,W,z,\mathrm{a}}(A*T)$, as though it is *derived relative dependent multinomial space*.

Note that if the representation of the *independent histogram*, A^{X}, has previously been computed, then a faster implementation of $I'^{*}_{z,\mathrm{a,l,r}}(A)(T)$ might be to *lift* rather than *derive*. That is, instead of the computation $I^{*}_{\mathrm{X}}(A * T) = (A * T)^{\mathrm{X}}$ only the computation $I^{*}_{*\mathrm{T}}((T, A^{\mathrm{X}})) = A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$ would be needed. The representation of the *independent histogram*, A^{X}, might be known, for example, if computed for a previous *transform*.

Consider the numerators of the *derived* computation, $\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)$ and $\hat{Q}_{\mathrm{m},U_A}((A*T)^{\mathrm{X}}, z)(A*T)$. The computation *times* of the numerators, at least $I^{\mathrm{t}}_{\mathrm{X}}(A)+I^{\mathrm{t}}_{\mathrm{m}}((A^{\mathrm{X}}, A))$ and at least $I^{\mathrm{t}}_{*\mathrm{T}}((T, A))+I^{\mathrm{t}}_{\mathrm{X}}(A*T)+I^{\mathrm{t}}_{\mathrm{m}}(((A*T)^{\mathrm{X}}, A*T))$, are similar, because the *time* to compute the *derived histogram*, $A * T$, in the *transformer*, $I^{\mathrm{t}}_{*\mathrm{T}}((T, A))$, is sometimes offset by the *time* to calculate the *generalised multinomial probability*, $\hat{Q}_{\mathrm{m},U_A}((A * T)^{\mathrm{X}}, z)(A * T)$, in the *multinomial space computer*, $I^{\mathrm{t}}_{\mathrm{m}}(((A*T)^{\mathrm{X}}, A*T))$, because of the possibly smaller *effective derived volume*, $|(A * T)^{\mathrm{F}}| \leq |A^{\mathrm{F}}|$.

Consider the denominators of the *derived* computation, $\sum \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)$:

$B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)$ and $\hat{Q}_{\mathrm{m},U_A}((A*T)^{\mathrm{X}},z)((A*T)^{\mathrm{X}})$. In contrast to the computation *times* of the numerators, the computation *time* of the *relative dependent* denominator, at least $I_{\mathrm{m}}^{\mathrm{t}}(((A*T)^{\mathrm{X}},(A*T)^{\mathrm{X}}))$, in $I'^{*}_{z,\mathrm{a,l,r}}(A)$ is less than the computation *time* of the *iso-transform-independent conditional dependent* denominator, at least $\sum I_{\mathrm{m}}^{\mathrm{t}}((A^{\mathrm{X}},B)) : B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)$, in $I^{*}_{z,\mathrm{y,l}}(A)$, if the set of *iso-transform-independents* is not singleton, $|\mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)| > 1$. Therefore, summing over all of the *literal substrate transforms*, $\mathcal{T}_{\mathrm{fa,j}}(A)$, the computation *time* of the *relative literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I'^{\mathrm{t}}_{z,\mathrm{a,l,r}}(A)$, must be less than that of the *literal alignment-bounded iso-transform space ideal transform inducer*, $I^{\mathrm{t}}_{z,\mathrm{y,l}}(A)$. That is,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{t}}_{z,\mathrm{a,l,r}}(A) < I^{\mathrm{t}}_{z,\mathrm{y,l}}(A))$$

The computation *time* of the *iso-transform-independent conditional dependent* denominator varies as the cardinality of the *integral iso-transform-independents*, $|\mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)|$. For comparison, the average cardinality of the *integral iso-independents* is

$$\frac{|\mathcal{A}_{U_A,\mathrm{i},V_A,z}|}{|\mathrm{ran}(Y_{U_A,\mathrm{i},V_A,z})|} = \frac{(z+v-1)!}{z!\ (v-1)!} / \prod_{w \in V_A} \frac{(z+|U_A(w)|-1)!}{z!\ (|U_A(w)|-1)!}$$

where *volume* $v = |V^{\mathrm{C}}|$. The cardinality of the *integral iso-transform-independents*, $|\mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)|$, is less than or equal to the cardinality of the *integal congruent support*

$$|\mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)| \leq |\mathcal{A}_{U_A,\mathrm{i},V_A,z}| = \frac{(z+v-1)!}{z!\ (v-1)!} = \frac{v}{z}\frac{z^{\overline{v}}}{v^{\underline{v}}} = \frac{v^{\overline{z}}}{z^{\underline{z}}}$$

If the *independent-sample-distributed iso-transform-independent conditional dependent multinomial space computer*, $I_{U,V,z,\mathrm{T,y}}$, is implemented such that the computations $\{I^{*}_{\mathrm{m}}((A^{\mathrm{X}},B)) : B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)\}$ are performed serially, the computation *time* complexity of $I_{U,V,z,\mathrm{T,y}}$ is at least exponential, $\mathrm{maximum}(z^v, v^z)$. If the computations are performed in parallel, then it is the computation *space* complexity which is at least exponential, $\mathrm{maximum}(z^v, v^z)$. The corresponding computation *time/space* complexity of the *alignment-bounded iso-transform space ideal transform inducer*, $I_{z,\mathrm{y,l}}$, which computes $\{I^{*}_{U_A,V_A,z,\mathrm{T,y}}((A,T)) : T \in \mathcal{T}_{\mathrm{fa,j}}(A)\}$, is therefore also at least exponential in *size*, z, or *substrate volume*, v. The *alignment-bounded iso-transform space ideal transform inducer*, $I_{z,\mathrm{y,l}}$, is therefore an *intractable inducer*.

The *derived alignment*, $\mathrm{algn}(A * T)$, is implemented above in the *relative literal derived alignment inducer*, $I'_{z,\mathrm{a,l,r}}$, as though it were *relative dependent multinomial space* for purposes of comparison. That is, by means of the *independent-sample-distributed relative dependent multinomial space computer*, $I_{U,V,z,\mathrm{a}}$, which in turn is implemented with the *multinomial space computer*, I_{m}. However, the *derived alignment* does not in fact depend on the *independent distribution histogram*, $(A*T)^{\mathrm{X}}$, and so a faster implementation is by means of an *alignmenter* instead. As shown in sections 'Computation of the application of a transform' and 'Computation of alignment', above, the computation of the *derived alignment* can be performed by the application of a *transformer*, $I_{*\mathrm{T}}$, followed by the application of an *alignmenter*, $I_{\mathrm{a}} = \mathrm{alignmenter}$, $I_{\mathrm{a}}^*(I_{*T}^*((T,A))) \approx \mathrm{algn}(A * T)$. The calculation of *alignment* in the *alignmenter* internally computes the *independent derived* in an *independenter*, $I_{\mathrm{X}}^*(A * T) = (A * T)^{\mathrm{X}}$. In the *alignmenter* implementation the computation *time* of *histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, is simplified to

$$I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A) > \sum I^{\mathrm{t}}_{*\mathrm{T}}((T,A)) + I^{\mathrm{t}}_{\mathrm{a}}(A * T) : T \in \mathcal{T}_{\mathrm{fa,j}}(A)$$

Thus the computation *time* of *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, is less than the computation *time* of the *relative literal derived alignment inducer*, $I'_{z,\mathrm{a,l,r}}$,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A) < I^{\mathrm{t}}_{z,\mathrm{a,l,r}}(A))$$

Therefore the computation *time* of the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A)$, must be less than that of the *literal alignment-bounded iso-transform space ideal transform inducer*, $I^{\mathrm{t}}_{z,\mathrm{y,l}}(A)$. That is,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A) < I^{\mathrm{t}}_{z,\mathrm{y,l}}(A))$$

The *time* of the *alignment* computation, $I^{\mathrm{t}}_{*\mathrm{T}}((T,A)) + I^{\mathrm{t}}_{\mathrm{a}}(A * T)$, depends on the representations of the *histogram* and *transform*, as well as the implementation of the *log factorialer*, $I_{\approx\ln!} = \mathrm{logfactorialer}$. If the *histogram*, A, is implemented in an *array histogram representation* and the *transform*, T, is implemented in a *binary map histogram representation*, both on *ordered list state representations*, then the *time* complexity is $\mathrm{maximum}(v \ln v, mw)$, where the *underlying dimension* $n = |V|$, the *underlying volume* $v = |V^{\mathrm{C}}|$, the *derived dimension* $m = |W|$, the *derived volume* $w = |W^{\mathrm{C}}|$ and the *derived variables* $W = \mathrm{der}(T)$. If the *histogram*, A, is implemented in a *binary map histogram representation*, then the *time* complexity is $\mathrm{maximum}(b \ln v, mw)$, where the *effective volume* $b = |A^{\mathrm{F}}|$. The *time* complexity of the *independenter* is mw.

Given that (i) the *independent substrate histogram* is *completely effective*, $A^{\mathrm{XF}} = V^{\mathrm{C}}$, and (ii) *literal substrate transforms*, $\mathcal{T}_{\mathrm{fa,j}}(A)$, are such that the *formal histogram* is *independent*, $A^{\mathrm{X}} * T \equiv (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$, then the *effective formal histogram* is a *cartesian sub-volume*, $(A^{\mathrm{X}} * T)^{\mathrm{F}} = (V^{\mathrm{C}} * T)^{\mathrm{F}} = (V^{\mathrm{C}} * T)^{\mathrm{XF}}$. The *derived variables* of the *transform*, T, are *partition variables*, $\mathrm{der}(T) \subseteq \mathrm{B}(V^{\mathrm{CS}})$, so the *cartesian sub-volume* must equal the *cartesian derived*, $(V^{\mathrm{C}} * T)^{\mathrm{XF}} = W^{\mathrm{C}}$. Therefore the *effective formal* equals the *cartesian derived*, $(A^{\mathrm{X}} * T)^{\mathrm{F}} = W^{\mathrm{C}}$. The *transform* is *functional*, $T \in \mathcal{T}_{\mathrm{f}}$, so the *derived volume* is no greater than the *underlying volume*, $w \leq v$. Thus the computation *time* complexity of the *alignmenter* implementation of the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I_{z,\mathrm{a,l}}^{'}(A)$, for a *literal substrate transform*, $T \in \mathcal{T}_{\mathrm{fa,j}}(A)$, is $\mathrm{maximum}(v \ln v, mv)$.

In the stricter case that the *transform* is *non-overlapping*, $\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}} * T \equiv (A^{\mathrm{X}} * T)^{\mathrm{X}}$, then the *derived dimension* is no greater than the *underlying dimension*, $m \leq n$. In this case, the computation *time* complexity of the *alignmenter* implementation of the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I_{z,\mathrm{a,l}}^{'}(A)$, for a *non-overlapping literal substrate transform*, $T \in \mathcal{T}_{\mathrm{fa,j}}(A) \cap \mathcal{T}_{U_A,V_A,\mathrm{n}}$, is at most log-linear in v, $\mathrm{maximum}(v \ln v, nv) = v \ln v$

$$
\begin{aligned}
&\forall z \in \mathbf{N}_{>0}\ \exists c \in \mathbf{N}_{>0} \\
&\quad (\{((A,T), I_{*\mathrm{T}}^{\mathrm{t}}((T,A)) + I_{\mathrm{a}}^{\mathrm{t}}(A * T)) : A \in \mathcal{A}_{z,\mathrm{xi}},\ T \in \mathcal{T}_{U_A,V_A,\mathrm{n}}\} \\
&\in \quad \mathrm{O}(\{((A,T), v \ln v) : A \in \mathcal{A}_{z,\mathrm{xi}},\ T \in \mathcal{T}_{U_A,V_A,\mathrm{n}},\ v = |V_A^{\mathrm{C}}|\}, c))
\end{aligned}
$$

The overall *non-overlapping* computation *time* complexity, $v \ln v$, is not limited by the computation *time* complexity of the *independenter*, mw. However, if the *independent histogram*, A^{X}, has previously been computed, then a faster implementation of $I_{z,\mathrm{a,l}}^{'*}(A)(T)$ might be to *lift* rather than *derive*. In this case the *alignmenter* would not need to compute $I_{\mathrm{X}}^{*}(A * T) = (A * T)^{\mathrm{X}}$ but merely apply the *log factorialer* to $I_{*\mathrm{T}}^{*}((T, A^{\mathrm{X}})) = A^{\mathrm{X}} * T$.

Contrast the *non-overlapping* computation *time* complexity, $v \ln v$, of the *alignmenter* implementation, in the *literal derived alignment inducer*, $I_{z,\mathrm{a,l}}^{'}$, to the serially implemented *independent-sample-distributed iso-transform-independent conditional dependent multinomial space computer*, $I_{U,V,z,\mathrm{T,y}}$, in the *literal alignment-bounded inducer*, $I_{z,\mathrm{y,l}}$, which has at least exponential computation *time* complexity in both v and z, maximum(z^v, v^z). Note that the *time* complexity of the *alignmenter* implementation in the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, whether *overlapping* or not, does not depend on z at all.

Consider a *literal derived alignment integral-independent substrate ideal formal-abstract fud inducer* $I'_{z,\mathrm{a,F,l}} \in \mathrm{inducers}(z)$ which has, as its subset of the *substrate models*, the *literal substrate fuds*,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (\mathrm{dom}(I'^{*}_{z,\mathrm{a,F,l}}(A)) = \mathcal{F}_{\mathrm{fa,j}}(A) \subset \mathcal{M}_{U_A,V_A})$$

where the *literal substrate fuds* is defined $\mathcal{F}_{\mathrm{fa,j}}(A) = \{F : F \in \mathcal{F}_{U_A,V_A},\ A^{\mathrm{X}} * F^{\mathrm{T}} = (A*F^{\mathrm{T}})^{\mathrm{X}},\ A = A*F^{\mathrm{T}}*F^{\mathrm{T}\dagger A}\} \subset \mathcal{F}_{U_A,V_A}$. The application of the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$, makes the same finite approximation of the *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$, as is made in the *literal transform inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{*}_{z,\mathrm{a,F,l}}(A) = \{(F, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algn}(A * F^{\mathrm{T}}))) : F \in \mathcal{F}_{\mathrm{fa,j}}(A)\})$$

So the *maximum transform* function of the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$, equals the *maximum transform* of the *literal derived alignment transform inducer*, $I'_{z,\mathrm{a,l}}$, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,l}} = \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}$. Therefore the correlation of the *maximum transform* function of the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$, with that of the *alignment-bounded iso-transform space ideal transform search set*, $X_{z,\mathrm{xi,T,y,fa,j}}$, equals the correlation of the *literal derived alignment transform inducer*, $I'_{z,\mathrm{a,l}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\ &\quad (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}) =\\ &\qquad\qquad \mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,l}}))\end{aligned}$$

The *literal substrate fuds* includes those *fuds* which consist solely of a *literal substrate transform*, $\{\{T\} : T \in \mathcal{T}_{\mathrm{fa,j}}(A)\} \subset \mathcal{F}_{\mathrm{fa,j}}(A)$. Therefore both the *time* and *space* of the *literal fud inducer*, $I'_{z,\mathrm{a,F,l}}$, are greater than the corresponding *time* and *space* of the *literal transform inducer*, $I'_{z,\mathrm{a,l}}$, whether the implementation is serial or parallel, $\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{t}}_{z,\mathrm{a,F,l}}(A) > I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A))$ and $\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{s}}_{z,\mathrm{a,F,l}}(A) > I'^{\mathrm{s}}_{z,\mathrm{a,l}}(A))$.

Similarly, consider a *literal derived alignment integral-independent substrate ideal formal-abstract decomposition inducer* $I'_{z,\mathrm{a,D,l}} \in \mathrm{inducers}(z)$ which has, as its subset of the *substrate models*, the *literal substrate decompositions*,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (\mathrm{dom}(I'^{*}_{z,\mathrm{a,D,l}}(A)) = \mathcal{D}_{\mathrm{fa,j}}(A) \subset \mathcal{M}_{U_A,V_A})$$

where the *literal substrate decompositions* is defined $\mathcal{D}_{\mathrm{fa,j}}(A) = \{D : D \in \mathcal{D}_{U_A,V_A},\ A^{\mathrm{X}} * D^{\mathrm{T}} = (A * D^{\mathrm{T}})^{\mathrm{X}},\ A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}\} \subset \mathcal{D}_{U_A,V_A}$. The application of the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$, makes the same finite approximation of the *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$, as is made in the *literal transform inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{*}_{z,\mathrm{a,D,l}}(A) = \{(D, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algn}(A * D^{\mathrm{T}}))) : D \in \mathcal{D}_{\mathrm{fa,j}}(A)\})$$

So the *maximum transform* function of the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$, equals the *maximum transform* of the *literal derived alignment transform inducer*, $I'_{z,\mathrm{a,l}}$, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}} = \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}$.

The *literal substrate decompositions* includes those *decompositions* which consist solely of a *literal substrate transform*, $\{\{((\emptyset, T), \emptyset)\} : T \in \mathcal{T}_{\mathrm{fa,j}}(A)\} \subset \mathcal{D}_{\mathrm{fa,j}}(A)$. Therefore both the *time* and *space* of the *literal decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$, are greater than the corresponding *time* and *space* of the *literal transform inducer*, $I'_{z,\mathrm{a,l}}$, whether the implementation is serial or parallel, $\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{t}}_{z,\mathrm{a,D,l}}(A) > I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A))$ and $\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{s}}_{z,\mathrm{a,D,l}}(A) > I'^{\mathrm{s}}_{z,\mathrm{a,l}}(A))$.

Finally, consider a *literal derived alignment integral-independent substrate ideal formal-abstract fud decomposition inducer* $I'_{z,\mathrm{a,D,F,l}} \in \mathrm{inducers}(z)$ which has, as its subset of the *substrate models*, the *literal substrate fud decompositions*,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (\mathrm{dom}(I'^{*}_{z,\mathrm{a,D,F,l}}(A)) = \mathcal{D}_{\mathrm{F,fa,j}}(A) \subset \mathcal{M}_{U_A,V_A})$$

where the *literal substrate fud decompositions* is defined $\mathcal{D}_{\mathrm{F,fa,j}}(A) = \{D : D \in \mathcal{D}_{\mathrm{F},U_A,V_A},\ A^{\mathrm{X}} * D^{\mathrm{T}} = (A * D^{\mathrm{T}})^{\mathrm{X}},\ A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}\} \subset \mathcal{D}_{\mathrm{F},U_A,V_A}$. The application of the *literal derived alignment fud decomposition inducer*, $I'_{z,\mathrm{a,D,F,l}}$, makes the same finite approximation of the *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$, as is made in the *literal transform inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{*}_{z,\mathrm{a,D,F,l}}(A) = \{(D, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algn}(A * D^{\mathrm{T}}))) : D \in \mathcal{D}_{\mathrm{F,fa,j}}(A)\})$$

So the *maximum transform* function of the *literal derived alignment fud decomposition inducer*, $I'_{z,\mathrm{a,D,F,l}}$, equals the *maximum transform* of the *literal derived alignment transform inducer*, $I'_{z,\mathrm{a,l}}$, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,F,l}} = \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}$.

The *literal substrate fud decompositions* includes those *decompositions* which consist solely of a *literal substrate fud*, $\{\{(\emptyset, F)\} : F \in \mathcal{F}_{\mathrm{fa,j}}(A)\} \subset \mathcal{D}_{\mathrm{F,fa,j}}(A)$. Therefore both the *time* and *space* of the *literal fud decomposition inducer*, $I'_{z,\mathrm{a,D,F,l}}$, are greater than the corresponding *time* and *space* of the *literal fud inducer*, $I'_{z,\mathrm{a,F,l}}$, whether the implementation is serial or parallel, $\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{t}}_{z,\mathrm{a,D,F,l}}(A) > I'^{\mathrm{t}}_{z,\mathrm{a,F,l}}(A))$ and $\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{s}}_{z,\mathrm{a,D,F,l}}(A) > I'^{\mathrm{s}}_{z,\mathrm{a,F,l}}(A))$.

4.21.2 Summation aligned decomposition inducers

Consider *non-literal inducers* which have, as their subset of the *substrate models*, the subset of the *substrate decompositions*, $\mathcal{D}_{U_A,V_A}$, that are also *summation aligned decompositions*, $\mathcal{D}_{\Sigma}(A)$, where $A \in \mathcal{A}_{z,\mathrm{xi}}$. The *summation aligned decompositions*, $\mathcal{D}_{\Sigma}(A)$, are defined in section 'Decomposition alignment' above. *Summation aligned decompositions* (a) are *well behaved distinct decompositions*, $\mathcal{D}_{\Sigma}(A) \subset \mathcal{D}_{\mathrm{w},U_A}$, (b) have no *variable symmetries*, $\{(w,(C,T)) : (C,T) \in \mathrm{cont}(D),\ w \in \mathrm{der}(T)\} \in \mathrm{der}(G) \to \mathrm{cont}(D)$, and (c) are subject to the constraints of (i) *contingent diagonalisation*, $\forall (C,T) \in \mathrm{cont}(D)\ (\mathrm{diagonal}(A*C*T))$, and (ii) *contingent formal-abstract equivalence*, $\forall (C,T) \in \mathrm{cont}(D)\ ((A*C*T)^{\mathrm{X}} = A^{\mathrm{X}}*C*T)$, with respect to the *histogram*, A, where $G = \mathrm{transforms}(D)$ and $\mathrm{cont} = \mathrm{elements} \circ \mathrm{contingents}$. *Summation aligned decompositions* are such that the *content alignment* equals the *summation alignment*,

$$\forall D \in \mathcal{D}_{\Sigma}(A)\ (\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) = \mathrm{alignmentSum}(A, D))$$

where D^{T} is the *nullable transform*, and the *summation alignment* is defined $\mathrm{alignmentSum} \in \mathcal{A} \times \mathcal{D} \to \mathbf{R}$ as

$$\mathrm{alignmentSum}(A, D) := \sum_{(C,T)\in\mathrm{cont}(D)} \mathrm{algn}(A * C * T)$$

In order to calculate the *summation alignment*, $\mathrm{alignmentSum}(A, D)$, only the *contingent alignments*, $\mathrm{algn}(A*C*T)$, of the recursive *contingents* tree, $\mathrm{contingents}(D) \in \mathrm{trees}(\mathcal{A} \times \mathcal{T}_{\mathrm{f}})$, need be computed. The *contingents* tree, $\mathrm{contingents}(D)$, does not depend on the *nullable fud*, $\mathrm{nullable}(U_A)(D)$, so there is no need to compute any of the *slice transforms* or their dependents. Thus the *nullable transform*, D^{T}, need not be computed by the *inducer*. However, first consider an *inducer* where the *nullable transform*, D^{T}, is computed.

Define the *derived alignment integral independent substrate summation aligned decomposition inducer* $I'_{z,\mathrm{a,D},\Sigma} \in \mathrm{inducers}(z)$ such that the application to a *substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ is the *nullable transform derived alignment* approximation function of the *substrate summation aligned decompositions*,

$$I'^{*}_{z,\mathrm{a,D},\Sigma}(A) = \{(D, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * D^{\mathrm{T}}))) : D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma}(A)\}$$

The *derived alignment summation aligned inducer*, $I'_{z,\mathrm{a,D},\Sigma}$, is defined with the *real approxer*, $I_{\approx\mathbf{R}}$, rather than the *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$, because in some cases the *abstract alignment* is not *integral*, $(A * D^{\mathrm{T}})^{\mathrm{X}} \notin \mathcal{A}_{\mathrm{i}}$. In these cases, the *derived alignment*, $\mathrm{algn}(A * D^{\mathrm{T}})$, must be computed in the *alignmenter*, I_{a}, by means of an implementation of the unit-translated gamma function, $\Gamma_!$, such as in the *log factorialer*, $I_{\approx\ln!}$, because the factorial computations are not always integral.

However, the application of the *real approxer* is a superset of the application of the *log-rational approxer*, $I^*_{\approx\ln\mathbf{Q}} \subset I^*_{\approx\mathbf{R}}$, and so there may exist an intersection between the application of the *derived alignment summation aligned inducer*, $I'_{z,\mathrm{a,D},\Sigma}$, and the application of the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$. That is, $|I'^*_{z,\mathrm{a,D},\Sigma}(A) \cap I'^*_{z,\mathrm{a,D,l}}(A)| \geq 0$.

In order to consider this intersection, compare the *substrate summation aligned decompositions*, $\mathcal{D}_{U_A,V_A} \cap \mathcal{D}_\Sigma(A)$, to the *literal substrate decompositions*, $\mathcal{D}_{\mathrm{fa,j}}(A) = \{D : D \in \mathcal{D}_{U_A,V_A},\ A^{\mathrm{X}}*D^{\mathrm{T}} = (A*D^{\mathrm{T}})^{\mathrm{X}},\ A = A*D^{\mathrm{T}}*D^{\mathrm{T}\dagger A}\}$. The *substrate summation aligned decompositions*, $\mathcal{D}_{U_A,V_A} \cap \mathcal{D}_\Sigma(A)$, are interesting because *inducers* having them as their set of *substrate models* are able to avoid the computation of the *nullable transform*, D^{T}. The *literal substrate decompositions*, $\mathcal{D}_{\mathrm{fa,j}}(A)$, are interesting because *inducers* having them as their set of *substrate models* have a *maximum transform* function of the *derived alignment* that is correlated with that of the *alignment-bounded iso-transform space ideal transform search set*, $X_{z,\mathrm{xi,T,y,fa,j}}$.

Intersecting *decompositions* $D \in \mathcal{D}_\Sigma(A) \cap \mathcal{D}_{\mathrm{fa,j}}(A)$, are such that they are (i) *contingently diagonalised*, $\forall (C,T) \in \mathrm{cont}(D)\ (\mathrm{diagonal}(A * C * T))$, (ii) *contingently formal-abstract equal*, $\forall (C,T) \in \mathrm{cont}(D)\ (A^{\mathrm{X}}*C*T = (A*C*T)^{\mathrm{X}})$, (iii) *formal-abstract equal*, $A^{\mathrm{X}} * D^{\mathrm{T}} = (A * D^{\mathrm{T}})^{\mathrm{X}}$, and (iv) *ideal*, $A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}$. For example, a *decomposition* consisting solely of a *root transform* $D = \{((\emptyset, T_{\mathrm{r}}), \emptyset)\}$ such that the *root transform*, T_{r}, is both *diagonal*, $\mathrm{diagonal}(A * T_{\mathrm{r}})$, and *ideal*, $A = A * T_{\mathrm{r}} * T_{\mathrm{r}}^{\dagger A}$. In this case, the *contingent formal-abstract equality* and *formal-abstract equality* constraints are equivalent, $A^{\mathrm{X}} * V_A^{\mathrm{C}} * T_{\mathrm{r}} = (A * V_A^{\mathrm{C}} * T_{\mathrm{r}})^{\mathrm{X}} \iff A^{\mathrm{X}} * T_{\mathrm{r}} = (A * T_{\mathrm{r}})^{\mathrm{X}}$, because $\mathrm{cont}(D) = \{(V_A^{\mathrm{C}}, T)\}$, so $\{((\emptyset, T_{\mathrm{r}}), \emptyset)\} \in \mathcal{D}_\Sigma(A) \cap \mathcal{D}_{\mathrm{fa,j}}(A)$.

As noted above, the set of *transforms* of the *substrate models* searched by an *inducer* need not intersect with the *literal substrate transforms.* All that is required of the *derived alignment summation aligned inducer*, $I'_{z,\mathrm{a,D},\Sigma}$, is that there is a positive correlation between the *maximum function* and the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^*_{z,\mathrm{a,D},\Sigma}) \geq 0$. Conjecture that this is indeed the case because of the positive, but not perfect, transitive correlation with the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$

$$\begin{aligned}
&\forall z \in \mathbf{N}_{>0} \\
&\quad (\mathrm{var}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}}) > 0 \wedge \mathrm{var}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D},\Sigma}) > 0 \implies \\
&\qquad\qquad 1 > \mathrm{corr}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D},\Sigma}) \geq 0)
\end{aligned}$$

As noted in section 'Tractable alignment-bounding', above, a *substrate decomposition* $D \in \mathcal{D}_{U_A,V_A}$ containing more than a *root transform*, $|G| > 1$ where $G = \mathrm{transforms}(D)$, is necessarily *overlapping*, $|G| > 1 \implies \mathrm{overlap}(D^{\mathrm{T}})$, and so in some cases has *non-independent formal histogram*, $A^{\mathrm{X}} * D^{\mathrm{T}} \neq (A^{\mathrm{X}} * D^{\mathrm{T}})^{\mathrm{X}}$. If this is the case, the *formal histogram* cannot be equal to the *abstract histogram*, $A^{\mathrm{X}} * D^{\mathrm{T}} \neq (A * D^{\mathrm{T}})^{\mathrm{X}}$, and so the *substrate decomposition*, D, cannot be a *literal substrate decomposition*, $D \notin \mathcal{D}_{\mathrm{fa,j}}(A)$.

If the *decomposition* is also a *summation aligned decomposition*, $D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma}(A)$, then the *formal histogram* is necessarily *non-independent*, $(D \in \mathcal{D}_{\Sigma}(A)) \wedge (|G| > 1) \implies A^{\mathrm{X}} * D^{\mathrm{T}} \neq (A^{\mathrm{X}} * D^{\mathrm{T}})^{\mathrm{X}}$, because the *skeletal contingent reduction* $D' \in \mathrm{reductions}(A, D)$ has *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * D'^{\mathrm{T}}) > 0$. The *derived alignment* of the *skeletal contingent reduction*, D', is purely *formal*, $A * D'^{\mathrm{T}} = A^{\mathrm{X}} * D'^{\mathrm{T}}$. The *derived alignment* of the *summation aligned decomposition*, D, equals that of the *skeletal contingent reduction* and the sum of the *contingent derived alignments*,

$$\mathrm{algn}(A * D^{\mathrm{T}}) = \mathrm{algn}(A * D'^{\mathrm{T}}) + \sum_{(C,T) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * T)$$

The *summation alignment* must be positive, $\mathrm{alignmentSum}(A, D) = \sum(\mathrm{algn}(A * C * T) : (C, T) \in \mathrm{cont}(D)) \geq 0$, because of *contingent diagonalisation*. Hence the *derived alignment* of a *multiple transform summation aligned decomposition* must be at least the *formal alignment* of the corresponding *skeletal contingent reduction*, $\mathrm{algn}(A * D^{\mathrm{T}}) \geq \mathrm{algn}(A^{\mathrm{X}} * D'^{\mathrm{T}}) > 0$. This is true even if the *summation alignment* is zero, $\sum(\mathrm{algn}(A * C * T) : (C, T) \in \mathrm{cont}(D)) = 0$. This constraint tends to reduce the correlation of the *derived alignment summation aligned inducer*, $I'_{z,\mathrm{a,D},\Sigma}$, with the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$.

Although the *slices* are *contingently formal-abstract equal*, $A^{\mathrm{X}} * C * T = (A * C * T)^{\mathrm{X}}$, and therefore *contingently independent-formal*, $A^{\mathrm{X}} * C * T = (A^{\mathrm{X}} * C * T)^{\mathrm{X}}$, they are not necessarily *independent-formal slices*, $(A * C)^{\mathrm{X}} * T = ((A * C)^{\mathrm{X}} * T)^{\mathrm{X}}$, and so the *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}})$, may be higher than would otherwise be the case. That is, an *independent slice*, $(A * C) = (A * C)^{\mathrm{X}}$, may have a *transform* T where the *derived alignment* is purely *formal*, $\mathrm{algn}(A * C * T) = \mathrm{algn}((A * C)^{\mathrm{X}} * T) > 0$. The overall *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}})$, is higher in the cases where the *slices* have *formal alignment*, not just where the *skeletal contingent reduction* has *formal alignment*.

Furthermore, given a *substrate summation aligned decomposition* $D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma}(A)$ that is *ideal*, $A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}$, there may exist a *super substrate summation aligned decomposition* $E \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma}(A)$, where $D \in \mathrm{subtrees}(E)$, having higher *derived alignment*, $\mathrm{algn}(A * E^{\mathrm{T}}) > \mathrm{algn}(A * D^{\mathrm{T}})$. This is because the *independent components* of the *partition*, D^{P}, are allowed purely *formal transforms*, $\mathrm{algn}(A * C * T) = \mathrm{algn}((A * C)^{\mathrm{X}} * T) > 0$, where $C^{\mathrm{S}} \in D^{\mathrm{P}}$.

Conversely, given a *substrate summation aligned decomposition* $D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma}(A)$ that is *ideal*, $A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}$, there may exist a *super substrate summation aligned decomposition* $E \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma}(A)$, where $D \in \mathrm{subtrees}(E)$, having the same *summation alignment*, $\mathrm{alignmentSum}(A, E) = \mathrm{alignmentSum}(A, D)$, but higher *derived alignment*, $\mathrm{algn}(A * E^{\mathrm{T}}) > \mathrm{algn}(A * D^{\mathrm{T}})$, because of higher *skeletal formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * E'^{\mathrm{T}}) > \mathrm{algn}(A^{\mathrm{X}} * D'^{\mathrm{T}})$. For example, if the *super-decomposition*, E, is an *immediate super-decomposition* having additional *slice* $\{(C, T)\} = \mathrm{cont}(E) \setminus \mathrm{cont}(D)$ which is such that $\mathrm{algn}(A * C * T) = 0$. Therefore these cases also reduce the correlation of the *derived alignment summation aligned inducer*, $I'_{z,\mathrm{a,D},\Sigma}$, with the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$.

Therefore consider the set of *idealising summation aligned decompositions*, $\mathcal{D}_{\Sigma,\mathrm{k}}(A) \subset \mathcal{D}_{\Sigma}(A)$, which are *summation aligned decompositions* that are subject to two additional constraints, (iii) *non-independent contingent derived histograms*, $\forall (C, T) \in \mathrm{cont}(D)\ (A * C * T \neq (A * C * T)^{\mathrm{X}})$ and (iv) *independent formal slice*, $\forall (C, T) \in \mathrm{cont}(D)\ ((A * C)^{\mathrm{X}} * T = ((A * C)^{\mathrm{X}} * T)^{\mathrm{X}})$. Define the *derived alignment integral-independent substrate idealising summation aligned decomposition inducer* $I'_{z,\mathrm{a,D},\Sigma,\mathrm{k}} \in \mathrm{inducers}(z)$ such that the application to a *non-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}} \setminus \{A^{\mathrm{X}}\}$ is the *nullable transform alignment* approximation function of the *substrate idealising summation aligned decompositions*,

$$I'^{*}_{z,\mathrm{a,D},\Sigma,\mathrm{k}}(A) = \{(D, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * D^{\mathrm{T}}))) : D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)\}$$

There are no *idealising summation aligned decompositions* of an *independent substrate histogram*, $\mathcal{D}_{\Sigma,\mathrm{k}}(A^{\mathrm{X}}) = \emptyset$, but the definition of an *inducer* requires a domain of at least the *integral-independent substrate histograms*, $\mathcal{A}_{z,\mathrm{xi}}$, so define

$$I'^{*}_{z,\mathrm{a,D},\Sigma,\mathrm{k}}(A^{\mathrm{X}}) = \{(D_{\mathrm{u}}, 0)\}$$

where $D_{\mathrm{u}} = \{((\emptyset, T_{\mathrm{u}}), \emptyset)\}$ and $T_{\mathrm{u}} = \{V_A^{\mathrm{CS}}\}^{\mathrm{T}}$.

An *idealising substrate summation aligned decomposition* $D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)$ that is *ideal*, $A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}$, has no *super idealising substrate summation aligned decomposition*, $\forall E \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ (D \notin \mathrm{subtrees}(E))$. Conjecture that the *derived alignment idealising summation aligned inducer*, $I'_{z,\mathrm{a,D},\Sigma,\mathrm{k}}$, is positively, but not perfectly, correlated with the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$

$$\begin{aligned}
&\forall z \in \mathbf{N}_{>0} \\
&\quad (\mathrm{var}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}}) > 0 \wedge \mathrm{var}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D},\Sigma,\mathrm{k}}) > 0 \implies \\
&\qquad\qquad 1 > \mathrm{corr}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D},\Sigma,\mathrm{k}}) \geq 0)
\end{aligned}$$

and that the correlation is greater than that for the *derived alignment summation aligned inducer*, $I'_{z,\mathrm{a,D},\Sigma}$,

$$\begin{aligned}
&\forall z \in \mathbf{N}_{>0} \\
&\quad (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D},\Sigma}) \leq \\
&\qquad\qquad \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D},\Sigma,\mathrm{k}}))
\end{aligned}$$

The correlation is increased by the additional *idealising* constraints which, while they do not increase the cardinality of the *literal* intersection, $|\mathcal{D}_{\Sigma,\mathrm{k}}(A) \cap \mathcal{D}_{\mathrm{fa,j}}(A)| \leq |\mathcal{D}_{\Sigma}(A) \cap \mathcal{D}_{\mathrm{fa,j}}(A)|$, remove *decompositions* from the *maximum* set, $\mathrm{maxd}(I'^{*}_{z,\mathrm{a,D},\Sigma}(A)) \subset \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma}(A)$, that are there because of unnecessary *formal alignment*. That is, $\mathrm{maxr}(I'^{*}_{z,\mathrm{a,D},\Sigma,\mathrm{k}}(A)) \leq \mathrm{maxr}(I'^{*}_{z,\mathrm{a,D},\Sigma}(A))$.

Now define the *content alignment integral-independent substrate idealising summation aligned decomposition inducer* $I'_{z,\mathrm{c,D},\Sigma,\mathrm{k}} \in \mathrm{inducers}(z)$ such that the application to a *non-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}} \setminus \{A^{\mathrm{X}}\}$ is the *summation alignment* approximation function of the *substrate idealising summation aligned decompositions*,

$$I'^{*}_{z,\mathrm{c,D},\Sigma,\mathrm{k}}(A) = \{(D, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algnSum}(A, D))) : D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)\}$$

where algnSum = alignmentSum. Define $I'^{*}_{z,\mathrm{c,D},\Sigma,\mathrm{k}}(A^{\mathrm{X}}) = \{(D_{\mathrm{u}}, 0)\}$. Note that the computation of *summation alignment*, $\mathrm{algnSum}(A, D)$, does not require the computation of non-integral factorials because the *abstract histogram* of the *slices* is *integral*, $(A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}) \wedge (A^{\mathrm{X}} * C * T = (A * C * T)^{\mathrm{X}}) \implies$

$(A * C * T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$. Therefore the *alignmenter*, I_{a}, need not be implemented by means of an implementation of the unit-translated gamma function, $\Gamma_!$, such as in the *log factorialer*, $I_{\approx\ln!}$. The *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$, can therefore be used in preference to the *real approxer*, $I_{\approx\mathbf{R}}$.

The *content alignment idealising summation aligned inducer*, $I'_{z,\mathrm{c,D},\Sigma,\mathrm{k}}$, avoids the computation of the *nullable transform*, D^{T}, but compromises by computing only the *content alignment*, $\mathrm{algn}(A*D^{\mathrm{T}})-\mathrm{algn}(A^{\mathrm{X}}*D^{\mathrm{T}}) = \mathrm{algnSum}(A, D)$, instead of the *derived alignment*, $\mathrm{algn}(A * D^{\mathrm{T}})$. Of course, the *literal substrate decompositions*, $\mathcal{D}_{\mathrm{fa,j}}(A)$, of the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$, have no *formal alignment*, $\forall E \in \mathcal{D}_{\mathrm{fa,j}}(A)\ (\mathrm{algn}(A^{\mathrm{X}}*E^{\mathrm{T}}) = 0)$, because *formal-abstract equality* implies *independent formal*, $A^{\mathrm{X}} * E^{\mathrm{T}} = (A * E^{\mathrm{T}})^{\mathrm{X}} \implies A^{\mathrm{X}} * E^{\mathrm{T}} = (A^{\mathrm{X}} * E^{\mathrm{T}})^{\mathrm{X}}$. Thus the intersecting *summation aligned decompositions*, $\mathcal{D}_{\Sigma}(A) \cap \mathcal{D}_{\mathrm{fa,j}}(A)$, are such that *content alignment* equals *derived alignment*. So the reduction in correlation is not necessarily as great as would otherwise be the case. Conjecture that the *content alignment idealising summation aligned inducer*, $I'_{z,\mathrm{c,D},\Sigma,\mathrm{k}}$, is positively correlated with the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D},\Sigma,\mathrm{k}}) \geq 0)$$

but that the correlation is lower than that for the *derived alignment idealising summation aligned inducer*, $I'_{z,\mathrm{a,D},\Sigma,\mathrm{k}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\ &\quad(\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D},\Sigma,\mathrm{k}}) \geq\\ &\qquad\qquad\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,D,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D},\Sigma,\mathrm{k}}))\end{aligned}$$

An *idealising substrate summation aligned decomposition* $D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)$ that is *ideal*, $A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}$, has no *super idealising substrate summation aligned decomposition*, $\forall E \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ (D \notin \mathrm{subtrees}(E))$. All of its *sub idealising substrate summation aligned decompositions* have lower *content alignment*, $\forall E \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ (E \in \mathrm{subtrees}(D) \implies \mathrm{algnSum}(A, E) < \mathrm{algnSum}(A, D))$. Therefore, the *maximum idealising substrate summation aligned decompositions* in the *content idealising inducer*, $I'_{z,\mathrm{c,D},\Sigma,\mathrm{k}}$, are all *ideal*, $\forall D \in \mathrm{maxd}(I'^{*}_{z,\mathrm{c,D},\Sigma,\mathrm{k}}(A))\ (\mathrm{ideal}(A, D^{\mathrm{T}}))$. Thus the *content idealising inducer*, $I'_{z,\mathrm{c,D},\Sigma,\mathrm{k}}$, is positively correlated with the *literal derived alignment decomposition inducer*, $I'_{z,\mathrm{a,D,l}}$, because the *maximum idealising substrate summation aligned decompositions* are all *ideal* even if not *formal-abstract equivalent*.

Define the *content alignment integral-independent substrate idealising summation aligned fud decomposition inducer* $I'_{z,\mathrm{c,D,F,\Sigma,k}} \in \mathrm{inducers}(z)$ such that the application to a *non-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}} \setminus \{A^{\mathrm{X}}\}$ is the *summation alignment* approximation function of the *substrate idealising summation aligned fud decompositions*,

$$I'^{*}_{z,\mathrm{c,D,F,\Sigma,k}}(A) =$$
$$\{(D, I^{*}_{\approx \ln \mathbf{Q}}(\mathrm{algnSum}(A,D))) : D \in \mathcal{D}_{\mathrm{F},U_A,V_A},\ D^{\mathrm{DV}_A} \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\}$$

where $D^{\mathrm{DV}} := \mathrm{map}(\mathrm{expand}(U,V) \circ \mathrm{transform}, D)$. Define $I'^{*}_{z,\mathrm{c,D,F,\Sigma,k}}(A^{\mathrm{X}}) = \{(D_{\mathrm{F,u}}, 0)\}$ where $D_{\mathrm{F,u}} = \{((\emptyset, \{T_{\mathrm{u}}\}), \emptyset)\}$ and $T_{\mathrm{u}} = \{V_A^{\mathrm{CS}}\}^{\mathrm{T}}$.

The *maximum transform* function of the *content alignment fud decomposition inducer* $I'_{z,\mathrm{c,D,F,\Sigma,k}}$, equals the *maximum transform* function of the *content alignment decomposition inducer* $I'_{z,\mathrm{c,D,\Sigma,k}}$, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D,F,\Sigma,k}} = \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D,\Sigma,k}}$, because the applications of the *inducers* can be mapped, $\{(D^{\mathrm{DV}_A}, c) : (D,c) \in I'^{*}_{z,\mathrm{c,D,F,\Sigma,k}}(A)\} = I'^{*}_{z,\mathrm{c,D,\Sigma,k}}(A)$.

4.21.3 Intractabilities

Although the computation *time* of the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I'_{z,\mathrm{a,l}}$, is lower than that of the *intractable literal alignment-bounded iso-transform space ideal transform inducer*, $I_{z,\mathrm{y,l}}$, $\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A) < I^{\mathrm{t}}_{z,\mathrm{y,l}}(A))$, the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, is also an *intractable inducer*. That is, either or both of (a) the computational *time* complexity, or (b) the representational *encoding space* complexity, is greater than polynomial with respect to some parameter. There are several reasons why this is the case, (i) intractable *substrate volume*, (ii) intractable *derived volume*, (iii) intractable *search set* domain, (iv) intractable *partition variables*, and (v) intractable *literal substrate model* inclusion.

4.21.4 Intractable substrate volume

The *literal substrate transforms* $\mathcal{T}_{\mathrm{fa,j}}(A) \subset \mathcal{T}_{U_A,V_A}$ of *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$, are not tractably computable. The application of the *transformer* $I_{*\mathrm{T}}$, in the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, to the *substrate histogram*, A, and any *literal substrate transform*, $T \in \mathcal{T}_{\mathrm{fa,j}}(A)$, is $I^{*}_{*\mathrm{T}}((T,A)) = A * T$. The representation *encoding space* complexity of the *one functional transform*, T, in the *transformer*, $I_{*\mathrm{T}}$, varies as the *underlying volume* $v = |V_A^{\mathrm{C}}|$. This is because the *histogram*, $\mathrm{his}(T) \in \mathcal{A}_{\mathrm{i}}$, of a *one functional transform*, $T \in \mathcal{T}_{U_A,\mathrm{f},1}$, has cardinality equal to the *underlying volume*, $|\mathrm{his}(T)| = v$. Hence the representation *space* of the *transform* in the *transformer* must be at least as large as the *underlying volume*, $I^{\mathrm{s}}_{*\mathrm{T}}((T, A)) > v$. All *substrate transforms* have the same *underlying volume*, $\forall T \in \mathcal{T}_{U_A,V_A}\ (|\mathrm{his}(T)| = v)$. The *volume* grows exponentially with *underlying dimension* $n = |V_A|$, and so the *space* complexity of the *transformer*, $I_{*\mathrm{T}}$, is exponential with respect to *underlying dimension*, n. For example, a *regular substrate* of *valency* d has *volume* $v = d^n$.

Although the *space* complexity of *substrate transforms* in the *transformer* is intractable, the *time* complexity of the application of the *transformer*, $I^{*}_{*\mathrm{T}}((T, A))$, is tractable. If the *transforms* are implemented in a *binary map histogram representation*, a lookup implemented by a *binary map getter*, $I_{\mathrm{B,g}}$, has *time* complexity of only $\ln v$. The overall *time* complexity of the application is then $b \ln v$ where $b = |A^{\mathrm{F}}|$, assuming that the *substrate histogram*, A, representation excludes *ineffective states*, for example in a *binary map histogram representation.*

Overall, intractable *substrate volume*, v, implies intractable *transformer*, $I_{*\mathrm{T}}$. To implement an *inducer* with tractable *substrate models*, consider subsets of the *substrate*, $\mathrm{P}(V_A)$. The cardinality of each of the *substrate* subsets can then be limited. For example, a *maximum underlying volume* limit of $\mathrm{xmax} \in \mathbf{N}_{\geq 4}$ could constrain *substrate* subset $K \subseteq V$ such that $|K^{\mathrm{C}}| \leq \mathrm{xmax}$. Another example is a *maximum underlying dimension* limit of $\mathrm{kmax} \in \mathbf{N}_{\geq 2}$ applied such that $|K| \leq \mathrm{kmax}$.

A *limited-underlying* subset of the *functional definition sets* $\mathcal{F}_{\mathrm{u}} \subset \mathcal{F}$ can be defined such that a *fud* $F \in \mathcal{F}_{\mathrm{u}}$ is such that its *transforms*, $F \subset \mathcal{T}$, are each tractably computable. Given *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$, a *limited-underlying substrate fud* $F \in \mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{u}}$ has possibly complete coverage of the *substrate*, $|\mathrm{und}(F)| \leq n$ where $n = |V_A|$, but is such that its *transforms*, $F \subset \mathcal{T}_{U_A,\mathrm{f},1}$, are each tractably computable. This is achieved by limiting the *underlying variables* of each *transform*. For example, a maximum *underlying volume* limit of xmax would constrain the *fud* $\forall T \in F\ \lozenge K = \mathrm{und}(T)\ (|K^{\mathrm{C}}| \leq \mathrm{xmax})$. If the *fud* is *non-overlapping*, $\neg\mathrm{overlap}(F)$, and the *fud* has a single *layer*, $\forall T \in F\ \lozenge K = \mathrm{und}(T)\ (K \subseteq V_A)$, then the components of the partition of the *substrate* each obey the limit. For example, $\forall u \in \mathrm{der}(F)\ \lozenge K = \mathrm{und}(\mathrm{dep}(F, \{u\}))\ (|K^{\mathrm{C}}| \leq \mathrm{xmax})$. The *limited-underlying fuds*, $\mathcal{F}_{\mathrm{u}}$, represents the class of subsets of the *functional definition sets* such that the application of the *fud* is tractable. That is, the *limited-underlying fuds*, $\mathcal{F}_{\mathrm{u}}$, here stands for one of the limiting methods, for example, *maximum underlying volume*, xmax, or *maximum underlying dimension*, kmax.

The *literal derived alignment integral-independent substrate ideal formal-abstract fud inducer*, $I'_{z,\mathrm{a,F,l}}$, has, as its subset of the *substrate models*, the *literal substrate fuds*,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (\mathrm{dom}(I'^{*}_{z,\mathrm{a,F,l}}(A)) = \mathcal{F}_{\mathrm{fa,j}}(A) \subset \mathcal{M}_{U_A,V_A})$$

As in the case of the *literal derived alignment transform inducer*, $I'_{z,\mathrm{a,l}}$, the serial *time* computation of the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$, is intractable with respect to *underlying volume*.

Consider a *limited-underlying derived alignment integral-independent substrate ideal formal-abstract fud inducer* $I'_{z,\mathrm{a,F,l,u}} \in \mathrm{inducers}(z)$ which has, as its subset of the *substrate models*, a *limited-underlying* subset of the *literal substrate fuds*,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (\mathrm{dom}(I'^{*}_{z,\mathrm{a,F,l,u}}(A)) = \mathcal{F}_{\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u}} \subset \mathcal{M}_{U_A,V_A})$$

The application of the *limited-underlying derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u}}$, is a subset of the application of the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$. That is, $I'^{*}_{z,\mathrm{a,F,l,u}}(A) \subseteq I'^{*}_{z,\mathrm{a,F,l}}(A)$. The *expanded transforms* of the domain of a *limited-underlying derived alignment fud inducer* is a subset of the *literal substrate transforms*, $\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (\{F^{\mathrm{T}V_A} : F \in \mathcal{F}_{\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u}}\} \subseteq \mathcal{T}_{\mathrm{fa,j}}(A))$. The application is implemented in a *fuder*, see below, but is otherwise equal to that of the *derived alignment inducer*, $\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ \forall F \in \mathrm{dom}(I'^{*}_{z,\mathrm{a,F,l,u}}(A))\ (I'^{*}_{z,\mathrm{a,F,l,u}}(A)(F) = I'^{*}_{z,\mathrm{a,l}}(A)(F^{\mathrm{T}V_A}))$. Conjecture that the *maximum transform* function of the *limited-underlying derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u}}$, is positively correlated with that of the *literal derived alignment transform inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,l,u}}) \geq 0)$$

The application of the *fud* $F \in \mathcal{F}_{\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u}}$ consists of the tractable sequential application of its *transforms*, $A * F^{\mathrm{T}} = \mathrm{apply}(F, A)$, where apply $\in \mathcal{F} \times \mathcal{A} \to \mathcal{A}$ is described in the section 'Functional Definition Sets'. This application is implemented in a *fuder* $I_{*\mathrm{F}} = \mathrm{fuder} \in \mathrm{computers}$, described in the section 'Computation of functional definition sets' above. The application of the *fuder* is $I^{*}_{*\mathrm{F}}((F, A)) = \mathrm{apply}(F, A)$. The representation *space* of the *substrate fud*, F, in the *fuder*, $I_{*\mathrm{F}}$, is tractable, because each of the *transformers* in the application is tractable. The *time* complexity is at most $ry \ln y$ where $r = |\mathrm{vars}(F)|$ and $y = |A^{\mathrm{F}}|$. Therefore both the *space* and *time* complexity of the *fuder* are tractable. There is no need to represent the intractable *expanded transform* of the *fud*, $F^{\mathrm{T}V_A} \in \mathcal{T}_{U_A,\mathrm{f},1}$, in order to compute the application.

Thus the computation of the *transformed histogram* in a *limited-underlying derived alignment fud inducer* is tractable despite intractable *substrate volume*, but at the cost of searching only the subset of the *literal substrate transforms* for which corresponding tractable *fuds* exist. The excluded intractable part of the *literal substrate transforms* is $\mathcal{T}_{\mathrm{fa,j}}(A) \setminus \{F^{\mathrm{T}V_A} : F \in \mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{u}}\}$. The exact composition of the subset of *literal substrate transforms*, $\{F^{\mathrm{T}V_A} : F \in \mathcal{F}_{\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u}}\}$, of a particular implementation of a *limited-underlying derived alignment fud inducer* depends on its definition of the limitations on the *underlying variables* in $\mathcal{F}_{\mathrm{u}}$. Thus there are multiple implementations of *limited-underlying derived alignment fud inducers*, $I'_{z,\mathrm{a,F,l,u}}$, for any given implementation of the *literal derived alignment transform inducer*, $I'_{z,\mathrm{a,l}}$.

4.21.5 Intractable derived volume

Given *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ and *literal substrate transform* $T \in \mathcal{T}_{\mathrm{fa,j}}(A)$, both the computation *time* and computation *space* of the *alignmenter* applied to the *transformed sample histogram*, $I^*_{\mathrm{a}}(A * T) \approx \mathrm{algn}(A * T)$, in the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, vary with the *derived volume*, $w = |W^{\mathrm{C}}|$, where $W = \mathrm{der}(T)$. That is, $I^{\mathrm{t}}_{\mathrm{a}}(A * T) > w$ and $I^{\mathrm{s}}_{\mathrm{a}}(A * T) > w$. This is because the calculation of *alignment* requires that the *independent derived* be computed by an *independenter*, $I^*_{\mathrm{X}}(A * T) = (A * T)^{\mathrm{X}}$, which has *time* and *space* complexities of at least w. Although the *formal histogram* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, and so the *derived volume* is no greater than the *underlying volume*, $w \leq v$, the *substrate histogram*, $A \in \mathcal{A}_{z,\mathrm{xi}}$, has a *completely effective independent*, $A^{\mathrm{XF}} = A^{\mathrm{C}}$, and the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$. Hence the *independent derived* is also *completely effective*, $(A^{\mathrm{F}} * T)^{\mathrm{XF}} = (A^{\mathrm{XF}} * T)^{\mathrm{XF}} = (A^{\mathrm{C}} * T)^{\mathrm{XF}} = W^{\mathrm{C}}$, and so both the computation *time* and *space* of the *alignmenter*, I_{a}, must be at least w. The *derived volume*, w, grows exponentially with *derived dimension* $m = |W|$ and so the *time* and *space* complexities are exponential, and therefore intractable, with respect to *derived dimension*, m. In the case of the *value full functional transform*, $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}\mathrm{T}} : w \in V_A\}^{\mathrm{T}} \in \mathcal{T}_{\mathrm{fa,j}}(A)$, the *derived dimension* equals the *underlying dimension*, $m = n$. In this case both complexities of the *alignmenter*, I_{a}, are also intractable with respect to *underlying dimension*, n.

So an implementation of the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I_{z,\mathrm{a,l}}'$, that uses the *alignmenter*, I_{a}, or any other computer that applies the *independenter*, I_{X}, to the *derived histogram*, $I_{\mathrm{X}}^{*}(A * T)$, must be intractable with respect to *derived dimension*, m. The *value full functional transform*, T_{s}, is a *literal substrate transform*, $T_{\mathrm{s}} \in \mathcal{T}_{\mathrm{fa,j}}(A)$, and so the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I_{z,\mathrm{a,l}}'$, is also intractable with respect to *underlying dimension*, n.

This is also the case for an implementation of a *limited-underlying derived alignment fud inducer*, $I_{z,\mathrm{a,F,l,u}}'$. Although the *fuder*, $I_{*\mathrm{F}}$, in the *limited-underlying derived alignment fud inducer* is tractable despite intractable *substrate volume*, because of the use of a tractable *fud*, $F \in \mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{u}}$, which allows tractable application, $A * F^{\mathrm{T}}$, the *limited-underlying derived alignment fud inducer*, $I_{z,\mathrm{a,F,l,u}}'$, must still compute $(A * F^{\mathrm{T}})^{\mathrm{X}}$ in an *independenter*, I_{X}, in order to compute *derived alignment*, $\mathrm{algn}(A * F^{\mathrm{T}})$. Thus *limited-underlying derived alignment fud inducer*, $I_{z,\mathrm{a,F,l,u}}'$, is intractable with respect to *derived dimension*, m. The *value full functional fud* is in the *substrate models* of the *limited-underlying derived alignment fud inducer* application, $F_{\mathrm{s}} = \{\{w\}^{\mathrm{CS\{\}T}} : w \in V_A\} \in \mathrm{dom}(I_{z,\mathrm{a,F,l,u}}'^{*}(A)) = \mathcal{F}_{\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u}}$, because it is practicable in the *fuder*. Hence the *limited-underlying derived alignment fud inducer*, $I_{z,\mathrm{a,F,l,u}}'$, is also intractable with respect to *underlying dimension*, n.

One possibility is to consider a further subset of the *literal substrate transforms*, $\mathcal{T}_{\mathrm{fa,j}}(A)$, which limits the *derived dimension*, m. For example, a *maximum derived volume* limit of $\mathrm{wmax} \in \mathbf{N}_{\geq 4}$ could constrain the subset to $\{T : T \in \mathcal{T}_{\mathrm{fa,j}}(A),\ W = \mathrm{der}(T),\ |W^{\mathrm{C}}| \leq \mathrm{wmax}\}$. Another example is a *maximum derived dimension* limit of $\mathrm{jmax} \in \mathbf{N}_{\geq 2}$ which could constrain the subset to $\{T : T \in \mathcal{T}_{\mathrm{fa,j}}(A),\ W = \mathrm{der}(T),\ |W| \leq \mathrm{jmax}\}$. Such a limit would exclude the *value full functional transform*, T_{s}, if, for example, $\mathrm{jmax} < n$.

A *limited-derived* subset of the *functional definition sets* $\mathcal{F}_{\mathrm{d}} \subseteq \mathcal{F}$ can be defined such that a *fud* $F \in \mathcal{F}_{\mathrm{d}}$ has tractably computable *independent derived*. Given *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$, a *limited-derived substrate fud* $F \in \mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{d}}$ has tractably computable *independent derived*. This is achieved by limiting the *derived variables* of the *fud*. For example, a *maximum derived volume* limit of wmax would constrain the *fud*

$|W^{\mathrm{C}}| \leq \mathrm{wmax}$ where $W = \mathrm{der}(F)$. The *limited-derived fuds*, $\mathcal{F}_{\mathrm{d}}$, represents the class of subsets of the *fuds* such that the *independent derived* of the *fud* is tractable. That is, the *limited-derived fuds* here stands for one of the limiting methods, for example, *maximum derived volume*, wmax, or *maximum derived dimension*, jmax.

Consider a *limited-variables derived alignment integral-independent substrate ideal formal-abstract fud inducer* $I'_{z,\mathrm{a,F,l,u,d}} \in \mathrm{inducers}(z)$, which has, as its subset of the *substrate models*, a *limited-underlying* and *limited-derived* subset of the *literal substrate fuds*,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (\mathrm{dom}(I'^{*}_{z,\mathrm{a,F,l,u,d}}(A)) = \mathcal{F}_{\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}} \subset \mathcal{M}_{U_A,V_A})$$

Then (i) the *fud*, F, in the *fuder*, $I_{*\mathrm{F}}$, is tractable because of the limited *underlying variables* of the *transforms* of the *fud*, and (ii) the *independent transformed histogram*, $(A * F^{\mathrm{T}})^{\mathrm{X}}$, in the *independenter*, I_{X}, is tractable because of the limited *derived variables* of the *fud*.

The application of the *limited-variables derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u,d}}$, is a subset of the application of the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$. That is, $I'^{*}_{z,\mathrm{a,F,l,u,d}}(A) \subseteq I'^{*}_{z,\mathrm{a,F,l}}(A)$. The excluded intractable part of the *literal substrate transforms* is $\mathcal{T}_{\mathrm{fa,j}}(A) \setminus \{F^{\mathrm{T}V_A} : F \in \mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}}\}$. The excluded set is a superset of the less limited *limited-underlying derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u}}$. Conjecture that the *maximum transform* function of the *limited-variables derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u,d}}$, is positively correlated with that of the *literal derived alignment transform inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,l,u,d}}) \geq 0)$$

but that the correlation is lower than that for the *limited-underlying derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\ &\quad(\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,l,u}}) \geq \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,l,u,d}}))\end{aligned}$$

The exact composition of the subset of *literal substrate transforms*, $\{F^{\mathrm{T}V_A} : F \in \mathcal{F}_{\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}}\}$, of a particular implementation of a *limited-variables derived alignment fud inducer* depends on its definition of the limitations on the *underlying variables*, $\mathcal{F}_{\mathrm{u}}$, and the limitations on the *derived variables*, $\mathcal{F}_{\mathrm{d}}$. Thus there are multiple implementations of *limited-variables derived alignment fud inducers*, $I'_{z,\mathrm{a,F,l,u,d}}$, for any given implementation of the *literal* *derived alignment transform inducer*, $I'_{z,\mathrm{a,l}}$.

The computation of *alignment* in such a *limited-variables derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u,d}}$, is therefore tractable despite both intractable *substrate volume* and intractable *derived volume*, because of tractable *fuder*, $I_{*\mathrm{F}}$, and *independenter*, I_{X}, respectively. However, although the coverage of the *inducer*, $\mathrm{und}(F) \subseteq V_A$, where $F \in \mathcal{F}_{\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}}$, is not constrained, and hence the *underlying volume*, $v = |V_A^{\mathrm{C}}|$, is not constrained, the *derived volume*, $w = |W^{\mathrm{C}}|$ where $W = \mathrm{der}(F)$, must be less than the *underlying volume*, $w < v$, if the *underlying volume* is impracticable. That is, the *underlying* freedoms are unlimited, but the *derived* freedoms are limited.

Another approach to the problem of intractable *derived volume* is to consider an *inducer* which has, as its subset of the *substrate models*, the set of *substrate summation aligned decompositions*, $\mathcal{D}_{U_A,V_A} \cap \mathcal{D}_\Sigma(A)$, where $A \in \mathcal{A}_{z,\mathrm{xi}}$. *Summation aligned decompositions* are *well behaved distinct decompositions* having no *variable symmetries* that are subject to the constraints of (i) *contingent diagonalisation* and (ii) *contingent formal-abstract equivalence*, with respect to the *histogram*, A. As described in section 'Summation aligned decomposition inducers', above, the computation of the *content alignment*, $\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}})$ where $D \in \mathcal{D}_\Sigma(A)$, does not require the computation of the *nullable transform*, D^{T}, because the *content alignment* equals the *summation alignment*, $\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) = \mathrm{alignmentSum}(A, D)$, where $\mathrm{alignmentSum}(A, D) = \sum \mathrm{algn}(A * C * T) : (C, T) \in \mathrm{cont}(D)$ and $\mathrm{cont} = \mathrm{elements} \circ \mathrm{contingents}$. Thus the intractable *derived volume*, $w = |N^{\mathrm{C}}|$ where $N = \mathrm{der}(D^{\mathrm{T}})$, need not be computed.

Define the *content alignment integral-independent substrate summation aligned decomposition inducer* $I'_{z,\mathrm{c,D},\Sigma} \in \mathrm{inducers}(z)$ such that the application to a *substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ is the *summation alignment* approximation function of the *substrate summation aligned decompositions*,

$$I'^{*}_{z,\mathrm{c,D},\Sigma}(A) = \{(D, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algnSum}(A, D))) : D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_\Sigma(A)\}$$

where algnSum = alignmentSum.

The computation of *summation alignment* in the *content summation aligned decomposition inducer*, $I'_{z,\mathrm{c,D},\Sigma}$, is not tractable because there is no constraint that the *contingent slice sizes* decrease. For example, a *summation aligned decomposition* $D \in \mathrm{dom}(I'^{*}_{z,\mathrm{c,D},\Sigma}(A))$ could consist of *mono-derived-variate transforms*, $\forall T \in \mathrm{transforms}(D)\ (|\mathrm{der}(T)| = 1)$. The computation of the *summation alignment* requires the *contingent* application of the *transformer* and *alignmenter* to each of the *transforms* requiring *time* of at least $\sum I^{\mathrm{t}}_{*\mathrm{T}}((T, A * C)) + I^{\mathrm{t}}_{\mathrm{a}}(A * C * T) : (C, T) \in \mathrm{cont}(D)$.

If the *contingent slice size* is constrained to decrease, then the longest path of the *decomposition* must be less or equal to the *size*, $\forall L \in \mathrm{paths}(D)\ (|L| \leq z)$, and the cardinality of the leaves must be less than the *size*, $\mathrm{leaves}(D) < z$. The cardinality of the *transforms* is, in this case, less than the square of the *size*, $|\mathrm{transforms}(D)| < z^2$, which has polynomial complexity.

Therefore consider the set of *idealising summation aligned decompositions*, $\mathcal{D}_{\Sigma,\mathrm{k}}(A) \subset \mathcal{D}_{\Sigma}(A)$, which are *summation aligned decompositions* that are subject to two additional constraints, (iii) *non-independent contingent derived histograms*, and (iv) *independent formal slice*. The *content alignment integral-independent substrate idealising summation aligned decomposition inducer* $I'_{z,\mathrm{c,D},\Sigma,\mathrm{k}} \in \mathrm{inducers}(z)$ is defined above such that the application to a *non-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}} \setminus \{A^{\mathrm{X}}\}$ is the *summation alignment* approximation function of the *substrate idealising summation aligned decompositions*,

$$I'^{*}_{z,\mathrm{c,D},\Sigma,\mathrm{k}}(A) = \{(D, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algnSum}(A, D))) : D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)\}$$

The *content alignment idealising summation aligned inducer*, $I'_{z,\mathrm{c,D},\Sigma,\mathrm{k}}$, is conjectured to be positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D},\Sigma,\mathrm{k}}) \geq 0)$$

The subset of *substrate models*, $\mathcal{M}_{U_A,V_A}$, of an implementation of an *idealising summation aligned decomposition inducer* cannot be *substrate decompositions* of *substrate transforms*, $\mathcal{D}_{U_A,V_A} \subset \mathrm{trees}(\mathcal{S} \times \mathcal{T}_{U_A,V_A})$, because the *substrate volume*, v, is still intractable even if the *derived volume*, w, is (indirectly) tractable. To be tractable with *limited-variables* methods the subset of *substrate models* must at least be a subset of *substrate fud decompositions*, $\mathcal{D}_{\mathrm{F},U_A,V_A}$. Define the *limited-variables content alignment integral-independent substrate idealising summation aligned fud decomposition inducer* $I'_{z,\mathrm{c,D,F},\Sigma,\mathrm{k,u,d}} \in \mathrm{inducers}(z)$ such that the application to a *non-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}} \setminus \{A^{\mathrm{X}}\}$ is the *summation alignment* function of the *limited-variables substrate idealising summation aligned fud decompo-*

sitions,

$$I'^{*}_{z,\mathrm{c,D,F,\Sigma,k,u,d}}(A) =$$
$$\{(D, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algnSum}(A,D))) :$$
$$D \in \mathcal{D}_{\mathrm{F},U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}})),\ D^{\mathrm{D}V_A} \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\}$$

Define $I'^{*}_{z,\mathrm{c,D,F,\Sigma,k,u,d}}(A^{\mathrm{X}}) = \{(D_{\mathrm{F,u}},0)\}$ where $D_{\mathrm{F,u}} = \{((\emptyset,\{T_{\mathrm{u}}\}),\emptyset)\}$ and *unary partition transform* $T_{\mathrm{u}} = \{V_A^{\mathrm{VC}}\}^{\mathrm{T}}$. Note that in the special case of *independent substrate histogram*, $A = A^{\mathrm{X}}$, the dummy *unary decomposition*, $D_{\mathrm{F,u}}$, is intractable because of intractable *substrate volume*, but, of course, the given *independent substrate histogram*, $A = A^{\mathrm{X}}$, is itself intractably computable for the same reason.

Here the $\mathcal{D}_{\mathrm{F},U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}}))$ stands for the class of subsets of the *substrate fuds*, the definitions of which depend in turn on the definitions of the *limited-underlying substrate fuds*, $\mathcal{F}_{\mathrm{u}}$, and the *limited-derived substrate fuds*, $\mathcal{F}_{\mathrm{d}}$.

Conjecture that the *limited-variables content alignment fud decomposition inducer*, $I'_{z,\mathrm{c,D,F,\Sigma,k,u,d}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D,F,\Sigma,k,u,d}}) \geq 0)$$

but that the correlation is lower than that for the *content content alignment fud decomposition inducer*, $I'_{z,\mathrm{c,D,F,\Sigma,k}}$, defined in section 'Summation aligned decomposition inducers', above,

$$\forall z \in \mathbf{N}_{>0}$$
$$(\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D,F,\Sigma,k}}) \geq$$
$$\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D,F,\Sigma,k,u,d}}))$$

The *idealising summation aligned decomposition* $D \in \mathrm{dom}(I'^{*}_{z,\mathrm{c,D,F,\Sigma,k,u,d}}(A))$ may have multiple *fuds*, so the *derived volume*, $w = |N^{\mathrm{C}}|$ where $N = \mathrm{der}(D^{\mathrm{T}})$, is not limited by the *derived dimension* of a *fud* such as is the case in the *limited-variables derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u,d}}$. However, because the correlation of the unlimited *content content alignment fud decomposition inducer*, $I'_{z,\mathrm{c,D,F,\Sigma,k}}$, with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, is not perfect, for the reasons described in section 'Summation aligned decomposition inducers', above, it is not obvious whether or not the correlation of the *limited-variables content alignment fud decomposition inducer*, $I'_{z,\mathrm{c,D,F,\Sigma,k,u,d}}$, $\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D,F,\Sigma,k,u,d}})$, is greater than that for the *limited-variables derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u,d}}$, $\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,l,u,d}})$.

4.21.6 Intractable search set elements

Given an *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$, the *literal substrate transforms*,

$$\mathcal{T}_{\mathrm{fa,j}}(A) = \mathrm{dom}(X_{z,\mathrm{xi,T,y,fa,j}}(A)) = \mathrm{dom}(X'_{z,\mathrm{xi,T,a,fa,j}}(A))$$

is defined

$$\mathcal{T}_{\mathrm{fa,j}}(A) = \{T : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}$$

The computation *time* of the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I'_{z,\mathrm{a,l}}$, must be at least as great as the cardinality of the *literal substrate transforms*, $I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A) > |\mathcal{T}_{\mathrm{fa,j}}(A)|$, if the computation of each *transform* is performed serially. If the computation is parallel, then it is the computation *space* which is at least as great as the cardinality of the *literal substrate transforms*, $I'^{\mathrm{s}}_{z,\mathrm{a,l}}(A) > |\mathcal{T}_{\mathrm{fa,j}}(A)|$. Consider the serial computation. In the *alignmenter* implementation of the *literal derived alignment inducer*, the computation *time* is such that

$$I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A) > \sum(I^{\mathrm{t}}_{*\mathrm{T}}((T, A)) + I^{\mathrm{t}}_{\mathrm{a}}(A * T) : T \in \mathcal{T}_{\mathrm{fa,j}}(A)) > |\mathcal{T}_{\mathrm{fa,j}}(A)|$$

However, to compute the *literal substrate transforms*, $\mathcal{T}_{\mathrm{fa,j}}(A)$, it is necessary to compute the entire superset of *substrate transforms*, $\mathcal{T}_{U_A,V_A} \supset \mathcal{T}_{\mathrm{fa,j}}(A)$, where $\mathcal{T}_{U_A,V_A} = \{F^{\mathrm{T}} : F \subseteq \{P^{\mathrm{T}} : P \in \mathrm{B}(V_A^{\mathrm{CS}})\}\}$. This is because tests of (i) *formal-abstract equality*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, and (ii) *ideality*, $\mathrm{ideal}(A, T)$, depend on the application of a *transform*, T, to the *substrate histogram*, A, and so all of the *substrate transforms*, $\mathcal{T}_{U_A,V_A}$, must be constructed before testing. Conjecture that there is no subset of the *multi-partition transforms*, $\mathcal{T}_{U,\mathrm{P}^*}$, that may be excluded for all *substrate histograms*, $\neg(\exists Q \in \mathrm{P}(\mathcal{T}_{U,\mathrm{P}^*}) \setminus \{\emptyset\}\ \forall A \in \mathcal{A}_{z,\mathrm{xi}}\ \forall T \in Q \cap \mathcal{T}_{U_A,V_A}\ ((A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}) \wedge (A = A * T * T^{\dagger A})))$. Therefore the computation *time* of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, must be at least as great as the cardinality of the *substrate histograms*, $I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A) > |\mathcal{T}_{U_A,V_A}|$. The cardinality of the *substrate transforms* is $|\mathcal{T}_{U_A,V_A}| = 2^{\mathrm{bell}(v)}$ where $v = |V_A^{\mathrm{C}}|$. Thus $I'^{\mathrm{t}}_{z,\mathrm{a,l}}(A) > 2^{\mathrm{bell}(v)}$. So the serial computation *time* complexity of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, is intractable with respect to *underlying volume*, v.

Consider the *derived alignment integral-independent substrate ideal formal-abstract non-overlapping transform inducer* $I'_{z,\mathrm{a,l,n}} \in \mathrm{inducers}(z)$ which is defined such that its application to an *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$ is a subset of the application of the *literal derived alignment inducer*, $I'^{*}_{z,\mathrm{a,l,n}}(A) = \{(T,a) : (T,a) \in I'^{*}_{z,\mathrm{a,l}}(A),\ \neg\mathrm{overlap}(T)\} \subset I'^{*}_{z,\mathrm{a,l}}(A)$. Thus the domain of the application is the *non-overlapping* subset of the *literal substrate transforms*, $\mathrm{dom}(I'^{*}_{z,\mathrm{a,l,n}}(A)) = \mathcal{T}_{\mathrm{fa,j}}(A) \cap \mathcal{T}_{U_A,V_A,\mathrm{n}}$ where $\mathcal{T}_{U,V,\mathrm{n}} = \{T : T \in \mathcal{T}_{U,V},\ \neg\mathrm{overlap}(T)\}$. A *non-overlapping transform* implies that the *formal histogram* is *independent*, $\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, which is also implied by *formal-abstract equality*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, so the domain of the application is not empty, $\mathrm{dom}(I'^{*}_{z,\mathrm{a,l,n}}(A)) \neq \emptyset$.

The *non-overlapping transform* limitation means that only the *non-overlapping substrate transforms set*, $\mathcal{T}_{U_A,V_A,\mathrm{n}} \subset \mathcal{T}_{U_A,V_A}$ need be constructed. This subset is independent of the *substrate histogram*, A. The serial computation *time* of the *non-overlapping derived alignment inducer*, $I'_{z,\mathrm{a,l,n}}$, is constrained $I'^{\mathrm{t}}_{z,\mathrm{a,l,n}}(A) > |\mathcal{T}_{U_A,V_A,\mathrm{n}}|$. The *non-overlapping substrate transforms set* can be constructed explicitly,

$$\mathcal{T}_{U_A,V_A,\mathrm{n}} = \{N^{\mathrm{T}V_A} : Y \in \mathrm{B}'(V_A),\ N \in \prod_{K \in Y} \mathrm{B}(K^{\mathrm{CS}})\} \cup \{(\emptyset,\emptyset)\}$$

As shown above, in section 'Substrate structures', the cardinality of this set is conjectured to be constrained $\mathrm{bell}(v) \leq |\mathcal{T}_{U_A,V_A,\mathrm{n}}| \leq 2 \times \mathrm{bell}(n) \times \mathrm{bell}(v) + 1$, where $n = |V_A|$ and $v = |V_A^{\mathrm{C}}|$. The complexity of the cardinality of the *non-overlapping substrate transforms set*, $\mathcal{T}_{U_A,V_A,\mathrm{n}}$, is therefore factorial on the *underlying volume*, $\mathrm{bell} \in \mathrm{O}(!)$. The serial computation *time* complexity of the *non-overlapping derived alignment inducer*, $I'_{z,\mathrm{a,l,n}}$, is still intractable with respect to *underlying volume*, v. Note that the *non-overlapping derived alignment inducer*, $I'_{z,\mathrm{a,l,n}}$, not only limits the *search set*, but also limits the *derived dimension*, $\neg\mathrm{overlap}(T) \implies m \leq n$, where $m = |\mathrm{der}(T)|$.

Consider the *literal derived alignment integral-independent substrate ideal formal-abstract fud inducer*, $I'_{z,\mathrm{a,F,l}}$, which has, as its subset of the *substrate models*, the *literal substrate fuds*, $\mathrm{dom}(I'^{*}_{z,\mathrm{a,F,l}}(A)) = \mathcal{F}_{\mathrm{fa,j}}(A)$ where $A \in \mathcal{A}_{z,\mathrm{xi}}$. The serial *time* computation of the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$, is intractable with respect to *underlying volume*. Similarly to the case of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, above, the entire set of *substrate fuds*, $\mathcal{F}_{U_A,V_A}$, must be constructed before testing for (i) *formal-abstract equality*, $A^{\mathrm{X}} * F^{\mathrm{T}} = (A * F^{\mathrm{T}})^{\mathrm{X}}$, and (ii) *ideality*, $\mathrm{ideal}(A, F^{\mathrm{T}})$. The cardinality of the *substrate fuds* is greater than that of the *substrate transforms*, $|\mathcal{F}_{U_A,V_A}| > |\mathcal{T}_{U_A,V_A}|$, so $I'^{\mathrm{t}}_{z,\mathrm{a,F,l}}(A) > 2^{\mathrm{bell}(v)}$ where $v = |V_A^{\mathrm{C}}|$.

Compare (i) the *literal derived alignment integral-independent substrate ideal formal-abstract fud inducer*, $I'_{z,\mathrm{a,F,l}}$, which has, as its subset of the *substrate models*, the *literal substrate fuds*, $\mathrm{dom}(I'^{*}_{z,\mathrm{a,F,l}}(A)) = \mathcal{F}_{\mathrm{fa,j}}(A)$, to (ii) the *limited-variables derived alignment integral-independent substrate ideal formal-abstract fud inducer*, $I'_{z,\mathrm{a,F,l,u,d}}$, which has, as its subset of the *substrate models*, a *limited-underlying* and *limited-derived* subset of the *literal substrate fuds*, $\mathrm{dom}(I'^{*}_{z,\mathrm{a,F,l,u,d}}(A)) = \mathcal{F}_{\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}}$. In the *limited-variables* case the definitions of the *limited-underlying*, $\mathcal{F}_{\mathrm{u}}$, and *limited-derived*, $\mathcal{F}_{\mathrm{d}}$, allow an explicit construction of a subset of *substrate fuds*, $\mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}}$, because they are independent of the *substrate histogram*, A. Thus the complexity of the cardinality of the *inducer's search models* is reduced by the limits, in a similar fashion to the reduction in complexity of the *non-overlapping transform* limitation, above. However, the *limited-variables derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u,d}}$, is still intractable with respect to *underlying dimension*, n, where $n = |V_A|$. This is for two reasons, (i) *fud flattening*, and (ii) *layer variables* cardinality.

The definition of the *substrate fud set*, $\mathcal{F}_{U_A,V_A}$, described in section 'Substrate structures', above, explicitly excludes duplicate nested *partitions* within the *fud*, $\forall F \in \mathcal{F}_{U_A,V_A}\ \forall u \in \mathrm{vars}(F) \setminus V_A\ \Diamond G = \mathrm{depends}(F,\{u\})\ \forall w \in \mathrm{vars}(G) \setminus V_A \setminus \{u\}\ \Diamond H = \mathrm{depends}(F,\{w\})\ (G^{\mathrm{TPV}_A} \neq H^{\mathrm{TPV}_A})$. The *substrate fud set* is a set of subsets of the *power fud*, $\mathcal{F}_{U_A,V_A} \subset \mathrm{P}(\mathrm{power}(U_A)(V_A)) \subset \mathcal{F}_{U_A,\mathrm{P}}$. The *power fud* is constructed recursively from the bottom *substrate layer* upwards by adding *layer partition transforms* which do not *flatten* to an existing *partition*. Without *flattening*, the *power fud* recursion would not terminate and so the *power fud* would have infinite *layers* and hence infinite cardinality. The *substrate fud set*, $\mathcal{F}_{U_A,V_A}$, is the set of subsets of the *power fud* such that the *fuds* have the same *underlying substrate variables*, $\mathcal{F}_{U_A,V_A} = \{F : F \subseteq \mathrm{power}(U_A)(V_A),\ \mathrm{und}(F) \subseteq V_A\}$, so the cardinality of the *substrate fud set* $\mathcal{F}_{U_A,V_A}$, would also be infinite without *flattening*. The explicit construction of the *substrate fud set*, $\mathcal{F}_{U_A,V_A}$, in an *inducer* application, such as in the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$, above, requires the computation of the *power fud*, and thus the *flattened partition* of each new *variable* must be computed. However this would imply intractable *underlying volume*. That is, the *space* of a *flattened partition transform* in *substrate fud* F would be at least equal to the *underlying volume*, $|\mathrm{his}(G^{\mathrm{TV}_A})| = v$ where $G = \mathrm{depends}(F,\{u\})$, $u \in \mathrm{vars}(F) \setminus V_A$ and $v = |V_A^{\mathrm{C}}|$, even in the serial case. So, for example, $I'^{\mathrm{s}}_{z,\mathrm{a,F,l}}(A) > v$. The *space* complexity is therefore intractable because it is exponential with respect to *underlying dimension*, n. Even an *inducer* that has, as its subset of the *substrate models*, the *limited-underlying* subset of the *substrate fuds*, $\mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{u}}$, such as the *limited-variables derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l,u,d}}$, above, must compute the *flattened partition transform*, $G^{\mathrm{T}V_A}$, and so must still have intractable *space* complexity, $I'^{\mathrm{s}}_{z,\mathrm{a,F,l,u,d}}(A) > v$. If the definition of the *search set* is relaxed to allow *fuds* containing duplicate *flattened partitions*, and thus doing away with the need to compute the *flattened partition transform*, some other method of limiting the cardinality of *layers* in the *fud* is required to prevent intractable infinite recursion.

Define the *infinite-layer substrate fud set* $\mathcal{F}_{\infty,U,V} \subset \mathcal{F}_{U,\mathrm{P}}$ as

$$\mathcal{F}_{\infty,U,V} = \{F : F \subseteq \mathrm{powinf}(U)(V,\emptyset),\ \mathrm{und}(F) \subseteq V\}$$

where U is the infinite *implied system*, $U = \mathrm{implied}(\mathrm{filter}(V,U))$, and the *infinite power fud* $\mathrm{powinf}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}} \to \mathcal{F}_{U,\mathrm{P}}$ is defined without termination

$$\begin{aligned}\mathrm{powinf}(U)(V,F) := F \cup G \cup \mathrm{powinf}(U)(V, F \cup G) :\\ G = \{T : K \subseteq \mathrm{vars}(F) \cup V,\ T \in F_{U,K}\}\end{aligned}$$

or explicitly,

$$\begin{aligned}\mathrm{powinf}(U)(V,F) := F \cup G \cup \mathrm{powinf}(U)(V, F \cup G) :\\ G = \{P^{\mathrm{T}} : K \subseteq \mathrm{vars}(F) \cup V,\ P \in \mathrm{B}(K^{\mathrm{CS}})\}\end{aligned}$$

The cardinality of the *infinite-layer substrate fud set* is infinite, $|\mathcal{F}_{\infty,U,V}| = \infty$. To implement an *inducer* with a tractable finite subset of the *infinite-layer substrate fud set*, $\mathcal{F}_{\infty,U_A,V_A}$, where $A \in \mathcal{A}_{z,\mathrm{xi}}$, consider a limit on the cardinality of the *layers* l, where $l = \mathrm{layer}(F, \mathrm{der}(F))$ and $F \in \mathcal{F}_{\infty,U_A,V_A}$. For example, a *maximum layer* limit of $\mathrm{lmax} \in \mathbf{N}_{>0}$ applied such that $l \leq \mathrm{lmax}$. Define a *limited-layer* subset of the *functional definition sets* $\mathcal{F}_{\mathrm{h}} \subset \mathcal{F}$ which represents the class of subsets of the *functional definition sets* such that the *layer* of the *fud* is limited. The cardinality of *limited-layer substrate fud set* is finite, $|\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{h}}| < \infty$. The *limited-layer substrate fud set* allows *fuds* containing duplicate *flattened partitions* and so is a superset of the intersection of the *substrate fud set* and the *limited-layer fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{h}} \supseteq \mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{h}}$.

Let the *literal substrate histogram search infinite-layer fuds* $\mathcal{F}_{\infty,\mathrm{fa,j}}(A) \subset \mathcal{F}_{\infty,U_A,V_A}$ be defined

$$\mathcal{F}_{\infty,\mathrm{fa,j}}(A) = \{F : F \in \mathcal{F}_{\infty,U_A,V_A},\ A^{\mathrm{X}} * F^{\mathrm{T}} = (A * F^{\mathrm{T}})^{\mathrm{X}},\ A = A * F^{\mathrm{T}} * F^{\mathrm{T}\dagger A}\}$$

The cardinality of the *literal substrate histogram search infinite-layer fuds* is infinite, $|\mathcal{F}_{\infty,\mathrm{fa,j}}(A)| = \infty$. The *literal substrate histogram search infinite-layer fuds* map to the *literal substrate transforms*, $\{F^{\mathrm{T}V_A} : F \in \mathcal{F}_{\infty,\mathrm{fa,j}}(A)\} = \mathcal{T}_{\mathrm{fa,j}}(A)$, and so is a subset of the *substrate models*, $\mathcal{F}_{\infty,\mathrm{fa,j}}(A) \subset \mathcal{M}_{U_A,V_A}$.

Define the *limited-layer limited-variables derived alignment integral-independent substrate ideal formal-abstract infinite-layer fud inducer* $I'_{z,\mathrm{a,F},\infty,\mathrm{l,u,d,h}} \in \mathrm{inducers}(z)$, which has, as its subset of the *substrate models*, a *limited-layer*, *limited-underlying* and *limited-derived* subset of the *literal substrate histogram search infinite-layer fuds*,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (\mathrm{dom}(I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{l,u,d,h}}(A)) = \mathcal{F}_{\infty,\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u,d,h}} \subset \mathcal{M}_{U_A,V_A})$$

where $\mathcal{F}_{\mathrm{u,d,h}} = \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}} \cap \mathcal{F}_{\mathrm{h}}$. The domain of the *limited-layer limited-variables derived alignment fud inducer* is finite, $|\mathrm{dom}(I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{l,u,d,h}}(A))| = |\mathcal{F}_{\infty,\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{u,d,h}}| < \infty$. The *space* complexity of the serially computed *limited-layer limited-variables derived alignment fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{l,u,d,h}}$, is tractable with respect to *flattening* because the *flattened partition transforms* are not computed.

The *space* required to construct a *substrate fud* depends on the sequence of the computation. If the computation is from the top *layer* downwards and definitions of (i) the *limited-derived*, $\mathcal{F}_{\mathrm{d}}$, is a *maximum derived dimension* limit of $\mathrm{jmax} \in \mathbf{N}_{\geq 2}$, (ii) the *limited-underlying*, $\mathcal{F}_{\mathrm{u}}$, is a *maximum underlying dimension* limit of $\mathrm{kmax} \in \mathbf{N}_{\geq 2}$, and (iii) the *limited-layer*, $\mathcal{F}_{\mathrm{h}}$, is a *maximum layer* limit of $\mathrm{lmax} \in \mathbf{N}_{>0}$, then there exists a *fud* $F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u,d,h}}$ having lmax *layers*, $\mathrm{layer}(F,\mathrm{der}(F)) = \mathrm{lmax}$, and such that the cardinality of the i-th *layer* from the top is

$$|\{u : u \in \mathrm{vars}(F),\ \mathrm{layer}(F,\{u\}) = \mathrm{lmax} - i\}| = \mathrm{jmax} \times \mathrm{kmax}^i$$

where $0 \leq i < \mathrm{lmax}$. If the *partition transforms* of the i-th *layer* of the *fud* are such that the *volume* equals the *maximum underlying volume* limit of $\mathrm{xmax} \in \mathbf{N}_{\geq 4}$, then the *space* of the *limited-variables derived alignment fud inducer* is such that,

$$I'^{\mathrm{s}}_{z,\mathrm{a,F},\infty,\mathrm{l,u,d,h}}(A) > \sum(\mathrm{xmax} \times \mathrm{jmax} \times \mathrm{kmax}^i : i \in \{0 \ldots \mathrm{lmax} - 1\})$$

which is tractable.

Similarly, if the implementation of the computation of the *fud* is from the bottom *layer* upwards with the same definitions of *limited-layer* and *limited-variables*, there exists a *fud* $F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u,d,h}}$ such that the cardinality of the i-th *layer* from the bottom is

$$|\{u : u \in \mathrm{vars}(F),\ \mathrm{layer}(F,\{u\}) = i\}| = n^{\mathrm{kmax}^i}$$

where $n = |V_A|$ and $0 \leq i < \mathrm{lmax}$. The *space* of the *limited-layer limited-variables derived alignment fud inducer* in the upwards sequence computation is such that,

$$I'^{\mathrm{s}}_{z,\mathrm{a,F},\infty,\mathrm{l,u,d,h}}(A) > \sum(\mathrm{xmax} \times n^{\mathrm{kmax}^i} : i \in \{1 \ldots \mathrm{lmax}\})$$

which is tractable because it is only polynomial in *underlying dimension*, n. In the *non-overlapping* case where the *underlying variables* are partitioned, there exists a *fud* F such that the cardinality of the i-th *layer* from the bottom is, with a certain abuse of notation,

$$|\{u : u \in \mathrm{vars}(F),\ \mathrm{layer}(F,\{u\}) = i\}| < n^{\underline{\mathrm{kmax}}^i}$$

Although in this implementation of the *inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{l,u,d,h}}$, the *space* complexity is tractable, the *time* complexity remains intractable with respect to *underlying dimension*, n. The cardinality of *fuds* having *space* as defined in the upwards computation above, and such that the *partition transforms* of the i-th *layer* each have *underlying volume* of xmax is

$$(\mathrm{bell}(\mathrm{xmax}))^{n^{\mathrm{kmax}^i}}$$

Thus the *time* complexity is at least exponential in *underlying dimension*, n,

$$|\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u,d,h}}| > (\mathrm{bell}(\mathrm{xmax}))^n \implies I'^{\mathrm{t}}_{z,\mathrm{a,F},\infty,\mathrm{l,u,d,h}}(A) > (\mathrm{bell}(\mathrm{xmax}))^n$$

To implement an *inducer* with tractable *time* complexity, consider limits on the cardinality of the *variables* in the *layers*. For example, a *maximum layer breadth* limit of $\mathrm{bmax} \in \mathbf{N}_{>0}$ could constrain a *fud* such that $\forall i \in \{1 \ldots l\}\ (|\{u : u \in \mathrm{vars}(F),\ \mathrm{layer}(F,\{u\}) = i\}| \leq \mathrm{bmax})$ where $l = \mathrm{layer}(F, \mathrm{der}(F))$. Define a *limited-breadth* subset of the *functional definition sets* $\mathcal{F}_{\mathrm{b}} \subset \mathcal{F}$ which represents the class of subsets of the *functional definition sets* such that the cardinality of the *variables* in any *layer* is limited.

Then define the *limited-models derived alignment integral-independent substrate ideal formal-abstract infinite-layer fud inducer* $I'_{z,\mathrm{a,F},\infty,\mathrm{l,q}} \in \mathrm{inducers}(z)$, which has, as its subset of the *substrate models*, a *limited-breadth*, *limited-layer*, *limited-underlying* and *limited-derived* subset of the *literal substrate histogram search infinite-layer fuds*,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (\mathrm{dom}(I^{'*}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}(A)) = \mathcal{F}_{\infty,\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{q}} \subset \mathcal{M}_{U_A,V_A})$$

where $\mathcal{F}_{\mathrm{q}} = \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}} \cap \mathcal{F}_{\mathrm{h}} \cap \mathcal{F}_{\mathrm{b}}$. The *substrate models* of the *limited-models derived alignment fud inducer*, $I^{'}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}$, are a subset of the *literal substrate histogram search infinite-layer fuds*, $\mathrm{dom}(I^{'*}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}(A)) \subset \mathcal{F}_{\infty,\mathrm{fa,j}}(A)$, but are not necessarily a subset of the *literal substrate fuds*, $|\mathrm{dom}(I^{'*}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}(A)) \setminus \mathcal{F}_{\mathrm{fa,j}}(A)| \geq 0$. This is because an *infinite-layer fud* $F \in \mathrm{dom}(I^{'*}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}(A))$ may contain duplicate *expanded partitions*. However, the corresponding *substate transforms* are all in the *literal substrate transforms*, $\{F^{\mathrm{T}V_A} : F \in \mathrm{dom}(I^{'*}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}(A))\} \subset \mathcal{T}_{\mathrm{fa,j}}(A)$.

Conjecture that the *limited-models derived alignment fud inducer*, $I^{'}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}$, is positively correlated with the *literal derived alignment transform inducer*, $I^{'}_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I^{'*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I^{'*}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}) \geq 0)$$

but that the correlation is lower than that for the *limited-variables derived alignment fud inducer*, $I^{'}_{z,\mathrm{a,F,l,u,d}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\ &\quad (\mathrm{cov}(z)(\mathrm{maxr} \circ I^{'*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I^{'*}_{z,\mathrm{a,F,l,u,d}}) \geq\\ &\qquad\qquad \mathrm{cov}(z)(\mathrm{maxr} \circ I^{'*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I^{'*}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}))\end{aligned}$$

The *limited-models derived alignment fud inducer*, $I^{'}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}$, has tractable *time* and *space* complexity with respect to the *search set* elements.

Similarly, a *tractable inducer*, with respect to the *search set* domain, may be defined for *idealising summation aligned decompositions*. The *substrate infinite-layer fud decompositions* $\mathcal{D}_{\mathrm{F},\infty,U,V}$ is defined, similarly to the *substrate fud decompositions* $\mathcal{D}_{\mathrm{F},U,V}$, in section 'Substrate structures', above, as

$$\begin{aligned}\mathcal{D}_{\mathrm{F},\infty,U,V} = \{&D : D \in \mathcal{D}_{\mathrm{F,d}},\ \mathrm{fuds}(D) \subseteq \mathcal{F}_{\infty,U,V},\\ &\forall L \in \mathrm{paths}(D)\ (\mathrm{maxr}(\mathrm{count}(\{(F,i) : (i,(\cdot,F)) \in L\})) = 1)\}\end{aligned}$$

Now define the *limited-models content alignment integral-independent substrate idealising summation aligned infinite-layer fud decomposition inducer*

$I'_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}} \in \mathrm{inducers}(z)$ such that the application to a *non-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}} \setminus \{A^{\mathrm{X}}\}$ is the *summation alignment* function of the *limited-models substrate idealising summation aligned fud decompositions*,

$$\begin{aligned}
&I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}(A) = \\
&\quad \{(D, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algnSum}(A,D))) : \\
&\qquad\qquad D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}}),\ D^{\mathrm{D}V_A} \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\}
\end{aligned}$$

Define $I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}(A^{\mathrm{X}}) = \{(D_{\mathrm{F,u}}, 0)\}$ where $D_{\mathrm{F,u}} = \{((\emptyset, \{T_{\mathrm{u}}\}), \emptyset)\}$ and *unary partition transform* $T_{\mathrm{u}} = \{V_A^{\mathrm{VC}}\}^{\mathrm{T}}$.

Conjecture that the *limited-models content alignment infinite-layer fud decomposition inducer*, $I'_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}) \geq 0)$$

but that the correlation is lower than that for the *limited-variables content alignment fud decomposition inducer*, $I'_{z,\mathrm{c,D,F},\Sigma,\mathrm{k,u,d}}$,

$$\begin{aligned}
&\forall z \in \mathbf{N}_{>0} \\
&\quad (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D,F},\Sigma,\mathrm{k,u,d}}) \geq \\
&\qquad\qquad \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}))
\end{aligned}$$

The *limited-models content alignment integral-independent substrate idealising summation aligned infinite-layer fud decomposition inducer* $I'_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}$, has tractable *time* and *space* complexity with respect to the *search set* elements.

4.21.7 Intractable partition variables

Given *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$, the *literal substrate transforms*, $\mathcal{T}_{\mathrm{fa,j}}(A)$, are the *substrate models* of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$. The *literal substrate transforms* are a subset of the *substrate transforms*, $\mathcal{T}_{\mathrm{fa,j}}(A) \subset \mathcal{T}_{U_A,V_A}$. The *substrate transforms* are defined in terms of *partition variables* of the *substrate*, $\mathcal{T}_{U,V} = \{F^{\mathrm{T}} : F \subseteq \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\}\}$. The *encoding space* of a *substrate transform partition variable* is at least $\mathrm{maximum}(\ln \mathrm{bell}(v), v)$ where $v = |V_A^{\mathrm{C}}|$. Other ways of *encoding* the *partition variable* must require greater *space*. For example, a nested *binary map* representation of sets of sets would have *space* complexity of at least $v \ln v$.

Furthermore, the *values* of *partition variables* are *components* of the *partition*, $(P, P) \in U_A$ where $P \in \mathrm{B}(V_A^{\mathrm{CS}})$, and so have *space* complexities equal to the *variables*. Of course, the *values* could be encoded as an index i of the cardinality of the *partition variable*, $1 \leq i \leq |P|$, which would have *space* of $\ln |P|$, but some order $M \in \mathrm{enums}(P)$ would then be required to list the *components*, $\mathrm{flip}(M) \in \mathcal{L}(P)$, and so either the *time* or *space* complexity to compute the index would be just as large.

Similarly, the *literal substrate fuds*, $\mathcal{F}_{\mathrm{fa,j}}(A)$, which are the *substrate models* of the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$, are a subset of the *substrate fuds*, $\mathcal{F}_{\mathrm{fa,j}}(A) \subset \mathcal{F}_{U_A,V_A}$. The *substrate fud set* is a set of subsets of the *power fud*, $\mathcal{F}_{U_A,V_A} \subset \mathrm{P}(\mathrm{power}(U_A)(V_A)) \subset \mathcal{F}_{U_A,\mathrm{P}}$, and so the *substrate fuds* consist of *partition transforms*. The *substrate fud set* contains the *base partition functional definition set*, $F_{U_A,V_A} \in \mathcal{F}_{U_A,V_A}$ where $F_{U,V} = \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\} \in \mathcal{F}_{U,\mathrm{P}}$. The *space* complexity of the *partitions* of the *partition functional definition set*, F_{U_A,V_A}, is, like the *space* complexity of the *substrate transform partition variables*, at least $\ln \mathrm{bell}(v)$. The *partition variables* in *layers* higher than the first are *partitions* of *states* of *partition variables*. For example, a *bivariate partition variable* $R \in \mathrm{B}(\{P, Q\}^{\mathrm{CS}})$ where $P, Q \in \mathrm{B}(V^{\mathrm{CS}})$, or a *partition* of the *self partition*, $R \in \mathrm{B}(\{V^{\mathrm{CS}\{\}}\}^{\mathrm{CS}})$. The *space* of *partition variables* therefore increases exponentially with *layer*.

Therefore both the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$ and the *literal derived alignment fud inducer*, $I'_{z,\mathrm{a,F,l}}$, are intractable because of exponential *space* complexity with respect to *dimension* n, where $n = |V_A|$. In addition, an unlimited *infinite-layer fud inducer* would be intractable because of exponential *space* complexity with respect to *fud layer*. However, the *limited-models infinite-layer fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{l,q}}$, which has, as its subset of the *substrate models*, a *limited-breadth*, *limited-layer*, *limited-underlying* and *limited-derived* subset of the *literal substrate histogram search infinite-layer fuds*, $\mathcal{F}_{\infty,\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{q}}$, is tractable with respect to *partition variables*. That is, the *partition variables* of *limited-models fuds*, $\mathcal{F}_{\mathrm{q}}$, must have no more than polynomial complexity for either *time* or *space*. This is because, as shown in section 'Intractable search set elements', above, (i) the *space* and *time* complexities of a *fud* F, considered separately from its *partition variables*, in the *limited-models substrate fuds*, $F \in \mathcal{F}_{\infty,\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{q}}$, are tractable, and (ii) the complexity of the cardinality of *limited-models substrate fuds*, $|\mathcal{F}_{\infty,\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{q}}|$ is tractable and therefore the *time complexity* of the serial *limited-models fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{l,q}}$, is also tractable. Tractable *fuds* imply tractable *partition variables* because the *partition variables* depend only on the *fuds*.

Note that a first *layer partition variable space* of at least xmax ln xmax, where the *maximum underlying volume* limit is $\mathrm{xmax} \in \mathbf{N}_{\geq 4}$, while tractable, may be impracticable. See section 'Practicable partition variables', below, for consideration of the cardinal number representation of *variables* in monadic *system inducers*.

4.21.8 Intractable literal substrate model inclusion

The computation of the set of *limited-models substrate infinite-layer fuds*, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{q}}$, in the *limited-models derived alignment integral-independent substrate ideal formal-abstract infinite-layer fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{l,q}}$, has tractable *time* and *space* complexity. Given an *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$, the computation of the *derived alignment* for each of the *fuds*, $\{(F, \mathrm{algn}(A*F^{\mathrm{T}})) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}\}$, would also be tractable. In fact, only the *derived alignments* of the *literal* subset of the *limited-models substrate infinite-layer fuds*, $\mathcal{F}_{\infty,\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{q}} \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}$, need be computed in the *inducer*, after testing for inclusion under the constraints of (i) *formal-abstract equality*, $A^{\mathrm{X}} * F^{\mathrm{T}} = (A * F^{\mathrm{T}})^{\mathrm{X}}$, and (ii) *ideality*, $\mathrm{ideal}(A, F^{\mathrm{T}})$. That is, if the computation is parallel, the process is (i) the *limited-models substrate infinite-layer fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}$, are computed, (ii) the inclusion tests are applied to construct the *limited-models literal substrate histogram search infinite-layer fuds*, $\mathcal{F}_{\infty,\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{q}} = \{F : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}},\ A^{\mathrm{X}} * F^{\mathrm{T}} = (A * F^{\mathrm{T}})^{\mathrm{X}},\ A = A * F^{\mathrm{T}} * F^{\mathrm{T}\dagger A}\}$, and (iii) the *derived alignment* is computed for each, $I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}(A) = \{(F, \mathrm{algn}(A * F^{\mathrm{T}})) : F \in \mathcal{F}_{\infty,\mathrm{fa,j}}(A) \cap \mathcal{F}_{\mathrm{q}}\}$. If the computation is serial, the process of (i) construction, (ii) inclusion testing, and (iii) *derived alignment* computation of each *fud* is performed one *fud* at a time.

However, both of these inclusion tests are intractable because of intractable *substrate volume*. The computation of the *independent histogram*, A^{X}, by an *independenter*, $I^{*}_{\mathrm{X}}(A) = A^{\mathrm{X}}$, requires *time* and *space* of at least v, where $v = |V_A^{\mathrm{CS}}|$, because the *substrate histogram*, A, has *completely effective independent*, $A^{\mathrm{XF}} = V_A^{\mathrm{C}}$. Thus the computation of the *independent histogram*, A^{X}, in the computation of the *formal histogram*, $A^{\mathrm{X}} * F^{\mathrm{T}}$, in the *formal-abstract equality* inclusion test, $A^{\mathrm{X}} * F^{\mathrm{T}} = (A * F^{\mathrm{T}})^{\mathrm{X}}$, is intractable with respect to *underlying dimension*, n, where $n = |V_A|$. Similarly, in the case where the *fud*, F, is equivalent to the *unary partition transform*, $F^{\mathrm{TV}_A} = T_{\mathrm{u}}$, where $T_{\mathrm{u}} = \{V_A^{\mathrm{CS}}\}^{\mathrm{T}}$, then the *idealisation* equals the *indepen-*

dent, $A * F^{\mathrm{T}} * F^{\mathrm{T}\dagger A} = A * T_{\mathrm{u}} * T_{\mathrm{u}}^{\dagger A} = A^{\mathrm{X}}$, and so the computation *time* and *space* of the *idealisation* in an *idealiser* $I_{\dagger} \in$ computers must be at least as great as that of the *independent* in the *independenter*, $I_{\dagger}^{\mathrm{s}}((A, T_{\mathrm{u}})) > I_{\mathrm{X}}^{\mathrm{s}}(A)$ and $I_{\dagger}^{\mathrm{t}}((A, T_{\mathrm{u}})) > I_{\mathrm{X}}^{\mathrm{t}}(A)$. Of course, the limits of the *limited-models fuds*, $\mathcal{F}_{\mathrm{q}}$, may exclude the *unary partition transform*, $T_{\mathrm{u}} \notin \{F^{\mathrm{T}V_A} : F \in \mathcal{F}_{\mathrm{q}},\ \mathrm{und}(F) \subseteq V_A\}$, and so the computation of the *space* and *time* required by the *ideality* inclusion test depends on the definition of the limits.

Consider the *formal-abstract equality* inclusion test in the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I_{z,\mathrm{a,l}}'$. Given an *integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}}$, the application of the *inducer* is defined

$$I_{z,\mathrm{a,l}}^{'*}(A) =$$
$$\{(T, I_{\approx\ln\mathbf{Q}}^{*}(\mathrm{algn}(A * T))) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}$$

Now replace the *formal-abstract equality* inclusion test, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, with the less strict *independent-formal* constraint, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, which is implied by *formal-abstract equality*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, to define the *derived alignment substrate ideal independent-formal transform inducer* $I_{z,\mathrm{a,fx,j}}' \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$I_{z,\mathrm{a,fx,j}}^{'*}(A) =$$
$$\{(T, I_{\approx\mathbf{R}}^{*}(\mathrm{algn}(A * T))) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}$$

The *independent-formal inducer*, $I_{z,\mathrm{a,fx,j}}'$, is defined with the *real approxer*, $I_{\approx\mathbf{R}}$, rather than the *log-rational approxer*, $I_{\approx\ln\mathbf{Q}}$, because in some cases the *abstract alignment* is not *integral*, $(A * T)^{\mathrm{X}} \notin \mathcal{A}_{\mathrm{i}}$. Also, there is no longer any need to constrain the domain of the *inducer* to the subset *integral-independent substrate histograms*, $\mathcal{A}_{z,\mathrm{xi}} \subset \mathcal{A}_z$.

The weaker constraint means that the application of the *independent-formal inducer*, $I_{z,\mathrm{a,fx,j}}'$, is a superset of that of the *formal-abstract inducer*, $I_{z,\mathrm{a,l}}'$, $I_{z,\mathrm{a,fx,j}}^{'*}(A) \supseteq I_{z,\mathrm{a,l}}^{'*}(A)$, in the case where the *substrate histogram* has *integral independent*, $A \in \mathcal{A}_{z,\mathrm{xi}}$. The *substrate models* are no longer constrained to be a subset of the *literal substrate models*, so that in some cases $\mathrm{dom}(I_{z,\mathrm{a,fx,j}}^{'*}(A)) \setminus \mathcal{T}_{\mathrm{fa,j}}(A) \neq \emptyset$. Conjecture that the *maximum transform* function of the *independent-formal inducer*, $I_{z,\mathrm{a,fx,j}}'$, is positively correlated with that of the *literal derived alignment inducer*, $I_{z,\mathrm{a,l}}'$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I_{z,\mathrm{a,l}}^{'*}, \mathrm{maxr} \circ I_{z,\mathrm{a,fx,j}}^{'*}) \geq 0)$$

The *independent-formal* inclusion test, $A^{\mathrm{X}}*T = (A^{\mathrm{X}}*T)^{\mathrm{X}}$, may be dropped altogether by altering the range of the application to be *content alignment*, $\mathrm{algn}(A * T) - \mathrm{algn}(A^{\mathrm{X}} * T)$, instead of *derived alignment*. Define the *content alignment substrate ideal transform inducer* $I'_{z,\mathrm{c,j}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$\begin{aligned}&I'^{*}_{z,\mathrm{c,j}}(A) = \\ &\quad \{(T, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * T) - \mathrm{algn}(A^{\mathrm{X}} * T))) : T \in \mathcal{T}_{U_A,V_A},\ A = A * T * T^{\dagger A}\}\end{aligned}$$

This still weaker constraint means that the set of *substrate models* of the *content inducer*, $I'_{z,\mathrm{c,j}}$, is a superset of that of those of the *independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$. That is, $\mathrm{dom}(I'^{*}_{z,\mathrm{c,j}}(A)) \supseteq \mathrm{dom}(I'^{*}_{z,\mathrm{a,fx,j}}(A)) \supseteq \mathrm{dom}(I'^{*}_{z,\mathrm{a,l}}(A))$. So the subset that is disjoint with the *literal substrate models*, $\mathrm{dom}(I'^{*}_{z,\mathrm{c,j}}(A)) \setminus \mathcal{T}_{\mathrm{fa,j}}(A)$, has possibly greater cardinality.

The *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * T)$, is zero when the *formal histogram* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}} \implies \mathrm{algn}(A^{\mathrm{X}} * T) = 0$. The *formal alignment* is always greater than or equal zero, where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies A^{\mathrm{X}} * T \in \mathcal{A}_{\mathrm{i}} \implies \mathrm{algn}(A^{\mathrm{X}} * T) \geq 0$. Conjecture that the *content alignment inducer*, $I'_{z,\mathrm{c,j}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,j}}) \geq 0)$$

but that the correlation is lower than that for the *independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0} \\ &\quad (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,fx,j}}) \geq \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{c,j}}))\end{aligned}$$

This is because in some cases the *maximum transforms*, $\mathrm{maxd}(I'_{z,\mathrm{c,j}}(A)) \subset \mathcal{T}_{U_A,V_A}$, are such that the *formal histogram* is not *independent*. That is, $A^{\mathrm{X}} * T_{\mathrm{c}} \neq (A^{\mathrm{X}} * T_{\mathrm{c}})^{\mathrm{X}}$, where $T_{\mathrm{c}} \in \mathrm{maxd}(I'_{z,\mathrm{c,j}}(A))$.

However, both (i) the *independent-formal* inclusion test, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, in the *independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$, and (ii) the *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * T)$, in the *content alignment inducer*, $I'_{z,\mathrm{c,j}}$, remain intractable because of intractable *substrate volume*. The *independent histogram*, A^{X}, must still be computed by the *independenter*, $I^{*}_{\mathrm{X}}(A) = A^{\mathrm{X}}$, requiring *time* and *space* of at least v, where $v = |V^{\mathrm{CS}}|$. The *independent-formal* constraint in the *derived alignment independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$, may be made more tractable by constraining the domain of the application more strictly to the *non-overlapping transforms*, $\mathcal{T}_{U_A,V_A,\mathrm{n}}$, where $\mathcal{T}_{U,V,\mathrm{n}} = \{T : T \in \mathcal{T}_{U,V},\ \neg\mathrm{overlap}(T)\}$. A *non-overlapping transform* implies that the *formal histogram* is *independent*, $\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$. Define the *derived alignment substrate ideal non-overlapping transform inducer* $I'_{z,\mathrm{a,n,j}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$I'^{*}_{z,\mathrm{a,n,j}}(A) =$$
$$\{(T, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * T))) : T \in \mathcal{T}_{U_A,V_A,\mathrm{n}},\ A = A * T * T^{\dagger A}\}$$

The set of *substrate models* of the *derived alignment non-overlapping inducer*, $I'_{z,\mathrm{a,n,j}}$, is a subset of that of the *derived alignment independent-formal inducer*, $\mathrm{dom}(I'^{*}_{z,\mathrm{a,n,j}}(A)) = \mathrm{dom}(I'^{*}_{z,\mathrm{a,fx,j}}(A)) \cap \mathcal{T}_{U_A,V_A,\mathrm{n}}$. The set of *substrate models* is neither a superset nor a subset of the set of the *literal substrate models*, $|\mathrm{dom}(I'^{*}_{z,\mathrm{a,n,j}}(A)) \setminus \mathcal{T}_{\mathrm{fa,j}}(A)| \geq 0$ and $|\mathcal{T}_{\mathrm{fa,j}}(A) \setminus \mathrm{dom}(I'^{*}_{z,\mathrm{a,n,j}}(A))| \geq 0$. This is because *independent-formal transforms* are not necessarily *non-overlapping*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}} \impliedby \neg\mathrm{overlap}(T)$, and therefore *transforms* that are subject to *formal-abstract equality*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, are not necessarily *non-overlapping*. However, the intersection with the *literal substrate models* is not empty, $\mathrm{dom}(I'^{*}_{z,\mathrm{a,n,j}}(A)) \cap \mathcal{T}_{\mathrm{fa,j}}(A) \neq \emptyset$.

Conjecture that the *derived alignment non-overlapping inducer*, $I'_{z,\mathrm{a,n,j}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,n,j}}) \geq 0)$$

but that the correlation is lower than that for the *derived alignment independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$,

$$\forall z \in \mathbf{N}_{>0}$$
$$(\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,fx,j}}) \geq \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,n,j}}))$$

because the intersection with the *literal substrate models* is sometimes smaller, $\mathrm{dom}(I'^{*}_{z,\mathrm{a,n,j}}(A)) \cap \mathcal{T}_{\mathrm{fa,j}}(A) \subseteq \mathrm{dom}(I'^{*}_{z,\mathrm{a,fx,j}}(A)) \cap \mathcal{T}_{\mathrm{fa,j}}(A)$.

In the *derived alignment non-overlapping inducer*, $I'_{z,\mathrm{a,n,j}}$, the inclusion test for *independent-formal*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, is replaced by a test for *non-overlapping transform*, $\neg\mathrm{overlap}(T)$. However, determining whether a *substrate transform* is *non-overlapping* or not requires *contracting* each of the *derived variables*, and then checking to see if the *contracted partition transforms* are disjoint, $\{\mathrm{vars}(P^{\%}) : w \in \mathrm{der}(T),\ P = (\mathrm{his}(T)\%(\mathrm{und}(T) \cup$

$\{w\}), \{w\})^{\mathrm{P}}\} \in \mathrm{B}(\mathrm{und}(T))$. The representation *space* of the *transform*, T, in the *inducer* must be at least as large as the *underlying volume*, $I'^{\mathrm{s}}_{z,\mathrm{a,n,j}}(A) > v$, and is therefore intractable. In section 'Intractable substrate volume', above, a similar intractability is addressed in the application of the *transform* to the *histogram* by the *transformer*, $I^{\mathrm{s}}_{*\mathrm{T}}((T,A)) > v$. Define the *derived alignment substrate ideal non-overlapping fud inducer* $I'_{z,\mathrm{a,F,n,j}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$I'^{*}_{z,\mathrm{a,F,n,j}}(A) = \{(F, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * F^{\mathrm{T}}))) : F \in \mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{n}},\ A = A * F^{\mathrm{T}} * F^{\mathrm{T}\dagger A}\}$$

where $\mathcal{F}_{\mathrm{n}} := \{F : F \in \mathcal{F},\ \neg\mathrm{overlap}(F)\}$.

A *fud* can be tractably tested for *overlap* by following the *depends* tree of the *partition transforms*, $\neg\mathrm{overlap}(F) \iff \{\mathrm{und}(\mathrm{depends}(F,\{w\})) : w \in \mathrm{der}(F)\} \in \mathrm{B}(\mathrm{und}(F))$.

The application of the *derived alignment ideal non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,n,j}}$, maps to that of the *derived alignment ideal non-overlapping inducer*, $I'_{z,\mathrm{a,n,j}}$, above, $\{(F^{\mathrm{T}V_A}, a) : (F,a) \in I'^{*}_{z,\mathrm{a,F,n,j}}(A)\} = I'^{*}_{z,\mathrm{a,n,j}}(A)$, so the correlations to the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, are equal

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,n,j}}) = \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,n,j}}))$$

Now constrain the *fud inducer* by *limited models*. Define the *limited-models derived alignment substrate ideal non-overlapping infinite-layer fud inducer* $I'_{z,\mathrm{a,F,\infty,n,q,j}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$I'^{*}_{z,\mathrm{a,F,\infty,n,q,j}}(A) = \{(F, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * F^{\mathrm{T}}))) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}},\ A = A * F^{\mathrm{T}} * F^{\mathrm{T}\dagger A}\}$$

Similarly to the *derived alignment ideal non-overlapping inducer*, $I'_{z,\mathrm{a,n,j}}$, above, the set of *substrate transforms* corresponding to the *substrate models* of the *limited-models derived alignment ideal non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q,j}}$, is neither a superset nor a subset of the set of the *literal substrate transforms*, $|\{F^{\mathrm{T}V_A} : F \in \mathrm{dom}(I'^{*}_{z,\mathrm{a,F,\infty,n,q,j}}(A))\} \setminus \mathcal{T}_{\mathrm{fa,j}}(A)| \geq 0$ and $|\mathcal{T}_{\mathrm{fa,j}}(A) \setminus \{F^{\mathrm{T}V_A} : F \in \mathrm{dom}(I'^{*}_{z,\mathrm{a,F,\infty,n,q,j}}(A))\}| \geq 0$. Therefore conjecture that the *limited-models derived alignment ideal non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q,j}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,\infty,n,q,j}}) \geq 0)$$

but that the correlation is lower than that for the *limited-models derived alignment ideal formal-abstract fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{l,q}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\&\quad(\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{l,q}}) \geq\\&\qquad\qquad \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{n,q,j}}))\end{aligned}$$

As shown in section 'Transform alignment', above, in the case of a *histogram-transform* pair $(A,T) \in \mathcal{O}_{U,z+y}$, where the *histogram*, A, is the *sum* of a *diagonal histogram* of *size* z and a *cartesian histogram* of *size* y, the *idealisation alignment* is approximately equal to the *derived alignment*, $\mathrm{algn}(A*T*T^{\dagger A}) \approx \mathrm{algn}(A*T)$. It is also conjectured that the *idealisation alignment* is always less than or equal to the *alignment* of the *histogram*, where the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies \mathrm{algn}(A*T*T^{\dagger A}) \leq \mathrm{algn}(A)$. The *histogram alignment*, $\mathrm{algn}(A)$, is constant, so at the maximum *idealisation alignment* the *transform* is *ideal*, $\mathrm{algn}(A*T*T^{\dagger A}) = \mathrm{algn}(A) \implies A \equiv A*T*T^{\dagger A}$. Therefore conjecture that the maximisation of the *derived alignment*, $\mathrm{algn}(A*T)$, weakly maximises the *idealisation alignment*, $\mathrm{algn}(A*T*T^{\dagger A})$, *idealising* the *transform*, $\mathrm{ideal}(A,T)$. In order to make an *inducer* computation tractable, the *ideality* inclusion test must be removed. Dropping the *ideality* inclusion test, $A = A*F^{\mathrm{T}}*F^{\mathrm{T}\dagger A}$, define the *limited-models derived alignment substrate non-overlapping infinite-layer fud inducer* $I'_{z,\mathrm{a,F},\infty,\mathrm{n,q}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$\begin{aligned}&I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{n,q}}(A) =\\&\quad\{(F, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A*F^{\mathrm{T}}))) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}\}\end{aligned}$$

Conjecture that the *derived alignment non-overlapping fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{n,q}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{n,q}}) \geq 0)$$

but that the correlation is lower than that for the *derived alignment ideal non-overlapping fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{n,q,j}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\&\quad(\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{n,q,j}}) \geq\\&\qquad\qquad \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{n,q}}))\end{aligned}$$

because the intersection of the set of *substrate transforms* corresponding to the *substrate models* of the *limited-models derived alignment non-overlapping* *fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q}}$, with the *literal substrate transforms*, $\{F^{\mathrm{T}V_A} : F \in \mathrm{dom}(I'^{*}_{z,\mathrm{a,F,\infty,n,q}}(A))\} \cap \mathcal{T}_{\mathrm{fa,j}}$, equals that of the *derived alignment ideal non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q,j}}$, but the set of *substrate models* is a superset, $I'^{*}_{z,\mathrm{a,F,\infty,n,q}}(A) \supseteq I'^{*}_{z,\mathrm{a,F,\infty,n,q,j}}(A)$. That is, the cardinality of the set of *non-literal substrate models*, $\{F : F \in \mathrm{dom}(I'^{*}_{z,\mathrm{a,F,\infty,n,q}}(A)),\ F^{\mathrm{T}V_A} \notin \mathcal{T}_{\mathrm{fa,j}}(A)\}$, is at least that of the *ideal inducer*.

The *derived alignment non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q}}$, is tractable in all respects.

Although the *derived alignment non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q}}$, is tractable, the *non-overlapping* constraint, $\neg\mathrm{overlap}(F)$, is weaker than a *formal-abstract equality* inclusion test of the *fud*, $A^{\mathrm{X}} * F^{\mathrm{T}} = (A * F^{\mathrm{T}})^{\mathrm{X}}$. To see how the *formal-abstract equality* condition might be adhered to more strictly, consider the *abstract-non-formal entropy substrate ideal independent-formal transform inducer* $I'_{z,\mathrm{e,fx,j}} \in \mathrm{inducers}(z)$. Given a *substrate histogram* $A \in \mathcal{A}_z$, the *abstract-non-formal entropy inducer* is defined as

$$\begin{aligned}
&I'^{*}_{z,\mathrm{e,fx,j}}(A) = \\
&\quad \{(T, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{entropy}((A * T)^{\mathrm{X}}) - \mathrm{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}}))) : \\
&\qquad\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}
\end{aligned}$$

The *abstract-non-formal entropy inducer*, $I'_{z,\mathrm{e,fx,j}}$, relaxes the *formal-abstract equality* inclusion test to constrain the *search transforms* to those that are such that the *formal histogram* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, which is implied by *formal-abstract equality*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$. The set of *substrate models* of the *abstract-non-formal entropy inducer*, $I'_{z,\mathrm{e,fx,j}}$, is the same as those of the *independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$, and therefore a superset of the *literal substrate models*, $\mathcal{T}_{\mathrm{fa,j}}(A)$. That is, $\mathrm{dom}(I'^{*}_{z,\mathrm{e,fx,j}}(A)) = \mathrm{dom}(I'^{*}_{z,\mathrm{a,fx,j}}(A)) \supseteq \mathrm{dom}(I'^{*}_{z,\mathrm{a,l}}(A))$. To compensate, the range of the application is the difference in the *entropy* of the *abstract histogram* and the *entropy* of the *formal histogram*, $\mathrm{entropy}((A * T)^{\mathrm{X}}) - \mathrm{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}})$. Although both the *formal histogram*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, and the *abstract histogram*, $(A * T)^{\mathrm{X}}$, are *independent*, it is conjectured that the *abstract histogram* tends to have lower *entropy* than the doubly *independent formal histogram*

$$\begin{aligned}
&\mathrm{average}(\{(T, \mathrm{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}}) - \mathrm{entropy}((A * T)^{\mathrm{X}})) : \\
&\qquad\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}) \geq 0
\end{aligned}$$

In particular, the *formal histogram* is *completely effective*, $(A^{\mathrm{XF}} * T)^{\mathrm{XF}} = (V_A^{\mathrm{C}} * T)^{\mathrm{XF}} = W^{\mathrm{C}}$, where $W = \mathrm{der}(T)$, whereas the *abstract histogram* is not necessarily *completely effective*, $(A * T)^{\mathrm{XF}} \leq W^{\mathrm{C}}$. For example, the *abstract histogram*, $(A * T)^{\mathrm{X}}$, may be a *cartesian sub-volume*.

Overall, the maximisation, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{e,fx,j}} \in \mathcal{A}_z \to \mathbf{Q}$, tends to equalise the *formal histogram*, $A^{\mathrm{X}} * T$ and the *abstract histogram*, $(A * T)^{\mathrm{X}}$.

The *abstract-non-formal entropy inducer*, $I'_{z,\mathrm{e,fx,j}}$, is properly considered to be an *inducer* because it is conjectured to obey the constraint on *inducers* that the maximum of the *inducer* application, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{e,fx,j}}$, is positively correlated with the finite *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$. That is,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{e,fx,j}}) \geq 0)$$

The *abstract-non-formal entropy inducer*, $I'_{z,\mathrm{e,fx,j}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{e,fx,j}}) \geq 0)$$

because, as shown in section 'Minimum alignment', above, the *derived alignment* approximates to the scaled difference between the *entropy* of the *abstract histogram* and the *entropy* of the *derived histogram*,

$$\mathrm{algn}(A * T) \quad \approx \quad z \times \mathrm{entropy}((A * T)^{\mathrm{X}}) - z \times \mathrm{entropy}(A * T)$$

The *abstract-non-formal entropy inducer*, $I'_{z,\mathrm{e,fx,j}}$, tends to maximise the first term, $\mathrm{entropy}((A * T)^{\mathrm{X}})$, relative, at least, to the *formal independent histogram entropy*, $\mathrm{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}})$, thereby weakly maximising the *derived alignment*, $\mathrm{algn}(A*T)$. The positive correlation with the *literal derived alignment inducer maximum function*, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}$, implies, transitively, a positive correlation with the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$.

A variation that more directly maximises the *derived alignment* is to replace the *sized abstract entropy* with the *derived alignment*,

$$\begin{aligned} & \mathrm{algn}(A * T) - z \times \mathrm{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \\ \approx \quad & \mathrm{algn}(A * T) - \mathrm{algn}(A^{\mathrm{X}} * T) + z \times \mathrm{entropy}(A^{\mathrm{X}} * T) \end{aligned}$$

That is, the maximisation of the *content alignment*, $\mathrm{algn}(A*T) - \mathrm{algn}(A^{\mathrm{X}}*T)$, plus the *sized formal entropy*, $z \times \mathrm{entropy}(A^{\mathrm{X}} * T)$.

This attempt to strengthen the *formal-abstract equality* may be taken a step further by considering the *actualisations*. If the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, and the *derived histogram* is *completely effective*, $(A * T)^{\mathrm{F}} = W^{\mathrm{C}}$, then the *contentisation* equals the *surrealisation*, $A^{\mathrm{X}} * T * T^{\odot A} = (A * T)^{\mathrm{X}} * T^{\odot A}$. Conjecture that if the inclusion tests are relaxed to constrain the *search transforms* to those that are such that the *formal histogram* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, which is implied by *formal-abstract equality*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, then the expected difference in the *alignments* of the *contentisation* and the *surrealisation* is negative,

$$\begin{aligned}\mathrm{average}(\{(T, \mathrm{algn}((A^{\mathrm{X}} * T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})) :\\ T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}) \leq 0\end{aligned}$$

This is because the *abstract histogram*, $(A * T)^{\mathrm{X}}$, tends to have lower *entropy* than the doubly *independent formal histogram*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, as noted above. The lower *entropy* of the *abstract histogram* in general means higher *alignment* of the *actualisation*. Thus the *surrealisation alignment* tends to be higher than the *contentisation alignment*.

Now weaken the inclusion testing by replacing the *formal-abstract equality* inclusion test, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, with the less strict *independent-formal* constraint, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, but compensate by altering the range of the application to be the difference in the *alignments* of the *contentisation* and the *surrealisation*, $\mathrm{algn}(A^{\mathrm{X}} * T * T^{\odot A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})$, to define the *contentised non-surrealised alignment substrate ideal independent-formal transform inducer* $I_{z,\mathrm{g,fx,j}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$\begin{aligned}&I^{*}_{z,\mathrm{g,fx,j}}(A) =\\ &\quad\{(T, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A^{\mathrm{X}} * T * T^{\odot A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}))) :\\ &\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}\end{aligned}$$

Note that the *contentisation* is not necessarily *size-conserving*. It is only *size-conserving* if the *derived histogram* is as *effective* as the *formal histogram*, $(A * T)^{\mathrm{F}} \geq (A^{\mathrm{X}} * T)^{\mathrm{F}}$, which requires that the *derived histogram* be *completely effective*, $(A * T)^{\mathrm{F}} = W^{\mathrm{C}}$. The *formal histogram* is *completely effective* because the *independent* is *completely effective* and *formal histogram* is *independent*, $(A^{\mathrm{XF}} = V_A^{\mathrm{C}}) \wedge (A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}) \implies (A^{\mathrm{X}} * T)^{\mathrm{F}} = (A^{\mathrm{XF}} * T)^{\mathrm{XF}} = (V_A^{\mathrm{C}} * T)^{\mathrm{XF}} = W^{\mathrm{C}}$. That is, the *contentisation* is *size-conserving* if the *derived histogram* is *completely effective*, $(A * T)^{\mathrm{F}} =$

$W^{\mathrm{C}} \implies \mathrm{size}(A^{\mathrm{X}} * T * T^{\odot A}) = z$. Similarly, the *surrealisation* is only *size-conserving* if the *derived histogram* is as *effective* as the *abstract histogram*, $(A * T)^{\mathrm{F}} \geq (A * T)^{\mathrm{XF}} \implies \mathrm{size}((A * T)^{\mathrm{X}} * T^{\odot A}) = z$. However, the *formal histogram* is at least as *effective* as the *abstract histogram*, $(A^{\mathrm{X}} * T)^{\mathrm{F}} = W^{\mathrm{C}} \geq (A * T)^{\mathrm{XF}}$. So in the cases where the *abstract histogram* is not *completely effective*, the *contentisation alignment* is sometimes less than the *surrealisation alignment* because the *contentisation size* is less than the *surrealisation size*.

Note, also, that the range of the application, $\mathrm{ran}(I^*_{z,\mathrm{g,fx,j}}(A))$, is not *derived* or *lifted*, unlike that, for example, of the *abstract-non-formal entropy inducer*, $I'_{z,\mathrm{e,fx,j}}$, so the *contentised non-surrealised derived alignment inducer*, $I_{z,\mathrm{g,fx,j}}$, is denoted without the prime embellishment.

As in the case of *substrate models* of the *abstract-non-formal entropy inducer*, $I'_{z,\mathrm{e,fx,j}}$, above, the set of *substrate models* of the *contentised non-surrealised derived alignment inducer*, $I_{z,\mathrm{g,fx,j}}$, is the same as those of the *independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$, and therefore a superset of the *literal substrate models*, $\mathcal{T}_{\mathrm{fa,j}}(A)$. That is, $\mathrm{dom}(I^*_{z,\mathrm{g,fx,j}}(A)) = \mathrm{dom}(I'^*_{z,\mathrm{a,fx,j}}(A)) \supseteq \mathrm{dom}(I'^*_{z,\mathrm{a,l}}(A))$. Conjecture that the *maximum transform* function of the *contentised non-surrealised derived alignment inducer*, $I_{z,\mathrm{g,fx,j}}$, is positively correlated with that of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^*_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I^*_{z,\mathrm{g,fx,j}}) \geq 0)$$

Overall, the maximisation, $\mathrm{maxr} \circ I^*_{z,\mathrm{g,fx,j}} \in \mathcal{A}_z \to \mathbf{Q}$, tends to equalise the *formal histogram*, $A^{\mathrm{X}} * T$ and the *abstract histogram*, $(A * T)^{\mathrm{X}}$, while still weakly maximising the *derived alignment*, $\mathrm{algn}(A * T)$, as in the *abstract-non-formal entropy inducer*, $I'_{z,\mathrm{e,fx,j}}$, above.

In section 'Transform alignment', above, it is conjectured that the *midisation alignment* varies with (a) the difference between the *alignments* of the *contentisation* and the *surrealisation*, and (b) the *midisation pseudo-alignment*,

$$\begin{aligned}
\mathrm{algn}(A^{\mathrm{M}(T)}) &\sim \mathrm{algn}(A^{\mathrm{X}} * T * T^{\odot A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) \\
&\sim \mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\odot A \dagger A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})
\end{aligned}$$

Define the *midisation pseudo-alignment substrate ideal independent-formal transform inducer* $I_{z,\mathrm{m,fx,j}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$I^*_{z,\mathrm{m,fx,j}}(A) = \\ \{(T, I^*_{\approx\mathbf{R}}(\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}))) : \\ T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}$$

Note that the *midisation pseudo-alignment inducer*, $I_{z,\mathrm{m,fx,j}}$, need not compute the *histogram alignment*, $\mathrm{algn}(A)$, because it is constant with regard to maximisation.

As in the case of *substrate models* of the *abstract-non-formal entropy inducer*, $I_{z,\mathrm{e,fx,j}}$, and the *contentised non-surrealised derived alignment inducer*, $I_{z,\mathrm{g,fx,j}}$, above, the set of *substrate models* of the *midisation pseudo-alignment ideal inducer*, $I_{z,\mathrm{m,fx,j}}$, is the same as those of the *independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$, and therefore a superset of the *literal substrate models*, $\mathcal{T}_{\mathrm{fa,j}}(A)$. That is, $\mathrm{dom}(I^*_{z,\mathrm{m,fx,j}}(A)) = \mathrm{dom}(I'^*_{z,\mathrm{a,fx,j}}(A)) \supseteq \mathrm{dom}(I'^*_{z,\mathrm{a,l}}(A))$. Conjecture that the *maximum transform* function of the *midisation ideal alignment inducer*, $I_{z,\mathrm{m,fx,j}}$, is positively correlated with that of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^*_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I^*_{z,\mathrm{m,fx,j}}) \geq 0)$$

However, the *ideality* inclusion test, $A = A*T*T^{\dagger A}$, implies that the *idealisation alignment* equals the *histogram alignment*, $\mathrm{algn}(A * T * T^{\dagger A}) = \mathrm{algn}(A)$, so in the case of *ideal transform*, the *midisation pseudo-alignment* equals the negative *surrealisation alignment*, $-\mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})$. In order to (i) improve the degree to which maximisation of *midisation pseudo-alignment* corresponds to the *formal-abstract equality* constraint and (ii) make the computation tractable, drop the *ideality* inclusion test, which constrains the *midisation pseudo-alignment*, defining the *midisation pseudo-alignment substrate independent-formal transform inducer* $I_{z,\mathrm{m,fx}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$I^*_{z,\mathrm{m,fx}}(A) = \\ \{(T, I^*_{\approx\mathbf{R}}(\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}))) : \\ T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}$$

This weaker constraint means that the set of *substrate models* of the *midisation pseudo-alignment inducer*, $I_{z,\mathrm{m,fx}}$, is a superset of those of the *midisation pseudo-alignment ideal inducer*, $I_{z,\mathrm{m,fx,j}}$, which equals those of the *independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$. That is, $\mathrm{dom}(I^*_{z,\mathrm{m,fx}}(A)) \supseteq \mathrm{dom}(I^*_{z,\mathrm{m,fx,j}}(A)) =$

$\mathrm{dom}(I'^{*}_{z,\mathrm{a,fx,j}}(A)) \supseteq \mathrm{dom}(I'^{*}_{z,\mathrm{a,l}}(A))$. So the subset that is disjoint with the *literal substrate models*, $\mathrm{dom}(I^{*}_{z,\mathrm{m,fx}}(A)) \setminus \mathcal{T}_{\mathrm{fa,j}}(A)$, has possibly greater cardinality. Conjecture that the *maximum transform* function of the *midisation pseudo-alignment inducer*, $I_{z,\mathrm{m,fx}}$, is positively correlated with that of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I^{*}_{z,\mathrm{m,fx}}) \geq 0)$$

Although the *non-literal substrate models*, $\mathrm{dom}(I^{*}_{z,\mathrm{m,fx}}(A)) \setminus \mathcal{T}_{\mathrm{fa,j}}(A)$, may have greater cardinality, it is not obvious whether or not the *midisation pseudo-alignment inducer*, $I_{z,\mathrm{m,fx}}$, has a lower correlation with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, than the *midisation ideal alignment inducer*, $I_{z,\mathrm{m,fx,j}}$. The cost of dropping the *ideality* inclusion test may be outweighed by the closer approximation to the *formal-abstract equality* inclusion test, in some cases.

In section 'Likely histograms', it is conjectured that there exists an intermediate *mid substrate transform* $T_{\mathrm{m}} \in \mathcal{T}_{U_A,V_A}$ which is neither *self* nor *unary*, $T_{\mathrm{m}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$, where the *formal* is *independent* and the *midisation entropy* is minimised,

$$T_{\mathrm{m}} \in \mathrm{mind}(\{(T, \mathrm{entropy}(A^{\mathrm{M}(T)})) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\})$$

Section 'Transform alignment', goes on to conjecture that an approximation to the *mid transform* may also be obtained by a maximisation of the *midisation pseudo-alignment*,

$$\begin{aligned} T_{\mathrm{m}} \ \in\ & \mathrm{maxd}(\{(T, \mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})) : \\ & \qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}) \end{aligned}$$

With the removal of the *ideality* inclusion test, the *maximum transform* function of the *midisation pseudo-alignment inducer*, $I_{z,\mathrm{m,fx}}$, is the *mid transform*, $T_{\mathrm{m}} \in \mathrm{maxd}(I^{*}_{z,\mathrm{m,fx}}(A))$. At the *mid transform* the *formal* tends to the *abstract*, $A^{\mathrm{X}} * T_{\mathrm{m}} \approx (A * T_{\mathrm{m}})^{\mathrm{X}}$, and the *mid component size cardinality relative entropy* is small,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V_A^{\mathrm{C}} * T_{\mathrm{m}}) \ \approx\ 0$$

The computation of the *midisation pseudo-alignment*, $\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A})$, requires at least the computation of the *idealisation*, $A * T * T^{\dagger A}$, and the *surrealisation*, $(A * T)^{\mathrm{X}} * T^{\odot A}$, at least one of which is intractable. Consider replacing *midisation pseudo-alignment* with *derived alignment valency density*. Section 'Transform alignment', above, describes the properties of *midisation*. Maximisation of *midisation* tends to move *component alignments* from *off-diagonal states* to *on-diagonal states*, balancing the high *derived alignment* of longer *diagonals* with the high *on-diagonal component alignments* of shorter *diagonals*. Thus the *midisation pseudo-alignment* varies with the *derived alignment valency density*,

$$\begin{aligned}\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A}) \sim \\ \mathrm{algn}(A * T)/\mathrm{capacityValency}(U)((A * T)^{\mathrm{FS}})\end{aligned}$$

where the *valency capacity*, $\mathrm{capacityValency}(U) \in \mathrm{capacities}$, is defined in terms of *geometry* as $\mathrm{capacityValency}(U)((A * T)^{\mathrm{FS}}) = w^{1/m}$, and $m = |W|$, $w = |W^{\mathrm{C}}|$ and $W = \mathrm{der}(T)$.

Define the *derived alignment valency-density substrate independent-formal transform inducer* $I'_{z,\mathrm{ad,fx}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$\begin{aligned}&I'^{*}_{z,\mathrm{ad,fx}}(A) = \\ &\quad \{(T, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * T)/w^{1/m})) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}\end{aligned}$$

where $m = |W|$, $w = |W^{\mathrm{C}}|$ and $W = \mathrm{der}(T)$.

The set of *substrate models* of the *derived alignment valency-density inducer*, $I'_{z,\mathrm{ad,fx}}$, equals those of the *midisation pseudo-alignment inducer*, $I_{z,\mathrm{m,fx}}$. That is, $\mathrm{dom}(I'^{*}_{z,\mathrm{ad,fx}}(A)) = \mathrm{dom}(I^{*}_{z,\mathrm{m,fx}}(A)) \supseteq \mathrm{dom}(I'^{*}_{z,\mathrm{a,fx,j}}(A)) \supseteq \mathrm{dom}(I'^{*}_{z,\mathrm{a,l}}(A))$. Conjecture that the *maximum transform* function of the *derived alignment valency-density inducer*, $I'_{z,\mathrm{ad,fx}}$, is positively correlated with that of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,fx}}) \geq 0)$$

In section 'Derived alignment and conditional probability', above, the *alignment-bounded lifted iso-transform error* is defined as the difference between the *alignment-bounded lifted iso-transform space* and the *derived alignment*

$$\ln \sum_{B' \in \mathcal{A}'_{U,\mathrm{i,y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A * T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}$$

The *alignment-bounded lifted iso-transform error ratio* is the *error* per *derived alignment*. Given *substrate histogram* $A \in \mathcal{A}_z$, let $\mathrm{erra}(A) \in \mathrm{P}(\mathcal{T}_{U_A,V_A}) \to$

$(\mathcal{T}_{U_A,V_A} \to \mathbf{R})$ be defined as the *alignment-bounded lifted iso-transform error ratios* of a set of *substrate transforms*,

$$\begin{aligned}
&\mathrm{erra}(A)(Q) := \\
&\quad \{(T, \left(\ln \sum_{B' \in \mathcal{A}'_{U_A,\mathrm{i,y},T,z}(A)} \frac{\prod_{R \in (A*T)^{\mathrm{XS}}} (A*T)^{\mathrm{X}}_R!}{\prod_{R \in B'^{\mathrm{S}}} B'_R!}\right) / \mathrm{algn}(A*T)) : \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad T \in Q,\ \mathrm{algn}(A*T) \neq 0\} \cup \\
&\quad \{(T, 0) : T \in Q,\ \mathrm{algn}(A*T) = 0\}
\end{aligned}$$

In that section it was shown that the *error ratio* varies as $\overline{w} \ln z / \mathrm{algn}(A*T)$ where the *size* is greater than the *derived volume*, $z > w$. This may be compared to the *derived alignment valency-density*, $\mathrm{algn}(A*T)/w^{1/m}$. Thus the *error ratio* varies against the *derived alignment valency-density*. Higher *capacities* such as *volume capacity*, $\mathrm{capacityVolume}(U)((A*T)^{\mathrm{FS}}) := w$, would vary inversely even more closely. In that case, the *derived alignment volume-density* would be $\mathrm{algn}(A*T)/w$. Therefore conjecture that the expected *error ratio* of the *derived alignment valency-density inducer*, $I'_{z,\mathrm{ad,fx}}$, is less than that of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\begin{aligned}
\forall z \in \mathbf{N}_{>0}\ (\mathrm{ex}(z)(\mathrm{maxr} \circ \mathrm{erra} \circ \mathrm{maxd} \circ I'^{*}_{z,\mathrm{a,l}}) \geq \\
\mathrm{ex}(z)(\mathrm{maxr} \circ \mathrm{erra} \circ \mathrm{maxd} \circ I'^{*}_{z,\mathrm{ad,fx}}))
\end{aligned}$$

Therefore conjecture that the *valency capacity* tends to increase the correlation of the *maximum function* of the *derived alignment valency-density inducer*, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,fx}}$, to the *alignment-bounded lifted iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,y,fa,j}}$, and thence transitively to the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$. In other words, the *derived alignment valency-density inducer*, $I'_{z,\mathrm{ad,fx}}$, has a higher *inducer correlation*, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,fx}})$, than might be expected, because maximisation of the *derived alignment valency-density*, $\mathrm{algn}(A*T)/w^{1/m}$, tends to shorten the *diagonals*, $w^{1/m}$, and reduce the *derived volume*, w, minimising the *alignment-bounded lifted iso-transform error ratio*. Note that the correlation is improved even though the overall *derived alignments* of the *valency-density inducer* are lower than the *literal inducer*. The correlation increases as the *size* exceeds the *derived volume*, $z > w$, because of the decreasing *error ratio* between the *expected alignment*, $\overline{w} \ln z / w \approx 1$, and the *maximum alignment*, $\overline{w} \ln z / z \ln w < 1$.

As is the case for the *derived alignment independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$, above, the *independent-formal* inclusion test, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, in the *valency-density inducer*, $I'_{z,\mathrm{ad,fx}}$, is intractable because of intractable *substrate volume*. Therefore replace the *independent-formal* inclusion test with the *non-overlapping transform* constraint, $\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$. Define the *derived alignment valency-density substrate non-overlapping transform inducer* $I'_{z,\mathrm{ad,n}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$\begin{aligned} &I'^{*}_{z,\mathrm{ad,n}}(A) = \\ &\quad \{(T, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * T)/w^{1/m})) : T \in \mathcal{T}_{U_A,V_A,\mathrm{n}}\} \end{aligned}$$

where $m = |W|$, $w = |W^{\mathrm{C}}|$ and $W = \mathrm{der}(T)$.

Conjecture that the *maximum transform* function of the *derived alignment non-overlapping valency-density inducer*, $I'_{z,\mathrm{ad,n}}$, is positively correlated with that of the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,n}}) \geq 0)$$

but that the correlation is lower than that for the *derived alignment independent-formal valency-density inducer*, $I'_{z,\mathrm{ad,fx}}$,

$$\begin{aligned} &\forall z \in \mathbf{N}_{>0} \\ &\quad (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,fx}}) \geq \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,n}})) \end{aligned}$$

because the intersection with the *literal substrate models* is sometimes smaller, $\mathrm{dom}(I'^{*}_{z,\mathrm{ad,n}}(A)) \cap \mathcal{T}_{\mathrm{fa,j}}(A) \subseteq \mathrm{dom}(I'^{*}_{z,\mathrm{ad,fx}}(A)) \cap \mathcal{T}_{\mathrm{fa,j}}(A)$.

Again, as is the case for the *derived alignment non-overlapping inducer*, $I'_{z,\mathrm{a,n,j}}$, above, determining whether a *substrate transform* is *non-overlapping* or not remains intractable. Also, the *limited-models* constraints are required for tractability. Define the *limited-models derived alignment valency-density substrate non-overlapping infinite-layer fud inducer* $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, as

$$\begin{aligned} &I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}(A) = \\ &\quad \{(F, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * F^{\mathrm{T}})/w^{1/m})) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}\} \end{aligned}$$

where $m = |W|$, $w = |W^{\mathrm{C}}|$ and $W = \mathrm{der}(F)$.

Conjecture that the *derived alignment valency-density non-overlapping fud* *inducer*, $I'_{z,\mathrm{ad,F,\infty,n,q}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F,\infty,n,q}}) \geq 0)$$

but that the correlation is lower than that for the *derived alignment non-overlapping valency-density inducer*, $I'_{z,\mathrm{ad,n}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\ &\quad (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,n}}) \geq \\ &\qquad\qquad \mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F,\infty,n,q}}))\end{aligned}$$

because of the additional *limited-models* constraints, $\mathcal{F}_{\mathrm{q}}$.

It is not obvious whether or not the *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F,\infty,n,q}}$, has a lower correlation, $\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F,\infty,n,q}})$, with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, than the *derived alignment non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q}}$, $\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,\infty,n,q}})$.

More importantly, it is not obvious whether or not the *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F,\infty,n,q}}$, has a lower *inducer correlation*, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F,\infty,n,q}})$, than the *derived alignment non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q}}$, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F,\infty,n,q}})$.

Like the *derived alignment non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q}}$, the *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F,\infty,n,q}}$, is tractable in all respects.

4.21.9 Tractable decomposition inducers

Both the *derived alignment non-overlapping fud inducer*, $I'_{z,\mathrm{a,F,\infty,n,q}}$, and the *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F,\infty,n,q}}$, are tractable. The *derived alignment inducer*, $I'_{z,\mathrm{a,F,\infty,n,q}}$, tends to be more *ideal*, $A \approx A * T * T^{\dagger A}$, than the *derived alignment valency-density inducer*, $I'_{z,\mathrm{ad,F,\infty,n,q}}$, which tends to be more *formal-abstract equivalent*, $A^{\mathrm{X}} * T \approx (A * T)^{\mathrm{X}}$. However, the lost *ideality* of the *valency-density fud inducer*, $I'_{z,\mathrm{ad,F,\infty,n,q}}$, can be partly recovered in a *valency-density decomposition inducer*. The search set of the *valency-density decomposition inducer* consists of *substrate fud decompositions* similar to *idealising summation aligned decompositions*, $\mathcal{D}_{\Sigma,\mathrm{k}}(A)$, but without the *idealising summation aligned decomposition* constraints in order to avoid intractable inclusion tests.

In section 'Intractable search set elements', above, the *limited-models content alignment integral-independent substrate idealising summation aligned infinite-layer fud decomposition inducer*, $I'_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}$, is defined such that the application to a *non-independent integral-independent substrate histogram* $A \in \mathcal{A}_{z,\mathrm{xi}} \setminus \{A^{\mathrm{X}}\}$ is the *summation alignment* function of the *limited-models substrate idealising summation aligned fud decompositions*,

$$\begin{aligned}
&I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}(A) = \\
&\quad \{(D, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algnSum}(A,D))) : \\
&\qquad\qquad D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}}),\ D^{\mathrm{D}V_A} \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)\}
\end{aligned}$$

Define $I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}(A^{\mathrm{X}}) = \{(D_{\mathrm{F,u}}, 0)\}$ where $D_{\mathrm{F,u}} = \{((\emptyset, \{T_{\mathrm{u}}\}), \emptyset)\}$ and *unary partition transform* $T_{\mathrm{u}} = \{V_A^{\mathrm{CS}}\}^{\mathrm{T}}$.

The *limited-models idealising fud decomposition inducer*, $I'_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}$, has tractable *time* and *space* complexity with respect to the *search set* elements. The *non-unary idealising summation aligned substrate infinite-layer fud decomposition* $D \in \mathrm{dom}(I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}(A)) \setminus \{D_{\mathrm{F,u}}\}$ is constrained (a) to be *well behaved*, $D \in \mathcal{D}_{\mathrm{F,w},U_A}$, (b) such that the *infinite-layer substrate fuds* appear no more than once in any path, $\forall L \in \mathrm{paths}(D)\ (\{(i,F) : (i,(\cdot,F)) \in L\} \in \mathbf{N} \leftrightarrow \mathcal{F}_{\infty,U_A,V_A})$, (c) to have no *variable symmetries*, $\{(w,(C,F)) : (C,F) \in \mathrm{cont}(D),\ w \in \mathrm{der}(F)\} \in \mathrm{der}(\bigcup G) \to \mathrm{cont}(D)$, which implies (b), and (d) such that the *fuds* have (i) *contingent diagonalisation*, $\forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{diagonal}(A * C * F^{\mathrm{T}}))$, (ii) *contingent formal-abstract equivalence*, $\forall (C,F) \in \mathrm{cont}(D)\ (A^{\mathrm{X}} * C * F^{\mathrm{T}} = (A * C * F^{\mathrm{T}})^{\mathrm{X}})$, (iii) *non-independent contingent derived histograms*, $\forall (C,F) \in \mathrm{cont}(D)\ (A * C * F^{\mathrm{T}} \neq (A * C * F^{\mathrm{T}})^{\mathrm{X}})$, and (iv) *independent formal slices*, $\forall (C,F) \in \mathrm{cont}(D)\ ((A * C)^{\mathrm{X}} * F^{\mathrm{T}} = ((A * C)^{\mathrm{X}} * F^{\mathrm{T}})^{\mathrm{X}})$, where $G = \mathrm{fuds}(D)$ and $\mathrm{cont} = \mathrm{elements} \circ \mathrm{contingents}$.

The *contingent formal-abstract equality* inclusion test, $A^{\mathrm{X}} * C * F^{\mathrm{T}} = (A * C * F^{\mathrm{T}})^{\mathrm{X}}$, is intractable because of intractable *substrate volume*. This is for the same reason that the *formal-abstract equality* inclusion test, $A^{\mathrm{X}} * F^{\mathrm{T}} = (A * F^{\mathrm{T}})^{\mathrm{X}}$, is intractable, as described in section 'Intractable literal substrate model inclusion', above. That is, the computation of the *independent histogram*, A^{X}, by an *independenter*, $I^{*}_{\mathrm{X}}(A) = A^{\mathrm{X}}$, requires *time* and *space* of at least v, where $v = |V_A^{\mathrm{CS}}|$, because the *substrate histogram*, A, has *completely effective independent*, $A^{\mathrm{XF}} = V_A^{\mathrm{C}}$. Thus the computation of the *independent histogram*, A^{X}, in the computation of the *contingent formal histogram*, $A^{\mathrm{X}} * C * F^{\mathrm{T}}$, in the *contingent formal-abstract equality* inclusion test, $A^{\mathrm{X}} * C * F^{\mathrm{T}} = (A * C * F^{\mathrm{T}})^{\mathrm{X}}$, is intractable with respect to *underlying dimension*, n, where $n = |V_A|$.

The *independent formal slice* inclusion test, $(A*C)^{\mathrm{X}} * F^{\mathrm{T}} = ((A*C)^{\mathrm{X}} * F^{\mathrm{T}})^{\mathrm{X}}$, also requires the computation of the *independent slice histogram*, $(A * C)^{\mathrm{X}}$, by an *independenter*, $I^{*}_{\mathrm{X}}(A * C) = (A * C)^{\mathrm{X}}$, in the same *substrate variables*, V_A, and hence is subject to the same intractability in the *root slice* at least.

Just as in the case of the *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$, above, the intractability of the *formal-abstract equality* inclusion test and *independent formal slice* inclusion test in the *limited-models idealising summation aligned infinite-layer fud decomposition inducer*, $I'_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}$, can be addressed by replacing them with (i) a *non-overlapping fud* constraint, $\neg\mathrm{overlap}(F)$, and (ii) the computation of the *summed derived alignment valency density*. Define the *summed derived alignment valency density* as $\mathrm{algnValDensSum}(U) \in \mathcal{A} \times \mathcal{D} \to \mathbf{R}$ as

$$\begin{aligned}
&\mathrm{algnValDensSum}(U)(A, D) := \\
&\qquad \sum_{(C,T) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * T)/\mathrm{capacityValency}(U)((A * C * T)^{\mathrm{FS}})
\end{aligned}$$

where the *valency capacity*, $\mathrm{capacityValency}(U) \in \mathrm{capacities}$, is defined in terms of *geometry* as $\mathrm{capacityValency}(U)((A * C * T)^{\mathrm{FS}}) = w^{1/m}$, and $m = |W|$, $w = |W^{\mathrm{C}}|$ and $W = \mathrm{der}(T)$. Then define the *limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer*, given *non-independent substrate histogram* $A \in \mathcal{A}_z \setminus \{A^{\mathrm{X}}\}$,

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A) = \\
&\quad \{(D, I^{*}_{\approx\mathbf{R}}(\mathrm{algnValDensSum}(U_A)(A, D^{\mathrm{D}}))) : \\
&\qquad\qquad D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})), \\
&\qquad\qquad\qquad \forall (C, F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A * C * F^{\mathrm{T}}) > 0)\}
\end{aligned}$$

Define $I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A^{\mathrm{X}}) = \{(D_{\mathrm{F,u}}, 0)\}$.

The *summed alignment valency-density aligned fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, constrains each of the *slices* $(C, F) \in \mathrm{cont}(D)$ such that the application of the *fud*, $A*C*F^{\mathrm{T}}$, has *derived alignment*, $\mathrm{algn}(A*C*F^{\mathrm{T}}) > 0$, where $A \neq A^{\mathrm{X}}$. Thus the *slice derived histogram*, $A * C * F^{\mathrm{T}}$, cannot be *independent* and the *idealising summation aligned decomposition non-independent contingent derived histograms* constraint, $\forall (C,F) \in \mathrm{cont}(D)\ (A * C * F^{\mathrm{T}} \neq (A * C * F^{\mathrm{T}})^{\mathrm{X}})$, is satisfied by the *inducer*. In addition, *fuds* are prevented from appearing more than once in any path of the *decomposition*, $\forall L \in \mathrm{paths}(D)\ (\{(i,F) : (i,(\cdot,F)) \in L\} \in \mathbf{N} \leftrightarrow G)$ where $G = \mathrm{fuds}(D)$, because a *fud* has zero *derived alignment* when constrained to its child *slice*. For example, if $(C_1,F),(C_2,F) \in \mathrm{steps}(\mathrm{contingents}(D))$ then $\mathrm{algn}(A * C_2 * F^{\mathrm{T}}) = 0$.

Variable symmetries are avoided only so far as the *fuds* are not repeated in a path, so in this respect the definition of the *substate models* of the *summed alignment valency-density non-overlapping fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is less strict than the definition of the *substrate idealising fud decompositions*, $|\mathrm{dom}(I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A)) \setminus \mathrm{dom}(I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}(A))| \geq 0$.

The *valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, does not test explicitly for *well behaved decomposition*, $\mathcal{D}_{\mathrm{F,w},U_A}$. Each of the *slices* has *derived alignment*, $\mathrm{algn}(A*C*F^{\mathrm{T}}) > 0$, and must therefore have non-zero *size*, $\mathrm{size}(A*C) > 0$ and so correspond to a non-empty *component*, $\mathrm{ran}(\mathrm{cont}(D)) \leftrightarrow \mathrm{elements}(\mathrm{components}(U)(D)) \setminus \mathrm{partition}(U)(D)$. That is, the *decomposition*, D, does not contain any contradictions and is therefore *well behaved*, $D \in \mathcal{D}_{\mathrm{F,w},U_A}$.

The *valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, does not test explicitly for *contingent diagonalisation*, $\forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{diagonal}(A * C * F^{\mathrm{T}}))$. The *maximum transform* function, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, optimises the *summed alignment valency density*, $\mathrm{algnValDensSum}(U_A)(A,D)$, which varies as the *derived alignments* of the *slices*, $\mathrm{algn}(A*C*F^{\mathrm{T}})$. As described in section 'Maximum alignment', above, maximum *alignment* is conjectured to occur when the *derived histogram* is *fully diagonalised*, $\mathrm{diagonalFull}(U_A)(A * C*F^{\mathrm{T}})$. Therefore higher *derived alignment* tends to *diagonalise* the *slice derived histogram*, $A * C * F^{\mathrm{T}}$, approximately satisfying the *summation aligned decomposition contingent diagonalisation* constraint.

In addition, *valency-density* tends to shorten *diagonals* and therefore the *derived slice* has fewer *ineffective states*, $|(A*C*F^{\mathrm{T}})^{\mathrm{C}} - (A*C*F^{\mathrm{T}})^{\mathrm{F}}| \approx (d-1)^m$ where $d = w^{1/m}$. Although the cardinality of strong compositions of the *di-*

agonal times the cardinality of subsets of the *volume* that are *diagonalised*,

$$\frac{(z-1)!}{(d-1)!\ (z-d)!}(d!)^{m-1} = dz^{\underline{d}}(d^{\underline{d}})^{m-2}/z$$

which is dominated by $z^{\underline{d}}$, increases with the *diagonal*, d, given *size* z, the ratio of this cardinality of the *diagonals* to the cardinality of weak compositions of the *volume*, $w = d^m$,

$$\frac{(z-1)!}{(d-1)!\ (z-d)!}(d!)^{m-1}\frac{(w-1)!\ z!}{(z+w-1)!} = dz^{\underline{d}}(d^{\underline{d}})^{m-2}w^{\underline{w}}/wz^{\overline{w}}$$

which is dominated by $1/z^{\overline{w}}$, decreases with the *diagonal*, d, where $d < w < z$. That is, the fraction of *derived histograms* of a given *derived geometry* that are *diagonalised* increases as the *diagonals* shorten.

This is in line with the maximisation of *midisation* of a *slice*, $\mathrm{algn}(A * C) - \mathrm{algn}((A * C * F^{\mathrm{T}})^{\mathrm{X}} * F^{\mathrm{T}\odot A*C}) - \mathrm{algn}(A * C * F^{\mathrm{T}} * F^{\mathrm{T}\dagger A*C})$, for which *slice derived valency-density*, $\mathrm{algn}(A * C * F^{\mathrm{T}})/w^{1/m}$, is a proxy. Maximisation of *midisation* tends to move *component alignments* within the *slice*, $A*C$, from *off-diagonal states*, $(A * C * F^{\mathrm{T}})^{\mathrm{CS}} \setminus (A * C * F^{\mathrm{T}})^{\mathrm{FS}}$, to *on-diagonal states*, $(A * C * F^{\mathrm{T}})^{\mathrm{FS}}$, balancing the high *derived alignment* of longer *diagonals* with the high *on-diagonal component alignments* of shorter *diagonals*. The lower *component alignments* of the *off-diagonal states* tend to make them less *effective*, while conversely the higher *component alignments* of *on-diagonal states* tend to make them more *effective*.

The *fuds* $G = \mathrm{fuds}(D)$ of the *decomposition* D are *non-overlapping*, $G \subset \mathcal{F}_{\mathrm{n}}$. So all *slices* have *independent formal*, $\forall (C, F) \in \mathrm{cont}(D)\ (\neg\mathrm{overlap}(F) \implies (A*C)^{\mathrm{X}}*F^{\mathrm{T}} = ((A*C)^{\mathrm{X}}*F^{\mathrm{T}})^{\mathrm{X}})$, satisfying the *idealising summation aligned decomposition independent formal slices* constraint. However, as noted for the *derived alignment non-overlapping inducer*, $I'_{z,\mathrm{a,n,j}}$, in section 'Intractable literal substrate model inclusion', above, *independent-formal transforms* are not necessarily *non-overlapping*, $A^{\mathrm{X}}*T = (A^{\mathrm{X}}*T)^{\mathrm{X}} \Longleftarrow \neg\mathrm{overlap}(T)$, and so the strict *non-overlapping* constraint sometimes excludes some *independent-formal overlapping fuds*. The set of *substrate models* of the *valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is therefore neither a superset nor a subset of the set of the *substrate idealising summation aligned infinite-layer fud decompositions*, $|\mathrm{dom}(I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A)) \setminus \mathrm{dom}(I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}(A))| \geq 0$ and $|\mathrm{dom}(I'^{*}_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}(A)) \setminus \mathrm{dom}(I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A))| \geq 0$.

Although both the (i) *non-overlapping fud* constraint, $\neg\mathrm{overlap}(F)$, and the

(ii) the *summed derived alignment valency density*, $\mathrm{algnValDensSum}(U_A)(A, D^{\mathrm{D}})$, maximisation, in the *valency-density decomposition inducer maximum transform* function, $\mathrm{maxr} \circ I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, tend to increase the adherence to the *formal-abstract equality* of individual *slices*, $(A * C)^{\mathrm{X}} * F^{\mathrm{T}} \approx (A * C * F^{\mathrm{T}})^{\mathrm{X}}$ where $(C, F) \in \mathrm{cont}(D)$, in a *decomposition* D, it is not necessarily the case that the *contingent formal-abstract equality* of the *slice* increases, $A^{\mathrm{X}} * C * F^{\mathrm{T}} \approx (A * C * F^{\mathrm{T}})^{\mathrm{X}}$, or that the *formal-abstract equality* of the *nullable transform* increases, $A^{\mathrm{X}} * D^{\mathrm{T}} \approx (A * D^{\mathrm{T}})^{\mathrm{X}}$. It is only in the case where the *independent slice* equals the *sliced independent*, $(A * C)^{\mathrm{X}} = A^{\mathrm{X}} * C$, that the *slice formal-abstract equality* and the *contingent formal-abstract equality* constraints are identical. This is always the case for the root *fud* of the *decomposition*, where $C = V^{\mathrm{C}} \implies A * C = A \implies (A * C)^{\mathrm{X}} = (A^{\mathrm{X}} * C)$.

Note that the purpose of the (i) *contingent diagonalisation* and (ii) *contingent formal-abstract equivalence* constraints in the *summation aligned decompositions*, $\mathcal{D}_{\Sigma}(A)$, is merely to allow the *nullable transform content alignment* to be computed without instantiating the intractable *nullable transform*. In section 'Decomposition alignment', above, it is shown that

$$\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) = \mathrm{alignmentSum}(A, D)$$

where these conditions are met. Contrast that to the purpose of the maximisation of *derived alignment valency density* in the *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$, which is to maximise the *formal-abstract equality* and hence increase the correlation, $\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}})$, to the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, and thence transitively to the increase the *inducer correlation* to the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}})$.

However, the goals of (a) computing the *content alignment* of the intractable *nullable transform*, D^{T}, by means of *contingent formal-abstract equivalence*, and (b) increasing the correlation of *summed alignment valency-density decomposition inducer*, $\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}})$, attributable to *formal-abstract equivalent slices*, can be made to converge by (i) choosing *components*, C, with high cardinality, and (ii) minimising the *formal alignment* of the *decomposition*, $\mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}})$.

In the first case, high cardinality *components*, C, tend to approximate better to the *cartesian volume*, $C \approx V_A^{\mathrm{CS}}$, and hence $A^{\mathrm{X}} * C \approx A^{\mathrm{X}}$ which implies $A^{\mathrm{X}} * C \approx (A * C)^{\mathrm{X}}$ because $(A * C)^{\mathrm{X}} \approx A^{\mathrm{X}}$. That is, shorter *diagonals* are preferable because the *component sizes* are larger. Maximisation of *derived alignment valency-density*, $\mathrm{algn}(A * C * T)/w^{1/m}$, tends to shorter *diagonals*, and therefore higher *component* cardinality of child *slices*.

In the second case, lower *formal alignment* of the *decomposition*, $\mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}})$, implies lower *contingent formal alignment* of the *slices*, $\mathrm{algn}(A^{\mathrm{X}} * C * F^{\mathrm{T}})$, and hence *contingent independent-formal equality*, $A^{\mathrm{X}} * C * F^{\mathrm{T}} \approx (A^{\mathrm{X}} * C * F^{\mathrm{T}})^{\mathrm{X}}$, which is implied by *contingent formal-abstract equality*, $A^{\mathrm{X}} * C * F^{\mathrm{T}} \approx (A * C * F^{\mathrm{T}})^{\mathrm{X}}$. Minimisation of the *formal alignment* of the *decomposition*, $\mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}})$, is desirable in any case because then the *summed derived alignment valency density*, $\mathrm{algnValDensSum}(U_A)(A, D)$, most approximates to the *nullable transform derived alignment valency density* rather than the *nullable transform content alignment valency density*

$$(\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}))/y^{1/p} \approx \mathrm{algnValDensSum}(U_A)(A, D)$$

where $p = |Y|$, $u = |Y^{\mathrm{C}}|$ and $Y = \mathrm{der}(D^{\mathrm{T}})$. That is, $\mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) = 0 \implies \mathrm{algn}(A * D^{\mathrm{T}})/y^{1/p} \approx \mathrm{algnValDensSum}(U_A)(A, D)$. Note that if the *decomposition*, D, contains more than one *fud*, $|G| > 1$, then the *volume* of the *derived variables* of the *nullable transform*, D^{T}, is greater than the *volume* of the *derived variables* of the *fuds*, $|\mathrm{der}(D^{\mathrm{T}})^{\mathrm{C}}| > |\mathrm{der}(\bigcup G)^{\mathrm{C}}|$. This is because the *nullable variables* that do not *originate* in the *root transform*, $\mathrm{nullables}(U)(D)$, have an additional *null value* with respect to their corresponding *originating variable*, $\exists(u, x) \in \mathrm{originals}(U)(D)\ (|U_u| = |U_x| + 1)$. So the incremented *valencies* lengthen the geometric average *diagonal* of the *nullable transform*, $y^{1/p}$.

As described in section 'Summation aligned decomposition inducers', above, part of the *formal alignment* of the *decomposition*, $\mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}})$, consists of the pure *formal alignment* of the *skeletal reduction*, $\mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) \geq \mathrm{algn}(A^{\mathrm{X}} * D'^{\mathrm{T}}) > 0$ where $D' \in \mathrm{reductions}(A, D)$, which is such that $\mathrm{skeletal}(A * D'^{\mathrm{T}})$. In section 'Skeletal alignment', above, it is shown that the *alignment* of a *uniform full regular skeleton histogram* is minimised for a given *size*, such that the *counts* are at least one, when the *regular skeleton* tree is a binary tree. That is, the pure *formal alignment* of the *skeletal reduction*, $\mathrm{algn}(A^{\mathrm{X}} * D'^{\mathrm{T}})$, is least when *bi-valent*, $d = 2$. Thus the shorter *diagonals*, $d = w^{1/m}$, of the maximisation of *summed derived alignment valency density*, tends to reduce the pure *formal alignment* of the *skeletal reduction*, if not the *contingent formal alignment* of the *slices*.

The *fuds* $G = \mathrm{fuds}(D)$ of a *decomposition* D are individually *non-overlapping*, $G \subset \mathcal{F}_{\mathrm{n}}$ or $\forall F \in G\ (\neg\mathrm{overlap}(F))$, but there is nothing to prevent *fuds* from *overlapping* with eachother, $\mathrm{overlap}(F_1 \cup F_2)$ where $F_1, F_2 \in G$. However, as was noted above, a *fud* cannot appear more than once in any path of the *decomposition* because it has zero *derived alignment* when constrained to its child *slice*. Similarly, if a pair of highly *overlapping fuds*, F_1, F_2, are in the same path, then the latter *derived alignment*, $\mathrm{algn}(A * C_2 * F_2^{\mathrm{T}})$, tends to be lower than if it were not constrained to be in a descendant *slice* of the former, $A{*}C_2 \subset A{*}C_1$, for example where $(C_1, F_1), (C_2, F_2) \in \mathrm{steps}(\mathrm{contingents}(D))$. Therefore, maximisation of *derived alignment* tends to reduce *overlapping* between *fuds* on the same path, and so reduces the overall *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}})$. In the case of *fuds* in separate paths of the *decomposition*, *overlap* merely allows the representation of *symmetries* without increasing *formal alignment*, albeit with creation of duplicate or highly similar *fuds*. For example, $\mathrm{algn}(A * C_1 * F^{\mathrm{T}}) + \mathrm{algn}(A * C_2 * F^{\mathrm{T}}) \approx \mathrm{algn}(A * (C_1 + C_2) * F^{\mathrm{T}})$ where $(C_1, F), (C_2, F) \in \mathrm{cont}(D)$.

Insofar as the *limited-models summed alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, adheres to the constraints of the *limited-models idealising summation aligned fud decomposition inducer*, $I'_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}$, the search set consists of *substrate fud decompositions* similar to *idealising summation aligned decompositions*, $\mathcal{D}_{\Sigma,\mathrm{k}}(A)$. The *limited-models idealising summation aligned fud decomposition inducer*, $I'_{z,\mathrm{c,D,F},\infty,\Sigma,\mathrm{k,q}}$, is derived from the *content alignment idealising summation aligned inducer*, $I'_{z,\mathrm{c,D},\Sigma,\mathrm{k}}$, described in section 'Summation aligned decomposition inducers', above. An *idealising substrate summation aligned decomposition* $D \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)$ that is *ideal*, $A = A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A}$, has no *super idealising substrate summation aligned decomposition*, $\forall E \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ (D \notin \mathrm{subtrees}(E))$. All of its *sub idealising substrate summation aligned decompositions* have lower *content alignment*, $\forall E \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ (E \in \mathrm{subtrees}(D) \implies \mathrm{algnSum}(A, E) < \mathrm{algnSum}(A, D))$. Therefore, the *maximum idealising substrate summation aligned decompositions* in the *content idealising inducer*, $I'_{z,\mathrm{c,D},\Sigma,\mathrm{k}}$, are all *ideal*, $\forall D \in \mathrm{maxd}(I'^{*}_{z,\mathrm{c,D},\Sigma,\mathrm{k}}(A))\ (\mathrm{ideal}(A, D^{\mathrm{T}}))$. The same reasoning applies to *summed alignment valency-density*, $\forall E \in \mathcal{D}_{U_A,V_A} \cap \mathcal{D}_{\Sigma,\mathrm{k}}(A)\ (E \in \mathrm{subtrees}(D) \implies \mathrm{algnValDensSum}(U_A)(A, E) < \mathrm{algnValDensSum}(U_A)(A, D))$. Thus the *substrate fud decompositions* of the *maximum function* of the *summed alignment valency-density fud decomposition inducer*, $\mathrm{maxd}(I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A)) \subset \mathcal{D}_{\mathrm{F},\infty,U_A,V_A}$, tend to be *ideal*, even though the non-leaf *fuds* of the *decompositions* are not themselves *ideal* with respect to their *slices*. This is the case even though the *summed alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is not subject to the intractabilities of the *idealising inducers*. In this way, the lost *ideality* of the *tractable derived alignment* *valency-density fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$, compared to the *tractable derived alignment fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{n,q}}$, can be restored to some extent.

This restoration of the *ideality* in the *limited-models summed alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, tends to increase the *component size cardinality relative entropy*, $\mathrm{entropyRelative}(A*D^{\mathrm{T}}, V_A^{\mathrm{C}}*D^{\mathrm{T}})$ where $D \in \mathrm{maxd}(I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A))$, in the case of *non-singleton decompositions*, $|\mathrm{nodes}(D)| > 1$.

In section 'Likely histograms', it is conjectured that there exists an intermediate *mid substrate transform* $T_{\mathrm{m}} \in \mathcal{T}_{U_A,V_A}$ which is neither *self* nor *unary*, $T_{\mathrm{m}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$, where the *formal* is *independent* and the *midisation entropy* is minimised,

$$T_{\mathrm{m}} \in \mathrm{mind}(\{(T, \mathrm{entropy}(A^{\mathrm{M}(T)})) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\})$$

At the *mid transform* the *formal* tends to the *abstract*, $A^{\mathrm{X}} * T_{\mathrm{m}} \approx (A * T_{\mathrm{m}})^{\mathrm{X}}$, and the *mid component size cardinality relative entropy* is small,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V_A^{\mathrm{C}} * T_{\mathrm{m}}) \quad \approx \quad 0$$

Section 'Transform alignment', goes on to conjecture that an approximation to the *mid transform* may also be obtained by a maximisation of the *midisation pseudo-alignment*,

$$\begin{aligned} T_{\mathrm{m}} \quad \in \quad & \mathrm{maxd}(\{(T, \mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})) : \\ & \qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}) \end{aligned}$$

Then it is shown that the *midisation pseudo-alignment* varies with the *derived alignment valency-density*,

$$\mathrm{algn}(A) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}) - \mathrm{algn}(A * T * T^{\dagger A}) \quad \sim \quad \mathrm{algn}(A * T)/w^{1/m}$$

where $m = |W|$, $w = |W^{\mathrm{C}}|$ and $W = \mathrm{der}(T)$. So the maximisation of the *derived alignment valency-density* in the *derived alignment valency-density inducer*, $I'_{z,\mathrm{ad,fx}}$, tends to *formal-abstract* equality, $A^{\mathrm{X}}*T_{\mathrm{m}} \approx (A*T_{\mathrm{m}})^{\mathrm{X}}$, allowing *lifting* and increasing the *inducer correlation*. Similarly, the maximisation of the *summed derived alignment valency density*, $\mathrm{algnValDensSum}(U_A)(A, D^{\mathrm{D}})$, in the *limited-models summed alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, tends to *formal-abstract* equality in each *slice*, $(A * C)^{\mathrm{X}} * F^{\mathrm{T}} \approx (A * C * F^{\mathrm{T}})^{\mathrm{X}}$ where $(C, F) \in \mathrm{cont}(D)$.

Section 'Likely histograms' goes on to show that the subsequent minimisation of the *idealisation entropy*, where the *mid idealisation* is *integral*, $A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A} \in \mathcal{A}_{\mathrm{i}}$, tends to increase the *mid component size cardinality relative entropy*,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V_A^{\mathrm{C}} * T_{\mathrm{m}}) \quad \sim \quad -\ \mathrm{entropy}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A})$$

In section 'Transform alignment', it is conjectured that subsequent maximisation of the *idealisation alignment* also tends to increase the *relative entropy*,

$$\mathrm{entropyRelative}(A * T_{\mathrm{m}}, V_A^{\mathrm{C}} * T_{\mathrm{m}}) \quad \sim \quad \mathrm{algn}(A * T_{\mathrm{m}} * T_{\mathrm{m}}^{\dagger A})$$

Let $L \in \mathrm{paths}(\mathcal{D}_{\Sigma,\mathrm{k}}(A))$ be a path of *idealising summation aligned decompositions* of *histogram* A such that (i) each *decomposition* is an immediate *super-decomposition* of the previous *decomposition*, $\forall i \in \{2 \dots l\}\ (L_{i-1} \in \mathrm{subtrees}(L_i)\ \wedge\ |\mathrm{nodes}(L_{i-1})| = |\mathrm{nodes}(L_i)| - 1)$, where $l = |L|$, and (ii) the last *decomposition* is *ideal*, $A * L_l^{\mathrm{T}} * L_l^{\mathrm{T}\dagger A} = A$. In section 'Decomposition alignment', above, it is shown that the *idealisation alignment* increases along the path,

$$\forall i \in \{2 \dots l\}\ (\mathrm{algn}(A * L_i^{\mathrm{T}} * L_i^{\mathrm{T}\dagger A}) > \mathrm{algn}(A * L_{i-1}^{\mathrm{T}} * L_{i-1}^{\mathrm{T}\dagger A}))$$

Consider the case where (i) the *root transform* is the *mid transform*, $L_1 = \{((\emptyset, T_{\mathrm{m}}), \emptyset)\}$, and (ii) the *idealisations* along the path are all *integral*, $\forall i \in \{1 \dots l\}\ (A * L_i^{\mathrm{T}} * L_i^{\mathrm{T}\dagger A} \in \mathcal{A}_{\mathrm{i}})$. In this case the *relative entropy* also increases along the path,

$$\begin{aligned}&\forall i \in \{2 \dots l\}\\ &\quad (\mathrm{entropyRelative}(A * L_i^{\mathrm{T}}, V_A^{\mathrm{C}} * L_i^{\mathrm{T}}) > \mathrm{entropyRelative}(A * L_{i-1}^{\mathrm{T}}, V_A^{\mathrm{C}} * L_{i-1}^{\mathrm{T}}))\end{aligned}$$

The first *decomposition*, L_1, which is a *sub-decomposition* of all subsequent, has the least *relative entropy*, $\mathrm{entropyRelative}(A * L_1^{\mathrm{T}}, V_A^{\mathrm{C}} * L_1^{\mathrm{T}}) \approx 0$. The last *decomposition*, L_l, which is a *super-decomposition* of all previous, has the greatest *relative entropy*, $\mathrm{entropyRelative}(A * L_l^{\mathrm{T}}, V_A^{\mathrm{C}} * L_l^{\mathrm{T}}) > 0$.

That is, an *idealising summation aligned decomposition* $D \in \mathcal{D}_{\Sigma,\mathrm{k}}(A)$ that (i) is *ideal*, $A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A} = A$, and (ii) is *rooted* in the *mid transform*, $D = \{((\emptyset, T_{\mathrm{m}}), \cdot)\}$, tends to increase *relative entropy* as the cardinality of *decomposition* nodes increases,

$$\mathrm{entropyRelative}(A * D^{\mathrm{T}}, V_A^{\mathrm{C}} * D^{\mathrm{T}}) \quad \sim \quad |\mathrm{nodes}(D)|$$

In the case where each *transform* is the *mid transform* for the *component*,

$$\begin{aligned}&\forall (C, T) \in \mathrm{cont}(D)\ (T \in \mathrm{mind}(\{(T', \mathrm{entropy}((A * C)^{\mathrm{M}(T')})) :\\ &\qquad\qquad T' \in \mathcal{T}_{U_A,V_A},\ (A * C)^{\mathrm{X}} * T' = ((A * C)^{\mathrm{X}} * T')^{\mathrm{X}}\}))\end{aligned}$$

then each *non-leaf decomposition node* $((\cdot,T),F) \in \text{nodes}(D)$, where $F \neq \emptyset$, forms a *child decomposition* $E = \{((\emptyset,T),F)\}$ in *slice* $A * C$ which is *rooted* in the *slice mid transform*, T, so that the *slice formal* approximates to the *slice abstract*, $(A*C)^{\mathrm{X}}*T \approx (A*C*T)^{\mathrm{X}}$, but the *child decomposition relative entropy*, $\text{entropyRelative}(A * C * E^{\mathrm{T}}, V_A^{\mathrm{C}} * E^{\mathrm{T}})$, is not necessarily small.

Therefore, insofar as the search set of the *tractable limited-models summed alignment valency-density fud decomposition inducer*,

$$\text{dom}(I'^{*}_{z,\text{Sd,D,F},\infty,\text{n,q}}(A)) \subset \mathcal{D}_{\text{F},\infty,U_A,V_A} \cap \text{trees}(\mathcal{S} \times (\mathcal{F}_{\text{n}} \cap \mathcal{F}_{\text{q}}))$$

consists of *substrate fud decompositions* similar to *idealising summation aligned decompositions*, $\mathcal{D}_{\Sigma,\text{k}}(A)$, the *component size cardinality relative entropy* of a maximal *model* may be expected to be (a) greater than that of the corresponding *model* in the *tractable derived alignment valency-density fud inducer*, $I'_{z,\text{ad,F},\infty,\text{n,q}}$,

$$\text{entropyRelative}(A * D^{\mathrm{T}}, V_A^{\mathrm{C}} * D^{\mathrm{T}}) \quad > \quad \text{entropyRelative}(A * F_{\text{ad}}^{\mathrm{T}}, V_A^{\mathrm{C}} * F_{\text{ad}}^{\mathrm{T}})$$

where $D \in \text{maxd}(I'^{*}_{z,\text{Sd,D,F},\infty,\text{n,q}}(A))$ and $F_{\text{ad}} \in \text{maxd}(I'^{*}_{z,\text{ad,F},\infty,\text{n,q}}(A))$, and (b) comparable to that of the corresponding *model* in the *tractable derived alignment fud inducer*, $I'_{z,\text{a,F},\infty,\text{n,q}}$,

$$\text{entropyRelative}(A * D^{\mathrm{T}}, V_A^{\mathrm{C}} * D^{\mathrm{T}}) \quad \approx \quad \text{entropyRelative}(A * F_{\text{a}}^{\mathrm{T}}, V_A^{\mathrm{C}} * F_{\text{a}}^{\mathrm{T}})$$

where $F_{\text{a}} \in \text{maxd}(I'^{*}_{z,\text{a,F},\infty,\text{n,q}}(A))$.

Note that, while the *idealisations* of the *sub-decompositions* are not necessarily *integral*, at least the *decomposition* itself is generally *ideal*, and therefore *integral*, $A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A} = A \in \mathcal{A}_{\text{i}}$.

That is, the *relative entropy* lost by maximisation of *midisation alignment* in the *fuds* can be restored to some extent by subsequent maximisation of *idealisation alignment* in the *decomposition*.

Insofar as the *limited-models summed alignment valency-density fud decomposition inducer*, $I'_{z,\text{Sd,D,F},\infty,\text{n,q}}$, adheres to the constraints of the *limited-models derived alignment valency-density fud inducer*, $I'_{z,\text{ad,F},\infty,\text{n,q}}$, the *alignment bounded lifted iso-transform error* of the *fuds* of the *search set substrate fud decompositions* is reduced. The *limited-models derived alignment valency-density fud inducer*, $I'_{z,\text{ad,F},\infty,\text{n,q}}$, is derived from the *derived alignment valency-density inducer*, $I'_{z,\text{ad,fx}}$, described in section 'Intractable literal substrate model inclusion', above. There it is conjectured that the *valency capacity* tends to increase the correlation of the *maximum function* of the *derived alignment valency-density inducer*, $\text{maxr} \circ I'^{*}_{z,\text{ad},\text{fx}}$, to the *alignment-bounded lifted iso-transform space ideal transform maximum function*, $\text{maxr} \circ X'_{z,\text{xi},\text{T},\text{y},\text{fa},\text{j}}$, and thence transitively to increase the *inducer correlation* with the *alignment-bounded iso-transform space ideal transform maximum function*, $\text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa},\text{j}}$. Although the *sizes* and *derived alignments* of the *slices* decrease along the paths of the *substrate fud decompositions* of the *maximum function* of the *summed alignment valency-density decomposition inducer*, $\text{maxd}(I'^{*}_{z,\text{Sd},\text{D},\text{F},\infty,\text{n},\text{q}}(A)) \subset \mathcal{D}_{\text{F},\infty,U_A,V_A}$, the *alignment-bounded lifted iso-transform error* of the *slice fuds* may be conjectured to be reduced nonetheless. Therefore the *nullable transform*, D^{T} where $D \in \text{maxd}(I'^{*}_{z,\text{Sd},\text{D},\text{F},\infty,\text{n},\text{q}}(A))$, may also have lower *alignment-bounded lifted iso-transform error*. Thus conjecture that the *valency capacity* tends to increase the *inducer correlation* of the *summed alignment valency-density fud decomposition inducer*, $I'_{z,\text{Sd},\text{D},\text{F},\infty,\text{n},\text{q}}$, with the *alignment-bounded iso-transform space ideal transform maximum function*, $\text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa},\text{j}}$.

The *limited-models summed alignment valency-density substrate aligned non overlapping infinite-layer fud decomposition inducer*, $I'_{z,\text{Sd},\text{D},\text{F},\infty,\text{n},\text{q}}$, is tractable in all respects.

Conjecture that the *summed alignment valency-density decomposition inducer*, $I'_{z,\text{Sd},\text{D},\text{F},\infty,\text{n},\text{q}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\text{a},\text{l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(\text{maxr} \circ I'^{*}_{z,\text{a},\text{l}}, \text{maxr} \circ I'^{*}_{z,\text{Sd},\text{D},\text{F},\infty,\text{n},\text{q}}) \geq 0)$$

but that the correlation is lower than that for the *limited-models idealising fud decomposition inducer*, $I'_{z,\text{c},\text{D},\text{F},\infty,\Sigma,\text{k},\text{q}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\ &\quad (\text{cov}(z)(\text{maxr} \circ I'^{*}_{z,\text{a},\text{l}}, \text{maxr} \circ I'^{*}_{z,\text{c},\text{D},\text{F},\infty,\Sigma,\text{k},\text{q}}) \geq\\ &\qquad\qquad \text{cov}(z)(\text{maxr} \circ I'^{*}_{z,\text{a},\text{l}}, \text{maxr} \circ I'^{*}_{z,\text{Sd},\text{D},\text{F},\infty,\text{n},\text{q}}))\end{aligned}$$

However, the correlation is higher than that for the *derived alignment valency-density fud inducer*, $I'_{z,\text{ad},\text{F},\infty,\text{n},\text{q}}$,

$$\begin{aligned}&\forall z \in \mathbf{N}_{>0}\\ &\quad (\text{cov}(z)(\text{maxr} \circ I'^{*}_{z,\text{a},\text{l}}, \text{maxr} \circ I'^{*}_{z,\text{ad},\text{F},\infty,\text{n},\text{q}}) \leq\\ &\qquad\qquad \text{cov}(z)(\text{maxr} \circ I'^{*}_{z,\text{a},\text{l}}, \text{maxr} \circ I'^{*}_{z,\text{Sd},\text{D},\text{F},\infty,\text{n},\text{q}}))\end{aligned}$$

This is because (i) the application of the *fud decomposition inducer* is a superset of that of the *fud inducer*, $\{(\{((\emptyset,F),\emptyset)\},a) : (F,a) \in I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}(A)\} \subset I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A)$, and (ii) *super-decompositions* must have higher *summed alignment valency-density*, $\forall E \in \mathrm{dom}(I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A))\ (D \in \mathrm{subtrees}(E) \implies I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A)(E) > I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A)(D))$ where $D = \{((\emptyset,F),\emptyset)\}$ and $F \in \mathrm{dom}(I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}(A))$. Therefore, in practice it is only necessary consider two *tractable inducers*, (i) the *derived alignment fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{n,q}}$, and (ii) the *summed alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$.

It is not obvious whether the *summed alignment valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, has a higher or lower *inducer correlation*, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}})$, than the *derived alignment fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{n,q}}$, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{n,q}})$.

4.22 Practicable alignment-bounding

As it is defined above the *summed alignment valency-density aligned fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is a *computer* that lacks explicit definition of (i) the *limited-models* constraints, $\mathcal{F}_{\mathrm{q}}$, (ii) the finite representations of the *substrate models* and their traversal, or (iii) the *alignmenter*, I_{a}, or *real approxer*, $I_{\approx\mathbf{R}}$. The following section, 'Substrate models computation', considers explicit definitions so that it can be determined whether an implementation of the *tractable inducer* can be shown to be *practicable* given particular computation *time* and *space* resources.

If it is the case that the computation resources are insufficient, section 'Optimisation' then goes on to consider *practicable inducers* and the constraints necessary to implement them. Consideration is given to the effects of the additional constraints on the correlation of the *maximum function* between the *practicable inducer* and its corresponding *tractable inducer*.

The theoretic optimisation definitions are then given an explicit example implementation in the next section, 'Implementation'. There the computation definition is less elegant, but more practical, because of (i) explicit recursion, (ii) defined ordering, (iii) caching of temporary values and structures, and (iv) the assignment of *variable* references or identifiers to replace *partition variables*.

4.22.1 Substrate models computation

The *summed alignment valency-density aligned fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is defined in section 'Tractable decomposition inducers', above, given *non-independent substrate histogram* $A \in \mathcal{A}_z \setminus \{A^{\mathrm{X}}\}$, as

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A) = \\
&\quad \{(D, I^{*}_{\approx\mathbf{R}}(\mathrm{algnValDensSum}(U_A)(A,D))) : \\
&\qquad\qquad D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})), \\
&\qquad\qquad\qquad \forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A * C * F^{\mathrm{T}}) > 0)\}
\end{aligned}$$

where $\mathrm{cont}(D) := \mathrm{elements}(\mathrm{contingents}(D))$ and the *summed derived alignment valency density*, $\mathrm{algnValDensSum}(U) \in \mathcal{A} \times \mathcal{D}_{\mathrm{F}} \to \mathbf{R}$, is defined

$$\begin{aligned}
&\mathrm{algnValDensSum}(U)(A,D) := \\
&\quad \sum(\mathrm{algn}(A * C * F^{\mathrm{T}})/\mathrm{cvl}(F) : (C,F) \in \mathrm{cont}(D))
\end{aligned}$$

where the *derived valency capacity* is $\mathrm{cvl}(F) := (w^{1/m} : W = \mathrm{der}(F),\ w = |W^{\mathrm{C}}|,\ m = |W|)$. The cardinality of the *substrate models*, $\mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}))$, has tractable *time* and *space* complexities, but may yet be *impracticable*.

The *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$, is defined above as

$$\begin{aligned}
&I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}(A) = \\
&\quad \{(F, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * F^{\mathrm{T}})/\mathrm{cvl}(F))) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}\}
\end{aligned}$$

The *fuds* of the *decompositions* of the *fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, are the *substrate fuds* of the *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$, $\bigcup\{\mathrm{fuds}(D) : D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}))\} = \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$.

Consider various ways in which the *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, may be constructed. The *infinite-layer substrate fud set* $\mathcal{F}_{\infty,U,V} \subset \mathcal{F}_{U,\mathrm{P}}$ is defined as

$$\mathcal{F}_{\infty,U,V} = \{F : F \subseteq \mathrm{powinf}(U)(V,\emptyset),\ \mathrm{und}(F) \subseteq V\}$$

where U is the infinite *implied system*, $U = \mathrm{implied}(\mathrm{filter}(V,U))$, and the *infinite power fud* $\mathrm{powinf}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}} \to \mathcal{F}_{U,\mathrm{P}}$ is defined without termination

$$\begin{aligned}\text{powinf}(U)(V,F) := F \cup G \cup \text{powinf}(U)(V, F \cup G) :\\ G = \{P^{\mathrm{T}} : K \subseteq \text{vars}(F) \cup V,\ P \in \mathrm{B}(K^{\mathrm{CS}})\}\end{aligned}$$

A tree of non-empty *infinite-layer substrate fuds* may be constructed such that successive *fuds* in a path have incremented *layer* cardinality. That is, the *fuds* are constructed from bottom-up. Define the infinite *partition infinite-layer fud tree* $\text{tfi}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}} \to \text{trees}(\mathcal{F}_{U,\mathrm{P}})$ as

$$\begin{aligned}&\text{tfi}(U)(V,F) :=\\ &\quad \{(F \cup G,\ \text{tfi}(U)(V, F \cup G)) :\\ &\qquad\qquad G \subseteq \{P^{\mathrm{T}} : K \subseteq \text{vars}(F) \cup V,\ (\text{der}(F) \neq \emptyset \implies K \cap \text{der}(F) \neq \emptyset),\\ &\qquad\qquad\qquad P \in \mathrm{B}(K^{\mathrm{CS}})\},\\ &\qquad\qquad G \neq \emptyset\}\end{aligned}$$

Let $\text{tfi}(U) \in \mathrm{P}(\mathcal{V}_U) \to \text{trees}(\mathcal{F}_{U,\mathrm{P}})$ be defined $\text{tfi}(U)(V) = \text{tfi}(U)(V, \emptyset)$.

The *fuds* of the tree are the *infinite-layer substrate fuds*

$$\mathcal{F}_{\infty,U,V} = \text{elements}(\text{tfi}(U)(V)) \cup \{\emptyset\}$$

In this construction the *fuds* are cumulative along the paths. That is, successive *fuds* are proper supersets,

$$\forall M \in \text{subpaths}(\text{tfi}(U)(V))\ \forall i \in \{2 \ldots |M|\}\ (M_{i-1} \subset M_i)$$

Moreover, the new *partition transforms*, $P \in \mathrm{B}(K^{\mathrm{CS}})$ where $K \subseteq \text{vars}(F) \cup V$, that are added to the next *fud* of a step, are constrained such that at least one *underlying variable* is in the highest *layer* of the previous *fud*, $\exists x \in K\ (x \in \text{der}(F))$. Thus

$$\forall M \in \text{subpaths}(\text{tfi}(U)(V))\ \forall i \in \{1 \ldots |M|\}\ (\text{layer}(M_i, \text{der}(M_i)) = i)$$

A *fud*, $F \in \text{elements}(\text{tfi}(U)(V))$, may appear more than once in the tree if there are multiple paths to its construction,

$$|\{M : M \in \text{subpaths}(\text{tfi}(U)(V)),\ M_{|M|} = F\}| \geq 1$$

For convenience, define the sets of *partition variables* in the next *layer* tuples $\in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}^*} \to \mathrm{P}(\mathrm{P}(\mathcal{V}_U))$ as

$$\text{tuples}(V,F) := \{K : K \subseteq \text{vars}(F) \cup V,\ (\text{der}(F) \neq \emptyset \implies K \cap \text{der}(F) \neq \emptyset)\}$$

The sets of *partition variables*, $K \in \text{tuples}(V,F)$, will be called *tuples* in the context of *practicable inducers*. Note that here a *tuple* is not an ordered list, although (i) in some implementations *tuples* have limited cardinalities, and (ii) when ordered they may be used to index an *array histogram representation*.

The *partition infinite-layer fud tree* may then be defined more succinctly in terms of *tuples* as

$$\begin{aligned}
&\text{tfi}(U)(V,F) := \\
&\quad \{(F \cup G,\ \text{tfi}(U)(V, F \cup G)) : \\
&\qquad\qquad G \subseteq \{P^{\text{T}} : K \in \text{tuples}(V,F),\ P \in \text{B}(K^{\text{CS}})\}, \\
&\qquad\qquad G \neq \emptyset\}
\end{aligned}$$

The set of next *layer fuds*, $\{G : G \subseteq \{P^{\text{T}} : K \in \text{tuples}(V,F),\ P \in \text{B}(K^{\text{CS}})\},\ G \neq \emptyset\} \subset \mathcal{F}_{U,\text{P}}$ may be defined in terms of a set of *partition-sets*, $\text{P}(\bigcup\{\text{B}(K^{\text{CS}}) : K \in \text{tuples}(V,F)\}) \setminus \{\emptyset\}$. For the first *layer*, $F = \emptyset$, the set of next *layer fuds* is the set of *fuds* of *partition transforms* of the non-empty *partition-sets* of the *substrate partition-sets set*,

$$\{\{P^{\text{T}} : P \in N\} : N \in \mathcal{N}_{U,V},\ N \neq \emptyset\} \subset \mathcal{F}_{U,\text{P}}$$

which has cardinality $|\mathcal{N}_{U,V}| - 1$. The *substrate partition-sets set* is defined

$$\mathcal{N}_{U,V} = \text{P}(\{P : K \subseteq V,\ P \in \text{B}(K^{\text{CS}})\})$$

The cardinality of the *substrate partition-sets set* is

$$|\mathcal{N}_{U,V}| = 2^c : c = \sum_{K \subseteq V} \text{bell}(|K^{\text{CS}}|)$$

In the case of *regular variables* V, having *valency* $\{d\} = \{|U_w| : w \in V\}$ and *dimension* $n = |V|$, the cardinality is

$$|\mathcal{N}_{U,V}| = 2^c : c = \sum_{k \in \{0 \ldots n\}} \binom{n}{k} \text{bell}(d^k)$$

For higher *layers*, $F \neq \emptyset$, the set of next *layer fuds* corresponds to the *intersecting substrate partition-sets set*,

$$\{\{P^{\text{T}} : P \in N\} : N \in \mathcal{N}_{U,W,X}\} \subset \mathcal{F}_{U,\text{P}}$$

where $W = \mathrm{vars}(F) \cup V$ and $X = \mathrm{der}(F)$. Note that the *partition infinite-layer fud tree*, $\mathrm{tfi}(U)(V)$, cannot contain non-empty *fuds* having empty *derived variables*, $F \neq \emptyset \implies \mathrm{der}(F) \neq \emptyset$. The cardinality of the set of higher *layer fuds* is $|\mathcal{N}_{U,W,X}|$. The *intersecting substrate partition-sets set*, $\mathcal{N}_{U,V,X}$, is defined,

$$\mathcal{N}_{U,V,X} = \mathrm{P}(\{P : K \subseteq V,\ K \cap X \neq \emptyset,\ P \in \mathrm{B}(K^{\mathrm{CS}})\})$$

The cardinality of the *intersecting substrate partition-sets set* is

$$|\mathcal{N}_{U,W,X}| = 2^c \ :\ c = \sum(\mathrm{bell}(|K^{\mathrm{CS}}|) : K \subseteq W,\ K \cap X \neq \emptyset)$$

In the case of *regular substrate variables* V and *regular fud variables* $\mathrm{vars}(F) \setminus V$, having *valency* d, *dimension* $q = |W|$ and *intersecting dimension* $x = |X|$, the cardinality is

$$|\mathcal{N}_{U,W,X}| = 2^c \ :\ c = \sum_{k \in \{1 \ldots q\}} \left(\binom{q}{k} - \binom{q-x}{k} \right) \mathrm{bell}(d^k)$$

where the binomial coefficient is defined such that $\forall a, b \in \mathbf{N}\ (b > a \implies \binom{a}{b} = 0)$.

The infinite *non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}}$, may be similarly constructed by selecting only those *fuds* which are *non-overlapping*,

$$\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} = \{F : F \in \mathrm{elements}(\mathrm{tfi}(U)(V)),\ \neg\mathrm{overlap}(F)\}$$

Now consider the construction of the finite *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, with particular definitions of the *limited-models* constraints. The *limited-models* subset of the *functional definition sets* $\mathcal{F}_{\mathrm{q}} = \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}} \cap \mathcal{F}_{\mathrm{h}} \cap \mathcal{F}_{\mathrm{b}} \subset \mathcal{F}$ represents the class of subsets of the *functional definition sets* that are (i) *limited-underlying*, $\mathcal{F}_{\mathrm{u}}$, (ii) *limited-derived*, $\mathcal{F}_{\mathrm{d}}$, (iii) *limited-layer*, $\mathcal{F}_{\mathrm{h}}$ and (iv) *limited-breadth*, $\mathcal{F}_{\mathrm{b}}$. Here the *limited-models* constraints are defined explicitly.

In order to be computable, the infinite *partition infinite-layer fud tree*, $\mathrm{tfi}(U)(V) \in \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$, may be made finite by limiting the path length. Define the *maximum layer limit* as $\mathrm{lmax} \in \mathbf{N}_{>0}$. Define the finite *limited-layer* *partition infinite-layer fud tree* $\mathrm{tfih}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}} \times \mathbf{N} \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$ as

$$\begin{aligned}
&\mathrm{tfih}(U)(V,F,h) := \\
&\quad \{(F \cup G,\ \mathrm{tfih}(U)(V, F \cup G, h+1)) : \\
&\qquad\qquad G \subseteq \{P^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ P \in \mathrm{B}(K^{\mathrm{CS}})\}, \\
&\qquad\qquad G \neq \emptyset, \\
&\qquad\qquad h \leq \mathrm{lmax}\}
\end{aligned}$$

Let $\mathrm{tfih}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$ be defined $\mathrm{tfih}(U)(V) = \mathrm{tfih}(U)(V, \emptyset, 1)$.

The *layer* cardinality increments at each step in the tree's path, so the limited path length constraint, $h \leq \mathrm{lmax}$, could equally well be defined as a *limited-layer* constraint, $\mathrm{layer}(F \cup G, \mathrm{der}(F \cup G)) \leq \mathrm{lmax}$, although this computation would be longer.

The *fuds* of the tree are the *limited-layer infinite-layer substrate fuds*,

$$\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{h}} = \mathrm{elements}(\mathrm{tfih}(U)(V)) \cup \{\emptyset\}$$

Define the finite *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree* $\mathrm{tfiubh}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}} \times \mathbf{N} \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$ as

$$\begin{aligned}
&\mathrm{tfiubh}(U)(V,F,h) := \\
&\quad \{(F \cup G,\ \mathrm{tfiubh}(U)(V, F \cup G, h+1)) : \\
&\qquad\qquad G \subseteq \{P^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\}, \\
&\qquad\qquad 1 \leq |G| \leq \mathrm{bmax}, \\
&\qquad\qquad h \leq \mathrm{lmax}\}
\end{aligned}$$

where the *maximum underlying volume limit* is $\mathrm{xmax} \in \mathbf{N}_{>0}$ and the *maximum breadth limit* is $\mathrm{bmax} \in \mathbf{N}_{>0}$. Let $\mathrm{tfiubh}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$ be defined $\mathrm{tfiubh}(U)(V) = \mathrm{tfiubh}(U)(V, \emptyset, 1)$.

The finite set of *limited-models non-overlapping infinite-layer substrate fuds* is

$$\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} = \{F : F \in \mathrm{elements}(\mathrm{tfiubh}(U)(V)),\ \mathrm{nd}(F)\}$$

where $\mathrm{nd} \in \mathcal{F} \to \mathbf{B}$ is defined as $\mathrm{nd}(F) = \neg\mathrm{overlap}(F) \wedge (|W^{\mathrm{C}}| \leq \mathrm{wmax} : W = \mathrm{der}(F))$, and the *maximum derived volume limit* is $\mathrm{wmax} \in \mathbf{N}_{>0}$.

Both the *non-overlapping* and *limited-derived* constraints must be tested after the construction of the *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubh}(U)(V)$. In order to be selected only the last *fud* in a sublist $M \in \mathrm{subpaths}(\mathrm{tfiubh}(U)(V))$ need be *non-overlapping* and *limited-derived*, $\forall i \in \{1 \ldots |M|\}\ (M_i \in \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}})$ and $M_{|M|} \in \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{d}}$. That is, an ancestor *fud*, M_i where $i < |M|$, need not be *non-overlapping* nor *limited-derived*, so these constraints cannot be applied when constructing the tree in $\mathrm{tfiubh}(U)(V)$.

As in the case of the *partition infinite-layer fud tree* above, a *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree fud* $F \in \mathrm{elements}(\mathrm{tfiubh}(U)(V))$ may appear more than once in the tree if there are multiple paths to its construction,

$$|\{M : M \in \mathrm{subpaths}(\mathrm{tfiubh}(U)(V)),\ M_{|M|} = F\}| \geq 1$$

The cardinality of the finite set of *limited-models non-overlapping* subpaths in the tree must be greater than or equal to the cardinality of *limited-models non-overlapping infinite-layer substrate fuds*

$$|\{M : M \in \mathrm{subpaths}(\mathrm{tfiubh}(U)(V)),\ \mathrm{nd}(M_{|M|})\}| \geq |\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}|$$

A finite *computer* $I_{\mathrm{tfinq}} \in \mathrm{computers}$ can be defined such that its application to the *substrate variables*, V, constructs the *limited-models non-overlapping infinite-layer substrate fuds*, $I^{*}_{\mathrm{tfinq}}(V) = \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, by traversing the finite *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubh}(U)(V)$. Let the finite traversal enumeration $P \in \mathrm{enums}(\mathrm{subpaths}(\mathrm{tfiubh}(U)(V)))$ be such that the paths of the tree are searched in sequence, $\forall L, M \in \mathrm{subpaths}(\mathrm{tfiubh}(U)(V))\ (L \subseteq M \implies P_L \leq P_M)$. An example of such an enumeration, P, would be a *breadth*-first traversal of the tree, $\forall i \in \{1 \ldots \mathrm{lmax} - 1\}\ (\mathrm{maxr}(\{(M,j) : (M,j) \in P,\ |M| = i\}) < \mathrm{minr}(\{(M,j) : (M,j) \in P,\ |M| = i+1\}))$. Then the finite search list is $N = \{(j, M_{|M|}) : (M,j) \in P\} \in \mathcal{L}(\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}})$. The finite set of *limited-models non-overlapping infinite-layer substrate fuds* is obtained by filtering the search list, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} = \mathrm{set}(\mathrm{filter}(\mathrm{nd}, N))$. Thus the cardinality of the searched list is greater than or equal to the cardinality of the set of *limited-models non-overlapping* nodes which in turn is greater than or equal to the cardinality of *limited-models non-overlapping infinite-layer substrate fuds*, $|N| \geq |\mathrm{filter}(\mathrm{nd}, N)| \geq |\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}|$. The computation *time* is therefore greater than the cardinality of the searched list, $I^{\mathrm{t}}_{\mathrm{tfinq}}(V) > |N|$. Strictly speaking, the *limited-derived non-overlapping* filtering can take place during the construction of the last *layer*, lmax, instead of subsequent to *fud* construction, because there are no descendant *fuds*. So cumulative computation *space* is only required for the *fuds* in all but the last *layer*. If the filtering takes place after searching, however, then the computation *space* is also greater than the cardinality of the searched list, $I^{\mathrm{s}}_{\mathrm{tfinq}}(V) > |N|$. Also note that children *fuds* need not copy the *transforms* of their parents, so the *space* required for a *fud* and a descendant *fud* is less than the sum of the *spaces*.

The set of next *layer fuds*, $\{G : G \subseteq \{P^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\},\ 1 \leq |G| \leq \mathrm{bmax}\} \subset \mathcal{F}_{U,\mathrm{P}}$ may be defined in terms of a set of *partition-sets*. For the first *layer*, $F = \emptyset$, the set of next *layer fuds* is the set of *fuds* of *partition transforms* of the non-empty *partition-sets* of the intersection of the *limited-underlying-volume substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{xmax}}$, and the *limited-breadth substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{bmax}}$,

$$\{\{P^{\mathrm{T}} : P \in N\} : N \in \mathcal{N}_{U,V,\mathrm{xmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}},\ N \neq \emptyset\} \subset \mathcal{F}_{U,\mathrm{P}}$$

which has cardinality $|\mathcal{N}_{U,V,\mathrm{xmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}}| - 1$. The *limited-underlying-volume substrate partition-sets set* is defined,

$$\mathcal{N}_{U,V,\mathrm{xmax}} = \mathrm{P}(\{P : K \subseteq V,\ |K^{\mathrm{CS}}| \leq \mathrm{xmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\})$$

The *limited-breadth substrate partition-sets set* is defined,

$$\mathcal{N}_{U,V,\mathrm{bmax}} = \{N : N \in \mathcal{N}_{U,V},\ |N| \leq \mathrm{bmax}\}$$

The cardinality of the intersection is

$$|\mathcal{N}_{U,V,\mathrm{xmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}}| = \left(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\right)\ :\ c = \sum(\mathrm{bell}(|K^{\mathrm{CS}}|) : K \subseteq V,\ |K^{\mathrm{CS}}| \leq \mathrm{xmax})$$

In the case of *pluri-valent regular variables* V, having *valency* $d > 1$ and *dimension* n, if the implied *underlying-dimension limit*, $\mathrm{kmax} = \ln \mathrm{xmax} / \ln d$, is integral, $\ln \mathrm{xmax} / \ln d \in \mathbf{N}$, then the cardinality is

$$|\mathcal{N}_{U,V,\mathrm{xmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}}| = \left(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\right)\ :\ c = \sum_{k \in \{0 \ldots \mathrm{kmax}\}} \binom{n}{k} \mathrm{bell}(d^k)$$

For higher *layers*, $F \neq \emptyset$, the set of next *layer fuds* corresponds to the intersection of the *intersecting substrate partition-sets set*, $\mathcal{N}_{U,W,X}$, the *limited-underlying-volume substrate partition-sets set*, $\mathcal{N}_{U,W,\mathrm{xmax}}$, and the *limited-breadth substrate partition-sets set* $\mathcal{N}_{U,W,\mathrm{bmax}}$

$$\{\{P^{\mathrm{T}} : P \in N\} : N \in \mathcal{N}_{U,W,X} \cap \mathcal{N}_{U,W,\mathrm{xmax}} \cap \mathcal{N}_{U,W,\mathrm{bmax}}\} \subset \mathcal{F}_{U,\mathrm{P}}$$

where $W = \text{vars}(F) \cup V$ and $X = \text{der}(F)$. The cardinality of the set of higher *layer fuds* is $|\mathcal{N}_{U,W,X} \cap \mathcal{N}_{U,W,\text{xmax}} \cap \mathcal{N}_{U,W,\text{bmax}}|$. The cardinality of the intersection is

$$|\mathcal{N}_{U,W,X} \cap \mathcal{N}_{U,W,\text{xmax}} \cap \mathcal{N}_{U,W,\text{bmax}}| =$$
$$\Big(\sum_{b \in \{0 \ldots \text{bmax}\}} \binom{c}{b} \Big) :$$
$$c = \sum (\text{bell}(|K^{\text{CS}}|) : K \subseteq W,\ K \cap X \neq \emptyset,\ |K^{\text{CS}}| \leq \text{xmax})$$

In the case of *regular substrate variables* V and *regular fud variables* $\text{vars}(F) \setminus V$, having *valency* d, *dimension* $q = |W|$ and *intersecting dimension* $x = |X|$, such that the implied *underlying-dimension limit* is integral, $\text{kmax} = \ln \text{xmax} \ / \ln d \in \mathbf{N}$, the cardinality is

$$|\mathcal{N}_{U,W,X} \cap \mathcal{N}_{U,W,\text{xmax}} \cap \mathcal{N}_{U,W,\text{bmax}}| =$$
$$\Big(\sum_{b \in \{0 \ldots \text{bmax}\}} \binom{c}{b} \Big) \ : \ c = \sum_{k \in \{1 \ldots \text{kmax}\}} \left(\binom{q}{k} - \binom{q-x}{k} \right) \text{bell}(d^k)$$

Consider the case where a *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree fud* $F \in \text{elements}(\text{tfiubh}(U)(V))$ is such that (i) it has lmax *layers*, $\text{layer}(F, \text{der}(F)) = \text{lmax}$, (ii) the first *layer* has *breadth* $\text{bmax}-n$, (iii) subsequent *layers* have *breadth* bmax, and (iv) the *variables* are *regular*, $\forall w \in \text{vars}(F)\ (|U_w| = d)$. In this case the cardinality of the *variables* is $|\text{vars}(F) \cup V| = \text{lmax} \times \text{bmax}$. The cardinality of the set of next *layer fuds* is

$$|\mathcal{N}_{U,W,X} \cap \mathcal{N}_{U,W,\text{xmax}} \cap \mathcal{N}_{U,W,\text{bmax}}| =$$
$$\Big(\sum_{b \in \{0 \ldots \text{bmax}\}} \binom{c}{b} \Big) :$$
$$c = \sum_{k \in \{1 \ldots \text{kmax}\}} \left(\binom{\text{lmax} \times \text{bmax}}{k} - \binom{(\text{lmax}-1) \times \text{bmax}}{k} \right) \text{bell}(d^k)$$

The cardinality of the selectable set, c, is therefore bounded

$$c < (\text{lmax} \times \text{bmax})^{\underline{\text{kmax}}} \times \text{bell}(\text{xmax})$$

This expression is dominated by the right-most term, bell(xmax), if $\text{lmax} \times \text{bmax} \leq \text{xmax}$, because $\text{kmax} < \text{xmax}$. The cardinality of the set of next *layer fuds* is also bounded,

$$|\mathcal{N}_{U,W,X} \cap \mathcal{N}_{U,W,\text{xmax}} \cap \mathcal{N}_{U,W,\text{bmax}}| < ((\text{lmax} \times \text{bmax})^{\underline{\text{kmax}}} \times \text{bell}(\text{xmax}))^{\underline{\text{bmax}}}$$

In case where the *maximum underlying volume* equals the *size*, $\mathrm{xmax} = z$, and the right-most term dominates, $\mathrm{lmax} \times \mathrm{bmax} \leq z$, the cardinality is comparable to z^{z^2}. Although finite and *tractable*, this cardinality may be *impracticable* if the computation *time* and *space* that it implies exceeds available resources.

In the case where the *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubh}(U)(V)$, is additionally constrained such that the *fuds* have *derived volume* less than or equal to the *maximum derived volume limit*, wmax, then a subset of the *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, may be constructed. Define the *limited-layer limited-derived-volume limited-underlying-volume limited-breadth partition infinite-layer fud tree* $\mathrm{tfiubhd}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}} \times \mathbf{N} \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$ as

$$\begin{aligned}
&\mathrm{tfiubhd}(U)(V,F,h) := \\
&\quad \{(F \cup G,\ \mathrm{tfiubhd}(U)(V, F \cup G, h+1)) : \\
&\qquad G \subseteq \{P^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\}, \\
&\qquad 1 \leq |G| \leq \mathrm{bmax}, \\
&\qquad W = \mathrm{der}(F \cup G),\ |W^{\mathrm{C}}| \leq \mathrm{wmax}, \\
&\qquad h \leq \mathrm{lmax}\}
\end{aligned}$$

Again, let $\mathrm{tfiubhd}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$ be defined $\mathrm{tfiubhd}(U)(V) = \mathrm{tfiubhd}(U)(V, \emptyset, 1)$.

The *limited-derived-volume* constraint, $|W^{\mathrm{C}}| \leq \mathrm{wmax}$, is applied at every *layer* of the *fud*, so only a subset of the *limited-models non-overlapping infinite-layer substrate fuds* is searched

$$\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} \supseteq \{F : F \in \mathrm{elements}(\mathrm{tfiubhd}(U)(V)),\ \neg\mathrm{overlap}(F)\}$$

For the first *layer*, $F = \emptyset$, the set of next *layer fuds* is the set of *fuds* of *partition transforms* of the non-empty *partition-sets* of the intersection of the *limited-underlying-volume substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{xmax}}$, the *limited-breadth substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{bmax}}$, and the *limited-derived-volume substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{wmax}}$,

$$\{\{P^{\mathrm{T}} : P \in N\} : N \in \mathcal{N}_{U,V,\mathrm{xmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}} \cap \mathcal{N}_{U,V,\mathrm{wmax}},\ N \neq \emptyset\} \subset \mathcal{F}_{U,\mathrm{P}}$$

which has cardinality $|\mathcal{N}_{U,V,\mathrm{xmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}} \cap \mathcal{N}_{U,V,\mathrm{wmax}}| - 1$.

The computation of the cardinality of the set of *fuds* in the higher next *layers* requires that the set itself be computed, because the *limited-derived-volume* constraint depends on both the given *fud*, F, and the next *layer fud*, G, for its determination. That is, in some cases the *derived variables* of the child *fud* intersect with the *derived variables* of the given *fud*, $|\mathrm{der}(F \cup G) \cap \mathrm{der}(F)| \geq 0$. However, the cardinality of the children *fuds* must be less than or equal to those of the *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubh}(U)(V)$,

$$\begin{aligned}
&|\{G : G \subseteq \{P^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\}, \\
&\qquad\qquad 1 \leq |G| \leq \mathrm{bmax},\ W = \mathrm{der}(F \cup G),\ |W^{\mathrm{C}}| \leq \mathrm{wmax}\}| \\
\leq\ &|\{G : G \subseteq \{P^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ P \in \mathrm{B}(K^{\mathrm{CS}})\}, \\
&\qquad\qquad 1 \leq |G| \leq \mathrm{bmax}\}|
\end{aligned}$$

The set of *substrate infinite-layer fud decompositions* $\mathcal{D}_{\mathrm{F},\infty,U,V}$ is defined such that all of the *fuds* are *infinite-layer substrate fuds* and none can appear more than once in a path

$$\begin{aligned}
\mathcal{D}_{\mathrm{F},\infty,U,V} = \{&D : D \in \mathcal{D}_{\mathrm{F,d}},\ \mathrm{fuds}(D) \subseteq \mathcal{F}_{\infty,U,V}, \\
&\forall L \in \mathrm{paths}(D)\ (\mathrm{maxr}(\mathrm{count}(\{(F,i) : (i,(\cdot,F)) \in L\})) = 1)\}
\end{aligned}$$

or equivalently,

$$\begin{aligned}
\mathcal{D}_{\mathrm{F},\infty,U,V} = \{&D : D \in \mathcal{D}_{\mathrm{F,d}},\ \mathrm{fuds}(D) \subseteq \mathcal{F}_{\infty,U,V}, \\
&\forall L \in \mathrm{paths}(D)\ (\{(i,F) : (i,(\cdot,F)) \in L\} \in \{1 \ldots |L|\} :\leftrightarrow: \mathrm{ran}(\mathrm{set}(L)))\}
\end{aligned}$$

The non-empty *infinite-layer substrate fud decompositions* of non-empty *infinite-layer substrate fuds* may be constructed by means of a tree of *immediate super-decompositions*. Define the infinite *infinite-layer fud decomposition tree* $\mathrm{tdfi}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{D}_{\mathrm{F,d}} \to \mathrm{trees}(\mathcal{D}_{\mathrm{F,d}})$ as

$$\begin{aligned}
&\mathrm{tdfi}(U)(V,D) := \\
&\quad \{(E,\ \mathrm{tdfi}(U)(V,E)) : \\
&\qquad\qquad Q = \mathrm{paths}(D),\ L \in Q,\ i \in \{1 \ldots |L|\}, \\
&\qquad\qquad (\cdot,F) = L_i,\ W = \mathrm{der}(F),\ S \in W^{\mathrm{CS}}, \\
&\qquad\qquad G \in \mathcal{F}_{\infty,U,V} \setminus (\mathrm{ran}(\mathrm{set}(L_{\{1 \ldots i\}})) \cup \{\emptyset\}), \\
&\qquad\qquad M = L_{\{1 \ldots i\}} \cup \{(i+1,(S,G))\}, \\
&\qquad\qquad E = \mathrm{tree}(Q \setminus \{L_{\{1 \ldots i\}}\} \cup \{M\}),\ E \in \mathcal{D}_{\mathrm{F,d}}\}
\end{aligned}$$

where $\mathrm{tdfi}(U)(V,\emptyset) := \{(E,\ \mathrm{tdfi}(U)(V,E)) : G \in \mathcal{F}_{\infty,U,V} \setminus \{\emptyset\},\ E = \{((\emptyset,G),\emptyset)\}\}$. Let $\mathrm{tdfi}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathcal{D}_{\mathrm{F,d}})$ be defined $\mathrm{tdfi}(U)(V) = \mathrm{tdfi}(U)(V,\emptyset)$.

The infinite *infinite-layer fud decomposition tree* is constrained to be a tree of *immediate super-decompositions*,

$$\begin{aligned}
&\text{tdfi}(U)(V,D) = \\
&\quad \{(E,\ \text{tdfi}(U)(V,E)) : \\
&\qquad\qquad E \in \mathcal{D}_{\text{F,d}},\ \text{fuds}(E) \subseteq \mathcal{F}_{\infty,U,V} \setminus \{\emptyset\}, \\
&\qquad\qquad \forall L \in \text{paths}(E)\ (L_{|L|} \notin \text{set}(L_{\{1\ldots|L|-1\}})), \\
&\qquad\qquad D \in \text{subtrees}(E),\ |\text{nodes}(E) \setminus \text{nodes}(D)| = 1\}
\end{aligned}$$

The *decompositions* of the tree are a subset of the *infinite-layer substrate fud decompositions*

$$\mathcal{D}_{\text{F},\infty,U,V} \supset \text{elements}(\text{tdfi}(U)(V))$$

The *decompositions* of the tree form a proper subset because the *empty fud* is excluded from the construction.

In this construction exactly one *fud* is added to the previous *decomposition* at each step. Thus the cardinality of the *fuds* equals the position in the path, $\forall L \in \text{paths}(\text{tdfi}(U)(V))\ \forall i \in \{1\ldots|L|\}\ (|\text{fuds}(L_i)| = i)$.

A constructed *decomposition*, $D \in \text{elements}(\text{tdfi}(U)(V))$, may appear more than once in the tree if there are multiple paths to its construction, $|\{M : M \in \text{subpaths}(\text{tdfi}(U)(V)),\ M_{|M|} = D\}| \geq 1$, because some *decompositions* have multiple *immediate sub-decompositions*.

The *limited-models infinite-layer substrate fud decompositions*, $\mathcal{D}_{\text{F},\infty,U_A,V_A} \cap \text{trees}(\mathcal{S} \times \mathcal{F}_{\text{q}})$, may also be constructed by means of a tree of *immediate super-decompositions*. Define the finite *limited-models infinite-layer fud decomposition tree* $\text{tdfiq}(U) \in \text{P}(\mathcal{V}_U) \times \mathcal{D}_{\text{F,d}} \to \text{trees}(\mathcal{D}_{\text{F,d}})$ as

$$\begin{aligned}
&\text{tdfiq}(U)(V,D) := \\
&\quad \{(E,\ \text{tdfiq}(U)(V,E)) : \\
&\qquad\qquad Q = \text{paths}(D),\ L \in Q,\ i \in \{1\ldots|L|\}, \\
&\qquad\qquad (\cdot,F) = L_i,\ W = \text{der}(F),\ S \in W^{\text{CS}}, \\
&\qquad\qquad G \in \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\text{q}} \setminus (\text{ran}(\text{set}(L_{\{1\ldots i\}})) \cup \{\emptyset\}), \\
&\qquad\qquad M = L_{\{1\ldots i\}} \cup \{(i+1,(S,G))\}, \\
&\qquad\qquad E = \text{tree}(Q \setminus \{L_{\{1\ldots i\}}\} \cup \{M\}),\ E \in \mathcal{D}_{\text{F,d}}\}
\end{aligned}$$

where $\text{tdfiq}(U)(V,\emptyset) := \{(E,\ \text{tdfiq}(U)(V,E)) : G \in \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\text{q}} \setminus \{\emptyset\},\ E = \{((\emptyset,G),\emptyset)\}\}$. Let $\text{tdfiq}(U) \in \text{P}(\mathcal{V}_U) \to \text{trees}(\mathcal{D}_{\text{F,d}})$ be defined $\text{tdfiq}(U)(V) =$

$\mathrm{tdfiq}(U)(V,\emptyset)$.

The *decompositions* of the tree are a subset of the *limited-models infinite-layer substrate fud decompositions*

$$\mathcal{D}_{\mathrm{F},\infty,U,V} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}}) \supset \mathrm{elements}(\mathrm{tdfiq}(U)(V))$$

The *limited-models infinite-layer fud decomposition tree*, $\mathrm{tdfiq}(U)(V)$, is finite. Therefore a *computer* $I_{\mathrm{tdfiq}} \in \mathrm{computers}$ that is defined such that its application to the *substrate variables*, V, constructs the *limited-models infinite-layer substrate fud decompositions*, $I_{\mathrm{tdfiq}}^{*}(V) \subset \mathcal{D}_{\mathrm{F},\infty,U,V} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}})$, by traversing the entire tree, $\mathrm{tdfiq}(U)(V)$, always terminates and is therefore also finite. That is, $\forall P \in \mathrm{enums}(\mathrm{tdfiq}(U)(V))\ (|P| < \infty)$ and so $|P| < I_{\mathrm{tdfiq}}^{\mathrm{t}}(V) < \infty$.

Similarly, the *limited-models non-overlapping infinite-layer substrate fud decompositions*, $\mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}))$, may also be constructed by the tree of *immediate super-decompositions* to contain only *non-overlapping fuds*. Define the finite *limited-models non-overlapping infinite-layer fud decomposition tree* $\mathrm{tdfinq}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{D}_{\mathrm{F,d}} \to \mathrm{trees}(\mathcal{D}_{\mathrm{F,d}})$ as

$$\begin{aligned}
&\mathrm{tdfinq}(U)(V,D) := \\
&\quad \{(E,\ \mathrm{tdfinq}(U)(V,E)) : \\
&\qquad\qquad Q = \mathrm{paths}(D),\ L \in Q,\ i \in \{1 \ldots |L|\}, \\
&\qquad\qquad (\cdot,F) = L_i,\ W = \mathrm{der}(F),\ S \in W^{\mathrm{CS}}, \\
&\qquad\qquad G \in \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} \setminus (\mathrm{ran}(\mathrm{set}(L_{\{1\ldots i\}})) \cup \{\emptyset\}), \\
&\qquad\qquad M = L_{\{1\ldots i\}} \cup \{(i+1,(S,G))\}, \\
&\qquad\qquad E = \mathrm{tree}(Q \setminus \{L_{\{1\ldots i\}}\} \cup \{M\}),\ E \in \mathcal{D}_{\mathrm{F,d}}\}
\end{aligned}$$

where $\mathrm{tdfinq}(U)(V,\emptyset) := \{(E,\ \mathrm{tdfinq}(U)(V,E)) : G \in \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} \setminus \{\emptyset\},\ E = \{((\emptyset,G),\emptyset)\}\}$. Let $\mathrm{tdfinq}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathcal{D}_{\mathrm{F,d}})$ be defined $\mathrm{tdfinq}(U)(V) = \mathrm{tdfinq}(U)(V,\emptyset)$.

The *decompositions* of the tree are a subset the *limited-models non-overlapping infinite-layer substrate fud decompositions*

$$\mathcal{D}_{\mathrm{F},\infty,U,V} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})) \supset \mathrm{elements}(\mathrm{tdfinq}(U)(V))$$

It is also the case that the *limited-models non-overlapping infinite-layer fud decomposition tree*, $\mathrm{tdfinq}(U)(V)$, is finite, because it is a subset of the finite *limited-models infinite-layer fud decomposition tree*, $\mathrm{tdfiq}(U)(V)$. That is,

$\text{elements}(\text{tdfinq}(U)(V)) \subset \text{elements}(\text{tdfiq}(U)(V))$.

A construction of a *fud decomposition tree* can be defined in terms of the finite *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\text{tfiubh}(U)(V)$. Define the finite *limited-derived non-overlapping limited-layer limited-underlying-volume limited-breadth infinite-layer fud decomposition tree* $\text{tdfiubhnd}(U) \in \text{P}(\mathcal{V}_U) \times \mathcal{D}_{\text{F,d}} \to \text{trees}(\mathbf{N}_{>0} \times \mathcal{D}_{\text{F,d}})$ as

$$\begin{aligned}
&\text{tdfiubhnd}(U)(V, D) := \\
&\quad \{((j, E),\ \text{tdfiubhnd}(U)(V, E)) : \\
&\qquad\qquad Q = \text{paths}(D),\ L \in Q,\ i \in \{1 \ldots |L|\}, \\
&\qquad\qquad (\cdot, F) = L_i,\ W = \text{der}(F),\ S \in W^{\text{CS}}, \\
&\qquad\qquad P \in \text{order}(D_{\text{tfiubh}}, \text{subpaths}(\text{tfiubh}(U)(V))), \\
&\qquad\qquad N = \{(j, M_{|M|}) : (M, j) \in P\}, \\
&\qquad\qquad j \in \{1 \ldots |N|\},\ N_j \notin \text{set}(L_{\{1 \ldots i\}}),\ \text{nd}(N_j), \\
&\qquad\qquad M = L_{\{1 \ldots i\}} \cup \{(i+1, (S, N_j))\}, \\
&\qquad\qquad E = \text{tree}(Q \setminus \{L_{\{1 \ldots i\}}\} \cup \{M\}),\ E \in \mathcal{D}_{\text{F,d}}\}
\end{aligned}$$

where

$$\begin{aligned}
&\text{tdfiubhnd}(U)(V, \emptyset) := \\
&\quad \{((j, E),\ \text{tdfiubhnd}(U)(V, E)) : \\
&\qquad\qquad P \in \text{order}(D_{\text{tfiubh}}, \text{subpaths}(\text{tfiubh}(U)(V))), \\
&\qquad\qquad N = \{(j, M_{|M|}) : (M, j) \in P\}, \\
&\qquad\qquad j \in \{1 \ldots |N|\},\ \text{nd}(N_j),\ E = \{((\emptyset, N_j), \emptyset)\}\}
\end{aligned}$$

Let $\text{tdfiubhnd}(U) \in \text{P}(\mathcal{V}_U) \to \text{trees}(\mathcal{D}_{\text{F,d}})$ be defined $\text{tdfiubhnd}(U)(V) = \text{tdfiubhnd}(U)(V, \emptyset)$.

Here search list $N \in \mathcal{L}(\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\text{u}} \cap \mathcal{F}_{\text{b}})$ is constructed given some order D_{tfiubh} on the subpaths of the finite *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\text{tfiubh}(U)(V)$, so that the search enumeration is $P \in \text{enums}(\text{subpaths}(\text{tfiubh}(U)(V)))$ and the finite search list is $N = \{(j, M_{|M|}) : (M, j) \in P\}$.

Again, the *decompositions* of the tree are a subset of the *limited-models non-overlapping infinite-layer substrate fud decompositions*

$$\mathcal{D}_{\text{F},\infty,U,V} \cap \text{trees}(\mathcal{S} \times (\mathcal{F}_{\text{n}} \cap \mathcal{F}_{\text{q}})) \supset \text{ran}(\text{elements}(\text{tdfiubhnd}(U)(V)))$$

but the cardinality *limited-derived non-overlapping limited-underlying limited-breadth infinite-layer fud decomposition tree* is greater than or equal to the cardinality of the *limited-models non-overlapping infinite-layer fud decomposition tree*, $|\text{tdfiubhnd}(U)(V)| \geq |\text{tdfinq}(U)(V)|$. Therefore a *computer* $I_{\text{tdfiubhnd}} \in \text{computers}$ that is defined such that its application to the *substrate variables*, V, constructs the *limited-models non-overlapping infinite-layer substrate fud decompositions*, $I^*_{\text{tdfiubhnd}}(V) \subset \mathcal{D}_{\text{F},\infty,U,V} \cap \text{trees}(\mathcal{S} \times (\mathcal{F}_{\text{n}} \cap \mathcal{F}_{\text{q}}))$, by traversing the *limited-derived non-overlapping limited-underlying limited-breadth infinite-layer fud decomposition tree*, $\text{tdfiubhnd}(U)(V)$, is such that $I^{\text{t}}_{\text{tdfiubhnd}}(V) > I^{\text{t}}_{\text{tdfinq}}(V)$.

Instead of constructing *non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\text{n}}$, from the *partition transforms* of *tuples*, consider constructing them from *contracted non-overlapping substrate transforms* of *tuples*. Define the infinite *contracted non-overlapping substrate transform infinite-layer fud tree* $\text{tfitn}(U) \in \text{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\text{P}^*} \to \text{trees}(\mathcal{F}_{U,\text{P}^*})$ as

$$\begin{aligned}
&\text{tfitn}(U)(V, F) := \\
&\quad \{(F \cup G,\ \text{tfitn}(U)(V, F \cup G)) : \\
&\qquad\qquad G \subseteq \{N^{\text{T}} : K \in \text{tuples}(V, F),\ N \in \mathcal{N}'_{U,K,\text{n}} \setminus \{\emptyset\}\}, \\
&\qquad\qquad G \neq \emptyset\}
\end{aligned}$$

Let $\text{tfitn}(U) \in \text{P}(\mathcal{V}_U) \to \text{trees}(\mathcal{F}_{U,\text{P}^*})$ be defined $\text{tfitn}(U)(V) = \text{tfitn}(U)(V, \emptyset)$.

Here the *weak non-overlapping substrate partition-sets set*, $\mathcal{N}'_{U,K,\text{n}}$, is defined

$$\mathcal{N}'_{U,K,\text{n}} = \{N : Y \in \text{B}'(N),\ N \in \prod_{J \in Y} \text{B}(J^{\text{CS}})\} \cup \{\emptyset\}$$

and the *non-overlapping substrate transforms set* is defined in terms of the *weak non-overlapping substrate partition-sets set*

$$\mathcal{T}_{U,K,\text{n}} = \{N^{\text{T}K} : N \in \mathcal{N}'_{U,K,\text{n}}\}$$

The tree is a tree of *multi-partition fuds*. The *partition fuds* are a subset of the *multi-partition fuds*, $\mathcal{F}_{U,\text{P}} \subset \mathcal{F}_{U,\text{P}^*}$, and the *non-overlapping substrate transforms* are a superset of the *partition transforms*, $\{P^{\text{T}} : P \in \text{B}(K^{\text{CS}})\} \subseteq \mathcal{T}_{U,K,\text{n}}$, so the infinite *non-overlapping infinite-layer substrate fuds* can be constructed from the *partition fuds* in the tree,

$$\begin{aligned}
&\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\text{n}} = \\
&\quad \{F : F \in \text{elements}(\text{tfitn}(U)(V)) \cap \mathcal{F}_{U,\text{P}},\ \neg\text{overlap}(F)\}
\end{aligned}$$

However, the infinite *non-overlapping infinite-layer substrate fuds* can also be constructed by *exploding* the *contracted non-overlapping substrate transforms* of the *multi-partition fuds*

$$\begin{aligned}
&\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} = \\
&\qquad \{F' : F \in \mathrm{elements}(\mathrm{tfitn}(U)(V)),\ F' = \mathrm{explode}(F),\ \neg\mathrm{overlap}(F')\}
\end{aligned}$$

where $\mathrm{explode}(F) := \{P^{\mathrm{T}} : (\cdot, W) \in F,\ P \in W\} \in \mathcal{F}_{U,\mathrm{P}}$. Note that, although *contracted non-overlapping substrate transforms* are being added, it is still necessary to explicitly test that the tree *exploded fuds* are *non-overlapping*, $\neg\mathrm{overlap}(\mathrm{explode}(F))$. The addition of *contracted non-overlapping substrate transforms* does not imply that ancestor *exploded fuds* are *non-overlapping*.

Although the resultant set of *non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}}$, is the same whether constructed with *partition transforms* or *contracted non-overlapping substrate transforms*, in the latter case more computation is required because $|\mathcal{T}_{U,K,\mathrm{n}}| \geq |\mathrm{B}(K^{\mathrm{CS}})|$. That is, the cardinality of possible construction paths may be greater,

$$\begin{aligned}
&|\{M : M \in \mathrm{subpaths}(\mathrm{tfitn}(U)(V)),\ M_{|M|} = F\}| \\
\geq\ &|\{M : M \in \mathrm{subpaths}(\mathrm{tfi}(U)(V)),\ M_{|M|} = \mathrm{explode}(F)\}|
\end{aligned}$$

where $F \in \mathrm{elements}(\mathrm{tfitn}(U)(V))$.

The cardinality of the *weak non-overlapping substrate partition-sets set* is twice that of the *non-overlapping substrate partition-sets set* plus one, $|\mathcal{N}'_{U,K,\mathrm{n}}| = 2 \times |\mathcal{N}_{U,K,\mathrm{n}}| + 1$. In the case of non-empty *tuple*, $K \neq \emptyset$, the cardinality of the *non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n}}$, is

$$|\mathcal{N}_{U,K,\mathrm{n}}| = \sum_{Y \in \mathrm{B}(K)} \prod_{J \in Y} |\mathrm{B}(J^{\mathrm{CS}})|$$

If the *underlying variables* are *regular*, having *dimension* $k = |K|$ and common *valency* d, $\{d\} = \{|U_x| : x \in K\}$, then the cardinality of the *non-overlapping substrate partition-sets set* is

$$|\mathcal{N}_{U,K,\mathrm{n}}| = \sum_{(L,c) \in \mathrm{bcd}(k)} \Big(c \prod_{(j,p) \in L} \mathrm{bell}(d^j)^p \Big)$$

where bcd = bellcd and the partition function cardinality function is $\mathrm{bellcd} \in \mathbf{N}_{>0} \to (\mathcal{L}(\mathbf{N}) \to \mathbf{N})$.

For the first *layer*, $F = \emptyset$, the cardinality of the set of next *layer fuds* is

$$\begin{aligned}&|\mathrm{P}(\{N^{\mathrm{T}} : K \subseteq V,\ N \in \mathcal{N}'_{U,K,\mathrm{n}},\ N \neq \emptyset\})| - 1 = \\ &\qquad 2^c - 1\ :\ c = \sum_{K \subseteq V} 2 \sum_{Y \in \mathrm{B}(K)} \prod_{J \in Y} \mathrm{bell}(|J^{\mathrm{C}}|)\end{aligned}$$

In the case of *regular substrate variables*, having *dimension* $k = |K|$ and *valency* d, the cardinality is

$$\begin{aligned}&|\mathrm{P}(\{N^{\mathrm{T}} : K \subseteq V,\ N \in \mathcal{N}'_{U,K,\mathrm{n}},\ N \neq \emptyset\})| - 1 = \\ &\qquad 2^c - 1\ :\ c = \sum_{k \in \{0\ldots n\}} 2\binom{n}{k} \sum_{(L,a) \in \mathrm{bcd}(k)} a \prod_{(j,p) \in L} \mathrm{bell}(d^j)^p\end{aligned}$$

For higher *layers*, $F \neq \emptyset$, the cardinality of the set of next *layer fuds* is

$$\begin{aligned}&|\mathrm{P}(\{N^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ N \in \mathcal{N}'_{U,K,\mathrm{n}},\ N \neq \emptyset\})| - 1 = \\ &\qquad 2^c - 1\ :\ c = \sum_{K \subseteq W,\ K \cap X \neq \emptyset} 2 \sum_{Y \in \mathrm{B}(K)} \prod_{J \in Y} \mathrm{bell}(|J^{\mathrm{C}}|)\end{aligned}$$

where $W = \mathrm{vars}(F) \cup V$ and $X = \mathrm{der}(F)$.

In the case of *regular substrate variables* V and *regular fud variables* $\mathrm{vars}(F) \setminus V$, having *valency* d, *dimension* $q = |W|$ and *intersecting dimension* $x = |X|$, the cardinality is

$$\begin{aligned}&|\mathrm{P}(\{N^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ N \in \mathcal{N}'_{U,K,\mathrm{n}},\ N \neq \emptyset\})| - 1 = \\ &\qquad 2^c - 1\ :\ c = \sum_{k \in \{1\ldots q\}} 2\left(\binom{q}{k} - \binom{q-x}{k}\right) \sum_{(L,a) \in \mathrm{bcd}(k)} a \prod_{(j,p) \in L} \mathrm{bell}(d^j)^p\end{aligned}$$

where the binomial coefficient is defined such that $\forall a, b \in \mathbf{N}\ (b > a \implies \binom{a}{b} = 0)$.

The construction of the finite *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, may also be made with *contracted non-overlapping substrate transforms* rather than *partition transforms*. Define the finite *limited-layer limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree* $\mathrm{tfitnubh}(U) \in \mathrm{P}(\mathcal{V}_U) \times$

$\mathcal{F}_{U,\mathrm{P}^*} \times \mathbf{N} \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ as

$$\begin{aligned}
&\mathrm{tfitnubh}(U)(V,F,h) := \\
&\quad \{(F \cup G,\ \mathrm{tfitnubh}(U)(V,F \cup G,h+1)) : \\
&\qquad\qquad G \subseteq \{N^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ N \in \mathcal{N}'_{U,K,\mathrm{n}} \setminus \{\emptyset\}\}, \\
&\qquad\qquad 1 \leq |\mathrm{explode}(G)| \leq \mathrm{bmax}, \\
&\qquad\qquad h \leq \mathrm{lmax}\}
\end{aligned}$$

Again, let $\mathrm{tfitnubh}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ be defined $\mathrm{tfitnubh}(U)(V) = \mathrm{tfitnubh}(U)(V,\emptyset,1)$.

The finite set of *limited-models non-overlapping infinite-layer substrate fuds* is

$$\begin{aligned}
&\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} = \\
&\quad \{F' : M \in \mathrm{subpaths}(\mathrm{tfitnubh}(U)(V)),\ F' = \mathrm{explode}(M_{|M|}),\ \mathrm{nd}(F')\}
\end{aligned}$$

Similarly to the case of the *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubh}(U)(V)$, above, a finite *computer* $I_{\mathrm{tfitnq}} \in \mathrm{computers}$ can be defined such that its application to the *substrate variables*, V, constructs the *limited-models non-overlapping infinite-layer substrate fuds*, $I^*_{\mathrm{tfitnq}}(V) = \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, by traversing the *limited-layer limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree*, $\mathrm{tfitnubh}(U)(V)$, such that all of the paths are searched in sequence. The cardinality of the *contracted non-overlapping substrate transform* searched list is greater than or equal to the cardinality of the *partition transform* searched list,

$$|\mathrm{subpaths}(\mathrm{tfitnubh}(U)(V))| \geq |\mathrm{subpaths}(\mathrm{tfiubh}(U)(V))|$$

so the computation *time* must be greater than or equal to that of the previous case, $I^{\mathrm{t}}_{\mathrm{tfitnq}}(V) \geq I^{\mathrm{t}}_{\mathrm{tfinq}}(V)$.

Given cardinality $b \in \mathbf{N}_{>0}$, the *fixed-breadth non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},b}$, applied to the *tuple*, K, is defined

$$\mathcal{N}_{U,K,\mathrm{n},b} = \{N : Y \in \mathrm{B}(K),\ N \in \prod_{J \in Y} \mathrm{B}(J^{\mathrm{CS}}),\ |N| = b\}$$

If the *underlying variables* are *regular*, having *dimension* $k = |K|$ and common *valency* d, $\{d\} = \{|U_x| : x \in K\}$, then the cardinality of the *fixed-breadth non-overlapping substrate partition-sets set* is

$$|\mathcal{N}_{U,K,\mathrm{n},b}| = \sum_{(L,c) \in \mathrm{sscd}(k,b)} \left(c \prod_{(j,p) \in L} \mathrm{bell}(d^j)^p\right)$$

where $\text{sscd} = \text{stircd}$ and the fixed cardinality partition function cardinality function is $\text{stircd} \in \mathbf{N}_{>0} \times \mathbf{N}_{>0} \to (\mathcal{L}(\mathbf{N}) \to \mathbf{N})$.

The cardinality of the set of next *layer fuds* is

$$\begin{aligned}
&|\{G : G \subseteq \{N^{\mathrm{T}} : K \in \text{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \text{xmax},\ N \in \mathcal{N}'_{U,K,\mathrm{n}},\ N \neq \emptyset\}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad 1 \leq |\text{explode}(G)| \leq \text{bmax}\}| \\
&\quad = \sum \Big(\prod 2|\mathcal{N}_{U,K,\mathrm{n},j}| : (K,j) \in Y,\ j > 0\Big) : \\
&\qquad\qquad b \in \{1 \ldots \text{bmax}\},\ X \in \mathrm{C}'(\{K : K \in \text{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \text{xmax}\}, b), \\
&\qquad\qquad Y \in X,\ \forall (K,j) \in Y\ (j \leq |K|))
\end{aligned}$$

where the weak composition function is $\mathrm{C}' \in \mathrm{P}(\mathcal{X}) \times \mathbf{N} \to \mathrm{P}(\mathcal{X} \to \mathbf{N})$. Note that the constraint $\forall (K,j) \in Y\ (j \leq |K|)$ is required because for some *tuples* the cardinality of *fixed-breadth non-overlapping substrate partition-sets set* is too small to admit all weak compositions, $|\mathcal{N}_{U,K,\mathrm{n},b}| < b$.

Consider the case where the *fuds* of the construction trees are constrained to consist of *recursively non-overlapping multi-partition transforms*. The construction of the finite *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, may also be made with *contracted non-overlapping substrate transforms* to form a tree of *recursively non-overlapping multi-partition transform fuds*. Define the finite *limited-layer limited-underlying-volume limited-breadth contracted recursively non-overlapping substrate transform infinite-layer fud tree* $\text{tfitrnubh}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}^*} \times \mathbf{N} \to \text{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ as

$$\begin{aligned}
&\text{tfitrnubh}(U)(V,F,h) := \\
&\quad \{(F \cup G,\ \text{tfitrnubh}(U)(V, F \cup G, h+1)) : \\
&\qquad\qquad G \subseteq \{N^{\mathrm{T}} : K \in \text{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \text{xmax},\ N \in \mathcal{N}'_{U,K,\mathrm{n}} \setminus \{\emptyset\}, \\
&\qquad\qquad\qquad \neg\text{overlap}(\text{depends}(\text{explode}(F \cup \{N^{\mathrm{T}}\}), N))\}, \\
&\qquad\qquad 1 \leq |\text{explode}(G)| \leq \text{bmax}, \\
&\qquad\qquad h \leq \text{lmax}\}
\end{aligned}$$

Let $\text{tfitrnubh}(U) \in \mathrm{P}(\mathcal{V}_U) \to \text{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ be defined $\text{tfitrnubh}(U)(V) = \text{tfitnubh}(U)(V, \emptyset, 1)$.

Now each *transform* in the *fud* is constrained to be *recursively non-overlapping*,

$$\forall F \in \text{elements}(\text{tfitrnubh}(U)(V))\ \forall (\cdot, W) \in F \\ (\neg\text{overlap}(\text{depends}(\text{explode}(F), W)))$$

but the *recursively non-overlapping multi-partition fuds* are a superset of the *partition fuds*, $\mathrm{elements}(\mathrm{tfiubh}(U)(V)) \subset \mathrm{elements}(\mathrm{tfitrnubh}(U)(V))$, because the *non-overlapping substrate transforms* are a superset of the *partition transforms*, $\{P^{\mathrm{T}} : P \in \mathrm{B}(K^{\mathrm{CS}})\} \subseteq \mathcal{T}_{U,K,\mathrm{n}}$, so the set of all of the *limited-models non-overlapping infinite-layer substrate fuds* is constructed,

$$\begin{aligned} &\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} = \\ &\qquad \{F' : F \in \mathrm{elements}(\mathrm{tfitrnubh}(U)(V)),\ F' = \mathrm{explode}(F),\ \mathrm{nd}(F')\} \end{aligned}$$

The cardinality of the *non-overlapping multi-partition fuds* is greater than or equal to the cardinality of the *recursively non-overlapping multi-partition fuds*,

$$|\mathrm{elements}(\mathrm{tfitnubh}(U)(V))| \geq |\mathrm{elements}(\mathrm{tfitrnubh}(U)(V))|$$

If the *recursively non-overlapping multi-partition fuds* are restricted to those that are *topped*, then only a subset of the *limited-models non-overlapping infinite-layer substrate fuds* is constructed,

$$\begin{aligned} \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} \supseteq \{&\mathrm{explode}(F) : F \in \mathrm{elements}(\mathrm{tfitrnubh}(U)(V,\emptyset)), \\ &(\exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))),\ W = \mathrm{der}(F),\ |W^{\mathrm{C}}| \leq \mathrm{wmax}\} \end{aligned}$$

The *topped recursively non-overlapping multi-partition fuds* are necessarily *non-overlapping* so there is no need to test $\neg\mathrm{overlap}(\mathrm{explode}(F))$.

The construction of a subset of the finite *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, may also be made with *contracted non-overlapping substrate transforms* to form a tree of *recursively non-overlapping pluri-derived-variate multi-partition transform fuds*. Define the finite *limited-layer limited-underlying-volume limited-breadth contracted recursively non-overlapping pluri-derived-variate substrate transform infinite-layer fud tree* $\mathrm{tfiptrnubh}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}^*} \times \mathbf{N} \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ as

$$\begin{aligned} &\mathrm{tfiptrnubh}(U)(V,F,h) := \\ &\quad \{(F \cup G,\ \mathrm{tfiptrnubh}(U)(V, F \cup G, h+1)) : \\ &\qquad\qquad G \subseteq \{N^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ N \in \mathcal{N}'_{U,K,\mathrm{n}}, \\ &\qquad\qquad\qquad |N| > 1, \\ &\qquad\qquad\qquad \neg\mathrm{overlap}(\mathrm{depends}(\mathrm{explode}(F \cup \{N^{\mathrm{T}}\}), N))\}, \\ &\qquad\qquad 1 \leq |\mathrm{explode}(G)| \leq \mathrm{bmax}, \\ &\qquad\qquad h \leq \mathrm{lmax}\} \end{aligned}$$

Let $\text{tfiptrnubh}(U) \in \mathrm{P}(\mathcal{V}_U) \to \text{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ be defined $\text{tfiptrnubh}(U)(V) = \text{tfitnubh}(U)(V, \emptyset, 1)$.

The *pluri-derived-variate* constraint means that only a subset of the *limited-models non-overlapping infinite-layer substrate fuds* is constructed,

$$\begin{aligned}&\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} \supseteq \\ &\qquad \{F' : F \in \text{elements}(\text{tfiptrnubh}(U)(V)),\ F' = \text{explode}(F),\ \text{nd}(F')\}\end{aligned}$$

The *limited-layer limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree*, $\text{tfitnubh}(U)(V)$, does not readily yield a polynomial-complexity computation of the *regular* cardinality set of next *layer fuds*. Consider a restricted variation that limits the *tuple derived dimension* to a parameter $\text{mmax} \in \mathbf{N}_{>0}$. The *tuple derived dimension limit* is also constrained such that the *breadth limit* is a multiple, $\text{bmax}/\text{mmax} \in \mathbf{N}_{>0}$. Define the finite *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree* $\text{tfitnmubh}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}^*} \times \mathbf{N} \to \text{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ as

$$\begin{aligned}&\text{tfitnmubh}(U)(V, F, h) := \\ &\quad \{(F \cup G,\ \text{tfitnmubh}(U)(V, F \cup G, h+1)) : \\ &\qquad\qquad G \subseteq \{N^{\mathrm{T}} : K \in \text{tuples}(V, F),\ |K^{\mathrm{C}}| \leq \text{xmax},\ N \in \mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}\}, \\ &\qquad\qquad 1 \leq |G| \leq \text{bmax}/\text{mmax}, \\ &\qquad\qquad h \leq \text{lmax}\}\end{aligned}$$

Again, let $\text{tfitnmubh}(U) \in \mathrm{P}(\mathcal{V}_U) \to \text{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ be defined $\text{tfitnmubh}(U)(V) = \text{tfitnmubh}(U)(V, \emptyset, 1)$.

Here the *limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}$, is defined as the *limited-breadth non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n},\mathrm{bmax}}$, applied to the *tuple*,

$$\mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}} = \{N : Y \in \mathrm{B}(K),\ |Y| \leq \text{mmax},\ N \in \prod_{J \in Y} \mathrm{B}(J^{\mathrm{CS}})\}$$

In the case where $\text{mmax} \leq |K|$,

$$\mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}} = \{N : m \in \{1 \ldots \text{mmax}\},\ Y \in \mathrm{S}(K, m),\ N \in \prod_{J \in Y} \mathrm{B}(J^{\mathrm{CS}})\}$$

In the case of *regular variables* K, having *valency* d and *dimension* k, the cardinality of the *limited-tuple-derived-dimension non-overlapping substrate partition-sets set* is

$$\begin{aligned}|\mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}| &= \sum \Big(\prod_{J \in Y} \mathrm{bell}(d^{|J|}) \Big) : m \in \{1 \dots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m) \\ &= \sum \Big(a \prod_{(j,p) \in L} \mathrm{bell}(d^j)^p \Big) : m \in \{1 \dots \mathrm{mmax}\},\ (L,a) \in \mathrm{sscd}(k,m)\end{aligned}$$

where sscd = stircd and the fixed cardinality partition function cardinality function is $\mathrm{stircd} \in \mathbf{N}_{>0} \times \mathbf{N}_{>0} \to (\mathcal{L}(\mathbf{N}) \to \mathbf{N})$.

The *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree* is defined with the *limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}$. This is a *strong partition-sets set* so it excludes the *empty transform*, $(\emptyset,\emptyset)$, and the *unary partition transform*, $\{\emptyset^{\mathrm{CS}}\}^{\mathrm{T}}$. The finite set of *strong limited-models non-overlapping infinite-layer substrate fuds* is

$$\begin{aligned}&\{F : F \in \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}},\ \{\emptyset^{\mathrm{CS}}\}^{\mathrm{T}} \notin F\} = \\ &\qquad \{F' : M \in \mathrm{subpaths}(\mathrm{tfitnmubh}(U)(V)),\ F' = \mathrm{explode}(M_{|M|}),\ \mathrm{nd}(F')\}\end{aligned}$$

The cardinality of the *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree* searched list is less than or equal to the cardinality of the *limited-layer limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree* searched list, because of the additional constraint,

$$|\mathrm{subpaths}(\mathrm{tfitnmubh}(U)(V))| \leq |\mathrm{subpaths}(\mathrm{tfitnubh}(U)(V))|$$

For the first *layer*, $F = \emptyset$, the cardinality of the set of next *layer fuds* is

$$\begin{aligned}&|\{G : G \subseteq \{N^{\mathrm{T}} : K \subseteq V,\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ N \in \mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}\}, \\ &\qquad\qquad 1 \leq |G| \leq \mathrm{bmax}/\mathrm{mmax}\}| \\ &= \Big(\sum_{b \in \{0 \dots \frac{\mathrm{bmax}}{\mathrm{mmax}}\}} \binom{c}{b} \Big) : \\ &\qquad c = \sum \Big(\prod_{J \in Y} \mathrm{bell}(|J^{\mathrm{C}}|) \Big) : K \subseteq V,\ |K^{\mathrm{CS}}| \leq \mathrm{xmax}, \\ &\qquad\qquad m \in \{1 \dots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m)\end{aligned}$$

In the case of *pluri-valent regular substrate variables*, having *dimension* $k = |K|$ and *valency* $d > 1$, if the implied *underlying-dimension limit*, $\mathrm{kmax} = \ln \mathrm{xmax} / \ln d$, is integral, $\ln \mathrm{xmax} / \ln d \in \mathbf{N}$, the cardinality is

$$|\{G : G \subseteq \{N^{\mathrm{T}} : K \subseteq V,\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ N \in \mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}\},\ 1 \leq |G| \leq \mathrm{bmax}/\mathrm{mmax}\}|$$

$$= \Bigg(\sum_{b \in \{0 \ldots \frac{\mathrm{bmax}}{\mathrm{mmax}}\}} \binom{c}{b}\Bigg) : c = \sum_{k \in \{0 \ldots \mathrm{kmax}\}} \binom{n}{k} \Big(\sum \Big(a \prod_{(j,p) \in L} \mathrm{bell}(d^j)^p\Big) : m \in \{1 \ldots \mathrm{mmax}\},\ (L,a) \in \mathrm{sscd}(k,m)\Big)$$

For higher *layers*, $F \neq \emptyset$, the cardinality of the set of next *layer fuds* is

$$|\{G : G \subseteq \{N^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ N \in \mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}\},\ 1 \leq |G| \leq \mathrm{bmax}/\mathrm{mmax}\}|$$

$$= \Bigg(\sum_{b \in \{0 \ldots \frac{\mathrm{bmax}}{\mathrm{mmax}}\}} \binom{c}{b}\Bigg) : c = \sum \Big(\prod_{J \in Y} \mathrm{bell}(|J^{\mathrm{C}}|)\Big) : K \subseteq W,\ K \cap X \neq \emptyset,\ |K^{\mathrm{CS}}| \leq \mathrm{xmax},\ m \in \{1 \ldots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m)$$

where $W = \mathrm{vars}(F) \cup V$ and $X = \mathrm{der}(F)$.

In the case of *regular substrate variables* V and *regular fud variables* $\mathrm{vars}(F) \setminus V$, having *valency* d, *dimension* $q = |W|$ and *intersecting dimension* $x = |X|$, the cardinality is

$$|\{G : G \subseteq \{N^{\mathrm{T}} : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ N \in \mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}\},\ 1 \leq |G| \leq \mathrm{bmax}/\mathrm{mmax}\}|$$

$$= \Bigg(\sum_{b \in \{0 \ldots \frac{\mathrm{bmax}}{\mathrm{mmax}}\}} \binom{c}{b}\Bigg) : c = \sum_{k \in \{1 \ldots \mathrm{kmax}\}} \Bigg(\binom{q}{k} - \binom{q-x}{k}\Bigg) \Big(\sum \Big(a \prod_{(j,p) \in L} \mathrm{bell}(d^j)^p\Big) : m \in \{1 \ldots \mathrm{mmax}\},\ (L,a) \in \mathrm{sscd}(k,m)\Big)$$

In section 'Substrate structures' above it is shown that the *strong non-overlapping substrate transforms set*, $\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} \subseteq \mathcal{T}_{U,V,\mathrm{n}}$, can be constructed explicitly in terms of *linear fuds* of a *strong non-overlapping substrate self-cartesian transform*, $\{N^{\mathrm{T}} : N \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n}}\}$, followed by a sequence of *strong self non-overlapping substrate decremented transforms*, $\{N^{\mathrm{T}} : N \in \mathcal{N}_{U,W,-} \cap \mathcal{N}_{U,W,\mathrm{n,s}}\}$. Let the finite set of *contracted decrementing linear non-overlapping fuds* $\mathcal{F}_{U,\mathrm{n},-,V} \subset \mathcal{F}_{U,\mathrm{P}*}$ be defined as

$$\begin{aligned}
&\mathcal{F}_{U,\mathrm{n},-,V} \\
&\quad= \{\{N^{\mathrm{T}} : (\cdot, N) \in L\} : M \in \mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,V,\mathrm{n}}, \\
&\qquad\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdec}(U)(M))\})\} \\
&\quad= \{\{N^{\mathrm{T}} : (\cdot, N) \in L\} : Y \in \mathrm{B}(V),\ M = \{K^{\mathrm{CS}\{\}} : K \in Y\}, \\
&\qquad\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdec}(U)(M))\})\}
\end{aligned}$$

where the tree of *self non-overlapping substrate decremented partition-sets* is defined $\mathrm{tdec}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathrm{P}(\mathcal{R}_U))$ as

$$\mathrm{tdec}(U)(M) := \{(N, \mathrm{tdec}(U)(N)) : N \in \mathcal{N}_{U,M,-} \cap \mathcal{N}_{U,M,\mathrm{n,s}}\}$$

and $\mathrm{tdec}(U)(\emptyset) := \emptyset$. Explicitly this is

$$\begin{aligned}
&\mathrm{tdec}(U)(M) := \{(N, \mathrm{tdec}(U)(N)) : \\
&\quad w \in M,\ Q \in \mathrm{decs}(\{w\}^{\mathrm{CS}\{\}}),\ N = \{Q\} \cup \{\{u\}^{\mathrm{CS}\{\}} : u \in M,\ u \neq w\}\}
\end{aligned}$$

where decs = decrements $\in \mathcal{R}_U \to \mathrm{P}(\mathcal{R}_U)$.

Then $\{N^{\mathrm{T}V} : N \in \mathcal{N}_{U,V,\mathrm{n}}\} = \{F^{\mathrm{T}V} : F \in \mathcal{F}_{U,\mathrm{n},-,V}\}$. Note that the *contracted decrementing linear non-overlapping fuds*, $\mathcal{F}_{U,\mathrm{n},-,V}$, are *multi-partition fuds*, $\mathcal{F}_{U,\mathrm{n},-,V} \subset \mathcal{F}_{U,\mathrm{P}^*}$, so are not necessarily *substrate fuds*, $\mathcal{F}_{U,V}$, because they do not necessarily consist of *partition transforms*, $\mathcal{F}_{U,V} \subset \mathcal{F}_{U,\mathrm{P}} \subset \mathcal{F}_{U,\mathrm{P}^*}$. The *transforms* are already *contracted* so the corresponding *substrate fuds* can be constructed $\{\mathrm{explode}(F) : F \in \mathcal{F}_{U,\mathrm{n},-,V}\} \subset \mathcal{F}_{U,V}$. Also, the computation of the *contracted decrementing linear non-overlapping fuds* would not need to check for *flattened partition transforms*, even if they were *substrate fuds*, because *partitions* in the *linear fuds* are necessarily distinct from all previous *partitions* in the sequence.

The cardinality of the *self non-overlapping substrate decremented partition-sets tree* may be computed by defining $\mathrm{tdeccd}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}))$ as

$$\mathrm{tdeccd}(U)(V) := \{((1, L), \mathrm{tdeccd}(1, L)) : L = \{(i, |U_v|) : (v, i) \in \mathrm{order}(D_{\mathrm{V}}, V)\}\}$$

where order D_{V} is such that $\mathrm{order}(D_{\mathrm{V}}, V) \in \mathrm{enums}(V)$, and $\mathrm{tdeccd} \in \mathbf{N} \times \mathcal{L}(\mathbf{N}) \to \mathrm{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}))$ as

$$\begin{aligned}&\mathrm{tdeccd}(k, L) := \{((m, M), \mathrm{tdeccd}(m, M)) :\\ &\quad i \in \{1 \ldots |L|\},\ L_i > 1,\ m = kL_i(L_i - 1),\ M = L \setminus \{(i, L_i)\} \cup \{(i, L_i - 1)\}\}\end{aligned}$$

The cardinalities of the nodes of the tree of *self non-overlapping substrate decremented partition-sets* is

$$|\mathrm{nodes}(\mathrm{tdec}(U)(V))| = \sum(p : L \in \mathrm{subpaths}(\mathrm{tdeccd}(U)(V)),\ (p, \cdot) = L_{|L|}) - 1$$

The cardinality of the *contracted decrementing linear non-overlapping fuds* is

$$\begin{aligned}&|\mathcal{F}_{U,\mathrm{n},-,V}|\\ &\quad = \sum_{Y \in \mathrm{B}(V)} (|\mathrm{nodes}(\mathrm{tdec}(U)(\{K^{\mathrm{CS}\{\}} : K \in Y\}))| + 1)\\ &\quad = \sum p : Y \in \mathrm{B}(V),\\ &\qquad\qquad L \in \mathrm{subpaths}(\mathrm{tdeccd}(U)(\{K^{\mathrm{CS}\{\}} : K \in Y\})),\ (p, \cdot) = L_{|L|}\end{aligned}$$

In the case of *regular substrate variables* of *valency* d and *dimension* n, the cardinality of the *contracted decrementing linear non-overlapping fuds* is

$$\begin{aligned}&|\mathcal{F}_{U,\mathrm{n},-,V}|\\ &\quad = \sum ap : (M, a) \in \mathrm{bcd}(n),\ R = \mathrm{reg}(d, M),\\ &\qquad\qquad L \in \mathrm{subpaths}(\{((1, R), \mathrm{tdeccd}(1, R))\}),\ (p, \cdot) = L_{|L|}\end{aligned}$$

where $\mathrm{reg} \in \mathbf{N} \times \mathcal{L}(\mathbf{N}) \to \mathcal{L}(\mathbf{N})$ converts a histogram of *regular* cardinalities to a list of *regular volumes*, $\mathrm{reg}(d, M) := \mathrm{concat}(\mathrm{flip}(\mathrm{order}(D_{\mathcal{L}(\mathbf{N})}, \{\{1 \ldots q\} \times \{d^j\} : (j, q) \in M\})))$, where $D_{\mathcal{L}(\mathbf{N})} \in \mathrm{enums}(\mathcal{L}(\mathbf{N}))$ is some order on integer lists.

Instead of constructing *non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}}$, from either (i) *partition transforms*, $\mathcal{T}_{U,\mathrm{P}}$, or (ii) *non-overlapping substrate transforms*, $\mathcal{T}_{U,V,\mathrm{n}}$, consider a construction with *contracted decrementing linear non-overlapping fuds*, $\mathcal{F}_{U,\mathrm{n},-,V}$. Define the infinite *contracted decrementing linear non-overlapping fuds infinite-layer fud tree* $\mathrm{tfifdn}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}^*} \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ as

$$\begin{aligned}&\mathrm{tfifdn}(U)(V, F) :=\\ &\quad \{(F \cup \bigcup Q,\ \mathrm{tfifdn}(U)(V, F \cup \bigcup Q)) :\\ &\qquad\qquad Q \subseteq \{H : K \in \mathrm{tuples}(V, F),\ H \in \mathcal{F}_{U,\mathrm{n},-,K}\},\\ &\qquad\qquad Q \neq \emptyset\}\end{aligned}$$

Note that, because the *contracted decrementing linear non-overlapping fuds* sometimes have more than one *layer*, the *layer* cardinality of the *fuds* no longer corresponds to the length of the construction path,

$$\forall L \in \text{subpaths}(\text{tfifdn}(U)(V))\ \forall i \in \{1 \ldots |L|\}\ (\text{layer}(L_i, \text{der}(L_i)) \geq i)$$

The *contracted decrementing linear non-overlapping fuds infinite-layer fud tree*, tfifdn$(U)(V)$, is constrained to contain only *strong fud* elements, so corresponding only to the *strong* subset of the *non-overlapping substrate fuds*

$$\begin{aligned}&\{F : F \in \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\text{n}},\ \{\emptyset^{\text{CS}}\}^{\text{T}} \notin F\} = \\ &\qquad \{F' : F \in \text{elements}(\text{tfifdn}(U)(V)),\ F' = \text{explode}(F),\ \neg\text{overlap}(F')\}\end{aligned}$$

Again note that, although *contracted decrementing linear non-overlapping fuds* are being added, it is still necessary to explicitly test that the tree *fuds* are *non-overlapping*, $\neg$overlap(explode(F)). The addition of *contracted decrementing linear non-overlapping fuds* does not imply that ancestor *fuds* are *non-overlapping*.

Even though only a *strong* subset of the *non-overlapping infinite-layer substrate fuds* is computed, the computation *time* is greatest when constructed with *contracted decrementing linear non-overlapping fuds*, rather than *partition transforms* or *contracted non-overlapping substrate transforms*, because $|\mathcal{F}_{U,\text{n},-,K}| \geq |\mathcal{T}_{U,K,\text{n}}| \geq |\text{B}(K^{\text{CS}})|$. There are sometimes multiple *contracted decrementing linear non-overlapping fuds* corresponding to a *non-overlapping substrate transform*, $\text{maxr}(\{(T, |Q|) : (T, Q) \in \{(F, F^{\text{TV}}) : F \in \mathcal{F}_{U,\text{n},-,K}\}^{-1}\}) \geq 1$, because of multiple *linear fud* paths to the same *non-overlapping substrate transform*. The cardinality of possible construction paths may be greater than when constructed with *partition transforms*,

$$\begin{aligned}&|\{M : M \in \text{subpaths}(\text{tfifdn}(U)(V)),\ M_{|M|} = F\}| \\ \geq\ &|\{M : M \in \text{subpaths}(\text{tfi}(U)(V)),\ M_{|M|} = \text{explode}(F)\}|\end{aligned}$$

where $F \in \text{elements}(\text{tfifdn}(U)(V))$.

The construction of a *strong* subset of the *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\text{n}} \cap \mathcal{F}_{\text{q}}$, may also be made with *contracted decrementing linear non-overlapping fuds*. Define the finite *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted decrementing linear non-overlapping fuds infinite-layer fud tree*

$\mathrm{tfifdnmubh}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,\mathrm{P}^*} \times \mathbf{N} \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ as

$$\begin{aligned}
&\mathrm{tfifdnmubh}(U)(V,F,h) := \\
&\quad \{(F \cup \bigcup Q,\ \mathrm{tfifdnmubh}(U)(V, F \cup \bigcup Q, h+1)) : \\
&\qquad\qquad Q \subseteq \{H : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ H \in \mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}\}, \\
&\qquad\qquad 1 \leq |Q| \leq \mathrm{bmax}/\mathrm{mmax}, \\
&\qquad\qquad h \leq \mathrm{lmax}\}
\end{aligned}$$

Again, let $\mathrm{tfifdnmubh}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}^*})$ be defined $\mathrm{tfifdnmubh}(U)(V) = \mathrm{tfifdnmubh}(U)(V, \emptyset, 1)$.

Here the finite set of *limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds* $\mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}$ is defined as

$$\begin{aligned}
&\mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}} \\
&\quad = \{\{N^{\mathrm{T}} : (\cdot, N) \in L\} : M \in \mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}, \\
&\qquad\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdec}(U)(M))\})\} \\
&\quad = \{\{N^{\mathrm{T}} : (\cdot, N) \in L\} : Y \in \mathrm{B}(K),\ |Y| \leq \mathrm{mmax},\ M = \{J^{\mathrm{CS}\{\}} : J \in Y\}, \\
&\qquad\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdec}(U)(M))\})\}
\end{aligned}$$

The set of *limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds* is defined with the *limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}$, which is defined as the *limited-breadth non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n},\mathrm{bmax}}$, applied to the *tuple*. In the case where $\mathrm{mmax} \leq k$, where $k = |K|$, the cardinality of the intersection of the *substrate self-cartesian partition-sets set* and the *limited-tuple-derived-dimension non-overlapping substrate partition-sets set* is

$$|\mathcal{N}_{U,V,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}| = \sum_{m \in \{1 \ldots \mathrm{mmax}\}} \mathrm{stir}(k, m)$$

where $\mathrm{stir} \in \mathbf{N}_{>0} \times \mathbf{N} \to \mathbf{N}_{>0}$ is the Stirling number of the second kind.

The cardinality of the *limited-tuple-derived-dimension contracted decrement-*

ing linear non-overlapping fuds is

$$\begin{aligned}
&|\mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}| \\
&\quad = \sum \big(|\mathrm{nodes}(\mathrm{tdec}(U)(\{J^{\mathrm{CS}\{\}} : J \in Y\}))| : \\
&\qquad\qquad m \in \{1 \dots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m)\big) + 1 \\
&\quad = \sum p : m \in \{1 \dots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m), \\
&\qquad\qquad L \in \mathrm{subpaths}(\mathrm{tdeccd}(U)(\{J^{\mathrm{CS}\{\}} : J \in Y\})),\ (p,\cdot) = L_{|L|}
\end{aligned}$$

In the case of *regular substrate variables* of *valency* d and *dimension* n, the cardinality of the *limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds* is

$$\begin{aligned}
&|\mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}| \\
&\quad = \sum ap : m \in \{1 \dots \mathrm{mmax}\},\ (M,a) \in \mathrm{sscd}(k,m),\ R = \mathrm{reg}(d,M), \\
&\qquad\qquad L \in \mathrm{subpaths}(\{((1,R),\mathrm{tdeccd}(1,R))\}),\ (p,\cdot) = L_{|L|}
\end{aligned}$$

where $k = |K|$.

The *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted decrementing linear non-overlapping fuds infinite-layer fud tree* $\mathrm{tfifdnmubh}(U)(V)$, is defined with the set of *strong limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds*, $\mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}$. These exclude the *empty transform*, $(\emptyset,\emptyset)$, and the *unary partition transform*, $\{\emptyset^{\mathrm{CS}}\}^{\mathrm{T}}$. The finite set of *strong limited-models non-overlapping infinite-layer substrate fuds* is

$$\begin{aligned}
&\{F : F \in \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}},\ \{\emptyset^{\mathrm{CS}}\}^{\mathrm{T}} \notin F\} = \\
&\quad \{F' : M \in \mathrm{subpaths}(\mathrm{tfifdnmubh}(U)(V)),\ F' = \mathrm{explode}(M_{|M|}),\ \mathrm{nd}(F')\}
\end{aligned}$$

The application of the *maximum layer limit* is applied to the position in the tree path, $h \leq \mathrm{lmax}$, rather than constraining the *layer* cardinality of the *fuds*, because the *contracted decrementing linear non-overlapping fuds* are purely a means of construction and so could be flattened to a single *layer*, $\mathrm{layer}(\{H^{\mathrm{T}}\},\mathrm{der}(\{H^{\mathrm{T}}\})) = 1$ where $H \in \mathcal{F}_{U,\mathrm{n},-,K}$. That is, the *decrementing* notionally takes place within each *layer*.

Similarly to the case (i) of the *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubh}(U)(V)$, and the case (ii) of the *limited-layer limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree*, $\mathrm{tfitnubh}(U)(V)$, above, a finite *computer* $I_{\mathrm{tfifdnq}} \in \mathrm{computers}$ can be defined such that its application to the *substrate variables*, V, constructs the *strong* subset of the *limited-models non-overlapping infinite-layer substrate fuds*, $I^*_{\mathrm{tfifdnq}}(V) \subseteq \mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, by traversing the finite *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted decrementing linear non-overlapping fuds infinite-layer fud tree*, $\mathrm{tfifdnmubh}(U)(V)$, such that all paths are searched in sequence. The cardinality of the *contracted decrementing linear non-overlapping fuds* searched list is greater than or equal to the cardinality of the *contracted non-overlapping substrate transform* searched list which in turn is greater than or equal to the cardinality of the *partition transform* searched list,

$$\begin{aligned}|\mathrm{subpaths}(\mathrm{tfifdnmubh}(U)(V))| &\geq |\mathrm{subpaths}(\mathrm{tfitnub}(U)(V))| \\ &\geq |\mathrm{subpaths}(\mathrm{tfiubh}(U)(V))|\end{aligned}$$

so the computation *time* must be greater than or equal to that of the previous cases, $I^{\mathrm{t}}_{\mathrm{tfifdnq}}(V) \geq I^{\mathrm{t}}_{\mathrm{tfitnq}}(V) \geq I^{\mathrm{t}}_{\mathrm{tfinq}}(V)$.

For the first *layer*, $F = \emptyset$, the cardinality of the set of next *layer fuds* is

$$\begin{aligned}&|\{Q : Q \subseteq \{H : K \subseteq V,\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ H \in \mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}\}, \\ &\hspace{20em} 1 \leq |Q| \leq \mathrm{bmax}/\mathrm{mmax}\}| \\ &= \Big(\sum_{b \in \{0 \ldots \frac{\mathrm{bmax}}{\mathrm{mmax}}\}} \binom{c}{b}\Big) : \\ &\qquad c = \sum p : K \subseteq V,\ |K^{\mathrm{CS}}| \leq \mathrm{xmax}, \\ &\qquad\qquad m \in \{1 \ldots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m), \\ &\qquad\qquad L \in \mathrm{subpaths}(\mathrm{tdeccd}(U)(\{J^{\mathrm{CS}\{\}} : J \in Y\})),\ (p,\cdot) = L_{|L|}\end{aligned}$$

In the case of *pluri-valent regular substrate variables*, having *dimension* $k = |K|$ and *valency* $d > 1$, if the implied *underlying-dimension limit*, $\mathrm{kmax} =$

$\ln \mathrm{xmax} \ / \ln d$, is integral, $\ln \mathrm{xmax} \ / \ln d \in \mathbf{N}$, the cardinality is

$$|\{Q : Q \subseteq \{H : K \subseteq V,\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ H \in \mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}\},$$
$$1 \leq |Q| \leq \mathrm{bmax}/\mathrm{mmax}\}|$$
$$= \Big(\sum_{b \in \{0 \ldots \frac{\mathrm{bmax}}{\mathrm{mmax}}\}} \binom{c}{b} \Big) :$$
$$c = \sum_{k \in \{0 \ldots \mathrm{kmax}\}} \binom{n}{k} \Big(\sum ap :$$
$$m \in \{1 \ldots \mathrm{mmax}\},\ (M,a) \in \mathrm{sscd}(k,m),\ R = \mathrm{reg}(d,M),$$
$$L \in \mathrm{subpaths}(\{((1,R),\mathrm{tdeccd}(1,R))\}),\ (p,\cdot) = L_{|L|} \Big)$$

For higher *layers*, $F \neq \emptyset$, the cardinality of the set of next *layer fuds* is

$$|\{Q : Q \subseteq \{H : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ H \in \mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}\},$$
$$1 \leq |Q| \leq \mathrm{bmax}/\mathrm{mmax}\}|$$
$$= \Big(\sum_{b \in \{0 \ldots \frac{\mathrm{bmax}}{\mathrm{mmax}}\}} \binom{c}{b} \Big) :$$
$$c = \sum p : K \subseteq W,\ K \cap X \neq \emptyset,\ |K^{\mathrm{CS}}| \leq \mathrm{xmax},$$
$$m \in \{1 \ldots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m),$$
$$L \in \mathrm{subpaths}(\mathrm{tdeccd}(U)(\{J^{\mathrm{CS}\{\}} : J \in Y\})),\ (p,\cdot) = L_{|L|}$$

where $W = \mathrm{vars}(F) \cup V$ and $X = \mathrm{der}(F)$.

In the case of *regular substrate variables* V and *regular fud variables* $\mathrm{vars}(F) \setminus V$, having *valency* d, *dimension* $q = |W|$ and *intersecting dimension* $x = |X|$, the cardinality is

$$|\{Q : Q \subseteq \{H : K \in \mathrm{tuples}(V,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ H \in \mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}\},$$
$$1 \leq |Q| \leq \mathrm{bmax}/\mathrm{mmax}\}|$$
$$= \Big(\sum_{b \in \{0 \ldots \frac{\mathrm{bmax}}{\mathrm{mmax}}\}} \binom{c}{b} \Big) :$$
$$c = \sum_{k \in \{1 \ldots \mathrm{kmax}\}} \left(\binom{q}{k} - \binom{q-x}{k} \right) \Big(\sum ap :$$
$$m \in \{1 \ldots \mathrm{mmax}\},\ (M,a) \in \mathrm{sscd}(k,m),\ R = \mathrm{reg}(d,M),$$
$$L \in \mathrm{subpaths}(\{((1,R),\mathrm{tdeccd}(1,R))\}),\ (p,\cdot) = L_{|L|} \Big)$$

4.22.2 Practicable shuffles

The application, $I^*_{z,\mathrm{p}}(A)$, of a *substrate histogram* $A \in \mathcal{A}_z$ in a *practicable inducer* $I_{z,\mathrm{p}} \in \mathrm{inducers}(z)$ requires that the *histogram* be practicably representable. For example, *substrate histogram*, A, may have a *binary map histogram representation* such that $I^{\mathrm{s}}_{z,\mathrm{p}}(A) \leq \mathrm{smax}$ where the *maximum space limit* is $\mathrm{smax} \in \mathbf{N}_{>0}$. In this case the representation *space* depends on the *effective volume*, $|A^{\mathrm{F}}|$.

In some cases the computation of the *independent histogram*, A^{X}, may be impracticable because the *effective volume*, $|A^{\mathrm{XF}}| = v = |V_A^{\mathrm{C}}|$, is too large for available resources, for example, if $v > \mathrm{smax}$. If this is the case, the computation of the *alignment* of the *histogram*, $\mathrm{algn}(A)$, is also impracticable, for example $I^{\mathrm{s}}_{\mathrm{a}}(A) > \mathrm{smax}$. Consider the case where any subset of the *cartesian*, A^{C}, of cardinality equal to the *size*, $\{B : B \subseteq A^{\mathrm{C}},\ \mathrm{size}(B) = z\}$, is practicably representable. In this case of practicable *size*, a *practicable inducer* may use a *shuffled histogram* as a proxy for the *independent histogram*. For example, given some *substrate transform* $T \in \mathcal{T}_{U_A,V_A}$, the computation of the *content alignment*, $\mathrm{algn}(A * T) - \mathrm{algn}(A^{\mathrm{X}} * T)$, may be approximated by the computation of $\mathrm{algn}(A * T) - \mathrm{algn}(B * T)$, where the *shuffled histogram*, B, approximates to the *independent*, $B \approx A^{\mathrm{X}}$, and the *effective volume* is practicable, $|B^{\mathrm{F}}| \leq z$.

Section 'Shuffled history', above, defines the function $\mathrm{shuffles} \in \mathcal{H} \to \mathrm{P}(\mathcal{H})$. Let the set of *shuffled histories* of the *substrate histogram* $A \in \mathcal{A}_z$ be $Q = \mathrm{shuffles}(\mathrm{history}(A)) \subset \mathcal{H}$. The *independent* of each of the *shuffled histories* is equal to that of the *independent histogram*, $\forall G \in Q\ \Diamond B = \mathrm{his}(G)\ (B^{\mathrm{X}} \equiv A^{\mathrm{X}})$, where $\mathrm{his} = \mathrm{histogram}$. If the *independent* is *integral*, $A \in \mathcal{A}_{z,\mathrm{xi}}$, there must exist *independent shuffles*, $\exists G \in Q\ \Diamond B = \mathrm{his}(G)\ (B = A^{\mathrm{X}})$. In this case, there exist *shuffles* having zero *alignment*, $\exists G \in Q\ \Diamond B = \mathrm{his}(G)\ (\mathrm{algn}(B) = 0)$.

As shown above in section 'Minimum alignment', the logarithm *expected exponential alignment* given *distribution histogram* of A^{X} is

$$\begin{aligned}\ln \mathrm{expected}(\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z))(\{(B, \exp(\mathrm{algn}(B))) : B \in \mathcal{A}_{U_A,\mathrm{i},V_A,z}\}) = \\ \ln \sum_{B \in \mathcal{A}_{U_A,\mathrm{i},V_A,z}} \mathrm{mpdf}(U_A)(A^{\mathrm{X}}, z)(B^{\mathrm{X}})\end{aligned}$$

and that the *expected alignment* in the case where the *independent* is *cartesian*, $A^{\mathrm{X}} = \mathrm{scalar}(z/v) * V_A^{\mathrm{C}}$, is such that

$$\begin{aligned}&\mathrm{expected}(\hat{Q}_{\mathrm{m},U_A}(V^{\mathrm{C}}, z))(\{(B, \mathrm{algn}(B)) : B \in \mathcal{A}_{U_A,\mathrm{i},V_A,z}\})\\ &\quad\leq\ \ln(z+v-1)! - \ln(v-1)! - v\ln(z/v)! - z\ln v\end{aligned}$$

where $v = |V_A^{\mathrm{CS}}|$. So conjecture that the *expected alignment* of the *shuffled histories* is also approximately subject to the same inequality,

$$\begin{aligned}&\mathrm{average}(\{(G, \mathrm{algn}(B)) : G \in Q,\ B = \mathrm{his}(G)\})\\ &\quad\leq\ \ln(z+v-1)! - \ln(v-1)! - v\ln(z/v)! - z\ln v\end{aligned}$$

If $z \ll v$ then the expression above approximates to $z\ln(v/z)$. In the case of a *regular volume* of *dimension* n and *valency* d, the *expected alignment* approximates to $zn\ln(d/z^{1/n})$. This may be compared to the *maximum alignment* which approximates to $z(n-1)\ln(d)$. So in this case, the inequality imposes little or no constraint. That is, if the *volume*, v, is impracticable, but the *size*, z, is practicable, the inequality above is not a practicable test of a randomly chosen *shuffle histogram* that ensures that its *alignment* does not exceed the *expected alignment*. In any case the computation of the *alignment* of the *shuffle histogram* is impracticable.

Even so, a *practicable inducer* may be implemented without guarantee that a randomly chosen *shuffle histogram* has an *alignment* that is small. Choose a *shuffle histogram* at random, L_r, where $X \in \mathrm{enums}(\mathrm{shuffles}(\mathrm{history}(A)))$, $L = \mathrm{map}(\mathrm{his}, \mathrm{flip}(X))$, $r \in \{1 \ldots z!^n\}$ and $n = |V_A|$. The randomly chosen *shuffle histogram*, L_r, is expected to have *alignment* that is near zero with respect to *maximum alignment*, $\mathrm{algn}(L_r) \approx \mathrm{expected}(\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z))(\{(B, \mathrm{algn}(B)) : B \in \mathcal{A}_{U_A,\mathrm{i},V_A,z}\}) \approx 0 = \mathrm{algn}(A^{\mathrm{X}})$. As the *size*, z, increases, the *alignment*, $\mathrm{algn}(L_r)$, decreases. Note that the computation of the *alignment* of the *shuffle histogram*, $\mathrm{algn}(L_r)$, is impracticable.

The confidence in the *shuffle histogram* may be increased in a *practicable inducer*, $I_{z,\mathrm{p}}$, by *scaling* the *sum* of *shuffle histograms*, $B = \mathrm{scalar}(1/|R|) * \sum_{r \in R} L_r$, where $R \subset \{1 \ldots z!^n\}$ and $R \neq \emptyset$. Note that the *scaled shuffle histogram*, B, is not necessarily *integral*. The *effective volume* of the *scaled shuffle histogram* is greater than or equal to the *effective volume* of the contributing *shuffle histograms*, $\forall r \in R\ (|B^{\mathrm{F}}| \geq |L_r^{\mathrm{F}}|)$. If all possible *shuffle histograms* are used the resultant *scaled shuffle histogram* is the *independent*, $\mathrm{scalar}(1/z!^n) * \sum_{r \in \{1 \ldots z!^n\}} L_r = A^{\mathrm{X}}$.

In addition, the *alignment* of the *scaled shuffle histogram*, B, may be tested for practicable subsets of the *substrate*. Choose the largest *substrate* subset cardinality $k \leq n$ such that the computation of the combinations, $\sum_{i \in \{1 \ldots k\}} n^{\underline{i}}/i! = |\{K : K \subseteq V,\ |K| \leq k\}|$, of *reduced alignments* is practicable. Then the highest *reduced alignment* is $\mathrm{maxr}(\{(K, \mathrm{algn}(B\%K)) : K \subseteq V,\ |K| \leq k\})$.

4.22.3 Optimisation

In order to find lower bounds on the computation *time* and *space* of implementations of the *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$, given *substrate histogram* $A \in \mathcal{A}_z$, section 'Substrate models computation', above, considers the computation of the *substrate models*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\mathrm{n} \cap \mathcal{F}_\mathrm{q}$, by explicitly defining the (i) *limited-models* constraints, and (ii) *layer*-ordered *limited-underlying limited-breadth infinite-layer substrate fuds trees*. Together these constrain the computation to be a two stage process of (i) computation of finite search lists of the *limited-layer limited-underlying limited-breadth infinite-layer substrate fuds* $N_A \in \mathcal{L}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\mathrm{u} \cap \mathcal{F}_\mathrm{b} \cap \mathcal{F}_\mathrm{h})$, where $\mathrm{flip}(N_A) \in \mathrm{enums}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\mathrm{u} \cap \mathcal{F}_\mathrm{b} \cap \mathcal{F}_\mathrm{h})$, and (ii) filtering subsequently applied to the search lists, $\mathrm{set}(\mathrm{filter}(\mathrm{nd}, N_A)) = \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\mathrm{n} \cap \mathcal{F}_\mathrm{q}$ where $\mathrm{nd}(F) = \neg\mathrm{overlap}(F) \wedge (|W^\mathrm{C}| \leq \mathrm{wmax} : W = \mathrm{der}(F))$.

In particular, the section 'Substrate models computation' defines these searches of the *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\mathrm{n} \cap \mathcal{F}_\mathrm{q}$: (i) the *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubh}(U)(V) \in \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$, which constructs the *layer*-ordered *fuds* from sets of *partition transforms* of the *tuple*, $\{P^\mathrm{T} : P \in \mathrm{B}(K^\mathrm{CS})\}$, and (ii) the *limited-layer limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree*, $\mathrm{tfitnubh}(U)(V) \in \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}*})$, which constructs the *fuds* with *non-overlapping substrate transforms* of the *tuples*, $\mathcal{T}_{U,K,\mathrm{n}}$.

Also defined are these *strong limited-models non-overlapping infinite-layer substrate fuds* searches: (iii) the *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree*, $\mathrm{tfitnmubh}(U)(V) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}*})$, for which there is a non-trivial computation of the *regular substrate* cardinalities, and (iv) the *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted decrementing linear non-overlapping fuds infinite-layer fud tree* $\mathrm{tfifdnmubh}(U)(V) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}*})$, which constructs the *fuds* with *strong limited-tuple-derived-dimension contracted decrementing linear* *non-overlapping fuds*, $\mathcal{F}_{U,\mathrm{n},-,K,\mathrm{mmax}}$, on the *tuples*.

The section also defines searches which are restricted to subsets of the *strong limited-models non-overlapping infinite-layer substrate fuds*: (v) the *limited-layer limited-derived-volume limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubhd}(U)(V) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$, which constructs the *layer*-ordered *fuds* from *tuple partition transforms* but constrains the *derived volume* of the *fuds*, and (vi) the *limited-layer limited-underlying-volume limited-breadth contracted recursively non-overlapping pluri-derived-variate substrate transform infinite-layer fud tree*, $\mathrm{tfiptrnubh}(U)(V) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}^*})$, in which the *fuds* are constructed with *non-overlapping substrate transforms* of the *tuples* that are *recursively non-overlapping* in the *dependent fud*.

In the case where the computation *time* and *space* requirements exceed available resources, a *practicable inducer* implementation of the *derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$, must choose a subset of the *substrate models*. That is, (i) computation *time* limits imply that only a searched selection $\mathrm{select}(T_A, N_A) \in \mathcal{L}(\mathrm{set}(N_A))$, where $T_A \subset \{1 \ldots |N_A|\}$, of the traversable list N_A may be computed, and (ii) computation *space* limits imply that only a further subset $\mathrm{select}(S_A(t), N_A)$, where $S_A(t) \subset T_A$, of these may be simultaneously represented at any step t of the computation.

Given that a selection of the traversable search list, N_A, is necessary in some circumstances, the choice of selection can be made according to various criteria. Let $P \in \mathcal{L}(\mathcal{X})$ be a tuple of parameters. Consider a *practicable derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q},P}$, which, given *substrate histogram* $A \in \mathcal{A}_z$, is defined such that the *substrate models* of its application is a subset of that of the *derived alignment valency-density non-overlapping fud inducer*, $\mathrm{dom}(I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q},P}(A)) \subseteq \mathrm{dom}(I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}(A))$. That is,

$$\begin{aligned}&I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q},P}(A) \subseteq \\ &\qquad \{(F, I^{*}_{\mathrm{a}}(A * F^{\mathrm{T}})/I^{*}_{\mathrm{cvl}}(F)) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}\}\end{aligned}$$

where $I^{*}_{\mathrm{cvl}}(F) := (I^{*}_{\approx\mathrm{pow}}((w, 1/m)) : W = \mathrm{der}(F),\ w = |W^{\mathrm{C}}|,\ m = |W|)$, and the *power approxer* $I_{\approx\mathrm{pow}} \in \mathrm{computers}$, is such that (i) $\mathrm{domain}(I_{\approx\mathrm{pow}}) = \mathbf{Q} \times \mathbf{Q}$, (ii) $\mathrm{range}(I_{\approx\mathrm{pow}}) = \mathbf{Q}$, and (iii) $I^{*}_{\approx\mathrm{pow}}((x, y)) \approx x^y$. The *practicable fud inducer* is defined in terms of the *alignmenter*, I_{a}, and the *power approxer*, $I_{\approx\mathrm{pow}}$, and so the application approximates to the application of the *tractable fud inducer*, $I^{*}_{\mathrm{a}}(A * F^{\mathrm{T}})/I^{*}_{\mathrm{cvl}}(F) \approx I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * F^{\mathrm{T}})/\mathrm{cvl}(F))$.

Let the *practicable inducer* have (i) computation *time* limit $I'^{\mathrm{t}}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q},P}(A) \leq \mathrm{tmax}$, where the *maximum time limit* is $\mathrm{tmax} \in \mathbf{N}_{>0}$, and (ii) computation *space* limit $I'^{\mathrm{s}}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q},P}(A) \leq \mathrm{smax}$, where the *maximum space limit* is $\mathrm{smax} \in \mathbf{N}_{>0}$. These limits are parameters of the *practicable inducer*, $\mathrm{tmax}, \mathrm{smax} \in \mathrm{set}(P)$.

In some cases the computation *time* and *space* limits will be such that the application of the *practicable inducer* will be a proper subset of the application of its corresponding *tractable inducer*, $|I'^{*}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q},P}(A)| < |I'^{*}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q}}(A)|$. This would be the case, for example, if $|I'^{*}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q}}(A)| > \mathrm{tmax}$. If it is also the case that the *maximum substrate models* are excluded from the *practicable inducer*, $\mathrm{dom}(I'^{*}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q},P}(A)) \cap \mathrm{maxd}(I'^{*}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q}}(A)) = \emptyset$, then the correlation between the *maximum functions* must be less than one, $\mathrm{corr}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q},P}) < 1$. The following sections consider various definitions of *practicable inducers* having different selection criteria and the effect on the correlation of the *maximum functions* between the *practicable inducer* and the corresponding *tractable inducer*.

Consider an implementation of a *practicable fud inducer*, $I'_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q},P}$, which, given *substrate histogram* $A \in \mathcal{A}_z$, optimises its subset of the *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, by first optimising the *limited-layer limited-underlying limited-breadth infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}$, and then filtering for the *limited-derived non-overlapping*, $\mathcal{F}_{\mathrm{d}} \cap \mathcal{F}_{\mathrm{n}}$. The optimisation is implemented by means of a *list maximiser*. See appendix 'Search and optimisation' for a definition of *list maximisers*. The maximisation is of a rational-valued left-total optimise function $X_{P,A,\mathrm{ad}}$ of the *limited-layer limited-underlying limited-breadth infinite-layer substrate fuds*, $X_{P,A,\mathrm{ad}} \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}} :\to \mathbf{Q}$. Given (i) *maximum optimise step cardinality* $\mathrm{omax} \in \mathbf{N}_{>0}$, such that $\mathrm{omax} < \mathrm{tmax}$, (ii) initial subset $R_{P,A,\mathrm{ad}} \subset X_{P,A,\mathrm{ad}}$, and (iii) neighbourhood function $P_{P,A,\mathrm{ad}} \in \mathrm{P}(X_{P,A,\mathrm{ad}}) \to \mathrm{P}(X_{P,A,\mathrm{ad}})$, the *list maximiser* $Z_{P,A,\mathrm{ad}} \in \mathrm{maximisers}(X_{P,A,\mathrm{ad}})$ is constructed

$$Z_{P,A,\mathrm{ad}} = \mathrm{maximiseLister}(X_{P,A,\mathrm{ad}}, P_{P,A,\mathrm{ad}}, \mathrm{top}(\mathrm{omax}), R_{P,A,\mathrm{ad}})$$

The cardinality of the elements of the *list maximiser* is constrained by the *maximum optimise step cardinality*,

$$|\mathrm{elements}(Z_{P,A,\mathrm{ad}})| \leq \mathrm{omax} \times |\mathrm{list}(Z_{P,A,\mathrm{ad}})|$$

where $\mathrm{elements}(Z) := \bigcup \mathrm{set}(\mathrm{list}(Z))$. Note that strictly speaking this is true only in the case where cardinality of the top(omax) function in each step is less than or equal to omax, $\forall Y \in \text{set}(\text{list}(Z_{P,A,\text{ad}}))\ (Y \in \text{dom}(X_{P,A,\text{ad}}) \leftrightarrow \mathbf{Q} \implies |Y| \leq \text{omax})$.

Given (iv) that the neighbourhood function, $P_{P,A,\text{ad}}$, is further constrained such that it terminates before the *maximum time limit*, the cardinality of the searched set is such that

$$|\text{elements}(Z_{P,A,\text{ad}})| \leq |\text{searched}(Z_{P,A,\text{ad}})| < \text{tmax}$$

where $\text{searched}(Z) := \bigcup\{P(Y) : Y \in \text{set}(\text{list}(Z))\} \cup R$.

The domain of the elements of the *list maximiser* is a subset of the *limited-layer limited-underlying limited-breadth infinite-layer substrate fuds*,

$$\text{dom}(\text{elements}(Z_{P,A,\text{ad}})) \subset \text{dom}(X_{P,A,\text{ad}}) = \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\text{u} \cap \mathcal{F}_\text{b} \cap \mathcal{F}_\text{h}$$

The subset of the *substrate fuds* is the further subset

$$\text{filter}(\text{nd}, \text{dom}(\text{elements}(Z_{P,A,\text{ad}}))) \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\text{n} \cap \mathcal{F}_\text{q}$$

So the domain of the searched set is the subset of the *limited-layer limited-underlying limited-breadth infinite-layer substrate fuds* search list,

$$\text{dom}(\text{searched}(Z_{P,A,\text{ad}})) = \text{set}(\text{select}(T_A, N_A))$$

where $\text{flip}(N_A) \in \text{enums}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\text{u} \cap \mathcal{F}_\text{b} \cap \mathcal{F}_\text{h})$ and $T \subset \{1 \ldots |N_A|\}$.

The *practicable inducer* is implemented

$$\begin{aligned} &I'^{*}_{z,\text{ad},\text{F},\infty,\text{n},\text{q},P}(A) = \\ &\quad \{(F, I^*_\text{a}(A * F^\text{T})/I^*_\text{cvl}(F)) : F \in \text{filter}(\text{nd}, \text{dom}(\text{elements}(Z_{P,A,\text{ad}})))\} \end{aligned}$$

The *fuds* optimise function, $X_{P,A,\text{ad}}$, cannot be simply a *derived alignment valency-density* function, $X_{P,A,\text{ad}} \neq \{(F, I^*_\text{a}(A*F^\text{T})/I^*_\text{cvl}(F)) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\text{u} \cap \mathcal{F}_\text{b} \cap \mathcal{F}_\text{h}\}$, because the *fuds* are not necessarily *non-overlapping* nor *limited-derived*, $\text{nd}(F)$, where $F \in \text{dom}(X_{P,A,\text{ad}})$. That is, if the *fud* is *overlapping* then the *derived alignment* may, for example, be purely *formal*, $\text{algn}(A * F^\text{T}) = \text{algn}(A^\text{X} * F^\text{T})$. This would be the case if the *fud* was *tautological*, $\text{tautology}(F^\text{T})$, which is allowed in the *infinite-layer fuds*, $\mathcal{F}_{\infty,U_A,V_A}$. Also, if the *derived volume* of the *fud* exceeds the *maximum derived volume limit*, $|W^\text{C}| > \text{wmax}$, the computation of the *independent derived*, $(A * F^\text{T})^\text{X}$, necessary to compute the *derived alignment*, $\mathrm{algn}(A * F^{\mathrm{T}})$, may be impracticable. However it is defined, the *fuds* optimise function, $X_{P,A,\mathrm{ad}}$, is constrained such that the *practicable fud inducer* is positively correlated with the *tractable fud inducer*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q},P}) \geq 0)$$

even if the *fuds* neighbourhood function, $P_{P,A,\mathrm{ad}}$, is purely arbitrary. In this way, the *practicable fud inducer* transitively satisfies the requirement that the *maximum functions* of *inducers* are positively correlated with the finite *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q},P}) \geq 0)$$

The *fuds* neighbourhood function, $P_{P,A,\mathrm{ad}}$, is considered to be arbitrary if the *maximiser* gain is zero, $\mathrm{optimum}(Z_{P,A,\mathrm{ad}}) = \mathrm{arbitrary}(Z_{P,A,\mathrm{ad}})$, where $\mathrm{optimum}(Z) := \mathrm{maxr}(\mathrm{elements}(Z))$ and $\mathrm{arbitrary}(Z) := \mathrm{average}(\{(Y, \mathrm{maxr}(Y)) : Y \subseteq X,\ |Y| = |\mathrm{searched}(Z)|\})$. If the gain of the neighbourhood function is greater than zero, $\mathrm{optimum}(Z_{P,A,\mathrm{ad}}) > \mathrm{arbitrary}(Z_{P,A,\mathrm{ad}})$, then the correlation of the *maximum functions*, $\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q},P})$, may be higher than would otherwise be the case. The discussion below considers various *limited-layer limited-underlying limited-breadth infinite-layer substrate fud* optimise functions and neighbourhood functions.

An optimisation of the two stage computation of the *substrate models*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, must (i) first compute a possibly *overlapping fuds* subset of the *limited-layer limited-underlying limited-breadth infinite-layer substrate fuds*, $\mathrm{select}(T_A, N_A) \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}$, and (ii) then compute a *non-overlapping fuds* subset of these by filtering, $\{F : F \in \mathrm{select}(T_A, N_A),\ \mathrm{nd}(F)\} \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$. In some definitions of the optimisation, computation *time* and *space* limits may constrain the cardinality of the possibly *overlapping fuds*, $|\mathrm{select}(T_A, N_A)|$, such that none are *non-overlapping*. That is, computational resources may be such that the filtered subset is empty, $\{F : F \in \mathrm{select}(T_A, N_A),\ \mathrm{nd}(F)\} \subset \{F : F \in \mathrm{select}(T_A, N_A),\ \neg\mathrm{overlap}(F)\} = \emptyset$. In this case, the *maximum function* of a *practicable inducer* would be undefined, $\max(I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q},P}(A)) = \emptyset$.

A *fud*, F, must be *pluri-derived-variate*, $|\mathrm{der}(F)| > 1$, if the *derived alignment* is non-zero, $\mathrm{algn}(A * F^{\mathrm{T}}) > 0$. Let $F \in \mathcal{F}_{U_A,\mathrm{P}*}$ be a *topped recursively non-overlapping contracted pluri-derived-variate multi-partition fud* such that its *explode* is a *non-overlapping substrate fud*, $\mathrm{explode}(F) \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}}$. That is, the *fud*, F, is subject to (i) *contracted transforms*, $\forall T \in F\ (T = T^{\%})$, (ii) *pluri-partition transforms*, $\forall T \in F\ (|\mathrm{der}(T)| > 1)$, (iii) *recursively non-overlapping*, $\forall T \in F\ (\neg\mathrm{overlap}(\mathrm{depends}(\mathrm{explode}(F), \mathrm{der}(T))))$, (iv) *topped*, $\exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$, and (v) *underlying variables* are in the *substrate*, $\mathrm{und}(F) \subseteq V_A$. Then the cardinality of the *substrate* $n = |V_A|$ implies a maximum *layer*, $\mathrm{layer}(F, \mathrm{der}(F)) \leq \lfloor \log_2(n) \rfloor$, and a minimum *substrate* cardinality, $n \geq 2$. Thus the *maximum layer limit* is itself limited, $\mathrm{lmax} \leq \lfloor \log_2(n) \rfloor$, where *maximum function* is non-zero and the *topped recursively non-overlapping fud* does not contain *mono-partition transforms*.

Let *fud*, F, be further subject to (vi) *bi-underlying-variate transforms*, $\forall T \in F\ (|\mathrm{und}(T)| = 2)$, (vii) *bi-partition transforms*, $\forall T \in F\ (|\mathrm{der}(T)| = 2)$, (viii) the *fud layer* is $l = \mathrm{layer}(F, \mathrm{der}(F)) = \log_2(n) \in \mathbf{N}_{>0}$, and (ix) the cardinality of the set of *bi-partition transforms* in each *layer* is $n/2^i$ where $i \in \{1 \ldots l\}$. The cardinality of all such *topped recursively non-overlapping bi-underlying-variate bi-partition linear fuds* is less than $n!^{l/2}$. The cardinality of possibly *overlapping linear fuds* similarly constrained is less than $n^{nl/2}$. The fraction that are *non-overlapping* may be compared to $(n!/n^n)^{l/2}$. This fraction is less than 10^{-6} where $n = 16$. In fact, the *limited-layer limited-underlying limited-breadth infinite-layer multi-partition fuds* are only constrained such that there no more than *maximum breadth limit*, bmax, *transforms* in each *layer*, rather than $n/2^i$, so the constraint on the cardinality is greater than $n^{\mathrm{bmax} \times l/2}$. That is, if $\mathrm{bmax} > n$, the fraction of the possibly *overlapping* optimised *fuds* that are *non-overlapping* may be compared to the smaller fraction, $(n!/n^{\mathrm{bmax}})^{l/2}$.

A possible solution is to constrain the optimisation to construct only *recursively non-overlapping pluri-partition fuds* corresponding to *limited-layer limited-underlying limited-breadth infinite-layer substrate fuds*, $\mathrm{select}(T_A, N_A) \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}$. For example, section 'Substrate models computation' defines the *limited-layer limited-underlying-volume limited-breadth contracted recursively non-overlapping pluri-derived-variate substrate transform infinite-layer fud tree*, $\mathrm{tfiptrnubh}(U)(V) \in \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}*})$, which is a tree of *recursively non-overlapping pluri-partition fuds* constructed from *contracted non-overlapping substrate transforms*. This tree is such that

$$\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}} \supseteq \{\mathrm{explode}(F) : F \in \mathrm{elements}(\mathrm{tfiptrnubh}(U_A)(V_A))\}$$

Each of the *recursively non-overlapping pluri-partition fuds*

$$F \in \mathrm{elements}(\mathrm{tfiptrnubh}(U_A)(V_A))$$

can be *topped* by choosing a *transform* T in the top *layer*, $\mathrm{der}(T) \subseteq \mathrm{der}(F)$, so that $\mathrm{top}(\mathrm{depends}(F, \mathrm{der}(T))) = T$. Any such *fud* is necessarily *non-overlapping* and is itself in the *fud* tree

$$\mathrm{depends}(F, \mathrm{der}(T)) \in \mathrm{elements}(\mathrm{tfiptrnubh}(U_A)(V_A))$$

Thus the *non-overlapping* subset of the selection is not empty, $\{F : F \in \mathrm{select}(T_A, N_A),\ \neg\mathrm{overlap}(F)\} \neq \emptyset$, if the selection contains *topped fuds*. In the case where the *maximum derived volume limit* is greater than or equal to the *maximum underlying volume limit*, $\mathrm{wmax} \geq \mathrm{xmax}$, the filtered subset is not empty, $\{F : F \in \mathrm{select}(T_A, N_A),\ \mathrm{nd}(F)\} \neq \emptyset$, because the *top transform* satisfies the *limited-derived* constraint.

However, in some cases only a proper subset of the *substrate models*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, can be constructed when constrained to a *top transform*,

$$\begin{aligned}\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}} \supseteq \{&\mathrm{explode}(F) : F \in \mathrm{elements}(\mathrm{tfiptrnubh}(U_A)(V_A)),\\ &(\exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))),\ W = \mathrm{der}(F),\ |W^{\mathrm{C}}| \leq \mathrm{wmax}\}\end{aligned}$$

In section 'Transform alignment', above, a definition of *degree of overlap* is $\mathrm{alignmentOverlap}(U)(T, z) := \mathrm{algn}(\mathrm{resize}(z, V^{\mathrm{C}}) * T)$, where $T \in \mathcal{T}_{U,\mathrm{f},1}$ and $V = \mathrm{und}(T)$. An optimisation method that contains only *topped recursively non-overlapping multi-partition fuds* may exclude a *fud* $F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}$ with a small *degree of overlap*, $\mathrm{algn}(\mathrm{resize}(z, V_A^{\mathrm{C}}) * F^{\mathrm{T}})$, but high *derived alignment*, $\mathrm{algn}(A * F^{\mathrm{T}})$. Also, some or all of the *derived variables* of an excluded *fud*, $F \notin \mathrm{elements}(\mathrm{tfiptrnubh}(U_A)(V_A))$, may form the lower *layers* of a descendant *non-overlapping fud* $G \supset F$, where $\neg\mathrm{overlap}(\mathrm{explode}(G))$, subsequent in the search path.

In section 'Intractable literal substrate model inclusion', above, it is shown that the *formal-abstract equality* inclusion test, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, is intractable in the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I'_{z,\mathrm{a},\mathrm{l}}$, which computes the *derived alignment*, $\mathrm{algn}(A * T)$, for each of the *formal-abstract-equal ideal substrate transforms*, $\{T : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}$. The discussion considers (i) first weakening it to the *independent-formal* constraint, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, in the *derived alignment substrate ideal independent-formal transform inducer*, $I'_{z,\mathrm{a},\mathrm{fx},\mathrm{j}}$, which computes the *derived alignment*, $\mathrm{algn}(A * T)$, for each of the *independent-formal ideal substrate transforms*, $\{T : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}$, and (ii) then dropping it altogether in the *content alignment substrate ideal transform inducer*, $I'_{z,\mathrm{c},\mathrm{j}}$, which computes the *content alignment*, $\mathrm{algn}(A * T) - \mathrm{algn}(A^{\mathrm{X}} * T)$, for each of the *ideal substrate transforms*, $\{T : T \in \mathcal{T}_{U_A,V_A},\ A = A * T * T^{\dagger A}\}$. However, both (i) the *independent-formal* inclusion test, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, in the *independent-formal inducer*, $I'_{z,\mathrm{a,fx,j}}$, and (ii) the *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * T)$, in the *content alignment inducer*, $I'_{z,\mathrm{c,j}}$, remain intractable because of intractable *substrate volume*, $|A^{\mathrm{XF}}| = |V_A^{\mathrm{C}}|$. The discussion then considers the *derived alignment substrate ideal non-overlapping transform inducer*, $I'_{z,\mathrm{a,n,j}}$, in which the *substrate models* consist only of *non-overlapping transforms*, $\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$. That is, the *derived alignment substrate ideal non-overlapping transform inducer*, $I'_{z,\mathrm{a,n,j}}$, computes the *derived alignment*, $\mathrm{algn}(A * T)$, for each of the *ideal non-overlapping substrate transforms*, $\{T : T \in \mathcal{T}_{U_A,V_A,\mathrm{n}},\ A = A * T * T^{\dagger A}\}$. The discussion then proceeds to drop the *ideality* inclusion test, $A = A * T * T^{\dagger A}$, and consider the *midisation pseudo-alignment substrate independent-formal transform inducer*, $I_{z,\mathrm{m,fx}}$, and the *derived alignment valency-density substrate non-overlapping transform inducer*, $I'_{z,\mathrm{ad,n}}$, in order to partly recover the *formal-abstract equality*. The discussion eventually defines the *tractable derived alignment valency-density non-overlapping fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$.

However, section 'Practicable shuffles', above, considers the use of a *shuffle histogram* as practicable approximation to the *independent*, A^{X}. The computation of an approximation to the *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * F^{\mathrm{T}})$, is then practicable. Although there is no guarantee that a randomly chosen *shuffle histogram* has an *alignment* that is small, consider a *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P}$, which, given *substrate histogram* $A \in \mathcal{A}_z$, is defined

$$I'^{*}_{z,\mathrm{csd,F},\infty,\mathrm{q},P}(A) \subseteq \{(F, (I^{*}_{\mathrm{a}}(A * F^{\mathrm{T}}) - I^{*}_{\mathrm{a}}(A_R * F^{\mathrm{T}}))/I^{*}_{\mathrm{cvl}}(F)) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}\}$$

The *scaled shuffle histogram*, A_R, is defined $A_R = \mathrm{scalar}(1/|R|) * \sum_{r \in R} L_r$ where $X \in \mathrm{enums}(\mathrm{shuffles}(\mathrm{history}(A)))$, $L = \mathrm{map}(\mathrm{his}, \mathrm{flip}(X))$, $R \subseteq \{1 \ldots z!^n\}$ and $n = |V_A|$. The *shuffle indices*, R, are in the *practicable parameters*, $R \in \mathrm{set}(P)$. The cardinality of the *shuffle indices*, $|R|$, is chosen such that the *effective volume* of the *scaled shuffle histogram*, $|A_R^{\mathrm{F}}| \leq |A^{\mathrm{XF}}|$, is practicable. In the case where the entire *volume* of the *independent*, $|A^{\mathrm{XF}}| = |V_A^{\mathrm{C}}|$, is practicable then $R = \{1 \ldots z!^n\}$ and $A_R = A^{\mathrm{X}}$. In this case the *shuffle content alignment* equals the *content alignment*.

The computation of the *substrate models*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}$, in the *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P}$, still takes place in two stages but the filtering need only test for *limited-derived*.

That is (i) first compute the *limited-layer limited-underlying limited-breadth infinite-layer substrate fuds*, $\text{select}(T_A, N_A) \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\text{u}} \cap \mathcal{F}_{\text{b}} \cap \mathcal{F}_{\text{h}}$, and (ii) then compute a *limited-derived* subset of these by filtering, $\{F : F \in \text{select}(T_A, N_A),\ W = \text{der}(F),\ |W^{\text{C}}| \leq \text{wmax}\} \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\text{q}}$. In terms of an implementation with a *list maximiser* define

$$Z_{P,A,A_R,\text{csd}} = \text{maximiseLister}(X_{P,A,A_R,\text{csd}}, P_{P,A,A_R,\text{csd}}, \text{top}(\text{omax}), R_{P,A,A_R,\text{csd}})$$

The *limited-derived* is $\{F : F \in \text{dom}(\text{elements}(Z_{P,A,A_R,\text{csd}})),\ W = \text{der}(F),\ |W^{\text{C}}| \leq \text{wmax}\} \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\text{q}}$. The *fuds* optimise function, $X_{P,A,A_R,\text{csd}}$, is related to *content alignment valency-density*, $(\text{algn}(A * F^{\text{T}}) - \text{algn}(A^{\text{X}} * F^{\text{T}}))/\text{cvl}(F)$. However it is defined, the *fuds* optimise function, $X_{P,A,A_R,\text{csd}}$, is constrained such that the *practicable fud inducer* is positively correlated with the *tractable fud inducer*,

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(\text{maxr} \circ I_{z,\text{ad},\text{F},\infty,\text{n},\text{q}}^{'*}, \text{maxr} \circ I_{z,\text{csd},\text{F},\infty,\text{q},P}^{'*}) \geq 0)$$

However, the correlation of the *maximum functions* between the *tractable derived alignment valency-density non-overlapping fud inducer*, $I_{z,\text{ad},\text{F},\infty,\text{n},\text{q}}^{'*}$, and an *intractable content alignment valency-density fud inducer*, $I_{z,\text{cd},\text{F},\infty,\text{q}}^{'*}$, is imperfect,

$$\forall z \in \mathbf{N}_{>0}\ (\text{corr}(z)(\text{maxr} \circ I_{z,\text{ad},\text{F},\infty,\text{n},\text{q}}^{'*}, \text{maxr} \circ I_{z,\text{cd},\text{F},\infty,\text{q}}^{'*}) < 1)$$

where the *content alignment valency-density fud inducer*, $I_{z,\text{cd},\text{F},\infty,\text{q}}^{'*}$, is defined

$$\begin{aligned} &I_{z,\text{cd},\text{F},\infty,\text{q}}^{'*}(A) = \\ &\quad \{(F, I_{\approx\mathbf{R}}^{*}((\text{algn}(A * F^{\text{T}}) - \text{algn}(A^{\text{X}} * F^{\text{T}}))/\text{cvl}(F))) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\text{q}}\} \end{aligned}$$

So, depending on available computational resources, the *practicable shuffle content alignment valency-density fud inducer*, $I_{z,\text{csd},\text{F},\infty,\text{q},P}^{'}$, may have lower correlation than the *practicable derived alignment valency-density non-overlapping fud inducer*, $I_{z,\text{ad},\text{F},\infty,\text{n},\text{q},P}^{'}$.

In some cases the computation *time* of the *overlap* test, $\neg\text{overlap}(F)$, consisting of set intersection and union operations on the *substrate variables*, V_A, may exceed the computation *time* of the *shuffle formal alignment*, $\text{algn}(A_R * F^{\text{T}})$. So a *practicable derived alignment valency-density non-overlapping fud inducer*, $I_{z,\text{ad},\text{F},\infty,\text{n},\text{q},P}^{'}$, which is additionally constrained to construct only *non-overlapped fuds*, may have a smaller searched set than the *practicable shuffle content alignment valency-density fud inducer*, $I_{z,\text{csd},\text{F},\infty,\text{q},P}^{'}$.

The discussion below considers the optimisation in a *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P}$. The implementation of the optimisation is not restricted to a single *optimiser* such as the notional *list maximiser* $Z_{P,A,A_R,\mathrm{csd}}$.

Consider the *limited-path-models tuple partition practicable shuffle content alignment valency-density fud inducer* $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{p}}$ which is implemented by means of a *list maximiser* $Z_{P,A,A_R,\mathrm{csd,p}}$ that has a neighbourhood function that constructs the *fuds* in the *layer* sequence of the paths of the *limited-layer limited-derived-volume limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubhd}(U)(V) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$, described in section 'Substrate models computation' above. The *fud tree* constructs the *fuds* from *tuple partition transforms* but constrains the *derived volume* of the *fuds* by applying the *limited-derived-volume* constraint, $|W^{\mathrm{C}}| \leq \mathrm{wmax}$, at every *layer*, so only a subset of the *limited-models infinite-layer substrate fuds* is searched

$$\mathcal{F}_{\infty,U,V} \cap \mathcal{F}_{\mathrm{q}} \supseteq \mathrm{elements}(\mathrm{tfiubhd}(U)(V))$$

The *limited-path-models tuple partition list maximiser*, $Z_{P,A,A_R,\mathrm{csd,p}}$, is constructed

$$\begin{aligned}&Z_{P,A,A_R,\mathrm{csd,p}} = \\ &\quad \mathrm{maximiseLister}(X_{P,A,A_R,\mathrm{csd,p}}, P_{P,A,A_R,\mathrm{csd,p}}, \mathrm{top}(\mathrm{omax}), R_{P,A,A_R,\mathrm{csd,p}})\end{aligned}$$

The optimise function $X_{P,A,A_R,\mathrm{csd,p}} \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}} :\to \mathbf{Q}$, is the rational approximation to the *shuffle content alignment valency-density* valued total function of the *limited-models infinite-layer substrate fuds*, defined

$$X_{P,A,A_R,\mathrm{csd,p}} = \{(F, I^*_{\mathrm{csd}}((A, A_R, F))) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}\}$$

where the *shuffle content alignment valency-density computer* $I_{\mathrm{csd}} \in \mathrm{computers}$ is defined as

$$I^*_{\mathrm{csd}}((A, A_R, F)) = (I^*_{\mathrm{a}}(A * F^{\mathrm{T}}) - I^*_{\mathrm{a}}(A_R * F^{\mathrm{T}}))/I^*_{\mathrm{cvl}}(F)$$

The neighbourhood function $P_{P,A,A_R,\mathrm{csd,p}} \in \mathrm{P}(X_{P,A,A_R,\mathrm{csd,p}}) \to \mathrm{P}(X_{P,A,A_R,\mathrm{csd,p}})$, derived from the *limited-layer limited-derived-volume limited-underlying-volume* *limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubhd}(U)(V)$, is defined

$$
\begin{aligned}
&P_{P,A,A_R,\mathrm{csd,p}}(Q) = \\
&\quad \{(F \cup G, I^{*}_{\mathrm{csd}}((A, A_R, F \cup G))) : \\
&\qquad (F, \cdot) \in Q,\ \mathrm{layer}(F, \mathrm{der}(F)) < \mathrm{lmax}, \\
&\qquad G \subseteq \{P^{\mathrm{T}} : K \in \mathrm{tuples}(V_A, F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax},\ P \in \mathrm{B}(K^{\mathrm{CS}}),\ |P| \geq 2\}, \\
&\qquad 1 \leq |G| \leq \mathrm{bmax}, \\
&\qquad W = \mathrm{der}(F \cup G),\ |W^{\mathrm{C}}| \leq \mathrm{wmax}\}
\end{aligned}
$$

The definition of the neighbourhood function is stricter than the definition of the *fud tree* because only *pluri-valent partitions* are allowed, $|P| \geq 2$. This avoids the increase in *capacity* caused by the addition of *mono-valent variables* which increase the *dimension* but do not affect the *alignment*. The *pluri-valent* constraint also excludes empty *tuples*, $K = \emptyset$, because the *unary partition variable* is *mono-valent*, $|\{\emptyset^{\mathrm{CS}}\}| = 1$.

The initial function, $R_{P,A,A_R,\mathrm{csd,p}} \subset X_{P,A,A_R,\mathrm{csd,p}}$, is a singleton of the *empty fud*, $R_{P,A,A_R,\mathrm{csd,p}} = \{(\emptyset, 0)\}$.

Then the *limited-path-models tuple partition practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{p}}$, is defined,

$$I'^{*}_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{p}}(A) = \mathrm{elements}(Z_{P,A,A_R,\mathrm{csd,p}}) \subseteq X_{P,A,A_R,\mathrm{csd,p}}$$

The *fud inducer* is defined without the need for filtering the elements of the *list maximiser* because (i) the maximisation of the *shuffle content alignment* tends to minimise the *degree of overlap*, and (ii) the *list maximiser* limits the *fud derived volume*.

The *limited-derived-volume* constraint means that in some cases the optimise function is not completely traversable, $\mathrm{dom}(\mathrm{traversable}(Z_{P,A,A_R,\mathrm{csd,p}})) \subseteq \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}$ where $\mathrm{traversable}(Z) := \mathrm{elements}(\mathrm{searchLister}(X, P, R))$, and so the searched set is sometimes a proper subset, depending on the *substrate* and *limits*,

$$
\begin{aligned}
\mathrm{dom}(\mathrm{elements}(Z_{P,A,A_R,\mathrm{csd,p}})) &\subseteq \mathrm{dom}(\mathrm{searched}(Z_{P,A,A_R,\mathrm{csd,p}})) \\
&\subseteq \mathrm{dom}(\mathrm{traversable}(Z_{P,A,A_R,\mathrm{csd,p}})) \\
&\subseteq \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}
\end{aligned}
$$

The post-application of the *limited-derived-volume* test implies that the computation of the cardinality of the first *layer fuds* and subsequent *layer fuds* of a *regular substrate* of *valency* d has non-deterministic *time*. That is, the computation requires that the set itself be explicitly constructed. For the first *layer*, $F = \emptyset$, the set of next *layer fuds* corresponds to the non-empty *pluri-valent partition-sets* of the intersection of (i) the *lower-limited-valency substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{umin}}$, where $\mathrm{umin} = 2$, (ii) the *limited-underlying-volume substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{xmax}}$, (iii) the *limited-breadth substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{bmax}}$, (iv) and the *limited-derived-volume substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{wmax}}$,

$$|P_{P,A,A_R,\mathrm{csd,p}}(\{(\emptyset,\cdot)\})| = \\ |\mathcal{N}_{U,V,\overline{2}} \cap \mathcal{N}_{U_A,V_A,\mathrm{xmax}} \cap \mathcal{N}_{U_A,V_A,\mathrm{bmax}} \cap \mathcal{N}_{U_A,V_A,\mathrm{wmax}}| - 1$$

The cardinality has upper bounds equal to the cardinality of the non-empty *pluri-valent partition-sets* subset of the *partition-sets* corresponding to the first *layer fuds* of the *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubh}(U)(V)$. That is, the non-empty *pluri-valent partition-sets* subset of the intersection of the *limited-underlying-volume substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{xmax}}$, and the *limited-breadth substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{bmax}}$,

$$|P_{P,A,A_R,\mathrm{csd,p}}(\{(\emptyset,\cdot)\})| \leq |\mathcal{N}_{U,V,\overline{2}} \cap \mathcal{N}_{U_A,V_A,\mathrm{xmax}} \cap \mathcal{N}_{U_A,V_A,\mathrm{bmax}}| - 1$$

In the case of *pluri-valent regular variables* V, having *valency* $d > 1$ and *dimension* n, if the implied *underlying-dimension limit*, $\mathrm{kmax} = \ln \mathrm{xmax} \,/\, \ln d$, is integral, $\ln \mathrm{xmax} \,/\, \ln d \in \mathbf{N}$, then the cardinality of the intersection is

$$|\mathcal{N}_{U,V,\overline{2}} \cap \mathcal{N}_{U,V,\mathrm{xmax}} \cap \mathcal{N}_{U,V,\mathrm{bmax}}| = \\ \left(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b} \right) \;:\; c = \sum_{k \in \{0 \ldots \mathrm{kmax}\}} \binom{n}{k} (\mathrm{bell}(d^k) - 1)$$

For higher *layers*, computation of the cardinality requires that the set itself be explicitly constructed for the additional reason that the constraint, $|W^{\mathrm{C}}| \leq \mathrm{wmax}$ where $W = \mathrm{der}(F \cup G)$, depends on both the given *fud*, F, and the next *layer fud*, G, for its determination. For higher *layers* where the list element of the *list maximiser* is a singleton of a *non-empty fud* $F \neq \emptyset$, for example if $\mathrm{omax} = 1$, the set of next *layer fuds* corresponds to a subset of (i) the *lower-limited-valency substrate partition-sets set* $\mathcal{N}_{U,V,\mathrm{umin}}$, where $\mathrm{umin} = 2$, (ii) the intersection of the *intersecting substrate partition-sets set*, $\mathcal{N}_{U,W,X}$, (iii) the *limited-underlying-volume substrate partition-sets set*, $\mathcal{N}_{U,W,\mathrm{xmax}}$, and (iv) the *limited-breadth substrate partition-sets set* $\mathcal{N}_{U,W,\mathrm{bmax}}$

$$|P_{P,A,A_R,\mathrm{csd,p}}(\{(F,\cdot)\})| \leq |\mathcal{N}_{U,V,\overline{2}} \cap \mathcal{N}_{U_A,W,X} \cap \mathcal{N}_{U_A,W,\mathrm{xmax}} \cap \mathcal{N}_{U_A,W,\mathrm{bmax}}|$$

where $W = \mathrm{vars}(F) \cup V_A$ and $X = \mathrm{der}(F)$. In the case of *regular substrate variables* V and *regular fud variables* $\mathrm{vars}(F) \setminus V$, having *valency* d, *dimension* $q = |W|$ and *intersecting dimension* $x = |X|$, such that the implied *underlying-dimension limit* is integral, $\mathrm{kmax} = \ln \mathrm{xmax} / \ln d \in \mathbf{N}$, the cardinality of the intersection is

$$|\mathcal{N}_{U,V,\overline{2}} \cap \mathcal{N}_{U,W,X} \cap \mathcal{N}_{U,W,\mathrm{xmax}} \cap \mathcal{N}_{U,W,\mathrm{bmax}}| =$$
$$\left(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\right) \; : \; c = \sum_{k \in \{1 \ldots \mathrm{kmax}\}} \left(\binom{q}{k} - \binom{q-x}{k}\right)(\mathrm{bell}(d^k) - 1)$$

The degree of constraint imposed by the *limited-underlying-volume* of the *tuple*, $|K^{\mathrm{C}}| \leq \mathrm{xmax}$, depends on the *maximum derived volume limit* of the *fud*, wmax. For example, the *tuple self-partition* may be excluded, $\mathrm{wmax} \leq |K^{\mathrm{CS}\{\}}| = |K^{\mathrm{CS}}| \leq \mathrm{xmax}$. Similarly the degree of constraint imposed by the *limited-breadth*, $|G| \leq \mathrm{bmax}$, also depends on the *maximum derived volume*, wmax. For example, for implied *valency* of at least two, $\mathrm{bmax} \leq \lfloor \ln \mathrm{wmax} / \ln 2 \rfloor$. The *limited-derived-volume* constraint is weakest when the *maximum derived volume* is much greater than the *maximum underlying volume*, $\mathrm{xmax} \ll \mathrm{wmax}$, and the *maximum breadth*, $2^{\mathrm{bmax}} \ll \mathrm{wmax}$. The weaker the *limited-derived-volume* constraint, the smaller the difference between the next *layer* cardinalities and the computed upper bounds. However, the *maximum derived volume* is sometimes restricted to be less than or equal to the *sample size*, $\mathrm{wmax} < z$ where $z = \mathrm{size}(A)$, in order to avoid arguments to the unit-translated gamma function, $\Gamma_!$, that are less than one.

In the cases where the *substrate* and *fud variables* are not *regular*, the cardinalities may estimated with a *regular valency* equal to the geometric product of the *valencies* of the *variables*. That is, for the first *layer*

$$d = \left(\prod_{v \in V_A} |U_A(v)|\right)^{1/n}$$

where $n = |V_A|$. For subsequent *layers*

$$d = \left(\prod_{w \in W} |U_A(w)|\right)^{1/q}$$

where $W = \mathrm{vars}(F) \cup V_A$ and $q = |W|$. The implied *maximum underlying dimension* is $\mathrm{kmax} = \lceil \ln \mathrm{xmax} / \ln d \rceil$. If $\mathrm{omax} > 1$, the total cardinality may be estimated by summing the estimates of the cardinalities of the *fuds* in the list element, Q, which is the argument to the neighbourhood function. Only the cardinalities of the next *layer fuds* may be estimated. The geometric average *valency* of successive *layers* is not necessarily constant, although *alignment valency-density* maximisation tends to shorter *diagonals*. However, in some cases the *inducer time* may be constrained to be less than the *maximum time limit*, $I'^{\mathrm{t}}_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{p}}(A) \leq \mathrm{tmax}$, at least for the next *layer fuds*, by reducing the other parameters to reduce the estimated cardinality. For example the *maximum optimise step cardinality* may be restricted, $\mathrm{omax} = 1$.

The *limited-path-models tuple partition practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{p}}$, is implemented with a *list maximiser*, $Z_{P,A,A_R,\mathrm{csd},\mathrm{p}}$, but a *tree maximiser* could also be used if *time* and *space* is available for the elements of the tree. Note that an implementation of a *list maximiser* can lazily evaluate the list, but an implementation of a *tree maximiser* must evaluate its tree strictly because the set operations require instantiation of the elements of the set. That is, a *list maximiser* need only evaluate the elements of the *layer*-ordered list in sequence as they are required. A lazy solution for *tree maximiser* would be to implement it a with a list tree (see appendix 'Trees'), but then it may contain multiple instances of the same *fud*.

The *limited-path-models tuple partition practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{p}}$, has some limitations. First, in some cases only a subset of the *limited-models infinite-layer substrate fuds* is traversable

$$\mathrm{dom}(I'^{*}_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{p}}(A)) \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}$$

This is because the *derived volumes* of the intermediate *layers* of the *fud* are limited, as well as the *derived volume* of the top *layer*.

Secondly, this limitation of the intermediate *layers* may restrict the *derived dimensions* of these *layers* to be considerably less than the *dimension* of the *substrate*. For example, in the case where the *maximum derived volume* equals the *sample size*, $\mathrm{wmax} = z$, then a *size* of 10000 implies a *maximum breadth* of at most $\mathrm{bmax} = \lfloor \ln \mathrm{wmax} / \ln 2 \rfloor = 13$, where the *derived valency* is at least two. *Maximum alignment* is approximately $z(n-1)\ln d$ and so the ratio of the *maximum alignment* of an intermediate *layer* to the *maximum alignment* of the *substrate* is roughly equal to the ratio of their *dimensions*. For example, if the *substrate dimension* is $n = 26$, then the ratio of the *maximum alignments* is roughly half. This is also true for *maximum alignment* *valency-density*. Of course, this is also the case for the top *layer*, not just intermediate *layers*, but the *summation alignment* of the *decomposition* partially compensates for this.

Thirdly, as described in section 'Substrate models computation' above, consider the special case of a *fud* $F \in \mathrm{dom}(\mathrm{elements}(Z_{P,A,A_R,\mathrm{csd},\mathrm{p}}))$ which is such that (i) the *underlying variables* equals the *substrate*, $\mathrm{und}(F) = V$, (ii) it has lmax *layers*, $\mathrm{layer}(F, \mathrm{der}(F)) = \mathrm{lmax}$, (iii) the first *layer* has *breadth* $\mathrm{bmax} - n$, (iv) subsequent *layers* have *breadth* bmax, and (v) the *variables* are *regular*, $\forall w \in \mathrm{vars}(F)\ (|U_w| = d)$. In this case the cardinality of the *variables* is $|\mathrm{vars}(F) \cup V| = \mathrm{lmax} \times \mathrm{bmax}$. The cardinality of the set of next *layer fuds* is bounded

$$\begin{aligned}
&|\mathcal{N}_{U,W,X} \cap \mathcal{N}_{U,W,\mathrm{xmax}} \cap \mathcal{N}_{U,W,\mathrm{bmax}}| = \\
&\quad \left(\sum_{b \in \{0 \ldots \mathrm{bmax}\}} \binom{c}{b}\right) : \\
&\quad c = \sum_{k \in \{1 \ldots \mathrm{kmax}\}} \left(\binom{\mathrm{lmax} \times \mathrm{bmax}}{k} - \binom{(\mathrm{lmax}-1) \times \mathrm{bmax}}{k}\right) \mathrm{bell}(d^k)
\end{aligned}$$

The cardinality of the selectable set, c, is therefore bounded

$$c < (\mathrm{lmax} \times \mathrm{bmax})^{\underline{\mathrm{kmax}}} \times \mathrm{bell}(\mathrm{xmax})$$

This expression is dominated by the right-most term, bell(xmax), if $\mathrm{lmax} \times \mathrm{bmax} \leq \mathrm{xmax}$, because $\mathrm{kmax} < \mathrm{xmax}$. The cardinality of the set of next *layer fuds* is bounded,

$$|\mathcal{N}_{U,W,X} \cap \mathcal{N}_{U,W,\mathrm{xmax}} \cap \mathcal{N}_{U,W,\mathrm{bmax}}| < ((\mathrm{lmax} \times \mathrm{bmax})^{\underline{\mathrm{kmax}}} \times \mathrm{bell}(\mathrm{xmax}))^{\underline{\mathrm{bmax}}}$$

and so the upper bound on the cardinality of the neighbourhood function is also bounded

$$|P_{P,A,A_R,\mathrm{csd},\mathrm{p}}(\{(F, X_{P,A,A_R,\mathrm{csd},\mathrm{p}}(F))\})| < ((\mathrm{lmax}\times\mathrm{bmax})^{\underline{\mathrm{kmax}}}\times\mathrm{bell}(\mathrm{xmax}))^{\underline{\mathrm{bmax}}}$$

In the case where $\mathrm{xmax} = \mathrm{wmax} = z$ the second term, bell(xmax), equals $\mathrm{bell}(z)$, which is impracticable in the example above where $z = 10000$. The first term, $(\mathrm{lmax} \times \mathrm{bmax})^{\underline{\mathrm{kmax}}}$, has complexity $(\ln z)!$, but the second term, bell(xmax), has complexity $z!$. So while the *limited-path-models tuple partition practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{p}}$, strongly constrains the intermediate *layer derived volume*, the *tuple partition* cardinality is only weakly constrained.

Given the limitations of the *limited-path-models tuple partition practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{p}}$, and given that optimisation in a *practicable inducer* is necessary because of limited *time* or *space*, consider an alternative method of optimisation that maximises *maximum function* correlation by optimising within the *tuple*.

The cardinality of the second term of the upper bound on the cardinality of the neighbourhood function of the *limited-path-models tuple partition inducer*, $|\mathrm{B}(K^{\mathrm{CS}})| \leq \mathrm{bell}(\mathrm{xmax})$ where $K \in \mathrm{tuples}(V, F)$ and $|K^{\mathrm{C}}| \leq \mathrm{xmax}$, can be addressed by choosing only a subset Q of the *partitions* of a *tuple*, $Q \subset \mathrm{B}(K^{\mathrm{CS}})$. The cardinality of the subset must be less than or equal to the *maximum optimise step cardinality* $|Q| \leq \mathrm{omax}$. If the subset, Q, is chosen arbitrarily then the *practicable inducer* correlation can be expected to be reduced. However, considered by itself a single *partition* of the subset $P \in Q$ cannot have an optimise function that depends on *alignment* because a *fud* G constructed from it must be *independent*, $A * G^{\mathrm{T}} = (A * G^{\mathrm{T}})^{\mathrm{X}} \implies \mathrm{algn}(A * G^{\mathrm{T}}) = 0$, where $G = \mathrm{depends}(F \cup \{P^{\mathrm{T}}\}, \{P\}) = \mathrm{depends}(F, K) \cup \{P^{\mathrm{T}}\}$. This is because the constructed *fud*, G, is *mono-derived-variate*, $|\mathrm{der}(G)| = 1$.

Section 'Substrate models computation' defines the *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree*, $\mathrm{tfiubh}(U)(V) \in \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}})$, which constructs the *layer*-ordered *fuds* from sets of *partition transforms* of the *tuple*, $\{P^{\mathrm{T}} : P \in \mathrm{B}(K^{\mathrm{CS}})\}$. The discussion then goes on to define the *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree*, $\mathrm{tfitnmubh}(U)(V) \to \mathrm{trees}(\mathcal{F}_{U,\mathrm{P}*})$, which constructs the *fuds* with *non-overlapping substrate transforms* of the *tuples*, $\mathcal{T}_{U,K,\mathrm{n}}$, such that the *regular substrate* cardinalities may be computed. The latter *fud tree* constructs *transforms* of the *limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}}$, which is defined as the *limited-breadth non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,V,\mathrm{n},\mathrm{bmax}}$, applied to the *tuple*,

$$\mathcal{N}_{U,K,\mathrm{n},\mathrm{mmax}} \;=\; \{N : Y \in \mathrm{B}(K),\ |Y| \leq \mathrm{mmax},\ N \in \prod_{J \in Y} \mathrm{B}(J^{\mathrm{CS}})\}$$

Consider a stricter set of *transforms* of the *partition-sets* on the *tuple* such that the *derived variables* are *pluri-variate* and *pluri-valent*, which avoid necessarily *independent derived histograms* that have zero *alignment*. The *pluri-valent pluri-limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$, is the intersection of the *lower-limited-valency substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{umin}}$, where $\mathrm{umin} = 2$, and the *range-limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\mathrm{mran}}$, where $\mathrm{mran} = (2, \mathrm{mmax})$ and $\mathrm{mmax} \geq 2$, is defined

$$\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}} = \{N : Y \in \mathrm{B}(K),\ 2 \leq |Y| \leq \mathrm{mmax},\ N \in \prod_{J \in Y} (\mathrm{B}(J^{\mathrm{CS}}) \setminus \{\{J^{\mathrm{CS}}\}\})\}$$

In the case where $\mathrm{mmax} \leq |K|$, the cardinality is

$$|\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}| = \sum \Big(\prod_{J \in Y} (\mathrm{bell}(|J^{\mathrm{CS}}|) - 1) \Big) : m \in \{2 \ldots \mathrm{mmax}\},\ Y \in \mathrm{S}(K, m)$$

In the case of *regular variables* K, having *valency* d and *dimension* k, the cardinality of the *pluri-valent pluri-limited-tuple-derived-dimension non overlapping substrate partition-sets set* is

$$|\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}| = \sum \Big(a \prod_{(j,p) \in L} (\mathrm{bell}(d^j) - 1)^p \Big) : m \in \{2 \ldots \mathrm{mmax}\},\ (L, a) \in \mathrm{stircd}(k, m)$$

where the fixed cardinality partition function cardinality function is $\mathrm{stircd} \in \mathbf{N}_{>0} \times \mathbf{N}_{>0} \to (\mathcal{L}(\mathbf{N}) \to \mathbf{N})$.

Therefore, given *pluri-variate tuple*, $|K| > 1$, consider instead a subset of the *transforms* of the *pluri-valent pluri-limited-tuple-derived-dimension non-overlapping substrate partition-sets set* of the *tuple* $Q \subset \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}\}$ such that $|Q| \leq \mathrm{pmax}$, where the *maximum tuple optimise limit* is $\mathrm{pmax} = \lfloor \mathrm{omax}/\mathrm{bmax} \rfloor \in \mathbf{N}_{>0}$ and $\mathrm{omax} \geq \mathrm{bmax}$. The *derived variables* are *pluri-variate* and *pluri-valent*, $\forall T \in Q\ (|\mathrm{der}(T)| > 1)$, and $\forall T \in Q\ \forall w \in \mathrm{der}(T)\ (|U_w| > 1)$. Each of these *transforms* has a corresponding *partition*, $\forall T \in Q\ (T^{\mathrm{P}K} \in \mathrm{B}(K^{\mathrm{CS}}))$, but now are not restricted to zero *derived alignment*. That is, in some cases, $A * G^{\mathrm{T}} \neq (A * G^{\mathrm{T}})^{\mathrm{X}}$, where $T \in Q$ and $G = \mathrm{depends}(F \cup \{T\}, \mathrm{der}(T))$.

The *tuple transform*, $T \in Q$, is *non-overlapping*, $\forall P_1, P_2 \in \mathrm{der}(T)\ (P_1 \neq P_2 \implies \mathrm{vars}(P_1) \cap \mathrm{vars}(P_2) = \emptyset)$. That is, there exists a partition Y of the *tuple*, $Y \in \mathrm{B}(K)$, which is such that $\exists M \in \mathrm{der}(T) :\leftrightarrow Y\ \forall (P, J) \in M\ (\mathrm{vars}(P) = J)$. However, the constructed *exploded fud* $G' = \mathrm{explode}(G)$ is not necessarily *non-overlapping* if overlap(depends(F, K)), so the *tuple transforms*, Q, are chosen by maximising the *shuffle content alignment valency-density* rather than *derived alignment valency-density*.

Together the optimised subset is defined

$$Q = \text{topd}(\text{pmax})(\{(N^{\text{T}}, (\text{algn}(A * G^{\text{T}}) - \text{algn}(A_R * G^{\text{T}}))/\text{cvl}(G)) : \\ N \in \mathcal{N}_{U,K,\text{n},\overline{\text{b}},\text{mmax},\overline{2}},\ G = \text{depends}(F \cup \{N^{\text{T}}\}, N)\})$$

Maximising the *shuffle content alignment*, $\text{algn}(A * G^{\text{T}}) - \text{algn}(A_R * G^{\text{T}})$, tends to minimise the *formal alignment*, $\text{algn}(A^{\text{X}} * G^{\text{T}})$. The *formal alignment* is zero when *non-overlapping*, $\neg\text{overlap}(G') \implies \text{algn}(A^{\text{X}} * G^{\text{T}}) = 0$.

As mentioned above, an alternative method to *shuffle content alignment* would be to simply exclude *overlapping* constructed *exploded fuds*, $\text{overlap}(G')$, from the optimisation, but note that each *fud*, G', must be tested because it is insufficient to exclude it on the basis that the *tuple*, K, is *overlapping*, $\text{overlap}(\text{depends}(F, K)) \Longleftarrow \text{overlap}(G')$.

As shown in section 'Substrate structures', above, the cardinality of the *strong non-overlapping substrate transforms set* is bounded

$$\text{bell}(|K^{\text{C}}|) \leq |\{N^{\text{T}K} : N \in \mathcal{N}_{U,K,\text{n}}\}| \leq \text{bell}(|K|) \times \text{bell}(|K^{\text{C}}|)$$

That is, to search for the optimised subset from the *strong non-overlapping substrate transforms set*, $Q \subset \{N^{\text{T}} : N \in \mathcal{N}_{U,K,\text{n}}\}$, would require computation *time* in some cases of $\text{bell}(\text{kmax}) \times \text{bell}(\text{xmax})$, if the *tuple volume* equals the *maximum underlying volume limit*, $|K^{\text{C}}| = \text{xmax}$. The lower bound is the cardinality of the *partitions* of the *tuple*, $\text{bell}(\text{xmax})$. So the search for *non-overlapping substrate transforms* of the *tuple* requires more computation *time* than the search for *partition transforms* of the *tuple*. As shown in section 'Substrate models computation', above, the cardinality of the searched list of the *limited-layer limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree* $\text{tfitnubh}(U)(V)$, is greater than or equal to the cardinality of the searched list of the *limited-layer limited-underlying-volume limited-breadth partition infinite-layer fud tree* $\text{tfiubh}(U)(V)$,

$$|\text{subpaths}(\text{tfitnubh}(U)(V))| \geq |\text{subpaths}(\text{tfiubh}(U)(V))|$$

However, only a subset of the *derived histograms* of the *strong non-overlapping substrate transforms* are *non-independent*, and hence *aligned*, so only this subset is searched. Even in the case where the *maximum derived dimension* equals the *tuple dimension*, $\text{mmax} = k$, the exclusion of the *partitions* of the *unary partition* of the *tuple*, $\text{B}(K^{\text{CS}})$, because of the *maximum derived* *dimension*, mmin $= 2$, means that the cardinality of the *pluri-valent pluri-limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$, is less than or equal to the cardinality of the *partitions* of the *tuple*, $|\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}| \leq |\mathrm{B}(K^{\mathrm{CS}})|$. This is because the Bell number is log-convex.

In the case where computational resources are still exceeded by this cardinality, $|\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}|$, consider searches restricted to subsets of the *transforms* of the *pluri-valent pluri-limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$.

Consider the subset of the *pluri-valent pluri-limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$, where the partition of the *tuple* is given, in the case where the *tuple* is at least *bi-variate*, $|K| > 1$. Let the given partition be $Y \in \mathrm{B}(K) \setminus \{\{K\}\}$. Then

$$\prod_{J \in Y} (\mathrm{B}(J^{\mathrm{CS}}) \setminus \{\{J^{\mathrm{CS}}\}\}) \subset \mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$$

The cardinality of this set is $\prod_{J \in Y} (\mathrm{bell}(|J^{\mathrm{CS}}|) - 1) < \mathrm{bell}(\mathrm{xmax})$. The computation *time* is comparable at least to $\mathrm{bell}(\mathrm{xmax}^{1/\mathrm{mmax}})^{\mathrm{mmax}} < \mathrm{bell}(\mathrm{xmax})$.

Consider an optimisation where the search is broken into two separate searches. The first search determines the partition of the *tuple*, Y, by searching the *transforms* of the intersection of the *substrate self-cartesian partition-sets set*, $\mathcal{N}_{U,V,\mathrm{c}}$, and the *pluri-limited-tuple-derived-dimension non-overlapping substrate transforms set*, $\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax}}$, where $2 \leq \mathrm{mmax} \leq |K|$, which is

$$\mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax}} = \{\{J^{\mathrm{CS}\{\}} : J \in Y\} : m \in \{2 \ldots \mathrm{mmax}\},\ Y \in \mathrm{S}(K, m)\}$$

The cardinality of the intersection is

$$|\mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax}}| = \sum_{m \in \{2 \ldots \mathrm{mmax}\}} \mathrm{stir}(|K|, m)$$

If the *tuple* is *bi-variate* the cardinality is $\mathrm{stir}(2, 2) = 1$, so in this case the *tuple* partition search need not be performed. The *tuple* partition search is optimised by maximising the *shuffle content alignment valency-density*,

$$\begin{aligned}
Y \in \mathrm{maxd}(\{(Z, (\mathrm{algn}(A * G^{\mathrm{T}}) - \mathrm{algn}(A_R * G^{\mathrm{T}}))/\mathrm{cvl}(G)) :\\
m \in \{2 \ldots \mathrm{mmax}\},\ Z \in \mathrm{S}(K, m),\\
N = \{J^{\mathrm{CS}\{\}} : J \in Z\},\ G = \mathrm{depends}(F \cup \{N^{\mathrm{T}}\}, N)\})
\end{aligned}$$

Note that if the *tuple* partition search, Y, had been optimised by maximising the *derived alignment valency-density*, $\mathrm{algn}(A * G^{\mathrm{T}})/\mathrm{cvl}(G)$, instead of the *shuffle content alignment valency-density*, $(\mathrm{algn}(A * G^{\mathrm{T}}) - \mathrm{algn}(A_R * G^{\mathrm{T}}))/\mathrm{cvl}(G)$, then (i) the *valency capacity*, $\mathrm{cvl}(G)$, would not need to be computed because the *derived alignment*, $\mathrm{algn}(A*G^{\mathrm{T}})$, and the *derived alignment valency-density*, $\mathrm{algn}(A * G^{\mathrm{T}})/\mathrm{cvl}(G)$, are monotonic functions with respect to their common domain, $\mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\bar{\mathrm{b}},\mathrm{mmax}}$, (see section 'Transform alignment', above), and (ii) there would be no need to compute *tuple* partitions of cardinality less than the *maximum tuple derived dimension*, $m < \mathrm{mmax}$, because parent partitions necessarily have lower or equal *alignment*. That is, the *derived alignment* valued function of *partition-sets* and the parent partition relation are monotonic, $\forall Y_1, Y_2 \in \mathrm{B}(K)\ (\mathrm{parent}(Y_1, Y_2) \implies \mathrm{algn}(A*\{J^{\mathrm{CS}\{\}} : J \in Y_1\}^{\mathrm{T}}) \leq \mathrm{algn}(A * \{J^{\mathrm{CS}\{\}} : J \in Y_2\}^{\mathrm{T}})$. So it would only be necessary to compute $\mathrm{S}(K, \mathrm{mmax})$. However, the optimisation depends on the *shuffle derived histogram*, $A_R * G^{\mathrm{T}}$, which in turn depends on the given *fud*, F, so the monotonicity does not necessarily hold for either pair of relations.

Then, given the partition, Y, of the *tuple*, the second search, optimised by *shuffle content alignment valency-density*, is

$$\begin{aligned} Q = \mathrm{topd}(\mathrm{pmax})(\{(N^{\mathrm{T}}, (\mathrm{algn}(A * G^{\mathrm{T}}) - \mathrm{algn}(A_R * G^{\mathrm{T}}))/\mathrm{cvl}(G)) :& \\ N \in \prod_{J \in Y} (\mathrm{B}(J^{\mathrm{CS}}) \setminus \{\{J^{\mathrm{CS}}\}\}),\ G = \mathrm{depends}(F \cup \{N^{\mathrm{T}}\}, N)\}) \end{aligned}$$

The expected cardinality of the second search is

$$\Big(1/ \sum_{m \in \{2 \ldots \mathrm{mmax}\}} \mathrm{stir}(|K|, m)\Big) \sum_{m \in \{2 \ldots \mathrm{mmax}\}} \sum_{Y \in \mathrm{S}(K,m)} (\prod_{J \in Y} (\mathrm{bell}(|J^{\mathrm{CS}}|) - 1)$$

In the case of a *regular tuple* of *dimension* $k = |K|$ and *valency* d, the expected cardinality of the second search is

$$\Big(1/ \sum_{m \in \{2 \ldots \mathrm{mmax}\}} \mathrm{stir}(|K|, m)\Big) \sum_{m \in \{2 \ldots \mathrm{mmax}\}} \sum_{(L,a) \in \mathrm{sscd}(k,m)} \Big(a \prod_{(j,p) \in L} (\mathrm{bell}(d^j) - 1)^p\Big)$$

where sscd = stircd and the fixed cardinality partition function cardinality function is $\mathrm{stircd} \in \mathbf{N}_{>0} \times \mathbf{N}_{>0} \to (\mathcal{L}(\mathbf{N}) \to \mathbf{N})$.

The overall computation *time* of the searches is at least

$$\sum_{m \in \{2 \ldots \mathrm{mmax}\}} (\mathrm{stir}(|K|, m)) + \prod_{J \in Y} (\mathrm{bell}(|J^{\mathrm{CS}}|) - 1)$$

The *tuple* partition search term is comparable to bell(kmax). The *transforms* search term is comparable to $\mathrm{bell}(\mathrm{xmax}^{1/\mathrm{mmax}})^{\mathrm{mmax}}$. The computation *time* of the *transforms* search, Q, dominates that of the *tuple* partition search Y.

Consider another subset of the *pluri-valent pluri-limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$. If the *tuple* is at least *bi-variate*, $|K| > 1$, the *binary non-overlapping substrate transforms set* of the *tuple*, $\mathcal{T}_{U,K,\mathrm{n,b}}$, is a proper subset of the *non-overlapping substrate transforms set*, $\mathcal{T}_{U,K,\mathrm{n,b}} \subset \mathcal{T}_{U,K,\mathrm{n}}$. This corresponds to the special case of the *pluri-valent pluri-limited-tuple-derived-dimension non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$, where the *maximum derived dimension* is two, $\mathrm{mmax} = 2$. The *pluri-valent binary non-overlapping substrate partition-sets set* $\mathcal{N}_{U,K,\mathrm{n,b},\overline{2}} \subseteq \mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$ is defined

$$\begin{aligned}\mathcal{N}_{U,K,\mathrm{n,b},\overline{2}} = \{\{P,Q\} : J \subset K,\ J \neq \emptyset,\ J \neq K,\\ P \in \mathrm{B}(J^{\mathrm{CS}}) \setminus \{\{J^{\mathrm{CS}}\}\},\ Q \in \mathrm{B}((K \setminus J)^{\mathrm{CS}}) \setminus \{\{(K \setminus J)^{\mathrm{CS}}\}\}\}\end{aligned}$$

The cardinality is

$$|\mathcal{N}_{U,K,\mathrm{n,b},\overline{2}}| = 1/2 \times \sum_{J \in \mathrm{P}(K) \setminus \{\emptyset, K\}} (\mathrm{bell}(|J^{\mathrm{CS}}|) - 1) \times (\mathrm{bell}(|(K \setminus J)^{\mathrm{CS}}|) - 1)$$

In the case of *regular variables* of *valency* d and *dimension* k, the cardinality is

$$|\mathcal{N}_{U,K,\mathrm{n,b},\overline{2}}| = 1/2 \times \sum_{j \in \{1 \dots k-1\}} \binom{k}{j} (\mathrm{bell}(d^j) - 1) \times (\mathrm{bell}(d^{k-j}) - 1)$$

The subset of the *transforms* of the *pluri-valent binary non-overlapping substrate partition-sets set* of the *tuple* $Q \subset \{N^{\mathrm{T}} : N \in \mathcal{N}_{U,K,\mathrm{n,b},\overline{2}}\}$, is such that the *derived variables* of the *transforms* are *bi-variate*, $\forall T \in Q\ (|\mathrm{der}(T)| = 2)$.

Again, the choice of *tuple transforms*, Q, can be made by maximising the *shuffle content alignment valency-density*,

$$\begin{aligned}Q = \mathrm{topd}(\mathrm{pmax})(\{(N^{\mathrm{T}}, (\mathrm{algn}(A * G^{\mathrm{T}}) - \mathrm{algn}(A_R * G^{\mathrm{T}}))/\mathrm{cvl}(G)) :\\ N \in \mathcal{N}_{U,K,\mathrm{n,b},\overline{2}},\ G = \mathrm{depends}(F \cup \{N^{\mathrm{T}}\}, N)\})\end{aligned}$$

A further subset of the *pluri-valent binary non-overlapping substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{n,b},\overline{2}}$, is where the binary partition of the *tuple* is given. Let $J \subset K$ be such that $\{J,\ K \setminus J\} \in \mathrm{B}(K)$ where $|K| > 1$. Then

$$\{\{P,Q\} : P \in \mathrm{B}(J^{\mathrm{CS}}) \setminus \{\{J^{\mathrm{CS}}\}\},\ Q \in \mathrm{B}((K \setminus J)^{\mathrm{CS}}) \setminus \{\{(K \setminus J)^{\mathrm{CS}}\}\}\} \subset \mathcal{N}_{U,K,\mathrm{n,b},\overline{2}}$$

The cardinality of this set is $(\text{bell}(|J^{\text{C}}|) - 1) \times (\text{bell}(|(K \setminus J)^{\text{C}}|) - 1) < \text{bell}(\text{xmax})$. The computation *time* is comparable at least to $\text{bell}(\text{xmax}^{1/2})^2 < \text{bell}(\text{xmax})$.

Again, the optimisation search can be broken into two separate searches. The first search determines the binary partition of the *tuple*, $\{J,\ K \setminus J\}$, by searching the intersection of the *substrate self-cartesian partition-sets set* and the *binary non-overlapping substrate partition-sets set* which is

$$\mathcal{N}_{U,K,\text{c}} \cap \mathcal{N}_{U,K,\text{n,b}} = \{\{J^{\text{CS}\{\}}, (K \setminus J)^{\text{CS}\{\}}\} : J \subset K,\ J \neq \emptyset,\ J \neq K\}$$

The *tuple* binary partition search is optimised by maximising the *shuffle content alignment*,

$$\begin{aligned} J \in \text{maxd}(\{(M, \text{algn}(A * G^{\text{T}}) - \text{algn}(A_R * G^{\text{T}})) :\ & \\ & M \subset K,\ M \neq \emptyset,\ M \neq K,\ N = \{M^{\text{CS}\{\}}, (K \setminus M)^{\text{CS}\{\}}\}, \\ & G = \text{depends}(F \cup \{N^{\text{T}}\}, N)\}) \end{aligned}$$

Note that in the case of binary *tuple* partition search, $\{J,\ K \setminus J\}$, it does not matter whether the optimisation maximises the *shuffle content alignment*, $\text{algn}(A*G^{\text{T}}) - \text{algn}(A_R*G^{\text{T}})$, or the *shuffle content alignment valency-density*, $(\text{algn}(A*G^{\text{T}}) - \text{algn}(A_R*G^{\text{T}}))/\text{cvl}(G)$. This is because the *valency capacity* is constant for all binary partitions of the *tuple*, $\forall J \subset K\ ((|J^{\text{C}}||(K \setminus J)^{\text{C}}|)^{1/2} = |K^{\text{C}}|^{1/2})$.

The intersection has cardinality $|\mathcal{N}_{U,K,\text{c}} \cap \mathcal{N}_{U,K,\text{n,b}}| = 2^{|K|-1} - 1$. Then, given the binary partition of the *tuple*, $\{J,\ K \setminus J\}$, the second search is the *transforms* search optimised by *shuffle content alignment valency-density*,

$$\begin{aligned} Q = \text{topd}(\text{pmax})(\{(N^{\text{T}}, (\text{algn}(A * G^{\text{T}}) - \text{algn}(A_R * G^{\text{T}}))/\text{cvl}(G)) :\ & \\ & P \in \text{B}(J^{\text{CS}}),\ R \in \text{B}((K \setminus J)^{\text{CS}}),\ N = \{P, R\}, \\ & G = \text{depends}(F \cup \{N^{\text{T}}\}, N)\}) \end{aligned}$$

The expected cardinality of the second search is

$$\frac{1}{2^{|K|} - 2} \sum_{J \in \text{P}(K) \setminus \{\emptyset, K\}} (\text{bell}(|J^{\text{CS}}|) - 1) \times (\text{bell}(|(K \setminus J)^{\text{CS}}|) - 1)$$

In the case of a *regular tuple* of *dimension* $k = |K|$ and *valency* d, the expected cardinality of the second search is

$$\frac{1}{2^k - 2} \sum_{j \in \{1 \ldots k-1\}} \binom{k}{j} (\text{bell}(d^j) - 1) \times (\text{bell}(d^{k-j}) - 1)$$

The overall computation *time* of the searches is at least

$$2^{|K|-1} - 1 + (\text{bell}(|J^{\text{C}}|) - 1) \times (\text{bell}(|(K \setminus J)^{\text{C}}|) - 1)$$

The *tuple* binary partition search term is comparable to 2^{kmax}. The *transforms* search term is comparable to $\text{bell}(\text{xmax}^{1/2})^2$. The computation *time* of the *transforms* search, Q, dominates that of the *tuple* binary partition search, $\{J,\ K \setminus J\}$.

Having discussed the *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted non-overlapping substrate transform infinite-layer fud tree*, $\text{tfitnmubh}(U)(V) \to \text{trees}(\mathcal{F}_{U,\text{P}^*})$, which constructs the *fuds* with *non-overlapping substrate transforms* of the *tuples*, $\mathcal{T}_{U,K,\text{n}}$, section 'Substrate models computation' goes on to consider the *limited-layer limited-tuple-derived-dimension limited-underlying-volume limited-breadth contracted decrementing linear non-overlapping fuds infinite-layer fud tree* $\text{tfifdnmubh}(U)(V) \to \text{trees}(\mathcal{F}_{U,\text{P}^*})$, which constructs the *fuds* with *strong limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds*, $\mathcal{F}_{U,\text{n},-,K,\text{mmax}}$, on the *tuples*, defined

$$\begin{aligned}
&\mathcal{F}_{U,\text{n},-,K,\text{mmax}} \\
&\quad = \{\{N^{\text{T}} : (\cdot, N) \in L\} : M \in \mathcal{N}_{U,K,\text{c}} \cap \mathcal{N}_{U,K,\text{n},\text{mmax}}, \\
&\qquad\qquad L \in \text{subpaths}(\{(M, \text{tdec}(U)(M))\})\} \\
&\quad = \{\{N^{\text{T}} : (\cdot, N) \in L\} : Y \in \text{B}(K),\ |Y| \leq \text{mmax},\ M = \{J^{\text{CS}\{\}} : J \in Y\}, \\
&\qquad\qquad L \in \text{subpaths}(\{(M, \text{tdec}(U)(M))\})\}
\end{aligned}$$

where the tree of *self non-overlapping substrate decremented partition-sets* is defined $\text{tdec}(U) \in \text{P}(\mathcal{V}_U) \to \text{trees}(\text{P}(\mathcal{R}_U))$ as

$$\text{tdec}(U)(M) := \{(N, \text{tdec}(U)(N)) : N \in \mathcal{N}_{U,M,-} \cap \mathcal{N}_{U,M,\text{n},\text{s}}\}$$

and $\text{tdec}(U)(\emptyset) := \emptyset$. Explicitly this is

$$\begin{aligned}
&\text{tdec}(U)(M) := \{(N, \text{tdec}(U)(N)) : \\
&\qquad w \in M,\ Q \in \text{decs}(\{w\}^{\text{CS}\{\}}),\ N = \{Q\} \cup \{\{u\}^{\text{CS}\{\}} : u \in M,\ u \neq w\}\}
\end{aligned}$$

where $\text{decs} = \text{decrements} \in \mathcal{R}_U \to \text{P}(\mathcal{R}_U)$.

Instead of a subset of the *transforms* of the *pluri-valent pluri-limited-tuple-derived-dimension non-overlapping substrate partition-sets* of the *tuple*, $Q \subset \{N^{\text{T}} : N \in \mathcal{N}_{U,K,\text{n},\overline{\text{b}},\text{mmax},\overline{2}}\}$, consider a subset of the *pluri-valent pluri-limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds* of the *tuple* $Q \subset \mathcal{F}_{U_A,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$ such that $|Q| \leq \mathrm{pmax}$. The *pluri-valent pluri-limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds* is defined

$$\begin{aligned}
&\mathcal{F}_{U,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}} \\
&\quad= \{\{N^{\mathrm{T}} : (\cdot,N) \in L\} : M \in \mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax}}, \\
&\qquad\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdecp}(U)(M))\})\} \\
&\quad= \{\{N^{\mathrm{T}} : (\cdot,N) \in L\} : Y \in \mathrm{B}(K),\ 2 \leq |Y| \leq \mathrm{mmax},\ M = \{J^{\mathrm{CS}\{\}} : J \in Y\}, \\
&\qquad\qquad L \in \mathrm{subpaths}(\{(M, \mathrm{tdecp}(U)(M))\})\}
\end{aligned}$$

where $\mathrm{mmax} \geq 2$ and the tree of *pluri-valent self non-overlapping substrate decremented partition-sets* is defined $\mathrm{tdecp}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathrm{P}(\mathcal{R}_U))$ as

$$\mathrm{tdecp}(U)(M) := \{(N, \mathrm{tdecp}(U)(N)) : N \in \mathcal{N}_{U,M,\overline{2}} \cap \mathcal{N}_{U,M,-} \cap \mathcal{N}_{U,M,\mathrm{n},\mathrm{s}}\}$$

and $\mathrm{tdecp}(U)(\emptyset) := \emptyset$. Explicitly this is

$$\begin{aligned}
\mathrm{tdecp}(U)(M) := \{(N, \mathrm{tdecp}(U)(N)) :& \\
& w \in M,\ |\{w\}^{\mathrm{C}}| > 2,\ Q \in \mathrm{decs}(\{w\}^{\mathrm{CS}\{\}}), \\
& N = \{Q\} \cup \{\{u\}^{\mathrm{CS}\{\}} : u \in M,\ u \neq w\}\}
\end{aligned}$$

The cardinality of the *pluri-valent pluri-limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds* is

$$\begin{aligned}
&|\mathcal{F}_{U,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}| \\
&\quad= \sum \big(|\mathrm{nodes}(\mathrm{tdecp}(U)(\{J^{\mathrm{CS}\{\}} : J \in Y\}))| : \\
&\qquad\qquad m \in \{2 \ldots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m)\big) + 1 \\
&\quad= \sum p : m \in \{2 \ldots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m), \\
&\qquad\qquad L \in \mathrm{subpaths}(\mathrm{tdecpcd}(U)(\{J^{\mathrm{CS}\{\}} : J \in Y\})),\ (p,\cdot) = L_{|L|}
\end{aligned}$$

In the case of *regular substrate variables* of *valency* d and *dimension* n, the cardinality of the *pluri-valent pluri-limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds* is

$$\begin{aligned}
&|\mathcal{F}_{U,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}| \\
&\quad= \sum ap : m \in \{2 \ldots \mathrm{mmax}\},\ (M,a) \in \mathrm{sscd}(k,m),\ R = \mathrm{reg}(d,M), \\
&\qquad\qquad L \in \mathrm{subpaths}(\{((1,R), \mathrm{tdeccpd}(1,R))\}),\ (p,\cdot) = L_{|L|}
\end{aligned}$$

where $k = |K|$ and $\mathrm{reg} \in \mathbf{N} \times \mathcal{L}(\mathbf{N}) \to \mathcal{L}(\mathbf{N})$ is defined $\mathrm{reg}(d,M) := \mathrm{concat}(\mathrm{flip}(\mathrm{order}(D_{\mathcal{L}(\mathbf{N})}, \{\{1 \ldots q\} \times \{d^j\} : (j,q) \in M\})))$.

The cardinality of the *pluri-valent self non-overlapping substrate decremented partition-sets tree* may be computed by defining $\mathrm{tdecpcd}(U) \in \mathrm{P}(\mathcal{V}_U) \to \mathrm{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}))$ as

$$\begin{aligned}&\mathrm{tdecpcd}(U)(V) := \\ &\quad \{((1,L),\mathrm{tdecpcd}(1,L)) : L = \{(i,|U_v|) : (v,i) \in \mathrm{order}(D_{\mathrm{V}},V)\}\}\end{aligned}$$

where order D_{V} is such that $\mathrm{order}(D_{\mathrm{V}},V) \in \mathrm{enums}(V)$, and $\mathrm{tdecpcd} \in \mathbf{N} \times \mathcal{L}(\mathbf{N}) \to \mathrm{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}))$ as

$$\begin{aligned}&\mathrm{tdecpcd}(k,L) := \{((m,M),\mathrm{tdecpcd}(m,M)) : \\ &\quad i \in \{1 \ldots |L|\},\ L_i > 2,\ m = kL_i(L_i-1),\ M = L \setminus \{(i,L_i)\} \cup \{(i,L_i-1)\}\}\end{aligned}$$

In the case of *regular substrate variables* of *valency* d and *dimension* n, the cardinality of the *pluri-valent self non-overlapping substrate decremented partition-sets cardinality tree* may be computed by defining $\mathrm{tdecpcd} \in \mathbf{N} \times \mathbf{N} \to \mathrm{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}))$ as

$$\mathrm{tdecpcd}(d,n) := \{((1,L),\mathrm{tdecpcd}(1,L)) : L = \{1 \ldots n\} \times \{d\}\}$$

If the *tuple* is *pluri-variate* and *pluri-valent*, $|K| > 1$, and $\forall v \in K\ (|U_v| > 1)$, then the *derived variables* are *pluri-variate* and *pluri-valent*, $\forall H \in Q\ (|\mathrm{der}(H)| > 1)$, and $\forall H \in Q\ \forall w \in \mathrm{der}(H)\ (|U_w| > 1)$. Each of these *fuds* has a corresponding *partition*, $\forall H \in Q\ (H^{\mathrm{TP}K} \in \mathrm{B}(K^{\mathrm{CS}}))$, but now are not restricted to zero *derived alignment*. That is, in some cases, $A * G^{\mathrm{T}} \neq (A * G^{\mathrm{T}})^{\mathrm{X}}$, where $H \in Q$ and $G = \mathrm{depends}(F \cup H, \mathrm{der}(H))$. The *tuple fud*, H, is *non-overlapping*, $\neg\mathrm{overlap}(H)$. However, the constructed *exploded fud* $G' = \mathrm{explode}(G)$ is not necessarily *non-overlapping* so the choice of *tuple fuds*, $Q \subset \mathcal{F}_{U_A,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$, is made by maximising the *shuffle content alignment valency-density*,

$$\begin{aligned}Q = \mathrm{topd}(\mathrm{pmax})(\{(H,(\mathrm{algn}(A * G^{\mathrm{T}}) - \mathrm{algn}(A_R * G^{\mathrm{T}}))/\mathrm{cvl}(G)) : \\ H \in \mathcal{F}_{U_A,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}},\ G = \mathrm{depends}(F \cup H, \mathrm{der}(H))\})\end{aligned}$$

The cardinality of the *contracted decrementing linear non-overlapping fuds* is greater than or equal to the cardinality of the *non-overlapping substrate transforms*, $|\mathcal{F}_{U,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}| \geq |\mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}|$, so the computation *time* is increased. However, consider the optimisation of the tree of *decrements*. Define the *contracted decrementing linear non-overlapping fuds list maximiser*

$$\begin{aligned}&Z_{P,A,A_R,F,\mathrm{n},-,K} = \\ &\quad \mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{n},-,K}, N_{P,A,A_R,F,\mathrm{n},-,K}, \mathrm{top}(\mathrm{pmax}), R_{P,A,A_R,F,\mathrm{n},-,K})\end{aligned}$$

where (i) the optimiser function is

$$X_{P,A,A_R,F,\mathrm{n},-,K} = \{(H, (I^*_{\mathrm{a}}(A * G^{\mathrm{T}}) - I^*_{\mathrm{a}}(A_R * G^{\mathrm{T}}))/I^*_{\mathrm{cvl}}(G)) : \\ H \in \mathcal{F}_{U_A,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}},\ G = \mathrm{depends}(F \cup H, \mathrm{der}(H))\}$$

(ii) the initial subset is

$$R_{P,A,A_R,F,\mathrm{n},-,K} = \{(\{M^{\mathrm{T}}\}, X_{P,A,A_R,F,\mathrm{n},-,K}(\{M^{\mathrm{T}}\})) : \\ Y \in \mathrm{B}(K),\ 2 \leq |Y| \leq \mathrm{mmax},\ M = \{J^{\mathrm{CS}\{\}} : J \in Y\}\}$$

and (iii) the neighbourhood function is

$$\begin{aligned} N_{P,A,A_R,F,\mathrm{n},-,K}(C) = \{(H \cup \{N^{\mathrm{T}}\}, X_{P,A,A_R,F,\mathrm{n},-,K}(H \cup \{N^{\mathrm{T}}\})) :& \\ (H,\cdot) \in C,\ M = \mathrm{der}(H),& \\ w \in M,\ |\{w\}^{\mathrm{C}}| > 2,\ Q \in \mathrm{decs}(\{w\}^{\mathrm{CS}\{\}}),& \\ N = \{Q\} \cup \{\{u\}^{\mathrm{CS}\{\}} : u \in M,\ u \neq w\}\}& \end{aligned}$$

Then the subset of the *decrementing linear non-overlapping fuds* is

$$\mathrm{dom}(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n},-,K})) \subset \mathcal{F}_{U_A,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$$

So the choice of *tuple fuds*, $Q \subset \mathcal{F}_{U_A,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}}$, can be made by maximising the *shuffle content alignment valency-density* of the elements of the *tree maximiser*, $Z_{P,A,A_R,F,\mathrm{n},-,K}$,

$$Q = \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n},-,K}))$$

The cardinality of the initial set is

$$\begin{aligned} |R_{P,A,A_R,F,\mathrm{n},-,K}| &= |\mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax}}| \\ &= \sum_{m \in \{2 \ldots \mathrm{mmax}\}} \mathrm{stir}(|K|, m) \\ &\leq \mathrm{bell}(|K|) - 1 \end{aligned}$$

For a given *tuple* partition $Y \in \mathrm{B}(K)$ the cardinality of the neighbourhood searched could be computed by constructing a tree of lists of *valencies*, $\mathrm{trees}(\mathcal{L}(\mathbf{N}_{>0}))$, congruent to the *pluri-valent self non-overlapping substrate decremented partition-sets cardinality tree*, $\mathrm{tdecpcd}(Y) \in \mathrm{trees}(\mathbf{N} \times \mathcal{L}(\mathbf{N}))$, and such that (i) at the root the list consists of the *volumes* of the components, $\{(i, |J^{\mathrm{C}}|) : (J, i) \in \mathrm{order}(D_{\mathrm{P}(\mathrm{P}(\mathcal{V}))}, Y)\}$, and (ii) at each step one of the *valencies* is decremented. The cardinality of the searched at each node L is then $\sum(c(c-1)/2 : (\cdot,c) \in L)$. However, the computation of the tree is exponential and so is impracticable as a measure of expected computation *time*.

Instead of finding the cardinalities of searched for all possible search paths, consider the cardinality of the searched in the worst case which is found by decrementing the head of a sorted list of component *volumes*. *Rolling* the shortest first is the worst case because the cardinality of the *decrements*, $c(c-1)/2$, is convex. Let $\mathrm{srch} \in \mathcal{L}(\mathbf{N}) \to \mathbf{N}$ be defined $\mathrm{srch}(L) := \sum(c(c-1)/2 : (\cdot,c) \in L,\ c > 2)$. Let $\mathrm{dec} \in \mathcal{K}(\mathbf{N}) \to \mathcal{K}(\mathbf{N})$ be defined $\mathrm{dec}((x,K)) := \mathrm{if}(x > 2, (x-1,K), \mathrm{dec}(K))$. Let $\mathrm{srchmax} \in \mathbf{N} \times \mathcal{L}(\mathbf{N}) \to \mathbf{N}$ be defined $\mathrm{srchmax}(m,\emptyset) := m$ and $\mathrm{srchmax}(m,L) := \mathrm{srchmax}(m + \mathrm{srch}(L), \mathrm{dec}(L))$. Let $\mathrm{srchmax} \in \mathrm{P}(\mathrm{P}(\mathcal{V}_U)) \to \mathbf{N}$ be defined $\mathrm{srchmax}(Y) := \mathrm{srchmax}(0, \mathrm{sort}(\{(i,|J^{\mathrm{C}}|) : (J,i) \in \mathrm{order}(D_{\mathrm{P}(\mathrm{P}(\mathcal{V}))}, Y)\}))$, where $\mathrm{sort}(L) := \{(i,a) : ((a,\cdot),i) \in \mathrm{order}(D_{\mathrm{N}^2}, \mathrm{flip}(L))\}$. Then $\mathrm{srchmax}(Y) \in \mathbf{N}$ is the greatest cardinality of the searched for the given partition, Y.

The expected cardinality of the searched excluding the initial set is

$$\frac{\mathrm{pmax}}{\sum_{m \in \{2\ldots\mathrm{mmax}\}} \mathrm{stir}(|K|,m)} \sum_{m \in \{2\ldots\mathrm{mmax}\}} \sum_{Y \in \mathrm{S}(K,m)} \mathrm{srchmax}(Y)$$

In the case of a *regular tuple* of *dimension* $k = |K|$ and *valency* d, the expected cardinality of the neighbourhood searched is

$$\frac{\mathrm{pmax}}{\sum_{m \in \{2\ldots\mathrm{mmax}\}} \mathrm{stir}(k,m)} \sum_{m \in \{2\ldots\mathrm{mmax}\}} \sum_{(L,a) \in \mathrm{sscd}(k,m)} a \times \mathrm{srchmax}(L)$$

where $\mathrm{srchmax}(L) := \mathrm{srchmax}(0, \mathrm{sort}(\mathrm{concat}(\{\{1\ldots p\} \times \{d^j\} : (j,p) \in L\})))$.

If it is the case that the topd(pmax) optimisation is such that each step of the list has no more than pmax *fuds*, $\forall (i,C) \in \mathrm{list}(Z_{P,A,A_R,F,\mathrm{n},-,K})\ (|C| \leq \mathrm{pmax})$, then the cardinality of the elements must be less than or equal to (a) the *maximum tuple optimise limit*, pmax, times (b) the length of the longest path of *decrements*, $|K^{\mathrm{C}}|$. Thus $|\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n},-,K})| \leq \mathrm{pmax} \times \mathrm{xmax}$. The cardinality of the searched must be less than or equal to (i) the cardinality of the initial set, $|R_{P,A,A_R,F,\mathrm{n},-,K}| = |\mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\overline{\mathrm{b}},\mathrm{mmax}}| \leq \mathrm{bell}(|K|) - 1$, plus (ii) (a) the cardinality of the elements, $|\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n},-,K})|$, times (b) the cardinality of the *decrements*, which is less than $|\mathcal{N}_{U,M,\overline{2}} \cap \mathcal{N}_{U,M,-} \cap \mathcal{N}_{U,M,\mathrm{n,s}}|$. If the *tuple* is *regular* having *dimension* $k = |K|$ and *valency* $d > 2$, then $|\mathcal{N}_{U,M,\overline{2}} \cap \mathcal{N}_{U,M,-} \cap \mathcal{N}_{U,M,\mathrm{n,s}}| = kd(d-1)/2$. Thus cardinality of the searched set is constrained

$$|\text{searched}(Z_{P,A,A_R,F,\text{n},-,K})| \leq \text{bell}(\text{kmax}) + \text{pmax} \times \text{kmax} \times \text{xmax}^2/2$$

If the *maximum tuple optimise limit* is one, $\text{pmax} = 1$, this cardinality is smaller than that of the *contracted non-overlapping substrate transforms*, $|\mathcal{T}_{U,K,\text{n}}| \leq \text{bell}(\text{kmax}) \times \text{bell}(\text{xmax})$.

Again, the cardinality of the searched set, $|\text{searched}(Z_{P,A,A_R,F,\text{n},-,K})|$, can be reduced further by restricting the initial set to *binary non-overlapping substrate transforms set*, $\mathcal{T}_{U,K,\text{n,b}} \subseteq \mathcal{T}_{U,K,\text{n}}$, which is the special case where $\text{mmax} = 2$. Define the *tuple-binary-partition contracted decrementing linear non-overlapping fuds list maximiser*

$$Z_{P,A,A_R,F,\text{n,b},-,K} =$$
$$\text{maximiseLister}(X_{P,A,A_R,F,\text{n},-,K}, N_{P,A,A_R,F,\text{n},-,K}, \text{top}(\text{pmax}), R_{P,A,A_R,F,\text{n,b},-,K})$$

where the initial set is

$$R_{P,A,A_R,F,\text{n,b},-,K} = \{(\{M^{\text{T}}\}, X_{P,A,A_R,F,\text{n},-,K}(\{M^{\text{T}}\})) : J \subset K,\ J \neq \emptyset,\ J \neq K,\ M = \{J^{\text{CS}\{\}}, (K \setminus J)^{\text{CS}\{\}}\}\}$$

In this case the choice of *tuple fuds* is

$$Q = \text{topd}(\text{pmax})(\text{elements}(Z_{P,A,A_R,F,\text{n,b},-,K}))$$

The cardinality of the initial set is

$$|R_{P,A,A_R,F,\text{n,b},-,K}| = |\mathcal{N}_{U_A,K,\text{c}} \cap \mathcal{N}_{U_A,K,\text{n,b}}| = 2^{|K|-1} - 1$$

The expected cardinality of the searched excluding the initial set is

$$\frac{\text{pmax}}{2^{|K|} - 2} \sum_{J \in \text{P}(K) \setminus \{\emptyset, K\}} \text{srchmax}(\{J, K \setminus J\})$$

In the case of a *regular tuple* of *dimension* $k = |K|$ and *valency* d, the expected cardinality of the neighbourhood searched is

$$\frac{\text{pmax}}{2^k - 2} \sum_{j \in \{1 \ldots k-1\}} \binom{k}{j} \times \text{srchmax}(0, \text{sort}(\{(1, d^j), (2, d^{k-j})\}))$$

The cardinality of the searched is constrained

$$|\text{searched}(Z_{P,A,A_R,F,\text{n,b},-,K})| \leq 2^{\text{kmax}} + \text{pmax} \times \text{kmax} \times \text{xmax}^2/2$$

In this case the latter term dominates because $2^{\mathrm{kmax}} \leq \mathrm{xmax}$.

A variation of the *contracted decrementing linear non-overlapping fuds list maximiser* is to restrict the neighbourhood optimisation to the maximum, $\mathrm{pmax} = 1$, and apply pmax only to the initial set. To do this a *tree tail maximiser* is used with an inclusion function of max and the initial set is explicitly optimised beforehand. Define the *maximum-roll contracted decrementing linear non-overlapping fuds tree maximiser*

$$\begin{aligned} &Z_{P,A,A_R,F,\mathrm{n},-,K,\mathrm{mr}} = \\ &\quad \mathrm{maximiseTailTreer}(X_{P,A,A_R,F,\mathrm{n},-,K}, N_{P,A,A_R,F,\mathrm{n},-,K}, \max, \\ &\qquad\qquad \mathrm{top}(\mathrm{pmax})(R_{P,A,A_R,F,\mathrm{n},-,K})) \end{aligned}$$

This restriction pushes the pmax path selection into the initial set rather than towards the end of the optimise path where sometimes different *decrementing linear fuds roll* to the same *derived partition variables*.

Another constraint that may be applied to reduce the cardinality of the searched set, $|\mathrm{searched}(Z_{P,A,A_R,F,\mathrm{n},-,K})|$, is to restrict the initial set such that the *volume* of each component of the *tuple* partition is limited. The *maximum valency* is $\mathrm{umax} \in \mathbf{N}_{>0}$. The initial *partition-set* forming the bottom *layer* of the *decrementing linear fud* is in the intersection of the *substrate self-cartesian partition-sets set*, the *pluri-limited-tuple-derived-dimension non-overlapping substrate transforms set* and the *limited-valency substrate partition-sets set*, $\mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\bar{\mathrm{b}},\mathrm{mmax}} \cap \mathcal{N}_{U,K,\mathrm{umax}}$. The *pluri-limited-valency pluri-limited-tuple-derived-dimension contracted decrementing linear non-overlapping fuds* is defined

$$\begin{aligned} &\mathcal{F}_{U,\mathrm{n},-,K,\bar{\mathrm{b}},\mathrm{mmax},\bar{2},\mathrm{umax}} \\ &\quad = \{\{N^{\mathrm{T}} : (\cdot,N) \in L\} : M \in \mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\bar{\mathrm{b}},\mathrm{mmax}} \cap \mathcal{N}_{U,K,\mathrm{umax}}, \\ &\qquad\qquad L \in \mathrm{subpaths}(\{(M,\mathrm{tdecp}(U)(M))\})\} \\ &\quad = \{\{N^{\mathrm{T}} : (\cdot,N) \in L\} : Y \in \mathrm{B}(K),\ 2 \leq |Y| \leq \mathrm{mmax}, \\ &\qquad\qquad (\forall J \in Y\ (|J^{\mathrm{C}}| \leq \mathrm{umax})),\ M = \{J^{\mathrm{CS}\{\}} : J \in Y\}, \\ &\qquad\qquad L \in \mathrm{subpaths}(\{(M,\mathrm{tdecp}(U)(M))\})\} \end{aligned}$$

Define the *limited-valency contracted decrementing linear non-overlapping fuds list maximiser*

$$\begin{aligned} &Z_{P,A,A_R,F,\mathrm{n},\mathrm{w},-,K} = \\ &\quad \mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{n},-,K}, N_{P,A,A_R,F,\mathrm{n},-,K}, \mathrm{top}(\mathrm{pmax}), R_{P,A,A_R,F,\mathrm{n},\mathrm{w},-,K}) \end{aligned}$$

where the initial set is

$$\begin{aligned} R_{P,A,A_R,F,\mathrm{n},\mathrm{w},-,K} = \{(\{M^{\mathrm{T}}\}, X_{P,A,A_R,F,\mathrm{n},-,K}(\{M^{\mathrm{T}}\})) :& \\ Y \in \mathrm{B}(K),\ 2 \leq |Y| \leq \mathrm{mmax},& \\ (\forall J \in Y\ (|J^{\mathrm{C}}| \leq \mathrm{umax})),\ M = \{J^{\mathrm{CS}\{\}} : J \in Y\}\}& \end{aligned}$$

In this case the choice of *tuple fuds* is

$$Q = \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n},\mathrm{w},-,K}))$$

The cardinality of the initial set now depends on the *system*,

$$\begin{aligned} |R_{P,A,A_R,F,\mathrm{n},\mathrm{w},-,K}| &= |\mathcal{N}_{U,K,\mathrm{c}} \cap \mathcal{N}_{U,K,\mathrm{n},\bar{\mathrm{b}},\mathrm{mmax}} \cap \mathcal{N}_{U,K,\mathrm{umax}}| \\ &= \sum_{m \in \{2 \ldots \mathrm{mmax}\}} |\{Y : Y \in \mathrm{S}(K,m),\ (\forall J \in Y\ (|J^{\mathrm{C}}| \leq \mathrm{umax}))\}| \\ &\leq \sum_{m \in \{2 \ldots \mathrm{mmax}\}} \mathrm{stir}(|K|, m) \end{aligned}$$

The expected cardinality of the searched excluding the initial set is

$$\begin{aligned} \mathrm{pmax} \quad \times \quad & \sum(\mathrm{srchmax}(Y) : m \in \{2 \ldots \mathrm{mmax}\},\ Y \in \mathrm{S}(K,m), \\ & \qquad (\forall J \in Y\ (|J^{\mathrm{C}}| \leq \mathrm{umax}))) \\ / \quad & \sum_{m \in \{2 \ldots \mathrm{mmax}\}} |\{Y : Y \in \mathrm{S}(K,m),\ (\forall J \in Y\ (|J^{\mathrm{C}}| \leq \mathrm{umax}))\}| \end{aligned}$$

In the case of a *regular tuple* of *dimension* $k = |K|$ and *valency* d, the expected cardinality of the neighbourhood searched is

$$\begin{aligned} \mathrm{pmax} \quad \times \quad & \sum(a \times \mathrm{srchmax}(L) : m \in \{2 \ldots \mathrm{mmax}\},\ (L,a) \in \mathrm{sscd}(k,m), \\ & \qquad (\forall (j,p) \in L\ (d^j \leq \mathrm{umax}))) \\ / \quad & \sum(a : m \in \{2 \ldots \mathrm{mmax}\},\ (L,a) \in \mathrm{sscd}(k,m), \\ & \qquad (\forall (j,p) \in L\ (q > 0 \implies d^j \leq \mathrm{umax}))) \end{aligned}$$

The effect of the *valency* limit is to make the partitions of the *tuple* larger and more *regular*. For example, a binary partition such that all but one of the *tuple variables*, having common *valency* d, is in one component has initial *valencies* of d^{k-1} and d. The cardinality of the searched approximately varies as the cube of the longest *valency*, so a binary *irregular* search may require considerably more computation *time* than a poly-component *regular* search in the same *tuple*.

Having addressed the cardinality of the second term of the upper bound on the cardinality of the neighbourhood function of the *limited-path-models tuple partition inducer*, $|\mathrm{B}(K^{\mathrm{CS}})| \leq \mathrm{bell}(\mathrm{xmax})$ where $K \in \mathrm{tuples}(V,F)$ and $|K^{\mathrm{C}}| \leq \mathrm{xmax}$, by optimising a subset Q of the *partitions* of a *tuple*, $Q \subset \mathrm{B}(K^{\mathrm{CS}})$, above, now consider the first term. A subset of the neighbourhood function of the *limited-path-models tuple partition inducer*, $P_{P,A,A_R,\mathrm{csd,p}}$, may be defined which allows only one *partition transform* for each *tuple*,

$$\begin{aligned}
&P_{P,A,A_R,\mathrm{csd,p,u}}(Q) = \\
&\quad \{(F \cup G, I^*_{\mathrm{csd}}((A, A_R, F \cup G))) : \\
&\qquad (F,\cdot) \in Q,\ \mathrm{layer}(F,\mathrm{der}(F)) < \mathrm{lmax}, \\
&\qquad B \subseteq \{K : K \in \mathrm{tuples}(V_A,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax}\}, \\
&\qquad 1 \leq |B| \leq \mathrm{bmax}, \\
&\qquad G \in \{\{P^{\mathrm{T}} : P \in N\} : N \in \prod_{K \in B} (\mathrm{B}(K^{\mathrm{CS}}) \setminus \{\{K^{\mathrm{CS}}\}\})\}, \\
&\qquad W = \mathrm{der}(F \cup G),\ |W^{\mathrm{C}}| \leq \mathrm{wmax}\}
\end{aligned}$$

In the case of the *fud* F, defined above, of *variable* cardinality $|\mathrm{vars}(F)| = \mathrm{lmax} \times \mathrm{bmax}$, the upper bound on the cardinality of the neighbourhood function is

$$\begin{aligned}
&|P_{P,A,A_R,\mathrm{csd,p,u}}(\{(F, X_{P,A,A_R,\mathrm{csd,p}}(F))\})| \\
&\qquad\qquad < ((\mathrm{lmax} \times \mathrm{bmax})^{\underline{\mathrm{kmax}}})^{\underline{\mathrm{bmax}}} \times \mathrm{bell}(\mathrm{xmax})^{\mathrm{bmax}}
\end{aligned}$$

The cardinality of the set of next *limited-underlying-volume limited-breadth layers* depends on (i) the cardinality of the set of next *limited-underlying-volume limited-breadth layer tuple sets*,

$$\begin{aligned}
&|\{B : B \subseteq \{K : K \in \mathrm{tuples}(V_A,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax}\},\ 1 \leq |B| \leq \mathrm{bmax}\}| \\
&\qquad\qquad < ((\mathrm{lmax} \times \mathrm{bmax})^{\underline{\mathrm{kmax}}})^{\underline{\mathrm{bmax}}}
\end{aligned}$$

and (ii) the product of the cardinalities of the sets of *partitions* of each *tuple*,

$$\prod_{K \in B} |\mathrm{B}(K^{\mathrm{CS}})| < \mathrm{bell}(\mathrm{xmax})^{\mathrm{bmax}}$$

within each *layer tuple set*, B.

The cardinality of the set of next *limited-underlying limited-breadth tuple sets*, $|\{B : B \subseteq \{K : K \in \mathrm{tuples}(V_A,F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax}\},\ 1 \leq |B| \leq \mathrm{bmax}\}|$, can be addressed by (i) constructing only a single *content alignment* optimised next *limited-underlying limited-breadth layer tuple set* $B_{\mathrm{B}} \subseteq \{K : K \in$

$\mathrm{tuples}(V_A, F),\ |K^{\mathrm{C}}| \leq \mathrm{xmax}\}$, and (ii) constructing the *tuples* of that *tuple set*, B_{B}, one *variable* at a time. The *tuple set*, B_{B}, consists of the *maximum breadth*, bmax, per *maximum tuple derived dimension*, mmax, top-most *tuples*, $|B_{\mathrm{B}}| = \lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor$, if the *tuples* are uniquely *aligned*, so that the cardinality of *derived variables* in the *layer* optimised by applying the *contracted decrementing linear non-overlapping fuds list maximiser*, $Z_{P,A,A_R,F,\mathrm{n},-,K}$, to each *tuple*, $K \in B_{\mathrm{B}}$, is no greater than the *maximum breadth*, bmax. Define the *limited-underlying tuple set list maximiser*

$$Z_{P,A,A_R,F,\mathrm{B}} = \mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{B}}, P_{P,A,A_R,F,\mathrm{B}}, \mathrm{top}(\mathrm{omax}), R_{P,A,A_R,F,\mathrm{B}})$$

where (i) the optimiser function is

$$\begin{aligned} &X_{P,A,A_R,F,\mathrm{B}} = \\ &\quad \{(K, I_{\mathrm{a}}^*(\mathrm{apply}(V_A, K, \mathrm{his}(F), A)) - I_{\mathrm{a}}^*(\mathrm{apply}(V_A, K, \mathrm{his}(F), A_R))) : \\ &\qquad\qquad\qquad\qquad\qquad\qquad K \in \mathrm{tuples}(V_A, F)\} \end{aligned}$$

where $\mathrm{his} = \mathrm{histograms} \in \mathcal{F} \to \mathrm{P}(\mathcal{A})$, $\mathrm{apply} \in \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{A}) \times \mathcal{A} \to \mathcal{A}$, and (ii) the neighbourhood function is

$$\begin{aligned} P_{P,A,A_R,F,\mathrm{B}}(B) = \{(J, X_{P,A,A_R,F,\mathrm{B}}(J)) :& \\ (K, \cdot) \in B,\ w \in \mathrm{vars}(F) \cup V_A \setminus K,&\ J = K \cup \{w\},\ |J^{\mathrm{C}}| \leq \mathrm{xmax}\} \end{aligned}$$

and (iii) the initial subset is

$$\begin{aligned} R_{P,A,A_R,\emptyset,\mathrm{B}} &= \{(\{w,u\}, X_{P,A,A_R,\emptyset,\mathrm{B}}(\{w,u\})) : \\ &\qquad w, u \in V_A,\ u \neq w,\ |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}\} \\ R_{P,A,A_R,F,\mathrm{B}} &= \{(\{w,u\}, X_{P,A,A_R,F,\mathrm{B}}(\{w,u\})) : \\ &\qquad w \in \mathrm{der}(F),\ u \in \mathrm{vars}(F) \cup V_A,\ u \neq w,\ |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}\} \end{aligned}$$

Then the *shuffle content alignment* optimised next *limited-underlying limited-breadth layer tuple set*, B_{B}, is

$$\begin{aligned} B_{\mathrm{B}} = \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{B}})) \in& \\ \{B : B \subseteq \{K : K \in \mathrm{tuples}(V_A, F),&\ |K^{\mathrm{C}}| \leq \mathrm{xmax}\}\} \end{aligned}$$

The *fud application*, $\mathrm{apply} \in \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{A}) \times \mathcal{A} \to \mathcal{A}$, traverses from the *susbtrate variables*, V_A, to the *tuple*, K, via the *histograms* of the *transforms* of the *fud*, $\mathrm{his}(F)$. The *fud application* is a tractable *application equivalent* to the *application* of the *fud's transform's histogram* followed by *reduction*, $\mathrm{apply}(V_A, K, \mathrm{his}(F), A) = A * \mathrm{histogram}(F^{\mathrm{T}})\ \%\ K$. The *shuffle content fud application alignment*, $\mathrm{algn}(\mathrm{apply}(V_A, K, \mathrm{his}(F), A)) -$

$\text{algn}(\text{apply}(V_A, K, \text{his}(F), A_R))$, resembles the *shuffle content alignment*, $\text{algn}(A * G^{\text{T}}) - \text{algn}(A_R * G^{\text{T}})$, where $G = \text{depends}(F, K)$, except that *fud variables* cannot be hidden by being nested in the higher *layer depends fud* of another *fud variable*. That is, in some cases $K \neq \text{der}(\text{depends}(F, K))$ and so $\text{apply}(V_A, K, \text{his}(F), A) \neq A * \text{depends}(F, K)^{\text{T}}$.

The *tuple set* search, $Z_{P,A,A_R,F,\text{B}}$, is optimised by maximising the *shuffle content fud application alignment*,

$$\text{algn}(\text{apply}(V_A, K, \text{his}(F), A)) - \text{algn}(\text{apply}(V_A, K, \text{his}(F), A_R))$$

not the *shuffle content fud application alignment valency-density*,

$$(\text{algn}(\text{apply}(V_A, K, \text{his}(F), A)) - \text{algn}(\text{apply}(V_A, K, \text{his}(F), A_R)))/y^{1/k}$$

where $y = |K^{\text{C}}|$ and $k = |K|$. The *maximum alignment* varies with *dimension*, for example a *regular histogram* of *dimension* k and *valency* d has approximate *maximum alignment* of $z(k-1)\ln d$. Therefore a *maximiser* function value of *shuffle content alignment* tends to select *tuples* of larger *dimension*. In contrast, the *maximum alignment valency-density* varies against geometric-average *valency*, approximately $z(k-1)(\ln d)/d$, and so tends to smaller *tuple dimension*, k, especially where the entropy of *valencies*, $\text{entropy}(\{(w, |U_A(w)|) : w \in K\})$, is low. The *contracted decrementing linear non-overlapping fuds* in subsequent applications of *contracted decrementing linear non-overlapping fuds list maximisers* to each *tuple* of the *tuple set*, $\bigcup\{\text{dom}(X_{P,A,A_R,F,\text{n},-,K}) : K \in B_{\text{B}}\} \subset \mathcal{F}_{U_A,\text{P}*}$, have *derived volume* less than or equal to the *tuple volume*, $|\text{der}(H)^{\text{C}}| \leq |K^{\text{C}}|$ where $H \in \mathcal{F}_{U_A,\text{n},-,K}$, so larger *tuples* can *value roll* down to smaller subsets of the *tuple*, but not the reverse. That is, a *maximiser* function value of *shuffle content alignment* increases the searched cardinality for each *tuple* making the neighbourhood function, $P_{P,A,A_R,\text{csd}}$, of the notional *list maximiser*, $Z_{P,A,A_R,\text{csd}}$, less arbitrary, so increasing the *maximum function* correlation of the *practicable inducer*, $I'_{z,\text{csd},\text{F},\infty,\text{q},P}$, to the *tractable inducer*, $I'_{z,\text{ad},\text{F},\infty,\text{n},\text{q}}$.

An upper bound on the expected cardinality of the searched may be computed given the *maximum underlying dimension*, kmax. The upper bound on the expected cardinality in the first *layer*, $F = \emptyset$, is

$$\sum_{k \in \{2 \ldots \min(\text{kmax}, n)\}} \binom{n}{k}$$

where $n = |V_A|$ and min = minimum. In subsequent *layers*, $F \neq \emptyset$, the upper bound on the expected cardinality is

$$\sum_{k \in \{2 \ldots \min(\text{kmax},q)\}} \binom{q}{k} - \binom{q-x}{k}$$

where $W = \text{vars}(F) \cup V_A$, $q = |W|$, $X = \text{der}(F)$ and $x = |X|$.

The *maximum underlying dimension*, kmax, may be approximated from the geometric average *valency* $d = |W^{\text{C}}|^{1/q}$, and the *maximum underlying volume*, xmax,

$$\text{kmax} = \left\lceil \frac{\ln \text{xmax}}{\ln d} \right\rceil$$

The *limited-underlying tuple set list maximiser*, $Z_{P,A,A_R,F,\text{B}}$, has an inclusion function defined $\text{top}(\text{omax}) \in \text{P}(X_{P,A,A_R,F,\text{B}}) \to \text{P}(X_{P,A,A_R,F,\text{B}})$. The application of the $\text{top}(n) \in \text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{X} \times \mathcal{Y})$ aggregation function with parameter n to a relation R may result in a subset of the relation having a cardinality greater than the given parameter, $|\text{top}(n)(R)| > n$, if there are duplicate range values at the n-th position of the ordered relation. This might be the case, for example, in a *tuple set list maximiser searched set*, $\text{searched}(Z_{P,A,A_R,F,\text{B}})$, that contains *tuples* which contain *variables* having a *partition* equal to the *self-partition* of a singleton *underlying variable*.

An implementation of the *tuple set list maximiser*, $Z_{P,A,A_R,F,\text{B}}$, that guarantees no more than omax *tuples* at each step of the *optimiser* list, $\forall B \in \text{set}(\text{list}(Z_{P,A,A_R,F,\text{B}}))\ (|B| \leq \text{omax})$, must have additional inclusion order criteria. An example is where the *tuples* are ordered first by ascending *alignment*, $X_{P,A,A_R,F,\text{B}}(K)$, and then by descending *sum derived variables layer*, $-\text{sumlayer}(F,K)$, where $\text{sumlayer} \in \mathcal{F} \times \text{P}(\mathcal{V}) \to \mathbf{N}$ is defined as

$$\text{sumlayer}(F,K) := \sum_{w \in K} \text{layer}(F,\{w\})$$

For example, the *tuples* $J, K \subset \text{vars}(F)$, such that *variable* $u \in K$, *self partition variable* $\{u\}^{\text{CS}\{\}} \in J$ and $J = K \setminus \{u\} \cup \{\{u\}^{\text{CS}\{\}}\}$, have the same *alignments*, $X_{P,A,A_R,F,\text{B}}(J) = X_{P,A,A_R,F,\text{B}}(K)$, but different *sum derived variables layers*, $\text{sumlayer}(F,J) = \text{sumlayer}(F,K) + 1$. Ordering by descending *sum derived variables layer* avoids the addition of extra *variables* to the *model* which are merely redundant *reframe variables*, where these *reframe variables* are at the inclusion boundary.

There are other order criteria including (i) descending *shuffle alignment*, $-\text{algn}(\text{apply}(V_A, K, \text{his}(F), A_R))$, and (ii) descending *tuple volume*, $|K^{\text{C}}|$. Ordering by descending *tuple volume* tends to prevent *tuples* from adding *mono-effective variables*, $|(A\%\{u\})^{\text{F}}| = 1 < |\{u\}^{\text{C}}|$.

Given the single *content alignment* optimised next *limited-underlying limited-breadth layer tuple set*, $B_{\text{B}} = \text{topd}(\lfloor \text{bmax}/\text{mmax} \rfloor)(\text{elements}(Z_{P,A,A_R,F,\text{B}}))$, from the *limited-underlying tuple set list maximiser*, $Z_{P,A,A_R,F,\text{B}}$, each *tuple* K of the *tuple set*, B_{B}, can be optimised in a *contracted decrementing linear non-overlapping fuds list maximiser*, $Z_{P,A,A_R,F,\text{n},-,K}$, to construct the single *content alignment* optimised next *limited-underlying limited-breadth layer*. The *limited-layer limited-underlying limited-breadth fud tree searcher* creates a path of *layer*-cumulative *fuds* of length lmax. Define the *limited-layer limited-underlying limited-breadth fud tree searcher*

$$Z_{P,A,A_R,\text{L}} = \text{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\text{u}} \cap \mathcal{F}_{\text{b}} \cap \mathcal{F}_{\text{h}}, P_{P,A,A_R,\text{L}}, \{\emptyset\})$$

where the neighbourhood function returns a singleton

$$\begin{aligned}
&P_{P,A,A_R,\text{L}}(F) = \{G : \\
&\quad G = F \cup \{T : K \in \text{topd}(\lfloor \text{bmax}/\text{mmax} \rfloor)(\text{elements}(Z_{P,A,A_R,F,\text{B}})), \\
&\qquad\qquad H \in \text{topd}(\text{pmax})(\text{elements}(Z_{P,A,A_R,F,\text{n},-,K})), \\
&\qquad\qquad w \in \text{der}(H),\ I = \text{depends}(\text{explode}(H), \{w\}),\ T = I^{\text{TPT}}\}, \\
&\quad \text{layer}(G, \text{der}(G)) \leq \text{lmax}\}
\end{aligned}$$

If the *substrate variables* are *pluri-variate*, $|V_A| > 1$, the tree of the *limited-layer limited-underlying limited-breadth fud tree searcher* has a single path, $|\text{paths}(\text{tree}(Z_{P,A,A_R,\text{L}}))| = 1$, and a single leaf, $|\text{leaves}(\text{tree}(Z_{P,A,A_R,\text{L}}))| = 1$.

Note that in some cases a *partition transform*, I^{TPT}, may already exist in the *fud*, F, because only one *variable* of the *tuple*, K, need be in the *fud derived variables*, $|K \cap \text{der}(F)| \geq 1$, and so some components of the partition of the *tuple* may consist of *variables* in lower *layers* of the *fud*, $J \cap \text{der}(F) = \emptyset$ where $J = \text{und}(I) \subset K$. Furthermore, if, after the first *layer*, a *partition* I_1^{TP} already exists in the *fud*, $I_1^{\text{TP}} \in \text{vars}(F)$, and is not a *derived variable*, $I_1^{\text{TP}} \notin \text{der}(F)$, it may sometimes be hidden by another *variable* I_2^{TP}. That is, $I_1^{\text{TP}} \in \text{vars}(\text{depends}(F, \text{und}(I_2)))$. It is therefore possible that the succeeding *fud*, G, may, in some cases, contain a single *derived variable*, $|\text{der}(G)| = 1$, and consequently be *independent*, $\text{algn}(A * G^{\text{T}}) = 0$. This limitation is due to the separation of the optimisation into two steps, (i) *tuple set list maximisation*, followed by (ii) *decrementing fuds list maximisation*.

Implementations of the neighbourhood function that do not use *partition variables*, $F \notin \mathcal{F}_{U,\mathrm{P}}$, must explicitly check for uniqueness, $I^{\mathrm{TP}} \notin \{T^{\mathrm{P}} : T \in F\}$.

If (i) each *content alignment* optimised next *limited-underlying limited-breadth layer tuple set* has cardinality less than or equal to the *maximum layer breadth* limit, $|B_{\mathrm{B}}| \leq \lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor$, and (ii) each *tuple* $K \in B_{\mathrm{B}}$ has *contracted decrementing linear non-overlapping fuds list maximiser* cardinality of less than or equal to the *maximum tuple optimise limit*, pmax, then the cardinality of each additional *layer* of the *fuds* in the path is less than or equal to the *maximum optimise step cardinality*, $\mathrm{omax} = \mathrm{bmax} \times \mathrm{pmax}$. That is, $|\mathrm{der}(F)| \leq \mathrm{omax}$ where $F \in \mathrm{elements}(Z_{P,A,A_R,\mathrm{L}})$.

If the *substrate variables* are *pluri-variate*, $|V_A| > 1$, the optimised *limited-layer limited-underlying limited-breadth fud* F_{L} of *layer* lmax is the leaf

$$\{F_{\mathrm{L}}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,\mathrm{L}})) \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}$$

If the optimised *limited-layer limited-underlying limited-breadth fud*, F_{L}, exists, it has at least two *variables*, $|\mathrm{vars}(F_{\mathrm{L}}) \setminus V_A| > 1$.

Now the filtering step is computed by constructing *pluri-partition transforms* of the *fud variables* one *variable* at a time. Define the *limited-derived derived variables set list maximiser*

$$Z_{P,A,A_R,F,\mathrm{D}} = \mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{D}}, P_{P,A,A_R,F,\mathrm{D}}, \mathrm{top}(\mathrm{omax}), R_{P,A,A_R,F,\mathrm{D}})$$

where (i) the optimiser function is

$$\begin{aligned} X_{P,A,A_R,F,\mathrm{D}} = \{(K, (I_{\mathrm{a}}^*(A * G^{\mathrm{T}}) - I_{\mathrm{a}}^*(A_R * G^{\mathrm{T}}))/I_{\mathrm{cvl}}^*(G)) :& \\ K \subseteq \mathrm{vars}(F),\ K \neq \emptyset,\ G = \mathrm{depends}(F,K)\}& \end{aligned}$$

(ii) the neighbourhood function is

$$\begin{aligned} P_{P,A,A_R,F,\mathrm{D}}(D) = \{(J, X_{P,A,A_R,F,\mathrm{D}}(J)) :& \\ &(K,\cdot) \in D,\ w \in \mathrm{vars}(F) \setminus V_A \setminus K, \\ &J = K \cup \{w\},\ |J^{\mathrm{C}}| \leq \mathrm{wmax},\ \mathrm{der}(\mathrm{depends}(F,J)) = J\} \end{aligned}$$

and (iii) the initial subset is

$$\begin{aligned} R_{P,A,A_R,F,\mathrm{D}} = \{(J, X_{P,A,A_R,F,\mathrm{D}}(J)) :& \\ &w, u \in \mathrm{vars}(F) \setminus V_A,\ u \neq w, \\ &J = \{w,u\},\ |J^{\mathrm{C}}| \leq \mathrm{wmax},\ \mathrm{der}(\mathrm{depends}(F,J)) = J\} \end{aligned}$$

The *limited-derived derived variables set list maximiser* has no elements if the *fud* is empty, $\mathrm{elements}(Z_{P,A,A_R,\emptyset,\mathrm{D}}) = \emptyset$, or if it consists of a single *partition transform*, $|F| = 1$.

The *derived variables* sets are such that none of the *derived variables* are nested in the *depends fud variables* of another *derived variable* in the same set, $\forall w, u \in J\ (w \neq u \implies u \notin \mathrm{vars}(\mathrm{depends}(F, \{w\})))$, so that $J = \mathrm{der}(\mathrm{depends}(F, J))$. This restriction prevents unnecessary searches in the *optimiser* where the *derived variables* of the *dependent fud* are a proper subset, $\mathrm{der}(\mathrm{depends}(F, J)) \subset J$. However, hidden *variables* that are excluded in a *fud* are not necessarily excluded in another lower *layer fud* that does not contain the *dependent variable*.

The *limited-derived derived variables set list maximiser* differs from the *limited-underlying tuple set list maximiser* in respect of nested *fud variables*. The *tuple set optimiser* allows nested *variables* in a *tuple* in prepartion for *rolling* in the subsequent application of the *contracted decrementing linear non-overlapping fuds list maximiser*, whereas the *depends fuds* of the *derived variables* of the *derived variables set optimiser* must be *limited-models fuds*, $\mathrm{depends}(F, K) \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}$.

An upper bound on the expected cardinality of the searched may be computed given the *maximum derived dimension*, jmax. The upper bound on the expected cardinality for a non-empty *fud*, $F \neq \emptyset$, is

$$\sum_{j \in \{2 \ldots \min(\mathrm{jmax}, r)\}} \binom{r}{j}$$

where $R = \mathrm{vars}(F) \setminus V_A$ and $r = |R|$ and min = minimum.

The *maximum derived dimension*, jmax, may be approximated from the geometric average *valency* $d = |R^{\mathrm{C}}|^{1/r}$, and the *maximum derived volume*, wmax,

$$\mathrm{jmax} = \left\lceil \frac{\ln \mathrm{wmax}}{\ln d} \right\rceil$$

Like the *limited-underlying tuple set list maximiser*, $Z_{P,A,A_R,F,\mathrm{B}}$, the *limited-derived derived variables set list maximiser*, $Z_{P,A,A_R,F,\mathrm{D}}$, has an inclusion function defined $\mathrm{top}(\mathrm{omax}) \in \mathrm{P}(X_{P,A,A_R,F,\mathrm{D}}) \to \mathrm{P}(X_{P,A,A_R,F,\mathrm{D}})$. An implementation of the *derived variables set list maximiser*, $Z_{P,A,A_R,F,\mathrm{D}}$, that guarantees no more than omax *derived variables sets* at each step of the *optimiser* list, $\forall D \in \mathrm{set}(\mathrm{list}(Z_{P,A,A_R,F,\mathrm{D}}))\ (|D| \leq \mathrm{omax})$, must also have additional inclusion order criteria such as descending *sum derived variables layer*, $-\text{sumlayer}(F, J)$.

The optimised *limited-model fuds* are

$$\begin{aligned}\{&\text{depends}(F_{\text{L}}, K) : \\ &\quad \{F_{\text{L}}\} = \text{leaves}(\text{tree}(Z_{P,A,A_R,\text{L}})), \\ &\quad K \in \text{maxd}(\text{elements}(Z_{P,A,A_R,F_{\text{L}},\text{D}}))\} \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\text{q}}\end{aligned}$$

The *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\text{csd},\text{F},\infty,\text{q},P}$, may then be implemented

$$\begin{aligned}&I'^{*}_{z,\text{csd},\text{F},\infty,\text{q},P}(A) = \\ &\quad \{(G, I^{*}_{\text{csd}}((A, A_R, G))) : \\ &\qquad |V_A| > 1,\ \{F_{\text{L}}\} = \text{leaves}(\text{tree}(Z_{P,A,A_R,\text{L}})), \\ &\qquad K \in \text{maxd}(\text{elements}(Z_{P,A,A_R,F_{\text{L}},\text{D}})),\ G = \text{depends}(F_{\text{L}}, K)\} \cup \\ &\quad \{(\emptyset, 0) : |V_A| \leq 1\}\end{aligned}$$

where the *shuffle content alignment valency-density computer* $I_{\text{csd}} \in \text{computers}$ is defined as

$$I^{*}_{\text{csd}}((A, A_R, F)) = (I^{*}_{\text{a}}(A * F^{\text{T}}) - I^{*}_{\text{a}}(A_R * F^{\text{T}}))/I^{*}_{\text{cvl}}(F)$$

In the case where the *substrate histogram*, A, is *scalar* or *mono-variate*, $|V_A| \leq 1$, the *practicable fud inducer* is stuffed with the *empty fud*, because the *contracted decrementing linear non-overlapping fuds list maximiser*, $Z_{P,A,A_R,F,\text{n},-,K}$, in the *limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,\text{L}}$, requires a *pluri-variate tuple*, $|K| > 1$, and so the *limited-underlying tuple set list maximiser*, $Z_{P,A,A_R,F,\text{B}}$, requires a *pluri-variate substrate*, $|V_A| > 1$.

A variation of this implementation of *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\text{csd},\text{F},\infty,\text{q},P}$, is (i) to constrain the *derived variables* to intersect with the highest *layer* of the *fud* and (ii) to terminate the *layer* search as soon as the *shuffle content alignment valency-density* decreases. Define the *highest-layer limited-derived derived variables set list maximiser*

$$Z_{P,A,A_R,F,\text{D},\text{d}} = \text{maximiseLister}(X_{P,A,A_R,F,\text{D}}, P_{P,A,A_R,F,\text{D}}, \text{top}(\text{omax}), R_{P,A,A_R,F,\text{D},\text{d}})$$

where the initial subset is

$$\begin{aligned}R_{P,A,A_R,F,\text{D},\text{d}} \;=\; &\{(J, X_{P,A,A_R,F,\text{D}}(J)) : \\ &\quad w \in \text{der}(F),\ u \in \text{vars}(F) \setminus V_A \setminus \text{vars}(\text{depends}(F, \{w\})), \\ &\qquad J = \{w, u\},\ |J^{\text{C}}| \leq \text{wmax}\}\end{aligned}$$

The upper bound on the expected cardinality for a non-empty *fud*, $F \neq \emptyset$, is

$$\sum_{j \in \{2 \ldots \min(\text{jmax}, r)\}} \binom{r}{j} - \binom{r-x}{j}$$

where $R = \text{vars}(F) \setminus V_A$ and $r = |R|$, $X = \text{der}(F)$ and $x = |X|$.

Define the *highest-layer limited-layer limited-underlying limited-breadth fud tree searcher*

$$Z_{P,A,A_R,\text{L},\text{d}} = \text{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\text{u} \cap \mathcal{F}_\text{b} \cap \mathcal{F}_\text{h}, P_{P,A,A_R,\text{L},\text{d}}, \{\emptyset\})$$

where the neighbourhood function returns a singleton

$$\begin{aligned}
&P_{P,A,A_R,\text{L},\text{d}}(F) = \{G : \\
&\quad G \in P_{P,A,A_R,\text{L}}(F), \\
&\quad (F \neq \emptyset \implies \text{maxr}(\text{el}(Z_{P,A,A_R,F,\text{D},\text{d}})) < \text{maxr}(\text{el}(Z_{P,A,A_R,G,\text{D},\text{d}})))\}
\end{aligned}$$

where el = elements.

The *practicable highest-layer shuffle content alignment valency-density fud inducer*, $I'_{z,\text{csd},\text{F},\infty,\text{q},P,\text{d}}$, may then be implemented

$$\begin{aligned}
&I'^{*}_{z,\text{csd},\text{F},\infty,\text{q},P,\text{d}}(A) = \\
&\quad \{(G, I^{*}_{\text{csd}}((A, A_R, G))) : \\
&\qquad\qquad |V_A| > 1,\ \{F_\text{L}\} = \text{leaves}(\text{tree}(Z_{P,A,A_R,\text{L},\text{d}})), \\
&\qquad\qquad K \in \text{maxd}(\text{elements}(Z_{P,A,A_R,F_\text{L},\text{D},\text{d}})),\ G = \text{depends}(F_\text{L}, K)\} \cup \\
&\quad \{(\emptyset, 0) : |V_A| \leq 1\}
\end{aligned}$$

The *practicable highest-layer shuffle content alignment valency-density fud inducer*, $I'_{z,\text{csd},\text{F},\infty,\text{q},P,\text{d}}$, assumes that there is one maximum along the *layer*-cumulative path of *fuds*. An advantage of the *highest-layer fud inducer* is that *fuds* containing *frame full functional transforms*, having exactly the same *alignment valency-density* of lower *layer fuds* excluding the *reframes*, will be excluded, avoiding the extra computation and reducing the cardinality of the maximum domain, $|\text{maxd}(\text{elements}(Z_{P,A,A_R,F_\text{L},\text{D},\text{d}}))|$. Note that a *computer* implementing the *highest-layer limited-derived derived variables set list maximiser* need not recompute the previous *layer* highest *shuffle content alignment valency-density*, $\text{maxr}(\text{elements}(Z_{P,A,A_R,F,\text{D},\text{d}}))$, but need only to carry it to this *layer*.

If the inclusion functions of the *tuple set list maximiser* and the *derived variables set list maximiser* are further ordered by descending *sum derived variables layer* the *highest-layer fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{d}}$, must be implemented with the *limited-derived derived variables set list maximiser*, $Z_{P,A,A_R,F,\mathrm{D}}$, rather than the *highest-layer limited-derived derived variables set list maximiser*, $Z_{P,A,A_R,F,\mathrm{D,d}}$. That is, $K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A,A_R,F_\mathrm{L},\mathrm{D}}))$. In this way *reframe variables* at the max inclusion boundary may be replaced by *variables* below the highest *layer*.

Another variation of the implementation of *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P}$, is to include the *tuple binary partition* constraint. Define

$$Z_{P,A,A_R,\mathrm{L,b}} = \mathrm{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\mathrm{u} \cap \mathcal{F}_\mathrm{b} \cap \mathcal{F}_\mathrm{h}, P_{P,A,A_R,\mathrm{L,b}}, \{\emptyset\})$$

where the neighbourhood function returns a singleton

$$\begin{aligned}
&P_{P,A,A_R,\mathrm{L,b}}(F) = \{G :\\
&\quad G = F \cup \{T : K \in \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{B}})),\\
&\qquad\qquad H \in \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n,b},-,K})),\\
&\qquad\qquad w \in \mathrm{der}(H),\ I = \mathrm{depends}(\mathrm{explode}(H),\{w\}),\ T = I^{\mathrm{TPT}}\},\\
&\quad \mathrm{layer}(G,\mathrm{der}(G)) \leq \mathrm{lmax}\}
\end{aligned}$$

The *practicable tuple-binary-partition shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{b}}$, may then be implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{b}}(A) =\\
&\quad \{(G, I^{*}_{\mathrm{csd}}((A,A_R,G))) :\\
&\qquad\qquad |V_A| > 1,\ \{F_\mathrm{L}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,\mathrm{L,b}})),\\
&\qquad\qquad K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A,A_R,F_\mathrm{L},\mathrm{D}})),\ G = \mathrm{depends}(F_\mathrm{L},K)\} \cup\\
&\quad \{(\emptyset,0) : |V_A| \leq 1\}
\end{aligned}$$

Another variation of the implementation of *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P}$, is to include the *maximum-roll* constraint by implementing with the *maximum-roll contracted decrementing linear non-overlapping fuds tree maximiser*, $Z_{P,A,A_R,F,\mathrm{n},-,K,\mathrm{mr}}$. Define

$$Z_{P,A,A_R,\mathrm{L,mr}} = \mathrm{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\mathrm{u} \cap \mathcal{F}_\mathrm{b} \cap \mathcal{F}_\mathrm{h}, P_{P,A,A_R,\mathrm{L,mr}}, \{\emptyset\})$$

where the neighbourhood function returns a singleton

$$\begin{aligned}
&P_{P,A,A_R,\mathrm{L},\mathrm{mr}}(F) = \{G : \\
&\quad G = F \cup \{T : K \in \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{B}})), \\
&\qquad\qquad H \in \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n},-,K,\mathrm{mr}})), \\
&\qquad\qquad w \in \mathrm{der}(H),\ I = \mathrm{depends}(\mathrm{explode}(H),\{w\}),\ T = I^{\mathrm{TPT}}\}, \\
&\quad \mathrm{layer}(G,\mathrm{der}(G)) \leq \mathrm{lmax}\}
\end{aligned}$$

The *practicable maximum-roll shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{mr}}$, may then be implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{mr}}(A) = \\
&\quad \{(G, I^{*}_{\mathrm{csd}}((A,A_R,G))) : \\
&\qquad\qquad |V_A| > 1,\ \{F_{\mathrm{L}}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,\mathrm{L},\mathrm{mr}})), \\
&\qquad\qquad K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A,A_R,F_{\mathrm{L}},\mathrm{D}})),\ G = \mathrm{depends}(F_{\mathrm{L}},K)\} \cup \\
&\quad \{(\emptyset,0) : |V_A| \leq 1\}
\end{aligned}$$

Another variation of the implementation of *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P}$, is to exclude *self partitions* from the *derived variables* of the *fuds* of the *limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,\mathrm{L}}$. Define the *excluded-self contracted decrementing linear non-overlapping fuds tree maximiser* as

$$Z_{P,A,A_R,\mathrm{L},\mathrm{xs}} = \mathrm{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}, P_{P,A,A_R,\mathrm{L},\mathrm{xs}}, \{\emptyset\})$$

where the neighbourhood function returns a singleton

$$\begin{aligned}
&P_{P,A,A_R,\mathrm{L},\mathrm{xs}}(F) = \{G : \\
&\quad G = F \cup \{P^{\mathrm{T}} : K \in \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{B}})), \\
&\qquad\qquad H \in \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n},-,K})), \\
&\qquad\qquad w \in \mathrm{der}(H),\ I = \mathrm{depends}(\mathrm{explode}(H),\{w\}), \\
&\qquad\qquad P = I^{\mathrm{TP}},\ P \neq (\cup P)^{\{\}}\}, \\
&\quad \mathrm{layer}(G,\mathrm{der}(G)) \leq \mathrm{lmax}\}
\end{aligned}$$

The rationale for excluding *self partition variables* is to reduce the computation necessary to process redundant *variables*, although note that the *self partition variables* will no longer appear in the top *layer*, $(\cup P)^{\{\}} \notin \mathrm{der}(G)$, and so cannot lift *variables* below during *tuple building*.

Also note that if all but one of the *derived variables* of the top *decrementing linear fuds* are *self partition variables* then the new *fud* G will have a single *derived variable*, $|\mathrm{der}(G)| = 1$, and hence have zero *alignment*, $\mathrm{algn}(A * G^{\mathrm{T}}) = 0$. If the top *decrementing linear fuds* contain only *self partition derived variables* then the neighbourhood function will return the given *fud* unchanged, $G = F$. That is, in this case no new *layer* is added.

The *practicable excluded-self shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{xs}}$, may then be implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{xs}}(A) = \\
&\quad \{(G, I^{*}_{\mathrm{csd}}((A, A_R, G))) : \\
&\qquad\qquad |V_A| > 1,\ \{F_{\mathrm{L}}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,\mathrm{L,xs}})), \\
&\qquad\qquad K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A,A_R,F_{\mathrm{L}},\mathrm{D}})),\ G = \mathrm{depends}(F_{\mathrm{L}}, K)\} \cup \\
&\quad \{(\emptyset, 0) : |V_A| \leq 1\}
\end{aligned}$$

Another variation of the implementation of *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P}$, is to include the *limited-valency* constraint by implementing with the *limited-valency contracted decrementing linear non-overlapping fuds list maximiser*, $Z_{P,A,A_R,F,\mathrm{n,w},-,K}$. Define

$$Z_{P,A,A_R,\mathrm{L,w}} = \mathrm{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}, P_{P,A,A_R,\mathrm{L,w}}, \{\emptyset\})$$

where the neighbourhood function returns a singleton

$$\begin{aligned}
&P_{P,A,A_R,\mathrm{L,w}}(F) = \{G : \\
&\quad G = F \cup \{T : K \in \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{B}})), \\
&\qquad\qquad H \in \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n,w},-,K})), \\
&\qquad\qquad w \in \mathrm{der}(H),\ I = \mathrm{depends}(\mathrm{explode}(H), \{w\}),\ T = I^{\mathrm{TPT}}\}, \\
&\quad \mathrm{layer}(G, \mathrm{der}(G)) \leq \mathrm{lmax}\}
\end{aligned}$$

The *practicable limited-valency shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{w}}$, may then be implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{w}}(A) = \\
&\quad \{(G, I^{*}_{\mathrm{csd}}((A, A_R, G))) : \\
&\qquad\qquad |V_A| > 1,\ \{F_{\mathrm{L}}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,\mathrm{L,w}})), \\
&\qquad\qquad K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A,A_R,F_{\mathrm{L}},\mathrm{D}})),\ G = \mathrm{depends}(F_{\mathrm{L}}, K)\} \cup \\
&\quad \{(\emptyset, 0) : |V_A| \leq 1\}
\end{aligned}$$

In some cases a *tuple* K returned by the *limited-underlying tuple set list maximiser*, $Z_{P,A,A_R,F,\mathrm{B}}$, will be rejected by the subsequent *limited-valency contracted decrementing linear non-overlapping fuds list maximiser*, $Z_{P,A,A_R,F,\mathrm{n,w},-,K}$, because there are no *limited-valency tuple partitions*, $\forall Y \in \mathrm{B}(K)\ ((|Y| > \mathrm{mmax}) \vee \neg(\forall M \in Y\ (|M^{\mathrm{C}}| \leq \mathrm{umax})))$. To avoid processing a *tuple* which is destined to fail the *limited-valency constraint*, a variation of the *limited-underlying tuple set list maximiser* checks to ensure there is at least one *limited-valency partition* of the *tuple*. Define the *checked-valency limited-underlying tuple set list maximiser*

$$\begin{aligned}&Z_{P,A,A_R,F,\mathrm{B,wc}} = \\ &\quad \mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{B}}, P_{P,A,A_R,F,\mathrm{B,wc}}, \mathrm{top}(\mathrm{omax}), R_{P,A,A_R,F,\mathrm{B,wc}})\end{aligned}$$

where the neighbourhood function is

$$\begin{aligned}P_{P,A,A_R,F,\mathrm{B,wc}}(B) = \{&(J, X_{P,A,A_R,F,\mathrm{B}}(J)) : \\ &(K,\cdot) \in B,\ w \in \mathrm{vars}(F) \cup V_A \setminus K,\ J = K \cup \{w\},\ |J^{\mathrm{C}}| \leq \mathrm{xmax}, \\ &\exists Y \in \mathrm{B}(J)\ ((|Y| \leq \mathrm{mmax}) \wedge (\forall M \in Y\ (|M^{\mathrm{C}}| \leq \mathrm{umax})))\}\end{aligned}$$

and the initial subset is

$$\begin{aligned}R_{P,A,A_R,\emptyset,\mathrm{B,wc}} &= \{(\{w,u\}, X_{P,A,A_R,\emptyset,\mathrm{B}}(\{w,u\})) : \\ &\qquad w,u \in V_A,\ u \neq w,\ |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}, \\ &\qquad |\{w\}^{\mathrm{C}}| \leq \mathrm{umax},\ |\{u\}^{\mathrm{C}}| \leq \mathrm{umax}\} \\ R_{P,A,A_R,F,\mathrm{B,wc}} &= \{(\{w,u\}, X_{P,A,A_R,F,\mathrm{B}}(\{w,u\})) : \\ &\qquad w \in \mathrm{der}(F),\ u \in \mathrm{vars}(F) \cup V_A,\ u \neq w,\ |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}, \\ &\qquad |\{w\}^{\mathrm{C}}| \leq \mathrm{umax},\ |\{u\}^{\mathrm{C}}| \leq \mathrm{umax}\}\end{aligned}$$

Another variation of the implementation of the *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P}$, is to add a cached or common *substrate fud* $F_{\mathrm{c}} \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}$. For example, a common *fud* may be supplied from the parent *slice fuds* of a *decomposition*.

The *common-fud limited-underlying tuple set list maximiser* $Z_{P,A,A_R,F_{\mathrm{c}},F,\mathrm{B}}$ can choose *tuples* from the *variables* of the *common fud*, $\mathrm{vars}(F_{\mathrm{c}})$, as well as from the *substrate variables*, V_A, and the *variables* of the given *fud*, $\mathrm{vars}(F)$. Define the *common-fud limited-underlying tuple set list maximiser*

$$\begin{aligned}&Z_{P,A,A_R,F_{\mathrm{c}},F,\mathrm{B}} = \\ &\quad \mathrm{maximiseLister}(X_{P,A,A_R,F_{\mathrm{c}},F,\mathrm{B}}, P_{P,A,A_R,F_{\mathrm{c}},F,\mathrm{B}}, \mathrm{top}(\mathrm{omax}), R_{P,A,A_R,F_{\mathrm{c}},F,\mathrm{B}})\end{aligned}$$

where (i) the optimiser function is

$$X_{P,A,A_R,F_c,F,\mathrm{B}} = \\ \{(K, I_{\mathrm{a}}^*(\mathrm{apply}(V_A, K, \mathrm{his}(F \cup F_c), A)) - I_{\mathrm{a}}^*(\mathrm{apply}(V_A, K, \mathrm{his}(F \cup F_c), A_R))) : \\ K \in \mathrm{tuples}(\mathrm{vars}(F_c) \cup V_A, F)\}$$

and (ii) the neighbourhood function is

$$P_{P,A,A_R,F_c,F,\mathrm{B}}(B) = \{(J, X_{P,A,A_R,F_c,F,\mathrm{B}}(J)) : \\ (K, \cdot) \in B,\ w \in \mathrm{vars}(F \cup F_c) \cup V_A \setminus K,\ J = K \cup \{w\},\ |J^{\mathrm{C}}| \leq \mathrm{xmax}\}$$

and (iii) the initial subset is

$$\begin{aligned} R_{P,A,A_R,F_c,\emptyset,\mathrm{B}} &= \{(\{w,u\}, X_{P,A,A_R,F_c,\emptyset,\mathrm{B}}(\{w,u\})) : \\ &\quad w,u \in \mathrm{vars}(F_c) \cup V_A,\ u \neq w,\ |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}\} \\ R_{P,A,A_R,F_c,F,\mathrm{B}} &= \{(\{w,u\}, X_{P,A,A_R,F_c,F,\mathrm{B}}(\{w,u\})) : \\ &\quad w \in \mathrm{der}(F),\ u \in \mathrm{vars}(F \cup F_c) \cup V_A,\ u \neq w, \\ &\quad |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}\} \end{aligned}$$

An upper bound on the expected cardinality of the searched may be computed given the *maximum underlying dimension*, kmax. The upper bound on the expected cardinality in the first *layer*, $F = \emptyset$, is

$$\sum_{k \in \{2 \ldots \min(\mathrm{kmax}, s)\}} \binom{s}{k}$$

where $s = |\mathrm{vars}(F_c) \cup V|$ and min = minimum. In subsequent *layers*, $F \neq \emptyset$, the upper bound on the expected cardinality is

$$\sum_{k \in \{2 \ldots \min(\mathrm{kmax}, t)\}} \binom{t}{k} - \binom{t-x}{k}$$

where $W = \mathrm{vars}(F \cup F_c) \cup V_A$, $t = |W|$, $X = \mathrm{der}(F)$ and $x = |X|$.

Define the *common-fud limited-layer limited-underlying limited-breadth fud tree searcher*

$$Z_{P,A,A_R,F_c,\mathrm{L}} = \mathrm{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}, P_{P,A,A_R,F_c,\mathrm{L}}, \{\emptyset\})$$

where the neighbourhood function returns a singleton

$$
\begin{aligned}
&P_{P,A,A_R,F_\mathrm{c},\mathrm{L}}(F) = \{G : \\
&\quad G = F \cup \bigcup \{\{T\} \cup \mathrm{depends}(F_\mathrm{c}, \mathrm{und}(T)) : \\
&\qquad\qquad K \in \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F_\mathrm{c},F,\mathrm{B}})), \\
&\qquad\qquad H \in \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F \cup F_\mathrm{c},\mathrm{n},-,K})), \\
&\qquad\qquad w \in \mathrm{der}(H),\ I = \mathrm{depends}(\mathrm{explode}(H), \{w\}),\ T = I^{\mathrm{TPT}}\}, \\
&\quad \mathrm{layer}(G, \mathrm{der}(G)) \leq \mathrm{lmax}\}
\end{aligned}
$$

Whereas in the *limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,\mathrm{L}}$, the *layers* of the *fud* increment at each step along the path, $\forall (i,G) \in L\ (\mathrm{layer}(G, \mathrm{der}(G)) = i)$ where $L \in \mathrm{paths}(\mathrm{tree}(Z_{P,A,A_R,\mathrm{L}}))$, in the *common-fud fud tree searcher*, $Z_{P,A,A_R,F_\mathrm{c},\mathrm{L}}$, there is no such guarantee.

Note that in some cases a *partition transform*, I^{TPT}, may already exist in the *common fud*, F_c. Just as in the case of the *fud tree searcher*, $Z_{P,A,A_R,\mathrm{L}}$, above, the *common-fud fud tree searcher*, $Z_{P,A,A_R,F_\mathrm{c},\mathrm{L}}$, may also hide *partition variables*, but in this case it may occur in the first step, because the *layer* need not correspond to the *common-fud fud tree searcher* path position.

Implementations of the neighbourhood function that do not use *partition variables*, $F \notin \mathcal{F}_{U,\mathrm{P}}$, must explicitly check for uniqueness, $I^{\mathrm{TP}} \notin \{T^{\mathrm{P}} : T \in F_\mathrm{c}\}$. If the *partition transform* is in the *common fud*, $T_\mathrm{c} \in F_\mathrm{c}$ where $T_\mathrm{c}^{\mathrm{P}} = I^{\mathrm{TP}}$, then the *common fud's transform* should be added to the given *fud* instead, $F \cup \{T_\mathrm{c}\}$.

The *practicable common-fud shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,F_\mathrm{c}}$, may then be implemented

$$
\begin{aligned}
&I'^{*}_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,F_\mathrm{c}}(A) = \\
&\quad \{(G, I^{*}_{\mathrm{csd}}((A, A_R, G))) : \\
&\qquad\qquad |V_A| > 1,\ \{F_\mathrm{L}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,F_\mathrm{c},\mathrm{L}})), \\
&\qquad\qquad K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A,A_R,F_\mathrm{L},\mathrm{D}})),\ G = \mathrm{depends}(F_\mathrm{L}, K)\} \cup \\
&\quad \{(\emptyset, 0) : |V_A| \leq 1\}
\end{aligned}
$$

Note that the addition of a *common fud* hint to the *common-fud fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,F_\mathrm{c}}$, will not necessarily produce the same *models* as the *fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P}$, without the hint. Nor is there necessarily an improvement in computation performance.

Another variation of the implementation of the *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P}$, is to explicitly specify the *substrate*. Rather than modelling with the given *substrate variables*, V_A, *level* modelling is parameterised by a pair of (i) a set of *variables* V_{g}, which is a subset of the *substrate variables*, $V_{\mathrm{g}} \subseteq V_A$, and (ii) a *level fud* $F_{\mathrm{g}} \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}$, which is such that its *underlying* is also a subset of the *substrate variables*, $\mathrm{und}(F_{\mathrm{g}}) \subseteq V_A$. Here only the union of (i) the *substrate variables* subset, V_{g}, and (ii) the *derived variables* of the given *level fud*, $\mathrm{der}(F_{\mathrm{g}})$, are visible to the *tuple maximiser*, so the *substrate variables*, V_A, are effectively replaced by the *level variables*, $V_{\mathrm{g}} \cup \mathrm{der}(F_{\mathrm{g}})$.

The *level fud inducer* allows multiple *levels* to be modelled in sequence, so, for example, large *substrates*, V_A, with large *underlying volumes*, $|V_A^{\mathrm{C}}|$, may be made practicable by (i) partitioning them into components, $V_{\mathrm{g}} \in P$, where $P \in \mathrm{B}(V_A)$, with smaller *underlying volumes*, $|V_{\mathrm{g}}^{\mathrm{C}}| < |V_A^{\mathrm{C}}|$, (ii) *inducing* a *level fud* on each component, V_{g}, of the *substrate* partition, and then (iii) combining these *level fuds* in a higher *level* to produce a model with coverage of the whole *substrate*, V_A. Another example is to use the *level fud inducer* in order to exclude *mono-valent substrate variables*, $V_{\mathrm{g}} = \{w : w \in V_A,\ |(A\%\{w\})^{\mathrm{F}}| > 1\}$, which might occur near the leaves of a *decomposition*. Note that higher *levels* do not necessarily require *non-overlapping level fuds*.

The *level limited-underlying tuple set list maximiser* $Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}}$ replaces the *substrate variables*, V_A, with the *level variables*, $V_{\mathrm{g}} \cup \mathrm{der}(F_{\mathrm{g}})$. Define the *level limited-underlying tuple set list maximiser*

$$\begin{aligned}
&Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}} = \\
&\quad \mathrm{maximiseLister}(X_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}}, P_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}}, \mathrm{top}(\mathrm{omax}), R_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}})
\end{aligned}$$

where (i) the optimiser function is

$$\begin{aligned}
&X_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}} = \\
&\quad \{(K, I_{\mathrm{a}}^*(\mathrm{apply}(V_A,K,\mathrm{his}(F \cup F_{\mathrm{g}}),A)) - I_{\mathrm{a}}^*(\mathrm{apply}(V_A,K,\mathrm{his}(F \cup F_{\mathrm{g}}),A_R))) : \\
&\qquad\qquad K \in \mathrm{tuples}(V_{\mathrm{g}} \cup \mathrm{der}(F_{\mathrm{g}}), F)\}
\end{aligned}$$

and (ii) the neighbourhood function is

$$\begin{aligned}
P_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}}(B) = \{&(J, X_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}}(J)) : \\
&(K,\cdot) \in B,\ w \in \mathrm{vars}(F) \setminus \mathrm{vars}(F_{\mathrm{g}}) \cup V_{\mathrm{g}} \cup \mathrm{der}(F_{\mathrm{g}}) \setminus K, \\
&J = K \cup \{w\},\ |J^{\mathrm{C}}| \leq \mathrm{xmax}\}
\end{aligned}$$

and (iii) the initial subset is

$$\begin{aligned}
R_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\emptyset,\mathrm{B}} &= \{(\{w,u\}, X_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\emptyset,\mathrm{B}}(\{w,u\})) : \\
&\quad w,u \in V_{\mathrm{g}} \cup \mathrm{der}(F_{\mathrm{g}}),\ u \neq w,\ |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}\} \\
R_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}} &= \{(\{w,u\}, X_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}}(\{w,u\})) : \\
&\quad w \in \mathrm{der}(F),\ u \in \mathrm{vars}(F) \setminus \mathrm{vars}(F_{\mathrm{g}}) \cup V_{\mathrm{g}} \cup \mathrm{der}(F_{\mathrm{g}}),\ u \neq w, \\
&\quad |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}\}
\end{aligned}$$

An upper bound on the expected cardinality of the searched may be computed given the *maximum underlying dimension*, kmax. The upper bound on the expected cardinality in the first *layer*, $F = \emptyset$, is

$$\sum_{k \in \{2 \ldots \min(\mathrm{kmax},s)\}} \binom{s}{k}$$

where $s = |V_{\mathrm{g}} \cup \mathrm{der}(F_{\mathrm{g}})|$ and min = minimum. In subsequent *layers*, $F \neq \emptyset$, the upper bound on the expected cardinality is

$$\sum_{k \in \{2 \ldots \min(\mathrm{kmax},t)\}} \binom{t}{k} - \binom{t-x}{k}$$

where $W = \mathrm{vars}(F) \setminus \mathrm{vars}(F_{\mathrm{g}}) \cup V_{\mathrm{g}} \cup \mathrm{der}(F_{\mathrm{g}})$, $t = |W|$, $X = \mathrm{der}(F)$ and $x = |X|$.

Define the *level limited-layer limited-underlying limited-breadth fud tree searcher*

$$Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}} = \mathrm{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}, P_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}}, \{\emptyset\})$$

where the neighbourhood function returns a singleton

$$\begin{aligned}
&P_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}}(F) = \{G : \\
&\quad G = F \cup \bigcup \{\{T\} \cup \mathrm{depends}(F_{\mathrm{g}}, \mathrm{und}(T)) : \\
&\qquad\qquad K \in \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},F,\mathrm{B}})), \\
&\qquad\qquad H \in \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F \cup F_{\mathrm{g}},\mathrm{n},-,K})), \\
&\qquad\qquad w \in \mathrm{der}(H),\ I = \mathrm{depends}(\mathrm{explode}(H), \{w\}),\ T = I^{\mathrm{TPT}}\}, \\
&\quad \mathrm{layer}(G, \mathrm{der}(G)) \leq \mathrm{lmax}\}
\end{aligned}$$

Note that the resultant *fud* of the *level fud tree searcher*, $Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}}$, has its *underlying variables* flattened to the *substrate*. That is, $\mathrm{und}(F_{\mathrm{L}}) \subseteq V_A$, where $\{F_{\mathrm{L}}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}}))$. So it is not necessary to supply

F_{g} along with F_{L}.

Whereas in the *limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,\mathrm{L}}$, the *layers* of the *fud* increment at each step along the path, $\forall(i,G) \in L\ (\mathrm{layer}(G,\mathrm{der}(G)) = i)$ where $L \in \mathrm{paths}(\mathrm{tree}(Z_{P,A,A_R,\mathrm{L}}))$, in the *level fud tree searcher*, $Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}}$, there is no such guarantee.

Define the *level limited-derived derived variables set list maximiser*

$$Z_{P,A,A_R,F_{\mathrm{g}},F,\mathrm{D}} = \mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{D}}, P_{P,A,A_R,F_{\mathrm{g}},F,\mathrm{D}}, \mathrm{top}(\mathrm{omax}), R_{P,A,A_R,F_{\mathrm{g}},F,\mathrm{D}})$$

where the neighbourhood function is

$$\begin{aligned}
P_{P,A,A_R,F_{\mathrm{g}},F,\mathrm{D}}(D) = \{&(J, X_{P,A,A_R,F,\mathrm{D}}(J)) :\\
&(K,\cdot) \in D,\ w \in \mathrm{vars}(F) \setminus V_A \setminus \mathrm{vars}(F_{\mathrm{g}}) \setminus K,\\
&J = K \cup \{w\},\ |J^{\mathrm{C}}| \leq \mathrm{wmax},\ \mathrm{der}(\mathrm{depends}(F,J)) = J\}
\end{aligned}$$

and the initial subset is

$$\begin{aligned}
R_{P,A,A_R,F_{\mathrm{g}},F,\mathrm{D}} = \{&(J, X_{P,A,A_R,F,\mathrm{D}}(J)) :\\
&w,u \in \mathrm{vars}(F) \setminus V_A \setminus \mathrm{vars}(F_{\mathrm{g}}),\ u \neq w,\\
&J = \{w,u\},\ |J^{\mathrm{C}}| \leq \mathrm{wmax},\ \mathrm{der}(\mathrm{depends}(F,J)) = J\}
\end{aligned}$$

The *practicable level shuffle content alignment valency-density fud inducer*, $I_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,V_{\mathrm{g}},F_{\mathrm{g}}}^{'}$, may then be implemented

$$\begin{aligned}
&I_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,V_{\mathrm{g}},F_{\mathrm{g}}}^{'*}(A) =\\
&\quad \{(G, I_{\mathrm{csd}}^{*}((A,A_R,G))) :\\
&\qquad\qquad |V_A| > 1,\ \{F_{\mathrm{L}}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}})),\\
&\qquad\qquad K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A,A_R,F_{\mathrm{g}},F_{\mathrm{L}},\mathrm{D}})),\ G = \mathrm{depends}(F_{\mathrm{L}},K)\} \cup\\
&\quad \{(\emptyset,0) : |V_A| \leq 1\}
\end{aligned}$$

Of the variations described above of the implementation of *practicable shuffle content alignment valency-density fud inducer*, $I_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P}^{'}$, only the *practicable limited-valency shuffle content alignment valency-density fud inducer*, $I_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{w}}^{'}$, is potentially unrestricted. That is, the *limited-valency inducer*, $I_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{w}}^{'}$, can perform the same search as the unlimited *inducer*, $I_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P}^{'}$, if the *maximum valency* is set equal to the *maximum underlying volume*, umax = xmax. The other variations all have restricted functionality with respect to the unlimited *inducer*, $I_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P}^{'}$, no matter what the parameters.

As shown above, the use of a *shuffle histogram*, A_R, is a practicable approximation to the *independent*, A^{X}, in the *practicable shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P}$. This allows optimisations to avoid a two stage (i) search of possibly *overlapping fuds*, $\mathrm{select}(T_A, N_A) \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}$, followed by (ii) filtering of *non-overlapping fuds*, $\{F : F \in \mathrm{select}(T_A, N_A),\ \mathrm{nd}(F)\} \subset \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$. The optimisers do this by maximisation of the *shuffle content alignment valency-density* to construct *fuds* that approximate loosely to *recursively non-overlapping pluri-partition fuds*, from which an approximately *non-overlapping top transform* can be chosen. The same reasoning may be extended to a *fud decomposition inducer*. Here a *shuffle histogram* is constructed from each of the *slices* of the *fuds* of the *decomposition*. Each *shuffle* approximates to the *independent* of the *contingent sample*. Redefine the *shuffle indices*, $R_A \subseteq \{1 \ldots z_A!^{n_A}\}$, where $A \in \mathcal{A}_{z(A)}$, $z_A = \mathrm{size}(A)$ and $n_A = |\mathrm{vars}(A)|$. Redefine the *scaled shuffle histogram*, $A_{R(A)} = \mathrm{scalar}(1/|R(A)|) * \sum_{r \in R(A)} L_A(r)$ where $X_A \in \mathrm{enums}(\mathrm{shuffles}(\mathrm{history}(A)))$ and $L_A = \mathrm{map}(\mathrm{his}, \mathrm{flip}(X_A))$. Then the *scaled contingent shuffle histogram* is $(A * C)_{R(A*C)} \approx (A * C)^{\mathrm{X}}$, where $(\cdot, C) \in \mathrm{cont}(D)$ and $D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A}$.

The *practicable summed shuffle content alignment valency-density fud decomposition inducer*, which, given *substrate histogram* $A \in \mathcal{A}_z$, is constrained

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P}(A) \subseteq \\
&\quad \{(D, I^{*}_{\mathrm{Scsd}}((A, D))) : \\
&\qquad\qquad D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}}), \\
&\qquad\qquad\quad \forall (C, F) \in \mathrm{cont}(D) \\
&\qquad\qquad\qquad (I^{*}_{\mathrm{a}}(A * C * F^{\mathrm{T}}) - I^{*}_{\mathrm{a}}((A * C)_{R(A*C)} * F^{\mathrm{T}}) > 0)\} \cup \\
&\quad \{(D_{\emptyset}, 0)\}
\end{aligned}$$

where $D_{\emptyset} = \{((\emptyset, \emptyset), \emptyset)\}$ and the *summed shuffle content alignment valency-density computer* $I_{\mathrm{Scsd}} \in \mathrm{computers}$ is defined as

$$\begin{aligned}
&I^{*}_{\mathrm{Scsd}}((A, D)) = \\
&\quad \sum (I^{*}_{\mathrm{a}}(A * C * F^{\mathrm{T}}) - I^{*}_{\mathrm{a}}((A * C)_{R(A*C)} * F^{\mathrm{T}}))/I^{*}_{\mathrm{cvl}}(F) : (C, F) \in \mathrm{cont}(D)
\end{aligned}$$

In some cases the *practicable inducer* optimisation may be empty. For example, the *independent substrate histogram*, A^{X}, cannot have non-zero positive *summation content alignment*. The *maximum function*, $\mathrm{maxr}(I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P}(A^{\mathrm{X}}))$, would therefore be undefined for some *inducer* domain *substrate histograms*. In order to have well defined *maximum correlation*, the *practicable inducer* is therefore stuffed with the *empty decomposition*, $D_\emptyset \in \mathcal{D}_{\mathrm{F}}$, in the case of empty optimisation.

Each non-zero positive *shuffle content alignment valency-density fud decomposition* of the application of the *practicable fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P}$, is related to the computation of *fuds* of the *slices* in the *practicable fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P}$, which is defined in terms of the *limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,\mathrm{L}}$, and the *limited-derived derived variables set list maximiser*, $Z_{P,A,A_R,F,\mathrm{D}}$,

$$\begin{aligned}
&\forall (D,a) \in I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P}(A)\ (a > 0 \implies \\
&\quad (\forall (C,F) \in \mathrm{cont}(D)\ \forall F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,A*C,(A*C)_{R(A*C)},\mathrm{L}})) \\
&\qquad \exists K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A*C,(A*C)_{R(A*C)},F_{\mathrm{L}},\mathrm{D}}))\ (F = \mathrm{depends}(F_{\mathrm{L}},K))))
\end{aligned}$$

Note that the *practicable summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P}$, is defined such that the *contingent histogram*, $A * C$, is *shuffled*, $(A * C)_{R(A*C)}$, rather taking the *contingent* of the *shuffled histogram*, $A_{R(A)} * C$. If the *contingent shuffle histogram* were used then any biases for or against the *alignment* in the *shuffle*, $A_{R(A)}$, in the *parent slice* would be safely removed. However, the *size* of the *contingent shuffle histogram* is not necessarily equal to that of the *contingent sample*, $\mathrm{size}(A_{R(A)} * C) \neq \mathrm{size}(A * C)$, and so *scaling* is often necessary. This would not be a disadvantage if it were not the case that typically the *entropy* of the *parent derived histogram* of the *contingent shuffle* is greater than that of the *sample*, and so the *contingent shuffle slice size* tends to zero much more quickly than the *sample slice size*. Therefore the *scaling* factor is often large, making the *contingent shuffle* less effective as an approximation to the *independent slice*, $(A * C)^{\mathrm{X}}$.

In section 'Substrate models computation' above, the finite *limited-models infinite-layer fud decomposition tree*, $\mathrm{tdfiq}(U) \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{D}_{\mathrm{F,d}} \to \mathrm{trees}(\mathcal{D}_{\mathrm{F,d}})$, is a tree of *immediate super-decompositions* of *limited-models infinite-layer substrate fuds*. The *decompositions* of the tree are a subset the *limited-models infinite-layer substrate fud decompositions*

$$\mathcal{D}_{\mathrm{F},\infty,U,V} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}}) \supset \mathrm{elements}(\mathrm{tdfiq}(U)(V))$$

The *limited-models infinite-layer substrate fud decompositions tree searcher* chooses a sublist of a path of *immediate super-decompositions* from the *limited-models infinite-layer fud decomposition tree*. Define the *limited-models infinite-layer substrate fud decompositions tree searcher*

$$Z_{P,A,\mathrm{D,F}} = \mathrm{searchTreer}(\mathcal{D}_{\mathrm{F},\infty,U,V} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}}), P_{P,A,\mathrm{D,F}}, R_{P,A,\mathrm{D,F}})$$

where the neighbourhood function returns a singleton

$$
\begin{aligned}
P_{P,A,\mathrm{D},\mathrm{F}}(D) = \{E :&\\
&(\cdot, S, G, L) \in \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}\times\mathrm{S}\times\mathcal{X}^2}, \{(\mathrm{size}(B), S, G, L) :\\
&\quad (L, Y) \in \mathrm{places}(D),\\
&\quad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)),\\
&\quad (\cdot, F) = L_{|L|},\ W = \mathrm{der}(F),\\
&\quad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)),\\
&\quad B = \mathrm{apply}(V_A, V_A, \mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\}, A),\ \mathrm{size}(B) > 0,\\
&\quad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,B,B_{R(B)},\mathrm{L}})),\\
&\quad (K, a) \in \max(\mathrm{elements}(Z_{P,B,B_{R(B)},F_{\mathrm{L}},\mathrm{D}})),\ a > 0,\\
&\quad G = \mathrm{depends}(F_{\mathrm{L}}, K)\})),\\
&M = L \cup \{(|L| + 1, (S, G))\},\\
&E = \mathrm{tree}(\mathrm{paths}(D) \setminus \{L\} \cup \{M\})\}
\end{aligned}
$$

where

$$
\begin{aligned}
R_{P,A,\mathrm{D},\mathrm{F}} = \{\{((\emptyset, G), \emptyset)\} :&\\
&G \in \mathrm{maxd}(\mathrm{order}(D_{\mathrm{F}}, \{G :\\
&\quad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_{R(A)},\mathrm{L}})),\\
&\quad (K, a) \in \max(\mathrm{elements}(Z_{P,A,A_{R(A)},F_{\mathrm{L}},\mathrm{D}})),\ a > 0,\\
&\quad G = \mathrm{depends}(F_{\mathrm{L}}, K)\}))\}
\end{aligned}
$$

The computation of the *slice* B is a tractable *fud application equivalent* to the *application* of the *fud's transforms' histograms*, $\mathrm{his}(H_L)$, *multiplied* by the *slice derived state*, $\{R_L \cup S\}^{\mathrm{U}}$, followed by *reduction* to the *substrate*, V_A,

$$\mathrm{apply}(V_A, V_A, \mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\}, A) = A * \prod \mathrm{his}(H_L) * \{R_L \cup S\}^{\mathrm{U}} \ \%\ V_A$$

The neighbourhood function $P_{P,A,\mathrm{D},\mathrm{F}}(D)$ returns an empty set or a singleton *super-decomposition* with an additional *slice* having non-zero positive *shuffle content alignment*. The order $D_{\mathbf{Q}\times\mathrm{S}\times\mathcal{X}^2}$ selects by *slice size* and then arbitrarily. Together the orders $D_{\mathbf{Q}\times\mathrm{S}\times\mathcal{X}^2}$ and D_{F} ensure that the *fud decomposition* is *distinct*. The tree of the *limited-models infinite-layer substrate fud decompositions tree searcher* has at most one path, $|\mathrm{paths}(\mathrm{tree}(Z_{P,A,\mathrm{D},\mathrm{F}}))| \leq 1$, and hence the tree has at most one leaf, $|\mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D},\mathrm{F}}))| \leq 1$. If the path exists, $\{L\} = \mathrm{paths}(\mathrm{tree}(Z_{P,A,\mathrm{D},\mathrm{F}}))$, it is in the *limited-models infinite-layer fud decomposition tree*, $L \in \mathrm{subpaths}(\mathrm{tdfiq}(U_A)(V_A, \emptyset))$.

The *practicable summed shuffle content alignment valency-density fud decomposition inducer* may then be implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_{\emptyset}, 0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D,F}})),\ \{D\} = Q
\end{aligned}$$

The *highest-layer limited-models infinite-layer substrate fud decompositions tree searcher*, $Z_{P,A,\mathrm{D,F,d}}$, is defined exactly the same as the *limited-models infinite-layer substrate fud decompositions tree searcher*, $Z_{P,A,\mathrm{D,F}}$, except that it depends instead on the *highest-layer limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,B,B_R,\mathrm{L,d}}$, and the *highest-layer limited-derived derived variables set list maximiser*, $Z_{P,B,B_R,F_\mathrm{L},\mathrm{D,d}}$. The *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer* is implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_{\emptyset}, 0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D,F,d}})),\ \{D\} = Q
\end{aligned}$$

The *common-fud limited-underlying tuple set list maximiser*, $Z_{P,A,A_R,F_\mathrm{c},F,\mathrm{B}}$, and the *common-fud limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,F_\mathrm{c},\mathrm{L}}$, of the *practicable common-fud shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd,F},\infty,\mathrm{q},P,F_\mathrm{c}}$, can be used to implement an accumulating *fud* along the path of *immediate super-decompositions* of the *limited-models infinite-layer fud decomposition tree*. An accumulated *fud* allows children *slice fuds* to incorporate the lower *layers* of parent *slice fuds* in a *decomposition* tree path, thus reducing computation in some cases. Implementations of *fud tree searchers* that do not use *partition variables*, $F \notin \mathcal{F}_{U,\mathrm{P}}$, also avoid unnecessary duplication of *partition variables*. These implementations can also detect *fud symmetries* in the *decomposition*, without the need to compare *fuds* in different paths explicitly.

An *accumulating-fud fud decompositions tree searcher* must carry around the *accumulating-fud* so far. So the *searcher* domain consists of pairs of *decompositions* and *common-fuds*. Define the *accumulating-fud limited-models infinite-layer substrate fud decompositions tree searcher*

$$\begin{aligned}
&Z_{P,A,\mathrm{D,F,c}} = \\
&\quad \mathrm{searchTreer}((\mathcal{D}_{\mathrm{F},\infty,U,V} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_\mathrm{q})) \times (\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\mathrm{u} \cap \mathcal{F}_\mathrm{b} \cap \mathcal{F}_\mathrm{h}), \\
&\qquad\qquad P_{P,A,\mathrm{D,F,c}}, R_{P,A,\mathrm{D,F,c}})
\end{aligned}$$

where the neighbourhood function returns a singleton

$$
\begin{aligned}
&P_{P,A,\mathrm{D},\mathrm{F},\mathrm{c}}((D,F_{\mathrm{c}})) = \{(E,G_{\mathrm{c}}) : \\
&\qquad (\cdot,S,G,L,G_{\mathrm{c}}) \in \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}\times\mathrm{S}\times\mathcal{X}^2\times\mathrm{F}},\{(\mathrm{size}(B),S,G,L,F_{\mathrm{c}}\cup F_{\mathrm{L}}) : \\
&\qquad\quad (L,Y) \in \mathrm{places}(D), \\
&\qquad\quad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad\quad (\cdot,F) = L_{|L|},\ W = \mathrm{der}(F), \\
&\qquad\quad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)), \\
&\qquad\quad B = \mathrm{apply}(V_A,V_A,\mathrm{his}(H_L)\cup\{\{R_L\cup S\}^{\mathrm{U}}\},A),\ \mathrm{size}(B)>0, \\
&\qquad\quad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,B,B_{R(B)},F_{\mathrm{c}},\mathrm{L}})), \\
&\qquad\quad (K,a) \in \max(\mathrm{elements}(Z_{P,B,B_{R(B)},F_{\mathrm{L}},\mathrm{D}})),\ a>0, \\
&\qquad\quad G = \mathrm{depends}(F_{\mathrm{L}},K)\})), \\
&\qquad M = L \cup \{(|L|+1,(S,G))\}, \\
&\qquad E = \mathrm{tree}(\mathrm{paths}(D)\setminus\{L\}\cup\{M\})\}
\end{aligned}
$$

where

$$
\begin{aligned}
&R_{P,A,\mathrm{D},\mathrm{F},\mathrm{c}} = \{(\{((\emptyset,G),\emptyset)\},F_{\mathrm{L}}) : \\
&\qquad (G,F_{\mathrm{L}}) \in \mathrm{maxd}(\mathrm{order}(D_{\mathrm{F}^2},\{(G,F_{\mathrm{L}}) : \\
&\qquad\quad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_{R(A)},\mathrm{L}})), \\
&\qquad\quad (K,a) \in \max(\mathrm{elements}(Z_{P,A,A_{R(A)},F_{\mathrm{L}},\mathrm{D}})),\ a>0, \\
&\qquad\quad G = \mathrm{depends}(F_{\mathrm{L}},K)\}))\}
\end{aligned}
$$

The *practicable accumulating-fud summed shuffle content alignment valency-density fud decomposition inducer* may then be implemented

$$
\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd},\mathrm{D},\mathrm{F},\infty,\mathrm{q},P,\mathrm{c}}(A) = \\
&\quad \mathrm{if}(Q\neq\emptyset,\{(D,I^{*}_{\mathrm{Scsd}}((A,D)))\},\{(D_{\emptyset},0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{c}})),\ \{(D,\cdot)\} = Q
\end{aligned}
$$

If the inclusion functions of the *tuple set list maximiser* and the *derived variables set list maximiser* in the *common-fud fud tree searcher*, $Z_{P,A,A_R,F_{\mathrm{c}},\mathrm{L}}$, are further ordered by descending *sum derived variables layer* in order to exclude redundant *reframe variables* at the inclusion boundaries, an implementation may also explicitly recursively exclude *reframe transforms* from the top *layer* of the *common fud*, G_{c}, where these do not also appear in the *decomposition fud*, G.

The *level limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}}$, of the *practicable level shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,V_{\mathrm{g}},F_{\mathrm{g}}}$, can be used to implement a *fud decomposition inducer* parameterised by a heritable tree of *levels*. Let $Z_{\mathrm{g}} \in \mathrm{trees}(\mathbf{N}_{>0} \times \mathrm{P}(\mathcal{V}) \times \mathcal{F})$ be the *level* hierarchy. A node $((\mathrm{wmax}_{\mathrm{g}}, V_{\mathrm{g}}, F_{\mathrm{g}}), X_{\mathrm{g}}) \in \mathrm{nodes}(Z_{\mathrm{g}})$ parameterises the node's *level fud tree searcher*, $Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}}$, with (i) the *maximum derived volume*, $\mathrm{wmax}_{\mathrm{g}}$, (ii) the subset of the *substrate*, V_{g}, and (iii) the union of (a) the *level fud*, F_{g}, and (b) the *level fuds* from the application of recursively parameterised *level fud tree searchers* of the children nodes, X_{g}. The *level* hierarchy has various uses including (i) the partitioning of a large *substrate*, for example into local regions implied by an external metric, so that the resultant *fud* has complete coverage, (ii) allowing multiple *overlapping* representations of a small *substrate*, (iii) hinting *derived variables* of the *substrate* that are externally known to be in *alignments*, and (iv) excluding *mono-valent substrate variables* that sometimes occur near the leaves of a *decomposition*.

Define the *level limited-models infinite-layer substrate fud decompositions tree searcher*

$$Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{g}} = \mathrm{searchTreer}(\mathcal{D}_{\mathrm{F},\infty,U,V} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}}), P_{P,A,\mathrm{D},\mathrm{F},\mathrm{g}}, R_{P,A,\mathrm{D},\mathrm{F},\mathrm{g}})$$

where the *parameters* includes the *level* hierarchy tree, $Z_{\mathrm{g}} \in \mathrm{set}(P)$, where $Z_{\mathrm{g}} \in \mathrm{trees}(\mathbf{N}_{>0} \times \mathrm{P}(V_A) \times (\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}))$. The neighbourhood function is defined

$$\begin{aligned}
P_{P,A,\mathrm{D},\mathrm{F},\mathrm{g}}(D) = \{E :\ & \\
& (\cdot, S, G, L) \in \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}\times\mathrm{S}\times\mathcal{X}^2}, \{(\mathrm{size}(B), S, G, L) : \\
& \quad (L, Y) \in \mathrm{places}(D), \\
& \quad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)), \\
& \quad (\cdot, F) = L_{|L|},\ W = \mathrm{der}(F), \\
& \quad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)), \\
& \quad B = \mathrm{apply}(V_A, V_A, \mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\}, A),\ \mathrm{size}(B) > 0, \\
& \quad G = \mathrm{level}(B, B_{R(B)})(Z_{\mathrm{g}}),\ G \neq \emptyset\})), \\
& M = L \cup \{(|L|+1, (S, G))\}, \\
& E = \mathrm{tree}(\mathrm{paths}(D) \setminus \{L\} \cup \{M\})\}
\end{aligned}$$

where

$$R_{P,A,\mathrm{D},\mathrm{F},\mathrm{g}} = \{\{((\emptyset, G), \emptyset)\} : G \in \mathrm{maxd}(\mathrm{order}(D_{\mathrm{F}}, \mathrm{level}(A, A_{R(A)})(Z_{\mathrm{g}})))\}$$

and $\mathrm{level}(A, A_R) \in \mathrm{trees}(\mathbf{N}_{>0} \times \mathrm{P}(V_A) \times (\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}})) \to (\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}})$ is defined

$$\begin{aligned}
&\mathrm{level}(A, A_R)(Z_{\mathrm{g}}) = \bigcup\{G : \\
&\quad ((\mathrm{wmax}_{\mathrm{g}}, V_{\mathrm{g}}, F_{\mathrm{g}}), X_{\mathrm{g}}) \in Z_{\mathrm{g}}, \\
&\quad F_{\mathrm{h}} = \mathrm{level}(A, A_R)(X_{\mathrm{g}}),\ \mathrm{wmax}_{\mathrm{g}} \in \mathrm{set}(P_{\mathrm{g}}), \\
&\quad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P_{\mathrm{g}},A,A_R,V_{\mathrm{g}},F_{\mathrm{g}} \cup F_{\mathrm{h}},\mathrm{L}})), \\
&\quad (K, a) \in \max(\mathrm{elements}(Z_{P_{\mathrm{g}},A,A_R,F_{\mathrm{g}} \cup F_{\mathrm{h}},F_{\mathrm{L}},\mathrm{D}})),\ a > 0, \\
&\quad G = \mathrm{depends}(F_{\mathrm{L}}, K)\}
\end{aligned}$$

The *practicable level summed shuffle content alignment valency-density fud decomposition inducer* may then be implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd},\mathrm{D},\mathrm{F},\infty,\mathrm{q},P,\mathrm{g}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A, D)))\}, \{(D_{\emptyset}, 0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{g}})),\ \{(D, \cdot)\} = Q
\end{aligned}$$

The *limited-nodes limited-models infinite-layer substrate fud decompositions tree searcher*, $Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{f}}$, is a variation of the *limited-models infinite-layer substrate fud decompositions tree searcher*, $Z_{P,A,\mathrm{D},\mathrm{F}}$, in which the cardinality of the *fuds* of the *decomposition* tree is limited to the *maximum fuds limit* $\mathrm{fmax} \in \mathbf{N}_{>0}$. The neighbourhood function $P_{P,A,\mathrm{D},\mathrm{F},\mathrm{f}}$ is modified

$$\begin{aligned}
&P_{P,A,\mathrm{D},\mathrm{F},\mathrm{f}}(D) = \{E : \\
&\quad |\mathrm{nodes}(D)| < \mathrm{fmax}, \\
&\quad (\cdot, S, G, L) \in \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q} \times \mathrm{S} \times \mathcal{X}^2}, \{(\mathrm{size}(B), S, G, L) : \\
&\qquad (L, Y) \in \mathrm{places}(D), \\
&\qquad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad (\cdot, F) = L_{|L|},\ W = \mathrm{der}(F), \\
&\qquad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)), \\
&\qquad B = \mathrm{apply}(V_A, V_A, \mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\}, A),\ \mathrm{size}(B) > 0, \\
&\qquad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,B,B_{R(B)},\mathrm{L}})), \\
&\qquad (K, a) \in \max(\mathrm{elements}(Z_{P,B,B_{R(B)},F_{\mathrm{L}},\mathrm{D}})),\ a > 0, \\
&\qquad G = \mathrm{depends}(F_{\mathrm{L}}, K)\})), \\
&\quad M = L \cup \{(|L| + 1, (S, G))\}, \\
&\quad E = \mathrm{tree}(\mathrm{paths}(D) \setminus \{L\} \cup \{M\})\}
\end{aligned}$$

The *practicable limited-nodes summed shuffle content alignment valency-density fud decomposition inducer* is implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{f}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_{\emptyset},0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D,F,f}})),\ \{D\} = Q
\end{aligned}$$

A further variation of the *limited-nodes limited-models infinite-layer substrate fud decompositions tree searcher*, $Z_{P,A,\mathrm{D,F,f}}$, is to modify the sequence of *fud* search and the termination condition in order to minimise *label entropy*. Let the query *variables* $V_{\mathrm{Q}} \subset V_A$ be a proper subset of the *substrate variables*, $V_{\mathrm{Q}} \neq V_A$. The difference forms the label *variables* $V_{\mathrm{L}} = V_A \setminus V_{\mathrm{Q}}$. Here the *modelling* is restricted to the query *variables*, V_{Q}, so that the *underlying variables* of the *decomposition* D are a subset, $\mathrm{und}(D) \subseteq V_{\mathrm{Q}}$. Given a query *histogram* $Q \in \mathcal{A}_U$ in the query *variables*, $\mathrm{vars}(Q) = V_{\mathrm{Q}}$, the *modelled transformed conditional product* is a *probability histogram* if $(Q * D^{\mathrm{T}})^{\mathrm{F}} \cap (A * D^{\mathrm{T}})^{\mathrm{F}} \neq \emptyset$,

$$(Q * D^{\mathrm{T}} * \mathrm{his}(D^{\mathrm{T}}) * A)^{\wedge}\ \%\ V_{\mathrm{L}} \in \mathcal{A} \cap \mathcal{P}$$

where his = histogram.

The *slice histogram* of the neighbourhood function is restricted to the query *variables*, $B\%V_{\mathrm{Q}}$, where the *slice histogram* is $B = \mathrm{apply}(V_A, V_A, \mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\}, A)$. The *sized label entropy* of the *slice* is defined as

$$\mathrm{size}(B) * \mathrm{entropy}(B\%V_{\mathrm{L}})$$

If the *slice* is an *effective singleton* in the label *variables*, $|(B\%V_{\mathrm{L}})^{\mathrm{F}}| = 1$, then the *sized label entropy* is zero, $\mathrm{entropy}(B\%V_{\mathrm{L}}) = 0$.

The *label-entropy limited-nodes limited-models infinite-layer substrate fud decompositions tree searcher*, $Z_{P,A,\mathrm{D,F,f,e},V_{\mathrm{L}}}$, is such that (i) the *limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,\mathrm{L}}$, and the *limited-derived derived variables set list maximiser*, $Z_{P,A,A_R,F,\mathrm{D}}$, are restricted to the query *variables*, V_{Q}, (ii) the order of *fud decomposition* is modified to maximise *slice label entropy* and then *slice size*, and (iii) the *decomposition* of a *slice* with zero *label entropy* terminates. The neighbourhood function

$P_{P,A,\mathrm{D},\mathrm{F},\mathrm{f},\mathrm{e},V_\mathrm{L}}$ is modified

$$\begin{aligned}
&P_{P,A,\mathrm{D},\mathrm{F},\mathrm{f},\mathrm{e},V_\mathrm{L}}(D) = \{E : \\
&\quad |\mathrm{nodes}(D)| < \mathrm{fmax}, \\
&\quad (\cdot,S,G,L) \in \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}^2 \times \mathrm{S} \times \mathcal{X}^2}, \{((e_B,z_B),S,G,L) : \\
&\qquad (L,Y) \in \mathrm{places}(D), \\
&\qquad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad (\cdot,F) = L_{|L|},\ W = \mathrm{der}(F), \\
&\qquad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)), \\
&\qquad B = \mathrm{apply}(V_A,V_A,\mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\},A), \\
&\qquad z_B = \mathrm{size}(B),\ e_B = z_B * \mathrm{entropy}(B\%V_\mathrm{L}),\ e_B > 0, \\
&\qquad B' = B\%(V_A \setminus V_\mathrm{L}),\ F_\mathrm{L} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,B',B'_{R(B')},\mathrm{L}})), \\
&\qquad (K,a) \in \max(\mathrm{elements}(Z_{P,B',B'_{R(B')},F_\mathrm{L},\mathrm{D}})),\ a > 0, \\
&\qquad G = \mathrm{depends}(F_\mathrm{L},K)\})), \\
&\quad M = L \cup \{(|L|+1,(S,G))\}, \\
&\quad E = \mathrm{tree}(\mathrm{paths}(D) \setminus \{L\} \cup \{M\})\}
\end{aligned}$$

The *practicable label-entropy limited-nodes summed shuffle content alignment valency-density fud decomposition inducer* is implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd},\mathrm{D},\mathrm{F},\infty,\mathrm{q},P,\mathrm{f},\mathrm{e},V_\mathrm{L}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_\emptyset,0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{f},\mathrm{e},V_\mathrm{L}})),\ \{D\} = Q
\end{aligned}$$

A similar method is to modify the sequence of *fud* search and the termination condition in order to minimise *non-modal label size.* The *non-modal label size* of the *slice* is defined as

$$\mathrm{size}(B) - \mathrm{maxr}(B\%V_\mathrm{L})$$

The neighbourhood function $P_{P,A,\mathrm{D,F,f,m},V_\mathrm{L}}$ is modified

$$\begin{aligned}
&P_{P,A,\mathrm{D,F,f,m},V_\mathrm{L}}(D) = \{E : \\
&\quad |\mathrm{nodes}(D)| < \mathrm{fmax}, \\
&\quad (\cdot,S,G,L) \in \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}^2 \times \mathrm{S} \times \mathcal{X}^2}, \{((m_B,z_B),S,G,L) : \\
&\qquad (L,Y) \in \mathrm{places}(D), \\
&\qquad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad (\cdot,F) = L_{|L|},\ W = \mathrm{der}(F), \\
&\qquad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)), \\
&\qquad B = \mathrm{apply}(V_A,V_A,\mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\},A), \\
&\qquad z_B = \mathrm{size}(B),\ m_B = z_B - \mathrm{maxr}(B\%V_\mathrm{L}),\ m_B > 0, \\
&\qquad B' = B\%(V_A \setminus V_\mathrm{L}),\ F_\mathrm{L} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,B',B'_{R(B')},\mathrm{L}})), \\
&\qquad (K,a) \in \max(\mathrm{elements}(Z_{P,B',B'_{R(B')},F_\mathrm{L},\mathrm{D}})),\ a > 0, \\
&\qquad G = \mathrm{depends}(F_\mathrm{L},K)\})), \\
&\quad M = L \cup \{(|L|+1,(S,G))\}, \\
&\quad E = \mathrm{tree}(\mathrm{paths}(D) \setminus \{L\} \cup \{M\})\}
\end{aligned}$$

The *practicable label-mode limited-nodes summed shuffle content alignment valency-density fud decomposition inducer* is implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{f,m},V_\mathrm{L}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_\emptyset,0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D,F,f,m},V_\mathrm{L}})),\ \{D\} = Q
\end{aligned}$$

The *delabelled limited-models infinite-layer substrate fud decompositions tree searcher*, $Z_{P,A,\mathrm{D,F,x},V_\mathrm{L}}$, is a variation of the *limited-models infinite-layer substrate fud decompositions tree searcher*, $Z_{P,A,\mathrm{D,F}}$, which allows the *fud* F to be constructed on the entire *substrate*, V_A, including label *variables*, $V_\mathrm{L} \subset V_A$, but then recursively removes all *variables* from the *fud* that directly or indirectly depend on the label *variables*, $F' = \mathrm{depends}(F, \{w : w \in \mathrm{der}(F),\ \mathrm{vars}(\mathrm{depends}(F,\{w\})) \cap V_\mathrm{L} = \emptyset\})$. If there are some *derived variables* of the *fud* that do not depend on the label *variables*, then the resultant *fud* is not empty, $\exists w \in \mathrm{der}(F)\ (\mathrm{vars}(\mathrm{depends}(F,\{w\})) \cap V_\mathrm{L} = \emptyset) \implies F' \neq \emptyset$. The *underlying variables* of the resultant *decomposition* do not include label *variables*, $\mathrm{und}(D) \cap V_\mathrm{L} = \emptyset$. This method allows the *fud tree searcher*, $Z_{P,A,A_R,\mathrm{L}}$, to detect *alignments* between label *variables* and non-label *variables*, but does not require queries to contain the label *variables*, for example by *expanding* with the *cartesian*, $V_\mathrm{L} \subset \mathrm{vars}(Q * V_\mathrm{L}^\mathrm{C})$.

The neighbourhood function $P_{P,A,\mathrm{D},\mathrm{F},\mathrm{x},V_{\mathrm{L}}}$ is modified

$$
\begin{aligned}
&P_{P,A,\mathrm{D},\mathrm{F},\mathrm{x},V_{\mathrm{L}}}(D) = \{E : \\
&\quad (\cdot, S, G, L) \in \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}\times\mathrm{S}\times\mathcal{X}^2}, \{(\mathrm{size}(B), S, G, L) : \\
&\qquad (L, Y) \in \mathrm{places}(D), \\
&\qquad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad (\cdot, F) = L_{|L|},\ W = \mathrm{der}(F), \\
&\qquad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)), \\
&\qquad B = \mathrm{apply}(V_A, V_A, \mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\}, A),\ \mathrm{size}(B) > 0, \\
&\qquad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,B,B_{R(B)},\mathrm{L}})), \\
&\qquad (K, a) \in \max(\mathrm{elements}(Z_{P,B,B_{R(B)},F_{\mathrm{L}},\mathrm{D}})),\ a > 0, \\
&\qquad G = \mathrm{depends}(F_{\mathrm{L}}, \{w : w \in K,\ \mathrm{vars}(\mathrm{depends}(F_{\mathrm{L}}, \{w\})) \cap V_{\mathrm{L}} = \emptyset\}), \\
&\qquad G \neq \emptyset\})), \\
&\quad M = L \cup \{(|L| + 1, (S, G))\}, \\
&\quad E = \mathrm{tree}(\mathrm{paths}(D) \setminus \{L\} \cup \{M\})\}
\end{aligned}
$$

The *practicable delabelled summed shuffle content alignment valency-density fud decomposition inducer* is implemented

$$
\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd},\mathrm{D},\mathrm{F},\infty,\mathrm{q},P,\mathrm{x},V_{\mathrm{L}}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A, D)))\}, \{(D_{\emptyset}, 0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{x},V_{\mathrm{L}}})),\ \{D\} = Q
\end{aligned}
$$

The *level limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}}$, of the *practicable level shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,V_{\mathrm{g}},F_{\mathrm{g}}}$, can also be used to implement a *supervised fud decomposition inducer* parameterised by (i) a tree of *level substrates* and (ii) a *goodness* function. Instead of creating a *fud* from a hierarchical tree of *levels*, as in the *practicable level summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd},\mathrm{D},\mathrm{F},\infty,\mathrm{q},P,\mathrm{g}}$, above, the *supervised fud decomposition inducer* uses the *goodness* function to select the maximum *goodness level substrate* and its corresponding *fud* from the maximum *goodness level substrate* path. Let $Z_{\mathrm{g}} \in \mathrm{trees}(\mathbf{N}_{>0} \times \mathrm{P}(\mathcal{V}) \times \mathcal{F})$ be the *level substrate* tree. A node $((\mathrm{wmax}_{\mathrm{g}}, V_{\mathrm{g}}, F_{\mathrm{g}}), X_{\mathrm{g}}) \in \mathrm{nodes}(Z_{\mathrm{g}})$ parameterises the node's *level fud tree searcher*, $Z_{P,A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}}$, with (i) the *maximum derived volume*, $\mathrm{wmax}_{\mathrm{g}}$, (ii) the subset of the *substrate*, V_{g}, and (iii) the *level fud*, F_{g}. Let $\mathrm{good}(U) \in \mathcal{A}_U \times \mathcal{A}_U \times \mathcal{F}_U \to \mathbf{Q}$ be some *goodness* function given in the *parameters*, good $\in \mathrm{set}(P)$.

Define the *goodness limited-models infinite-layer substrate fud decompositions tree searcher*

$$Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{p}} = \mathrm{searchTreer}(\mathcal{D}_{\mathrm{F},\infty,U,V} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}}), P_{P,A,\mathrm{D},\mathrm{F},\mathrm{p}}, R_{P,A,\mathrm{D},\mathrm{F},\mathrm{p}})$$

where the *parameters* includes the *level substrate* tree, $Z_{\mathrm{g}} \in \mathrm{set}(P)$, where $Z_{\mathrm{g}} \in \mathrm{trees}(\mathbf{N}_{>0} \times \mathrm{P}(V_A) \times (\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}))$. The neighbourhood function is defined

$$\begin{aligned}
&P_{P,A,\mathrm{D},\mathrm{F},\mathrm{p}}(D) = \{E : \\
&\qquad (\cdot, S, G, L) \in \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}\times\mathrm{S}\times\mathcal{X}^2}, \{(\mathrm{size}(B), S, G, L) : \\
&\qquad\qquad (L, Y) \in \mathrm{places}(D), \\
&\qquad\qquad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad\qquad (\cdot, F) = L_{|L|},\ W = \mathrm{der}(F), \\
&\qquad\qquad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)), \\
&\qquad\qquad B = \mathrm{apply}(V_A, V_A, \mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\}, A),\ \mathrm{size}(B) > 0, \\
&\qquad\qquad (\cdot, G) = \mathrm{best}(B, B_{R(B)})(Z_{\mathrm{g}}),\ G \neq \emptyset\})), \\
&\qquad M = L \cup \{(|L|+1, (S, G))\}, \\
&\qquad E = \mathrm{tree}(\mathrm{paths}(D) \setminus \{L\} \cup \{M\})\}
\end{aligned}$$

where

$$R_{P,A,\mathrm{D},\mathrm{F},\mathrm{p}} = \{\{((\emptyset, G), \emptyset)\} : (\cdot, G) = \mathrm{best}(A, A_{R(A)})(Z_{\mathrm{g}}),\ G \neq \emptyset\}$$

and $\mathrm{best}(A, A_R) \in \mathrm{trees}(\mathbf{N}_{>0} \times \mathrm{P}(V_A) \times (\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}})) \to (\mathbf{Q} \times (\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}))$ is defined

$$\begin{aligned}
&\mathrm{best}(A, A_R)(Z_{\mathrm{g}}) = \mathrm{if}(Q \neq \emptyset, \mathrm{if}((G' \neq \emptyset) \wedge (g' > g), (g', G'), (g, G)), (0, \emptyset)) \\
&\qquad Q = \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}\times\mathcal{F}\times\mathcal{X}}, \\
&\qquad\qquad \{(\mathrm{good}(U_A)(A, A_R, G), G, X_{\mathrm{g}}) : ((\mathrm{wmax}_{\mathrm{g}}, V_{\mathrm{g}}, F_{\mathrm{g}}), X_{\mathrm{g}}) \in Z_{\mathrm{g}}, \\
&\qquad\qquad\qquad \mathrm{wmax}_{\mathrm{g}} \in \mathrm{set}(P_{\mathrm{g}}), \\
&\qquad\qquad\qquad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P_{\mathrm{g}},A,A_R,V_{\mathrm{g}},F_{\mathrm{g}},\mathrm{L}})), \\
&\qquad\qquad\qquad (K, a) \in \max(\mathrm{elements}(Z_{P_{\mathrm{g}},A,A_R,F_{\mathrm{g}},F_{\mathrm{L}},\mathrm{D}})),\ a > 0, \\
&\qquad\qquad\qquad G = \mathrm{depends}(F_{\mathrm{L}}, K)\})) \\
&\qquad \{(g, G, X_{\mathrm{g}})\} = Q, \\
&\qquad (g', G') = \mathrm{best}(A, A_R)(X_{\mathrm{g}})
\end{aligned}$$

The *practicable goodness summed shuffle content alignment valency-density fud decomposition inducer* may then be implemented

$$I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{p}}(A) =$$
$$\mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_{\emptyset},0)\}) :$$
$$Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D,F,p}})),\ \{(D,\cdot)\} = Q$$

An example of the *goodness* function is simply *shuffle content alignment valency-density computer* application,

$$\mathrm{good}(U)(A, A_R, F) = I^{*}_{\mathrm{csd}}((A, A_R, F))$$

Another example is a *goodness* function that minimises *label entropy.* Let the query *variables* $V_{\mathrm{Q}} \subset V_A$ be a proper subset of the *substrate variables*, $V_{\mathrm{Q}} \neq V_A$. The difference forms the label *variables* $V_{\mathrm{L}} = V_A \setminus V_{\mathrm{Q}}$. Here the *level substrates* are restricted to the query *variables*, $\forall(\cdot, V_{\mathrm{g}}, \cdot) \in \mathrm{elements}(Z_{\mathrm{g}})\ (V_{\mathrm{g}} \subseteq V_{\mathrm{Q}})$, and $\forall(\cdot,\cdot,F_{\mathrm{g}}) \in \mathrm{elements}(Z_{\mathrm{g}})\ (\mathrm{und}(F_{\mathrm{g}}) \subseteq V_{\mathrm{Q}})$. The *label entropy* is the sum of the *component label entropies*,

$$\sum_{(\cdot,C)\in(F^{\mathrm{T}})^{-1}} \mathrm{size}(A * C) \times \mathrm{entropy}(A * C\ \%\ V_{\mathrm{L}})$$

and the *goodness* function is the negative *label entropy* computed by *histogram application*

$$\mathrm{good}(U)(A,\cdot,F) = I^{*}_{\approx\mathbf{R}}(-\sum \mathrm{size}(B) \times \mathrm{entropy}(B\%V_{\mathrm{L}}) :$$
$$R \in (A * F^{\mathrm{T}})^{\mathrm{FS}},\ B = \mathrm{apply}(V_A, V_A, \mathrm{his}(F) \cup \{\{R\}^{\mathrm{U}}\}, A))$$

A similar example is a *goodness* function that minimises *non-modal label size.* The *non-modal label size* is the sum of the *component non-modal label sizes*,

$$\sum_{(\cdot,C)\in(F^{\mathrm{T}})^{-1}} \mathrm{size}(A * C) - \mathrm{maxr}(A * C\ \%\ V_{\mathrm{L}})$$

and the *goodness* function is the negative *non-modal label size* computed by *histogram application*

$$\mathrm{good}(U)(A,\cdot,F) = -\sum \mathrm{size}(B) - \mathrm{maxr}(B\%V_{\mathrm{L}}) :$$
$$R \in (A * F^{\mathrm{T}})^{\mathrm{FS}},\ B = \mathrm{apply}(V_A, V_A, \mathrm{his}(F) \cup \{\{R\}^{\mathrm{U}}\}, A)$$

4.22.4 Implementation

The implementation of *practicable inducers* must exclude *partition variables* because they are impracticable. In the following *computers* none of the instantiated *variables* are *partition variables.*

In order to implement the *limited-underlying tuple set list maximiser*, $Z_{P,A,A_R,F,\mathrm{B}}$, define the *limited-underlying tuple set list builder* $I_{P,U,\mathrm{B}} \in \mathrm{computers}$ such that (i) the domain is $\mathrm{domain}(I_{P,U,\mathrm{B}}) = \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,1} \times \mathcal{A}_U \times \mathcal{A}_U$, (ii) the range is $\mathrm{range}(I_{P,U,\mathrm{B}}) = \mathrm{P}(\mathcal{V}_U) \times \mathcal{A}_U \times \mathcal{A}_U \to \mathbf{Q}$, and (iii) the application is

$$\begin{aligned}
I^*_{P,U,\mathrm{B}}((V,\emptyset,X,X_R)) &= \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{buildb}(V,X,X_R,\mathrm{init}(V),\emptyset)) \\
I^*_{P,U,\mathrm{B}}((V,F,X,X_R)) &= \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{buildb}(\mathrm{vars}(F) \cup V, X, X_R, \\
&\qquad\qquad \mathrm{init}(\mathrm{der}(F)),\emptyset))
\end{aligned}$$

where $\mathrm{init}(V) := \{((((\{w\},\emptyset,\emptyset),0),0) : w \in V\}$, $\mathrm{buildt} = ((\mathrm{P}(\mathcal{V}_U) \times \mathcal{A}_U \times \mathcal{A}_U) \times \mathbf{Q}) \times (\mathbf{Q} \times \mathbf{N} \times \mathbf{Q} \times \mathbf{N} \times \mathbf{N})$ and $\mathrm{buildb} \in \mathrm{P}(\mathcal{V}_U) \times \mathcal{A}_U \times \mathcal{A}_U \times \mathrm{buildt} \times \mathrm{buildt} \to \mathrm{buildt}$ is defined

$$\begin{aligned}
&\mathrm{buildb}(W,X,X_R,Q,N) = \\
&\quad \mathrm{if}(M \neq \emptyset, \mathrm{buildb}(W,X,X_R,M,N \cup M), \mathrm{final}(N)) : \\
&\qquad P = \{J : (((K,\cdot,\cdot),\cdot),\cdot) \in Q,\ w \in W \setminus K,\ J = K \cup \{w\}\}, \\
&\qquad M = \mathrm{top}(\mathrm{omax})(\{((((J,B,B_R),a_1 - b_1), \\
&\qquad\qquad (a_1 - a_2 - b_1 + b_2,\ -l,\ -b_1 + b_2,\ -u,\ D_{\mathcal{X}}(J))) : \\
&\qquad\quad J \in P,\ u = |J^{\mathrm{C}}|,\ u \leq \mathrm{xmax},\ l = \mathrm{sumlayer}(F,J), \\
&\qquad\quad B = I^*_{\%}((J,X)),\ B_R = I^*_{\%}((J,X_R)), \\
&\qquad\quad a_1 = I^*_{\mathrm{S}\approx\ln!}(B),\ a_2 = I^*_{\mathrm{S}\approx\ln!}(I^*_{\mathrm{X}}(B)), \\
&\qquad\quad b_1 = I^*_{\mathrm{S}\approx\ln!}(B_R),\ b_2 = I^*_{\mathrm{S}\approx\ln!}(I^*_{\mathrm{X}}(B_R))\})
\end{aligned}$$

where $\mathrm{final}(N) := \{((((K,A,B),y),a) : (((K,A,B),y),a) \in N,\ |K| > 1\}$, $D_{\mathcal{X}} \in \mathrm{enums}(\mathcal{X})$ is an arbitrary order, $\mathrm{sumlayer} \in \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathbf{N}$, the *reducer* $I_{\%} = \mathrm{reducer} \in \mathrm{computers}$, the *independenter* $I_{\mathrm{X}} = \mathrm{independenter} \in \mathrm{computers}$ is such that $I^*_{\mathrm{X}}(A) = A^{\mathrm{X}}$, and $I_{\mathrm{S}\approx\ln!} = \mathrm{sumlogfactorialer} \in \mathrm{computers}$ is defined

$$I^*_{\mathrm{S}\approx\ln!}(A) = \sum_{S \in A^{\mathrm{FS}}} I^*_{\approx\ln!}(A_S)$$

where $I_{\approx\ln!} = \mathrm{logfactorialer} \in \mathrm{computers}$ is defined $I^*_{\approx\ln!}(x) \approx \ln \Gamma_! x$.

The *tuples* of the *limited-underlying tuple set list builder* $I_{P,U,\mathrm{B}}$, are *pluri-variate*, $\forall((K,\cdot,\cdot),\cdot) \in I^*_{P,U,\mathrm{B}}((V,F,X,X_R))\ (|K| > 1)$.

The buildb function argument *histograms*, X, X_R, have the same *variables*, $\mathrm{vars}(X) = \mathrm{vars}(X_R)$. The argument *variables*, W, are a subset of the argument *histograms variables*, $W \subseteq \mathrm{vars}(X)$. It is sufficient that the *system*, U, contains the *variables* of the argument *histograms*, $\mathrm{vars}(X) \subseteq \mathrm{vars}(U)$.

The argument *histograms*, X, X_R, are in a *list representation* or *binary map representation*, because in some cases the *volume*, $|\mathrm{vars}(X)^{\mathrm{C}}|$, is impracticably large for an *array representation*. The resultant *histograms*, B, B_R, may be in *array representation* because their *volume* cannot exceed the given limit, $|B^{\mathrm{C}}| \leq \mathrm{xmax}$, which is chosen to be practicable.

The *limited-underlying tuple set list builder* implements the *limited-underlying tuple set list maximiser*, insofar as the inclusion boundaries are the same, $\mathrm{dom}(I^*_{P,U,\mathrm{B}}((V, F, A, A_R))) \subseteq \mathrm{top}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{B}}))$.

The *tuple set list builder* never returns more than $\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor$ *tuples*, because of the trailing arbitrary ordering of the *tuples*, $D_{\mathcal{X}}(J)$.

Although *inducers* are defined only for *substrate histograms*, $A \in \mathcal{A}_z$, which are constrained such that the *independent sample histogram* is *completely effective*, $A^{\mathrm{XF}} = A^{\mathrm{C}}$, the implementation of *induction computers* here only requires that the argument *histogram*, X, be in the *system* U, $X \in \mathcal{A}_U$.

The *limited-underlying tuple set list builder* returns the *non-independent content sum factorial*, $a_1 - b_1 = I^*_{\mathrm{S}\approx\ln!}(B) - I^*_{\mathrm{S}\approx\ln!}(B_R)$, to avoid unnecessary recomputation subsequently in the *partitioner*.

The computation of $\mathrm{sumlayer}(F, J)$ is costly so some implementations may exclude it, especially as it only affects *tuples* in the inclusion boundary. If the *layerer* (see later) is subject to the *excluded-self* restriction then it is less likely that there will be duplicate *tuple alignments*, so the inclusion boundary is more likely to be a singleton.

Another performance improvement is to restrict the *builder* to *variables* that are *multi-effective*, $\{u : u \in V,\ |(X\%\{u\})^{\mathrm{F}}| > 1\}$. This prevents some *variables* from being included in the *tuple* that are necessarily *independent* of the other *variables* in the *tuple*.

To implement the *highest-layer limited-derived derived variables set list maximiser*, $Z_{P,A,A_R,F,\mathrm{D,d}}$, define the *highest-layer limited-derived derived variables set builder* $I_{P,U,\mathrm{D,d}} \in \mathrm{computers}$ such that (i) the domain is $\mathrm{domain}(I_{P,U,\mathrm{D,d}}) = \mathrm{P}(\mathcal{V}_U) \times \mathcal{F}_{U,1} \times \mathcal{A}_U \times \mathcal{A}_U$, (ii) the range is $\mathrm{range}(I_{P,U,\mathrm{D,d}}) = (\mathrm{P}(\mathcal{V}_U) \times \mathcal{A}_U \times$

$\mathcal{A}_U) \to \mathbf{Q}$, and (iii) the application is

$$I^*_{P,U,\mathrm{D,d}}((V,F,X,X_R)) = \mathrm{maxd}(\mathrm{buildd}(\mathrm{vars}(F) \setminus V, X, X_R, \mathrm{init}(\mathrm{der}(F)), \emptyset))$$

where buildd $\in \mathrm{P}(\mathcal{V}_U) \times \mathcal{A}_U \times \mathcal{A}_U \times \mathrm{buildt} \times \mathrm{buildt} \to \mathrm{buildt}$ is defined

$$\begin{aligned}
&\mathrm{buildd}(W,X,X_R,Q,N) = \\
&\quad \mathrm{if}(M \neq \emptyset, \mathrm{buildb}(W,X,X_R,M,N \cup M), \mathrm{final}(N)) : \\
&\qquad P = \{J : (((K,\cdot,\cdot),\cdot),\cdot) \in Q,\ w \in W \setminus K,\ J = K \cup \{w\}\}, \\
&\qquad M = \mathrm{top}(\mathrm{omax})(\{(((J,B,B_R),(a-b)/c), \\
&\qquad\qquad\qquad ((a-b)/c,\ \ -l,\ \ -b/c,\ \ -u,\ D_{\mathcal{X}}(J))) : \\
&\qquad\quad J \in P,\ u = |J^{\mathrm{C}}|,\ u \leq \mathrm{wmax},\ \mathrm{der}(\mathrm{depends}(F,J)) = J, \\
&\qquad\quad m = |J|,\ l = \mathrm{sumlayer}(F,J), \\
&\qquad\quad B = I^*_{\%}((J,X)),\ B_R = I^*_{\%}((J,X_R)), \\
&\qquad\quad a = I^*_{\mathrm{a}}(B),\ b = I^*_{\mathrm{a}}(B_R),\ c = I^*_{\approx\mathrm{pow}}((u,1/m))\})
\end{aligned}$$

where the *power approxer* $I_{\approx\mathrm{pow}} \in$ computers is such that $I^*_{\approx\mathrm{pow}}((x,y)) \approx x^y$, and the *alignmenter* $I_{\mathrm{a}} =$ alignmenter $\in$ computers is such that $I^*_{\mathrm{a}}(A) \approx \mathrm{algn}(A)$.

The *tuples* of the *highest-layer limited-derived derived variables set builder* $I_{P,U,\mathrm{D,d}}$, are *pluri-variate*, $\forall((K,\cdot,\cdot),\cdot) \in I^*_{P,U,\mathrm{D,d}}((V,F,X,X_R))\ (|K| > 1)$.

The buildd function argument *histograms*, X, X_R, have the same *variables*, $\mathrm{vars}(X) = \mathrm{vars}(X_R)$. The argument *variables*, W, are a subset of the argument *histograms variables*, $W \subseteq \mathrm{vars}(X)$. It is sufficient that the *system*, U, contains the *variables* of the argument *histograms*, $\mathrm{vars}(X) \subseteq \mathrm{vars}(U)$.

The argument *histograms*, X, X_R, are in a *list representation* or *binary map representation*, because in some cases the *volume*, $|\mathrm{vars}(X)^{\mathrm{C}}|$, is impracticably large for an *array representation*. The resultant *histograms*, B, B_R, may be in *array representation* because their *volume* cannot exceed the given limit, $|B^{\mathrm{C}}| \leq \mathrm{wmax}$, which is chosen to be practicable.

If the *fud* is a non-empty *substrate fud*, $F \in \mathcal{F}_{U_A,V_A} \setminus \{\emptyset\}$, the *highest-layer limited-derived derived variables set builder* implements the *highest-layer limited-derived derived variables set list maximiser*, insofar as the inclusion boundaries are the same,

$$\mathrm{dom}(I^*_{P,U,\mathrm{D,b}}((V,F,A,A_R))) \subseteq \mathrm{top}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{D,b}}))$$

Similarly to the *tuple builder* above, some implementations may drop the computation of $\mathrm{sumlayer}(F, J)$, especially if the *layerer* is subject to the *excluded-self* restriction.

Also, some implementations may drop the exclusion of hidden *variables* $J = \mathrm{der}(\mathrm{depends}(F, J))$. This computation is costly, but dropping may in some cases lead to *tuple rolls* that result in a single *derived variable*. However, this is also true of the *excluded-self* restriction which is applied in the *layerer* (see later).

In order to implement the *limited-valency contracted decrementing linear non-overlapping fuds list maximiser initial set*, $R_{P,A,A_R,F,\mathrm{n,w},-,K}$, the *tuple partitioner* $I_{P,U,\mathrm{K}} \in$ computers is defined such that (i) $\mathrm{domain}(I_{P,U,\mathrm{K}}) = (\mathrm{P}(\mathcal{V}_U) \times \mathcal{A}_U \times \mathcal{A}_U) \times \mathbf{Q}$, (ii) $\mathrm{range}(I_{P,U,\mathrm{K}}) = \mathrm{P}(\mathcal{L}(\mathcal{S}_U \to \mathbf{N}) \times \mathcal{A}_U \times \mathcal{A}_U)$, and (iii) the application is

$$\begin{aligned}
&I^*_{P,U,\mathrm{K}}(((K, B, B_R), y_1)) = \\
&\quad \mathrm{topd}(\mathrm{pmax})(\{((N, C, C_R),\ ((y_1 - a_2 + b_2)/c,\ b_2/c,\ -m,\ D_{\mathcal{X}}(J))) : \\
&\qquad m \in \{2 \ldots \mathrm{mmax}\},\ Y \in \mathrm{S}(K, m),\ (\forall J \in Y\ (|J^{\mathrm{C}}| \leq \mathrm{umax})), \\
&\qquad N = \{(i, \mathrm{order}(D_{\mathrm{S}}, J^{\mathrm{CS}})) : (J, i) \in \mathrm{order}(D_{\mathrm{P}(\mathcal{V})}, Y)\}, \\
&\qquad T = (\{\bigcup\{S \cup \{(w, u)\} : (w, (S, u)) \in L\} : L \in \prod N\}^{\mathrm{U}}, \{1 \ldots m\}), \\
&\qquad C = I^*_{*\mathrm{T}}((T, B)),\ C_R = I^*_{*\mathrm{T}}((T, B_R)), \\
&\qquad a_2 = I^*_{\mathrm{S}\approx\ln!}(I^*_{\mathrm{X}}(C)),\ b_2 = I^*_{\mathrm{S}\approx\ln!}(I^*_{\mathrm{X}}(C_R)),\ c = I^*_{\approx\mathrm{pow}}((v, 1/m))\})
\end{aligned}$$

where $v = |K^{\mathrm{C}}|$, $\mathrm{vars}(U) \cap \mathbf{N} = \emptyset$, $D_{\mathrm{S}} \in \mathrm{enums}(\mathcal{S}_U)$, $D_{\mathrm{P}(\mathcal{V})} \in \mathrm{enums}(\mathrm{P}(\mathcal{V}_U))$, and the *transformer* $I_{*\mathrm{T}} =$ transformer $\in$ computers is such that $I^*_{*\mathrm{T}}((T, A)) = A * T$.

The *tuple partitioner* has non-empty application if $|K| \geq 2$. The resultant *histograms*, C, C_R where $(\cdot, C, C_R) \in I^*_{P,U,\mathrm{K}}((K, B, B_R, y_1))$, should be in *array representation*, suitable for succeeding *value roll computers*. The *tuple partitioner* assumes that the array index *variables* are not *system variables*, $\mathrm{vars}(U) \cap \mathbf{N} = \emptyset$.

Because (i) the *alignmenter* equals the difference in the *non-independent sum log factorialer* and the *independent sum log factorialer*, $I^*_{\mathrm{a}}(A) = I^*_{\mathrm{S}\approx\ln!}(A) - I^*_{\mathrm{S}\approx\ln!}(A^{\mathrm{X}})$, and (ii) the *non-independent* terms are constant, $\sum_{S \in C^{\mathrm{S}}} \ln \Gamma_! C_S = \sum_{S \in B^{\mathrm{S}}} \ln \Gamma_! B_S$, so only the *independent* terms, $\sum_{S \in C^{\mathrm{XS}}} \ln \Gamma_! C^{\mathrm{X}}_S$, need be computed for each of the possible partitions. Thus the *non-independent* part of the computation of the difference in *alignments*, $I_{\mathrm{a}}^*(C) - I_{\mathrm{a}}^*(C_R)$, need not be re-computed, but can be carried from the *tuple builder*.

After constructing the initial set in a *tuple partitioner*, $I_{P,U,\mathrm{K}}$, the remainder of the *limited-valency contracted decrementing linear non-overlapping fuds list maximiser*, $Z_{P,A,A_R,F,\mathrm{n},\mathrm{w},-,K}$, is implemented by means of *value roll computers*, defined in section 'Value roll computers', above. The *tuple-partition value roller* $I_{P,U,\mathrm{R}} \in$ computers is defined such that (i) the domain is $\mathrm{domain}(I_{P,U,\mathrm{R}}) = \mathrm{P}(\mathcal{L}(\mathcal{S}_U \to \mathbf{N}) \times \mathcal{A}_U \times \mathcal{A}_U)$, (ii) the range is $\mathrm{range}(I_{P,U,\mathrm{R}}) = \mathrm{P}(\mathcal{L}(\mathcal{S}_U \to \mathbf{N}))$, and (iii) the application is

$$\begin{aligned}
&I_{P,U,\mathrm{R}}^*(Q) = \\
&\quad \{N' : \\
&\qquad M = \{((N, R_A, R_B), (a-b)/c) : \\
&\qquad\qquad (N, A, B) \in Q, \\
&\qquad\qquad a = I_{\mathrm{a}}^*(A),\ b = I_{\mathrm{a}}^*(B), \\
&\qquad\qquad w = \prod_{(\cdot, I) \in N} |\mathrm{ran}(I)|,\ m = |N|,\ c = I_{\approx\mathrm{pow}}^*((w, 1/m)), \\
&\qquad\qquad R_A = (a, A, I_{\mathrm{X}}^*(A)),\ R_B = (b, B, I_{\mathrm{X}}^*(B))\}, \\
&\qquad (N', \cdot, \cdot) \in \mathrm{topd}(\mathrm{pmax})(\mathrm{rollb}(M, M))\}
\end{aligned}$$

where $\mathrm{rollb} \in \mathrm{rollbt} \times \mathrm{rollbt} \to \mathrm{rollbt}$, where $\mathrm{rollbt} = \mathcal{L}(\mathcal{S}_U \to \mathbf{N}) \times (\mathbf{Q} \times \mathcal{A}_U \times \mathcal{A}_U)^2 \to \mathbf{Q}$, is defined

$$\begin{aligned}
&\mathrm{rollb}(Q, P) = \\
&\quad \mathrm{if}(M \neq \emptyset, \mathrm{rollb}(M, P \cup M), P) : \\
&\qquad M = \mathrm{top}(\mathrm{pmax})(\{((N', R'_A, R'_B), (a'-b')/c') : \\
&\qquad\quad ((N, R_A, R_B), \cdot) \in Q, \\
&\qquad\quad V = \mathrm{dom}(N),\ (\cdot, A, A_{\mathrm{X}}) = R_A,\ (\cdot, B, B_{\mathrm{X}}) = R_B, \\
&\qquad\quad Y_A = \mathrm{rals}(N, A, A_{\mathrm{X}}),\ Y_B = \mathrm{rals}(N, B, B_{\mathrm{X}}), \\
&\qquad\quad (v, I) \in N,\ |\mathrm{ran}(I)| > 2,\ s, t \in \mathrm{ran}(I),\ s > t, \\
&\qquad\quad N' = N \setminus \{(v, I)\} \cup \{(v, \{(s, t)\} \circ I)\}, \\
&\qquad\quad R'_A = I_{\mathrm{R,a}}^*(((V, v, s, t), Y_A, R_A)), \\
&\qquad\quad R'_B = I_{\mathrm{R,a}}^*(((V, v, s, t), Y_B, R_B)), \\
&\qquad\quad (a', \cdot, \cdot) = R'_A,\ (b', \cdot, \cdot) = R'_B, \\
&\qquad\quad w = \prod_{(\cdot, I') \in N'} |\mathrm{ran}(I')|,\ m = |V|,\ c' = I_{\approx\mathrm{pow}}^*((w, 1/m))\})
\end{aligned}$$

where $I_{\mathrm{R,a}} = \mathrm{rollValueAlignmenter} \in \mathrm{computers}$ and $\mathrm{rals} \in \mathcal{L}(\mathcal{S}_U \to \mathbf{N}) \times \mathcal{A} \times \mathcal{A} \to (\mathcal{V} \to (\mathcal{S} \to \mathbf{Q}))$ is defined as

$$\begin{aligned}
&\mathrm{rals}(N, A, A_{\mathrm{X}}) := \\
&\quad \{(w, \{(S, \sum(I^{*}_{\approx\ln!}(A(T)) : T \in A^{\mathrm{S}},\ T \supseteq S) - \\
&\qquad\qquad \sum(I^{*}_{\approx\ln!}(A_{\mathrm{X}}(T)) : T \in A_{\mathrm{X}}^{\mathrm{S}},\ T \supseteq S)) : \\
&\qquad\qquad\quad u \in \mathrm{ran}(N_w),\ S \in \{(w,u)\}\}) : w \in \mathrm{dom}(N)\}
\end{aligned}$$

The *roll value alignmenter*, $I_{\mathrm{R,a}}$, requires that all *histograms* are implemented in *array histogram representations* on *ordered list state representations.*

The *value roll* compositions do not necessarily lead to a contiguous set, so in some cases $\mathrm{ran}(I) \neq \{1 \ldots |\mathrm{ran}(I)|\}$. In some implementations, however, a source *value* may be completely removed from the *representation*, rather than simply zeroed out. In these cases the *value roll* compositions must also *value roll* by one all *values* higher than the source *values.* That is, instead of $\{(s,t)\} \circ I$ the composition is $\{(r, r-1) : r \in \mathrm{ran}(I),\ r > s\} \circ \{(s,t)\} \circ I$.

The operation to take the top(pmax) at each step requires that the *value roll list* composition, I, be computed for each *value roll* because different *value roll lists* can have the same composition. However, the computation is costly, so some implementations may simply take the top *value roll lists* rather than the top *value roll list* compositions. The functionality is only equivalent with respect to *value roll list* compositions when $\mathrm{pmax} = 1$. An alternative is to implement the *limited-valency maximum-roll contracted decrementing linear non-overlapping fuds tree maximiser*, $Z_{P,A,A_R,F,\mathrm{n,w},-,K,\mathrm{mr}}$, which only applies the pmax parameter to the initial set. The *tuple partitioner*, $I_{P,U,\mathrm{K}}$, is unchanged and the *tuple-partition value roller*, $I_{P,U,\mathrm{R}}$, is modified to use the max inclusion function instead of top(pmax). The *maximum-roll tuple-partition value roller* $I_{P,U,\mathrm{R,mr}} \in \mathrm{computers}$ is defined such that (i) the domain is $\mathrm{domain}(I_{P,U,\mathrm{R,mr}}) = \mathcal{L}(\mathcal{S}_U \to \mathbf{N}) \times \mathcal{A}_U \times \mathcal{A}_U$, (ii) the range is range$(I_{P,U,\mathrm{R,mr}}) = \mathrm{P}(\mathcal{L}(\mathcal{S}_U \to \mathbf{N}))$, and (iii) the application is

$$\begin{aligned}
&I^*_{P,U,\mathrm{R,mr}}((N,A,B)) = \\
&\quad \{N' : \\
&\qquad M = \{((N,R_A,R_B),(a-b)/c) : \\
&\qquad\qquad a = I^*_{\mathrm{a}}(A),\ b = I^*_{\mathrm{a}}(B), \\
&\qquad\qquad w = \prod_{(\cdot,I) \in N} |\mathrm{ran}(I)|,\ m = |N|,\ c = I^*_{\approx\mathrm{pow}}((w,1/m)), \\
&\qquad\qquad R_A = (a,A,I^*_{\mathrm{X}}(A)),\ R_B = (b,B,I^*_{\mathrm{X}}(B))\}, \\
&\qquad (N',\cdot,\cdot) \in \mathrm{maxd}(\mathrm{rollb}(M,M))\}
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{rollb}(Q,P) = \\
&\quad \mathrm{if}(M \neq \emptyset, \mathrm{rollb}(M, P \cup M), P) : \\
&\qquad M = \max(\{((N',R'_A,R'_B),(a'-b')/c') : \\
&\qquad\quad ((N,R_A,R_B),\cdot) \in Q, \\
&\qquad\quad V = \mathrm{dom}(N),\ (\cdot,A,A_{\mathrm{X}}) = R_A,\ (\cdot,B,B_{\mathrm{X}}) = R_B, \\
&\qquad\quad Y_A = \mathrm{rals}(N,A,A_{\mathrm{X}}),\ Y_B = \mathrm{rals}(N,B,B_{\mathrm{X}}), \\
&\qquad\quad (v,I) \in N,\ |\mathrm{ran}(I)| > 2,\ s,t \in \mathrm{ran}(I),\ s > t, \\
&\qquad\quad N' = N \setminus \{(v,I)\} \cup \{(v,\{(s,t)\} \circ I)\}, \\
&\qquad\quad R'_A = I^*_{\mathrm{R,a}}(((V,v,s,t),Y_A,R_A)), \\
&\qquad\quad R'_B = I^*_{\mathrm{R,a}}(((V,v,s,t),Y_B,R_B)), \\
&\qquad\quad (a',\cdot,\cdot) = R'_A,\ (b',\cdot,\cdot) = R'_B, \\
&\qquad\quad w = \prod_{(\cdot,I') \in N'} |\mathrm{ran}(I')|,\ m = |V|,\ c' = I^*_{\approx\mathrm{pow}}((w,1/m))\})
\end{aligned}$$

Next, the functionality of (i) the *highest-layer limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,\mathrm{L,d}}$, and (ii) the *highest-layer limited-derived derived variables set list maximiser*, $Z_{P,A,A_R,F,\mathrm{D,d}}$, is implemented in the *highest-layer layerer* $I_{P,U,\mathrm{L,d}} \in$ computers, which is defined such that (i) the domain is domain$(I_{P,U,\mathrm{L,d}}) = \mathrm{P}(\mathcal{V}_U) \times \mathcal{A}_U \times \mathcal{A}_U \times \mathbf{N}$, (ii) the range is range$(I_{P,U,\mathrm{L,d}}) = \mathcal{U} \times \mathcal{F} \times (\mathrm{P}(\mathcal{V}) \to \mathbf{Q})$, and (iii) the application is

$$I^*_{P,U,\mathrm{L,d}}((V,A,A_R,f)) = \mathrm{layer}(V,U,\emptyset,\emptyset,A,A_R,f,1)$$

where layer $\in \mathrm{P}(\mathcal{V}) \times \mathcal{U} \times \mathcal{F} \times ((\mathrm{P}(\mathcal{V}) \times \mathcal{A} \times \mathcal{A}) \to \mathbf{Q}) \times \mathcal{A} \times \mathcal{A} \times \mathbf{N} \times \mathbf{N} \to (\mathcal{U} \times \mathcal{F} \times (\mathrm{P}(\mathcal{V}) \to \mathbf{Q}))$ is defined

$$\begin{aligned}
&\mathrm{layer}(V, U, F, M, X, X_R, f, l) = \\
&\quad \mathrm{if}((l \leq \mathrm{lmax}) \wedge (H \neq \emptyset) \wedge (M \neq \emptyset \implies \mathrm{maxr}(M') > \mathrm{maxr}(M)), \\
&\qquad\qquad \mathrm{layer}(V, U', F \cup H, M', X', X'_R, f, l+1), \\
&\qquad\qquad (U, F, M)) : \\
&\qquad L = \{(b, (T, (w, \mathrm{ran}(I)))) : \\
&\qquad\qquad ((\cdot, I), b) \in \mathrm{order}(D_{\mathrm{L}}, \{(v, I) : \\
&\qquad\qquad\qquad Q \in I^*_{P,U,\mathrm{B}}((V, F, X, X_R)), \\
&\qquad\qquad\qquad N \in I^*_{P,U,\mathrm{R}}(I^*_{P,U,\mathrm{K}}(Q)), \\
&\qquad\qquad\qquad (v, I) \in N\}), \\
&\qquad\qquad w = (f, l, b),\ T = (\{S \cup \{(w, k)\} : (S, k) \in I\}, \{w\})\}, \\
&\qquad H = \mathrm{dom}(\mathrm{set}(L)),\ U' = U \cup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad X' = I^*_{*\mathrm{X}}((H, X)),\ X'_R = I^*_{*\mathrm{X}}((H, X_R)), \\
&\qquad M' = I^*_{P,U',\mathrm{D,d}}((V, F \cup H, X', X'_R))
\end{aligned}$$

where $I_{*\mathrm{X}} = \mathrm{applier} \in \mathrm{computers}$.

Here the *order* D_{L} is some enumeration of the *layer fud* representation, $D_{\mathrm{L}} \in \mathrm{enums}(\mathbf{N} \times (\mathcal{S}_U \to \mathbf{N}))$.

The new *variable*, $w = (f, l, b)$, is constructed from the *fud* identifier, f, the *layer* identifier, l, and the position within the breadth of the *layer fud*, b. The new *variable*, w, is added to a new *system*, U'. The *values* of the new *variable* are cardinal numbers, $U'(w) \subset \mathbf{N}_{>0}$, such that $1 \in U'(w)$. The *values* are not necessarily contiguous, unless the implementation completely removes source *values*, in which case $U'(w) = \{1 \ldots |U'(w)|\}$.

A variation of the *highest-layer layerer*, $I_{P,U,\mathrm{L,d}}$, is to implement the *limited-valency maximum-roll contracted decrementing linear non-overlapping fuds tree maximiser*, $Z_{P,A,A_R,F,\mathrm{n,w},-,K,\mathrm{mr}}$, by means of the *maximum-roll tuple-partition value roller*, $I_{P,U,\mathrm{R,mr}}$. The *highest-layer maximum-roll layerer* $I_{P,U,\mathrm{L,mr,d}} \in$ computers is defined such that the application is

$$I^*_{P,U,\mathrm{L,mr,d}}((V, A, A_R, f)) = \mathrm{layer}(V, U, \emptyset, \emptyset, A, A_R, f, 1)$$

and

$$
\begin{aligned}
&\mathrm{layer}(V,U,F,M,X,X_R,f,l) = \\
&\quad \mathrm{if}((l \leq \mathrm{lmax}) \wedge (H \neq \emptyset) \wedge (M \neq \emptyset \implies \mathrm{maxr}(M') > \mathrm{maxr}(M)), \\
&\qquad\qquad \mathrm{layer}(V,U',F \cup H,M',X',X'_R,f,l+1), \\
&\qquad\qquad (U,F,M)) : \\
&\qquad L = \{(b,(T,(w,\mathrm{ran}(I)))) : \\
&\qquad\qquad ((\cdot,I),b) \in \mathrm{order}(D_{\mathrm{L}},\{(v,I) : \\
&\qquad\qquad\quad Q \in I^{*}_{P,U,\mathrm{B}}((V,F,X,X_R)), \\
&\qquad\qquad\quad P \in I^{*}_{P,U,\mathrm{K}}(Q),\ N \in I^{*}_{P,U,\mathrm{R},\mathrm{mr}}(P), \\
&\qquad\qquad\quad (v,I) \in N\}), \\
&\qquad\qquad w = (f,l,b),\ T = (\{S \cup \{(w,k)\} : (S,k) \in I\},\{w\})\}, \\
&\qquad H = \mathrm{dom}(\mathrm{set}(L)),\ U' = U \cup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad X' = I^{*}_{*\mathrm{X}}((H,X)),\ X'_R = I^{*}_{*\mathrm{X}}((H,X_R)), \\
&\qquad M' = I^{*}_{P,U',\mathrm{D},\mathrm{d}}((V,F \cup H,X',X'_R))
\end{aligned}
$$

A further variation of the *highest-layer maximum-roll layerer* $I_{P,U,\mathrm{L},\mathrm{mr},\mathrm{d}}$, is to add the functionality of the *excluded-self contracted decrementing linear non-overlapping fuds tree maximiser*, $Z_{P,A,A_R,\mathrm{L},\mathrm{xs}}$, by excluding *self partitions*. The *highest-layer excluded-self maximum-roll layerer* $I_{P,U,\mathrm{L},\mathrm{mr},\mathrm{xs},\mathrm{d}} \in$ computers is defined such that the application is

$$
I^{*}_{P,U,\mathrm{L},\mathrm{mr},\mathrm{xs},\mathrm{d}}((V,A,A_R,f)) = \mathrm{layer}(V,U,\emptyset,\emptyset,A,A_R,f,1)
$$

and

$$
\begin{aligned}
&\mathrm{layer}(V,U,F,M,X,X_R,f,l) = \\
&\quad \mathrm{if}((l \leq \mathrm{lmax}) \wedge (H \neq \emptyset) \wedge (M \neq \emptyset \implies \mathrm{maxr}(M') > \mathrm{maxr}(M)), \\
&\qquad\qquad \mathrm{layer}(V,U',F \cup H,M',X',X'_R,f,l+1), \\
&\qquad\qquad (U,F,M)) : \\
&\qquad L = \{(b,(T,(w,\mathrm{ran}(I)))) : \\
&\qquad\qquad ((\cdot,I),b) \in \mathrm{order}(D_{\mathrm{L}},\{(v,I) : \\
&\qquad\qquad\quad Q \in I^{*}_{P,U,\mathrm{B}}((V,F,X,X_R)), \\
&\qquad\qquad\quad P \in I^{*}_{P,U,\mathrm{K}}(Q),\ N \in I^{*}_{P,U,\mathrm{R},\mathrm{mr}}(P), \\
&\qquad\qquad\quad (v,I) \in N,\ |\mathrm{ran}(I)| < |I|\}), \\
&\qquad\qquad w = (f,l,b),\ T = (\{S \cup \{(w,k)\} : (S,k) \in I\},\{w\})\}, \\
&\qquad H = \mathrm{dom}(\mathrm{set}(L)),\ U' = U \cup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad X' = I^{*}_{*\mathrm{X}}((H,X)),\ X'_R = I^{*}_{*\mathrm{X}}((H,X_R)), \\
&\qquad M' = I^{*}_{P,U',\mathrm{D},\mathrm{d}}((V,F \cup H,X',X'_R))
\end{aligned}
$$

The functionality of the *practicable highest-layer shuffle content alignment valency-density fud inducer*, $I_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{d}}'$, is implemented in the *highest-layer fud induction computer* $I_{P,U,\mathrm{Z,F,d}} \in \mathrm{computers}$, which is defined such that (i) the domain is $\mathrm{domain}(I_{P,U,\mathrm{Z,F,d}}) = \mathrm{P}(\mathcal{V}_U) \times \mathcal{A}_U \times \mathcal{A}_U \times \mathbf{N}$, (ii) the range is $\mathrm{range}(I_{P,U,\mathrm{Z,F,d}}) = \mathcal{F} \to \mathbf{Q}$, and (iii) the application is

$$\begin{aligned}
&I_{P,U,\mathrm{Z,F,d}}^*((V,A,A_R,f)) = \\
&\quad\quad \{(G,a) : \\
&\quad\quad\quad (\cdot,F,N) = I_{P,U,\mathrm{L,d}}^*((V,A,A_R,f)), \\
&\quad\quad\quad (K,a) \in N,\ G = \mathrm{depends}(F,K)\}
\end{aligned}$$

The *fud* identifier, f, is used to construct the new *variables* in the new *system*. The *fud* identifier should be such that $\forall i,j \in \mathbf{N}_{>0}\ ((f,i,j) \notin \mathrm{vars}(U))$.

The *fuds* of the *fud induction computer*, $I_{P,U,\mathrm{Z,F,d}}$, are not *partition fuds*, $\forall (U',F,\cdot) \in I_{P,U,\mathrm{L,d}}^*((V,A,A_R,f))\ (F \notin \mathcal{F}_{U',\mathrm{P}})$, so the *fud induction computer* is not a literal implementation of the *fud inducer*, $I_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{d}}'$. However, the *fuds* flatten to the same *substrate partition transforms*, $\forall A \in \mathcal{A}_z\ (\{(F^{\mathrm{TPT}},a) : (F,a) \in I_{P,U_A,\mathrm{Z,F,b,d}}^*((V_A,A,A^{\mathrm{X}},f))\} = \{(F^{\mathrm{TPT}},a) : (F,a) \in I_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{d}}'^*(A)\} \subset (\mathcal{T}_{U_A,V_A} \to \mathbf{Q}))$. Thus, an *inducer* implemented with the *fud induction computer*, $I_{P,U,\mathrm{Z,F,d}}$, after suitable conversion of the *fuds* to *substrate models*, would have the same *inducer correlation* as the *fud inducer*, $I_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{d}}'$. That is, $\mathrm{maxr}(I_{P,U,\mathrm{Z,F,d}}^*((V,A,A^{\mathrm{X}},f))) = \mathrm{maxr}(I_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{d}}'^*(A))$.

Finally, the even more restricted functionality of the *practicable highest-layer excluded-self maximum-roll shuffle content alignment valency-density fud inducer*, $I_{z,\mathrm{csd,F},\infty,\mathrm{q},P,\mathrm{mr,xs,d}}'$, is implemented in the *highest-layer maximum-roll excluded-self fud induction computer* $I_{P,U,\mathrm{Z,F,mr,xs,d}} \in \mathrm{computers}$, which is defined such that the application is

$$\begin{aligned}
&I_{P,U,\mathrm{Z,F,mr,xs,d}}^*((V,A,A_R,f)) = \\
&\quad\quad \{(G,a) : \\
&\quad\quad\quad (\cdot,F,N) = I_{P,U,\mathrm{L,mr,xs,d}}^*((V,A,A_R,f)), \\
&\quad\quad\quad (K,a) \in N,\ G = \mathrm{depends}(F,K)\}
\end{aligned}$$

Now consider the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}'$, implemented by means of *induction computers*. The implementation of the *practicable fud decomposition inducer* in terms of *optimisers* is described above in section 'Optimisation'. The functionality of the *highest-layer limited-models infinite-layer substrate fud decompositions tree searcher*, $Z_{P,A,\mathrm{D,F,d}}$, is implemented in the *highest-layer fud decomper* $I_{P,U,\mathrm{D,F,d}} \in \mathrm{computers}$, which is defined such that (i) the domain is $\mathrm{domain}(I_{P,U,\mathrm{D,F,d}}) = \mathrm{P}(\mathcal{V}_U) \times \mathcal{A}_U$, (ii) the range is $\mathrm{range}(I_{P,U,\mathrm{D,F,d}}) = \mathcal{U} \times \mathcal{D}_{\mathrm{F,d}}$, and (iii) the application is

$$I^*_{P,U,\mathrm{D,F,d}}((V,A)) = \mathrm{decomp}(V,A,U,\emptyset,1,\emptyset)$$

where $\mathrm{decomp} \in \mathrm{P}(\mathcal{V}) \times \mathcal{A} \times \mathcal{U} \times \mathcal{D}_{\mathrm{F,d}} \times \mathbf{N} \times \mathrm{P}(\mathcal{L}(\mathcal{S} \times \mathcal{F}) \times \mathcal{S}) \to (\mathcal{U} \times \mathcal{D}_{\mathrm{F,d}})$ is defined as

$$\begin{aligned}
&\mathrm{decomp}(V,A,U,D,f,Z) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \\
&\qquad\quad \mathrm{if}(N \neq \emptyset \wedge \mathrm{maxr}(N) > 0, \\
&\qquad\qquad \mathrm{decomp}(V,A,U',E,f+1,Z'), \\
&\qquad\qquad \mathrm{decomp}(V,A,U,D,f,Z')), \\
&\qquad\quad (U,D)) : \\
&\qquad Q = \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}\times\mathrm{S}\times\mathcal{X}^3}, \{(z_B,S,L,B) : \\
&\qquad\quad (L,Y) \in \mathrm{places}(D), \\
&\qquad\quad (\cdot,F) = L_{|L|},\ W = \mathrm{der}(F), \\
&\qquad\quad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)), \\
&\qquad\quad (L,S) \notin Z, \\
&\qquad\quad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad\quad B = I^*_{\%}((V, I^*_*((I^*_{*\mathrm{X}}((H_L,A)), \{R_L \cup S\}^{\mathrm{U}})))), \\
&\qquad\quad z_B = \mathrm{size}(B),\ z_B > 0\})), \\
&\qquad \{(\cdot,S,L,B)\} = Q, \\
&\qquad Z' = Z \cup \{(L,S)\}, \\
&\qquad (U',F,N) = I^*_{P,U,\mathrm{L,d}}((V,B,B_{R(B)},f)), \\
&\qquad \{K\} = \mathrm{maxd}(\mathrm{order}(D_{\mathrm{K}}, \mathrm{dom}(N))), \\
&\qquad G = \mathrm{depends}(F,K), \\
&\qquad M = L \cup \{(|L|+1,(S,G))\}, \\
&\qquad E = \mathrm{tree}(\mathrm{paths}(D) \setminus \{L\} \cup \{M\})\}
\end{aligned}$$

and

$$
\begin{aligned}
&\mathrm{decomp}(V,A,U,\emptyset,f,Z) = \\
&\quad \mathrm{if}(N \neq \emptyset \wedge \mathrm{maxr}(N) > 0, \\
&\qquad\qquad \mathrm{decomp}(V,A,U',D,f+1,\emptyset), \\
&\qquad\qquad (U,D_{\emptyset})) : \\
&\qquad (U',F,N) = I^{*}_{P,U,\mathrm{L,d}}((V,A,A_{R(A)},f)), \\
&\qquad \{K\} = \mathrm{maxd}(\mathrm{order}(D_{\mathrm{K}},\mathrm{dom}(N)))), \\
&\qquad G = \mathrm{depends}(F,K), \\
&\qquad D = \{((\emptyset,G),\emptyset)\}
\end{aligned}
$$

The *fuds* of the *decomposition* in the *fud decomposition induction computer*, $I_{P,U,\mathrm{D,F,d}}$, are not *partition fuds*, $\forall F \in \mathrm{fuds}(D)\ (F \notin \mathcal{F}_{U',\mathrm{P}})$ where $(U',D) = \mathrm{decomp}(V,A,U,\emptyset,1,\emptyset)$, so the *fud decomposition induction computer* is not a literal implementation of the *fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}$. However, the *fuds* flatten to the same *substrate partition transforms*, so an *inducer* implemented with the *fud decomposition induction computer*, $I_{P,U,\mathrm{D,F,d}}$, after suitable conversion of the *fuds* to *substrate models*, would have the same *maximum function correlation* as the *fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}$. That is, $I^{*}_{\mathrm{Scsd}}((A,D)) = \mathrm{maxr}(I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}(A))$, where $D = I^{*}_{P,U,\mathrm{D,F,d}}((V,A))$.

The *practicable highest-layer excluded-self maximum-roll summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{mr,xs,d}}$, is implemented in the *highest-layer excluded-self maximum-roll fud decomper* $I_{P,U,\mathrm{D,F,mr,xs,d}} \in \mathrm{computers}$ exactly as in the *highest-layer fud decomper* $I_{P,U,\mathrm{D,F,d}}$, above, except that the *highest-layer maximum-roll excluded-self fud induction computer*, $I_{P,U,\mathrm{Z,F,mr,xs,d}}$, replaces the *highest-layer fud induction computer*, $I_{P,U,\mathrm{Z,F,d}}$.

5 Induction

This section considers how *tractable* and *practicable induction* is related to (a) *structure* and *compression*, and (b) *likelihood* and *sensitivity*.

A variable or structure is defined as *known* below if (a) its type or containing class is defined, and (b) its instance of the type is specified. For example, it is *known* if (i) it is finite and can be explicitly constructed in a first order formula, or (ii) it is countably enumerable and can be defined recursively/algorithmically. An *unknown* variable or structure may be subject to *known* constraints. That is, the variable or structure is partially *known*. At the least, the type of an *unknown* is usually defined.

First review the definitions of the *degree of structure* and *compression*.

Let U be a non-empty finite *system*, $U \in \mathcal{U} = \mathcal{V} \to (\mathrm{P}(\mathcal{W}) \setminus \{\emptyset\})$ such that $0 < |U| < \infty$ and $\forall (\cdot, W) \in U\ (|W| < \infty)$. Let $X \subset \mathcal{X}$ be a non-empty *unknown* finite set of *event identifiers*, $0 < |X| < \infty$. Let $\mathcal{H}_{U,X}$ be the non-empty *unknown* finite set of *histories* in *system* U having *event identifiers* X,

$$\mathcal{H}_{U,X} = \bigcup \{X \to W^{\mathrm{CS}} : W \subseteq \mathrm{vars}(U)\} \subset \mathcal{H}_U \subset \mathcal{X} \to \mathcal{S}_U$$

and $0 < |\mathcal{H}_{U,X}| < \infty$.

Let $P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ be an *unknown history probability function* in the *histories* $\mathcal{H}_{U,X}$. The expected *space* of a *history coder* $C \in \mathrm{coders}(\mathcal{H}_{U,X})$ is greater than or equal to the *entropy* of the *history probability function*,

$$\begin{aligned}
\mathrm{expected}(P)(C^{\mathrm{s}}) &= \sum_{H \in \mathcal{H}_{U,X}} P_H \times C^{\mathrm{s}}(H) \\
&\geq -\sum (P_H \ln P_H : H \in \mathcal{H}_{U,X},\ P_H > 0) \\
&= \mathrm{entropy}(P)
\end{aligned}$$

The expected *space* of a *history coder* $C \in \mathrm{coders}(\mathcal{H}_{U,X})$ may also be compared to the *canonical space* which is the lesser *space* of the *canonical history coders*, (i) *index history coder*, C_{H}, and (ii) *classification history coder*, C_{G}. The expected *canonical space* is defined $\mathrm{canonical}(U, X) \in ((\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}) \to \mathbf{Q}_{\geq 0} \ln \mathbf{N}_{>0}$ as

$$\mathrm{canonical}(U, X)(P) := \mathrm{expected}(P)(\mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}, C^{\mathrm{s}}_{\mathrm{G}}))$$

The expected *canonical space* is also always greater than or equal to the *entropy* of the *history probability function*, $\mathrm{canonical}(U, X)(P) \geq \mathrm{entropy}(P)$.

The *degree of structure* is defined for *probability function* $P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ with respect to a *history coder* $C \in \mathrm{coders}(\mathcal{H}_{U,X})$ in terms of a relationship between (i) the *expected space*, $\mathrm{expected}(P)(C^{\mathrm{s}})$, (ii) the expected *canonical space*, $\mathrm{canonical}(U, X)(P)$, and (iii) the *entropy*, $\mathrm{entropy}(P)$. The *degree of structure* is defined in section 'Derived history space', $\mathrm{structure}(U, X) \in ((\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}) \times \mathrm{coders}(\mathcal{H}_{U,X}) \to \mathbf{Q} \ln \mathbf{Q}_{>0} / \ln \mathbf{Q}_{>0}$ as

$$\mathrm{structure}(U, X)(P, C) := \frac{\mathrm{canonical}(U, X)(P) - \mathrm{expected}(P)(C^{\mathrm{s}})}{\mathrm{canonical}(U, X)(P) - \mathrm{entropy}(P)}$$

The *compression* of *coder* C with respect to *probability function* P is a synonym for the *degree of structure* of *probability function* P with respect to the *coder* C.

The *degree of structure*, or *compression*, is defined for any *history coder*, $\mathrm{coders}(\mathcal{H}_{U,X})$. The *derived history coders* are a special case of *history coders*. Given a *transform* T, the *expanded specialising derived history coder* $C_{\mathrm{G,T,H}}(T) \in \mathrm{coders}(\mathcal{H}_{U,X})$ is derived from the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T,H}}$. It *expands* the *transform* to the *history variables*, V_H, where the set of *history variables* is a superset of the *underlying variables*, $V = \mathrm{und}(T)$, and otherwise defaults to an *index coder*,

$$\begin{aligned} C_{\mathrm{G,T,H}}(T)^{\mathrm{s}}(H) &= (C_{\mathrm{G},V_H,\mathrm{T,H}}(T^{\mathrm{P}V_H\mathrm{T}})^{\mathrm{s}}(H) + s_{|V_H|} \ : \ V_H \supseteq V) \ + \\ &\quad (C^{\mathrm{s}}_{\mathrm{H}}(H) \ : \ V_H \not\supseteq V) \end{aligned}$$

where $s_n = \mathrm{spaceVariables}(U)(n)$ and the *specialising derived substrate history coder* is

$$C_{\mathrm{G},V,\mathrm{T,H}}(T) \ = \ \mathrm{coderHistorySubstrateDerivedSpecialising}(U, X, T, D_{\mathrm{S}}, D_{\mathrm{X}})$$

The *specialising degree of structure* of the *probability function* P with respect to the *expanded specialising derived history coder* for some *transform* T is

$$\mathrm{structure}(U, X)(P, C_{\mathrm{G,T,H}}(T)) \in \mathbf{Q} \ln \mathbf{Q}_{>0} / \ln \mathbf{Q}_{>0}$$

In section 'Derived history space', above, the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{T,H}}(T)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, of *history* $H \in \mathcal{H}_{U,X}$ is characterised for given *transform* T in terms of (i) the *component size cardinality relative entropy*,

$$\mathrm{entropyRelative}(A_H * T, V^{\mathrm{C}} * T)$$

(ii) the *possible derived volume space*, w', where $w' = |T^{-1}|$, and (iii) the *expected component entropy*,

$$\mathrm{entropyComponent}(A_H, T)$$

The *specialising-canonical space difference* is minimised by varying the *transform* such that (i) the *derived entropy* is low, (ii) the *possible derived volume* is small, (iii) the *underlying components* have high *entropy* and (iv) high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*. The *canonical space* terms, $C^{\mathrm{s}}_{\mathrm{H},V}(H)$ and $C^{\mathrm{s}}_{\mathrm{G},V}(H)$, do not depend on the *transform*, T, and so the minimisation of the *specialising-canonical space difference* is also the minimisation of the *specialising derived substrate history coder space*, $C_{\mathrm{G},V,\mathrm{T,H}}(T)^{\mathrm{s}}(H)$.

Now review *substrate structure alignment.*

The *substrate transforms set* in *system* U and *variables* V is defined

$$\mathcal{T}_{U,V} = \{F^{\mathrm{T}} : F \subseteq \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\}\}$$

The set of *complete congruent integral substrate histograms* of *size* z is defined

$$\mathcal{A}_{U,\mathrm{i},V,z} = \{A : A \in \mathcal{A}_{U,\mathrm{i}},\ A^{\mathrm{U}} = V^{\mathrm{C}},\ \mathrm{size}(A) = z\}$$

Given $T \in \mathcal{T}_{U,V}$, the *integral iso-transform-independents* is derived from the *formal-abstract* pair valued function of the *complete congruent integral substrate histograms*,

$$Y_{U,\mathrm{i},T,z} = \{(A, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The *integral iso-transform-independents* given *transform* $T \in \mathcal{T}_{U,V}$ for *integral substrate histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ are abbreviated

$$\begin{aligned}\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) &= Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \\ &= \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}\end{aligned}$$

The *integral iso-transform-independents*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)$, are equivalently the subset of *integral substrate histograms* which are both *iso-formal* and *iso-abstract* with respect to *substrate histogram* A, $\forall B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)\ ((B^{\mathrm{X}} * T = A^{\mathrm{X}} * T) \wedge ((B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}))$.

The *generalised multinomial probability distribution* of *draw* $(E, z) \in \mathcal{A}_{U,V,z_E} \times \mathbf{N}$ is defined $\hat{Q}_{\mathrm{m},U}(E, z) \in (\mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{Q}_U \cap \mathcal{Q}_z \cap \mathcal{P}$. The *generalised multinomial probability* for *integral substrate histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ is

$$\hat{Q}_{\mathrm{m},U}(E, z)(A) = \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S}$$

The set of *sized cardinal substrate histograms* $\mathcal{A}_z$ is the finite set of *complete integral cardinal substrate histograms* of *size* z and *dimension* less than or equal to the *size* such that the *independent* is *completely effective*,

$$\mathcal{A}_z = \{A : A \in \mathcal{A}_{\mathrm{c}} \cap \mathcal{A}_{\mathrm{i}},\ \mathrm{size}(A) = z,\ |V_A| \leq z,\ A^{\mathrm{U}} = A^{\mathrm{XF}} = A^{\mathrm{C}}\}$$

where $A^{\mathrm{CS}} = \mathrm{cartesian}(U_A)(V_A)$ and $U_A = \mathrm{implied}(\mathrm{implied}(A))$ and $V_A = \mathrm{vars}(A)$.

The subset of the *sized cardinal substrate histograms*, $\mathcal{A}_z$, for which the *independent*, A^{X}, is *integral*, is the set of *integral-independent substrate histograms*,

$$\mathcal{A}_{z,\mathrm{xi}} = \{A : A \in \mathcal{A}_z,\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\}$$

The *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*, also known as the *alignment-bounded iso-transform space ideal transform search set*, is defined $X_{z,\mathrm{xi,T,y,fa,j}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$ as

$$\begin{aligned} X_{z,\mathrm{xi,T,y,fa,j}}(A) &= \{(T, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\sum_{B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)} \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B)}) : \\ &\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\} \end{aligned}$$

The *derived alignment integral-independent substrate ideal formal-abstract transform search set* is defined $X'_{z,\mathrm{xi,T,a,fa,j}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$ as

$$\begin{aligned} X'_{z,\mathrm{xi,T,a,fa,j}}(A) &= \{(T, \mathrm{algn}(A * T)) : \\ &\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\} \end{aligned}$$

In section 'Substrate structures alignment', above, it is conjectured that the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$, is correlated with the *derived alignment integral-independent substrate ideal formal-abstract transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}) \geq 0)$$

where $\mathrm{cov}(z)(F,G) := \mathrm{covariance}(\hat{R}_z)(F,G)$ and the *renormalised geometry-weighted probability function* is $\hat{R}_z = \mathrm{normalise}(\{(A, 1/(|V_A|! \prod_{w \in V_A} |U_A(w)|!)) : A \in \mathrm{dom}(F)\})$.

Now review the definition of *induction*. First *inducers* and *literal inducers* are defined.

The set of *inducers* is defined in section 'Tractable alignment-bounding', above. The *inducers* are *computers* $I_z \in \mathrm{inducers}(z) \subset \mathrm{computers}$ such that (i) the domain is a set of *substrate histograms* which are at least a superset of the *integral-independent substrate histograms*, $\mathcal{A}_{z,\mathrm{xi}} \subseteq \mathrm{domain}(I_z) \subseteq \mathcal{A}_z$, (ii) the finite *time* and *space* application returns a rational-valued function of the *substrate models set*, $I_z^*(A) \in \mathcal{M}_{U_A,V_A} \to \mathbf{Q}$, and (iii) the maximum of the *inducer* application, $\mathrm{maxr} \circ I_z^*$, is positively correlated with the finite *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I_z^*) \geq 0)$$

That is, the *induction correlation* of *inducer* I_z is positive.

The *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer* $I'_{z,\mathrm{a,l}} \in \mathrm{inducers}(z)$ is a literal finite approximation to the *derived alignment integral-independent substrate ideal formal-abstract transform search set*, $X'_{z,\mathrm{xi,T,a,fa,j}}(A)$,

$$\begin{aligned} I'^{*}_{z,\mathrm{a,l}}(A) \ &= \ \{(T, I^*_{\approx\ln\mathbf{Q}}(\mathrm{algn}(A * T))) : \\ &\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\} \end{aligned}$$

The *induction correlation* of the *literal derived alignment inducer* is conjectured to be positive,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}) \geq 0)$$

Now consider the definition of the class of *tractable inducers*.

Although the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, is finitely computable and faster than a literal implementation of the *alignment-bounded iso-transform space ideal transform search set*, $X_{z,\mathrm{xi,T,y,fa,j}}$, it is nonetheless intractable. Section 'Tractable alignment-bounding' discusses the various intractabilities and the classes of limits and constraints on the structures of more tractable *inducers*.

First, the *substrate volume* is intractable. The application of the *transformer* $I_{*\mathrm{T}}$, in the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, to a *substrate histogram*, A, and a *substrate transform*, $T \in \mathcal{T}_{U_A,V_A}$, is $I^*_{*\mathrm{T}}((T, A)) = A * T$. The *volume*, $|V_A^{\mathrm{C}}|$, grows exponentially with *underlying dimension* $n = |V_A|$, and so the *space* complexity of the *transformer*, $I_{*\mathrm{T}}$, is exponential with respect to *underlying dimension*, n. To address this (i) the *inducer models* of the *literal derived alignment inducer* are expanded from *substrate transforms*, $\mathcal{T}_{U_A,V_A}$, to *substrate fuds*, $\mathcal{F}_{U_A,V_A}$, and (ii) the *substrate fuds* are then limited by intersecting with one of the class of *limited-underlying* subsets of the *functional definition sets* $\mathcal{F}_{\mathrm{u}} \subset \mathcal{F}$. A set of *limited-underlying fuds*, $\mathcal{F}_{\mathrm{u}}$, is defined such that a *fud* $F \in \mathcal{F}_{\mathrm{u}}$ is such that its *transforms*, $F \subset \mathcal{T}$, are each tractably computable. For example the *underlying volume* of the *transforms* may be limited by a *maximum underlying volume* limit $\mathrm{xmax} \in \mathbf{N}_{\geq 4}$. The set of *inducer models* is $\mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{u}}$.

Next, the *derived volume* is intractable. Both the computation *time* and computation *space* of the *alignmenter* applied to the *transformed sample histogram*, $I_{\mathrm{a}}^*(A * T) \approx \mathrm{algn}(A * T)$, in the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$, vary with the *derived volume*, $w = |W^{\mathrm{C}}|$, where $W = \mathrm{der}(T)$. The *derived volume*, w, grows exponentially with *derived dimension* $m = |W|$ and so the *time* and *space* complexities are exponential, and therefore intractable, with respect to *derived dimension*, m. This is also the case where the implementation uses a *fuder*, $I_{*\mathrm{F}}$, in a *limited-underlying derived alignment fud inducer*, because the application of a *fud* $F \in \mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{u}}$ must still compute $(A * F)^{\mathrm{X}}$ in an *independenter*, I_{X}, in order to compute *derived alignment*, $\mathrm{algn}(A * F)$. So a further compromise is made by intersecting the *substrate fuds* with one of the class of *limited-derived* subsets of the *functional definition sets* $\mathcal{F}_{\mathrm{d}} \subset \mathcal{F}$. A set of *limited-derived fuds*, $\mathcal{F}_{\mathrm{d}}$, is defined such that a *fud* $F \in \mathcal{F}_{\mathrm{d}}$ is such that the *independent derived* of the *fud* is tractable. For example the *derived volume* of the *fud* may be limited by a *maximum derived volume* limit of $\mathrm{wmax} \in \mathbf{N}_{\geq 4}$.

Although the *limited-variables substrate fuds*, $\mathcal{F}_{U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}}$, has coverage of the entire *substrate* even when the *substrate volume*, v, is greater than the *underlying volume limit*, for example $v > \mathrm{xmax}$, the *derived volume* is still strictly limited, $w \leq \mathrm{wmax}$. In section 'Summation aligned decomposition inducers', above, it is conjectured that a *summation aligned decomposition* $D \in \mathcal{D}_{\Sigma}(A)$ is such that the *content alignment* equals the *summation alignment*, $\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) = \mathrm{alignmentSum}(A, D)$, where $\mathrm{alignmentSum}(A, D) = \sum \mathrm{algn}(A * C * T) : (C, T) \in \mathrm{cont}(D)$ and $\mathrm{cont} = \mathrm{elements} \circ \mathrm{contingents}$. Thus, insofar as the *content alignment* approximates to the *derived alignment*, summing the *derived alignments* of the *contingent fuds* avoids the computation of the *nullable transform*, D^{T}, which may have intractable *derived volume*, for example $w > \mathrm{wmax}$, where $w = |W^{\mathrm{C}}|$ and $W = \mathrm{der}(D^{\mathrm{T}})$. Just as above where the set of *inducer models* is increased from *substrate transforms*, $\mathcal{T}_{U_A,V_A}$, to *substrate fuds*, $\mathcal{F}_{U_A,V_A}$, the *inducer models* is again expanded to the *substrate fud decompositions* $\mathcal{D}_{\mathrm{F},U_A,V_A}$. The set of *inducer models* is then the *limited-variables substrate fud decompositions*, $\mathcal{D}_{\mathrm{F},U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}}))$. Given a *limited-variables substrate fud decomposition* $D \in \mathcal{D}_{\mathrm{F},U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}}))$, the *inducer* computes the tractable sum of the *contingent derived alignments*,

$\sum \text{algn}(A * C * F^{\text{T}}) : (C, F) \in \text{cont}(D)$.

Third, the computation of the *search set models* is intractable for two reasons, (i) *fud flattening*, and (ii) *layer variables* cardinality. The computation of the finite *substrate fud set*, $\mathcal{F}_{U_A,V_A}$, requires the exclusion of duplicate nested *partitions*. This is done by checking for the uniqueness of the *flattened partitions*. This check is intractable so the *substrate fud set*, $\mathcal{F}_{U_A,V_A}$, is replaced by the intersection of (i) the *infinite-layer substrate fud set* $\mathcal{F}_{\infty,U_A,V_A}$, which dispenses with the check, and (ii) one of the class of the sets of *limited-layer fuds*, $\mathcal{F}_{\text{h}}$. The *limited-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\text{h}}$, places a limit on the number of *layers*, for example, a *maximum layer* limit of $\text{lmax} \in \mathbf{N}_{>0}$.

The *layer variables* cardinality intractability is because of exponential *time* complexity of the computation of the *layers* of the *fuds*. This is addressed by defining one of the class of the sets of *limited-breadth fuds*, $\mathcal{F}_{\text{b}}$. For example a *maximum layer breadth* limit of $\text{bmax} \in \mathbf{N}_{>0}$. Together the classes of limits are intersected together to form the class of *limited-models* $\mathcal{F}_{\text{q}} = \mathcal{F}_{\text{u}} \cap \mathcal{F}_{\text{d}} \cap \mathcal{F}_{\text{h}} \cap \mathcal{F}_{\text{b}}$. The set of *inducer models* is then the *limited-models infinite-layer fud decompositions*, $\mathcal{D}_{\text{F},\infty,U_A,V_A} \cap \text{trees}(\mathcal{S} \times \mathcal{F}_{\text{q}})$.

Last, the computation of the *literal substrate model inclusion* is intractable. The derivation of the conjectured *induction correlation* of the *literal derived alignment inducer*,

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(\text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa},\text{j}}, \text{maxr} \circ I'^{*}_{z,\text{a},\text{l}}) \geq 0)$$

is described in section 'Derived alignment and conditional probability'. The derivation imposes several constraints, (i) *integral independent histogram*, $A \in \mathcal{A}_{\text{xi}} \implies A^{\text{X}} \in \mathcal{A}_{\text{i}}$, (ii) the *formal histogram* equals the *abstract histogram*, $A^{\text{X}} * T = (A * T)^{\text{X}}$, and (iii) the *transform* is *ideal*, $A = A * T * T^{\dagger A}$. *Formal-abstract equality* implies *independent formal*, $A^{\text{X}} * T = (A * T)^{\text{X}} = (A^{\text{X}} * T)^{\text{X}}$. Together with *integral independent histogram*, $A^{\text{X}} \in \mathcal{A}_{\text{i}}$, this implies that the *independent* is an *integral iso-transform-independent*,

$$A^{\text{X}} \in \mathcal{A}_{U,\text{i},\text{y},T,z}(A) = Y^{-1}_{U,\text{i},T,z}(((A^{\text{X}} * T), (A * T)^{\text{X}}))$$

and the *lifted integral iso-transform-independents* contains the *abstract histogram*

$$(A * T)^{\text{X}} \in \mathcal{A}'_{U,\text{i},\text{y},T,z}(A) = \{B * T : B \in Y^{-1}_{U,\text{i},T,z}(((A^{\text{X}} * T), (A * T)^{\text{X}}))\}$$

Then, given the *minimum alignment conjecture*, it can be shown that the *alignment-bounded lifted iso-transform space* is bounded by the *derived align-*

ment,

$$\begin{aligned}
&\mathrm{algn}(A * T) \\
\leq\ & \Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(A * T)}{\sum_{B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}} * T, z)(B')} : \\
&\quad E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}},\ E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \Bigg) \\
\leq\ & \mathrm{algn}(A * T) + \ln |\mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)|
\end{aligned}$$

The corresponding *alignment-bounded iso-transform space* is

$$\begin{aligned}
&\Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(B)} : \\
&\quad E^{\mathrm{X}} * T = (E^{\mathrm{X}} * T)^{\mathrm{X}},\ E^{\mathrm{XF}} \geq A^{\mathrm{XF}},\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \Bigg) \in \ln \mathbf{Q}_{>0}
\end{aligned}$$

The *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, and so each *derived histogram* maps to exactly one set of *iso-transform-independents*,

$$\begin{aligned}
&\{(A * T, Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})))) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\} \\
\in\ & \mathcal{A}_{U,\mathrm{i},W,z} \to \mathrm{P}(\mathcal{A}_{U,\mathrm{i},V,z})
\end{aligned}$$

Thus the *alignment-bounded lifted iso-transform space* is correlated with the *alignment-bounded iso-transform space*.

The derivation goes on to conjecture that when *independent-sample distributed*, $E^{\mathrm{X}} = A^{\mathrm{X}}$, the correlation is highest when the *transform* is *ideal*, $A = A * T * T^{\dagger A}$. That is, the *alignment-bounded lifted iso-transform space* is correlated with the *alignment-bounded iso-transform idealisation space*,

$$\begin{aligned}
&\Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(A * T * T^{\dagger A})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \hat{Q}_{\mathrm{m},U}(A^{\mathrm{X}}, z)(B)} : \\
&\quad A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A * T * T^{\dagger A} \in \mathcal{A}_{\mathrm{i}} \Bigg) \in \ln \mathbf{Q}_{>0}
\end{aligned}$$

Therefore the *derived alignment*, $\mathrm{algn}(A * T)$, is conjectured to be correlated with the *alignment-bounded iso-transform idealisation space*, hence the *literal derived alignment inducer correlation*.

The constraint that the *independent histogram* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, is sometimes not the case if the *sample histogram*, A, is a given. As noted in section 'Tractable alignment-bounding', the *inducer correlation*, $\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I_z^*)$, is restricted to the intersection of the domains of the argument functions, which is the *integral-independent substrate histograms*, $\mathcal{A}_{z,\mathrm{xi}}$. However, in the case of *non-integral-independent substrate histograms*, $\mathcal{A}_z \setminus \mathcal{A}_{z,\mathrm{xi}}$, an *inducer* defined in terms of the *generalised multinomial probability distribution* $\hat{Q}_{\mathrm{m},U}(E,z) \in \mathcal{A}_{U,\mathrm{i},V,z} \to \mathbf{Q}_{\geq 0}$ can be extended by interpolating instead with the *multinomial probability density function*, $\mathrm{mpdf}(U)(E,z) \in \mathcal{A}_{U,V,z} \to \mathbf{R}_{\geq 0}$,

$$\mathrm{mpdf}(U)(E,z)(A) := \frac{\Gamma_! z}{\prod_{S \in A^{\mathrm{S}}} \Gamma_! A_S} \prod_{S \in A^{\mathrm{S}}} \left(\frac{E_S}{z_E}\right)^{A_S}$$

The *multinomial probability density function* is defined in terms of the unit-translated gamma function, $\Gamma_! \in \mathbf{R} \to \mathbf{R}$. An *inducer* defined in terms of *alignment* is already extended to the *non-integral-independent substrate histograms*, $\mathrm{algn} \in \mathcal{A} \to \mathbf{R}$ is defined as $\mathrm{algn}(A) := \sum_{S \in A^{\mathrm{S}}} \ln \Gamma_! A_S - \sum_{S \in A^{\mathrm{XS}}} \ln \Gamma_! A^{\mathrm{X}}_S$.

The *substrate histograms*, $\mathcal{A}_z$, defined in section 'Substrate structures alignment', are such that the *independent* is *completely effective*, $A^{\mathrm{XF}} = A^{\mathrm{C}}$. Hence, if the *size*, z, is less than the *volume*, $v = |A^{\mathrm{C}}|$, the *independent* is necessarily *non-integral*, $z < v \implies A^{\mathrm{X}} \notin \mathcal{A}_{\mathrm{i}}$. For this reason, any *volume* limits, for example, $\mathrm{xmax} \in \mathbf{N}_{\geq 4}$, should be chosen such that they are less than or equal to the *size*, $\mathrm{xmax} \leq z$. This is also more likely to avoid the region of negative logarithm unit-translated gamma function, $\forall x \in \mathbf{R}\ (0 < x < 1 \implies 0 > \ln \Gamma_! x < 0)$.

Consider the remaining two constraints, (i) *formal-abstract equality*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, and (ii) *ideal transform*, $A = A * T * T^{\dagger A}$. As described in section 'Intractable literal substrate model inclusion', both of these inclusion tests of the *model* are intractable with respect to *substrate volume*. The section considers how *inducers* can be made tractable while adhering to these constraints as closely as possible.

First, the *formal-abstract equality* is weakened to the *independent-formal* constraint, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, in the *derived alignment substrate ideal independent-formal transform inducer*, $I'_{z,\mathrm{a,fx,j}}$. This constraint is still intractable, so it is replaced by constraining the *transforms* to be *non-overlapping*,

$\neg\mathrm{overlap}(T) \implies A^{\mathrm{X}}*T = (A^{\mathrm{X}}*T)^{\mathrm{X}}$, in the *derived alignment substrate ideal non-overlapping transform inducer*, $I'_{z,\mathrm{a,n,j}}$. If the *ideality* inclusion test is dropped and the *inducer model* set of *substrate transforms*, $\mathcal{T}_{U_A,V_A}$, is replaced by the *limited-models infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}$, then the tractable *limited-models derived alignment substrate non-overlapping infinite-layer fud inducer* $I'_{z,\mathrm{a,F},\infty,\mathrm{n,q}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, can be defined

$$\begin{aligned}&I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{n,q}}(A) = \\ &\quad \{(F, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * F^{\mathrm{T}}))) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}\}\end{aligned}$$

Although this *inducer* is tractable, the *non-overlapping* constraint is weaker than the *formal-abstract equality* constraint and the *ideality* constraint has been dropped altogether. The *entropy* of the doubly-*independent formal independent histogram*, $\mathrm{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}})$, is expected to be greater than the *entropy* of the *abstract histogram*, $\mathrm{entropy}((A * T)^{\mathrm{X}})$, whereas if the *formal-abstract equality* constraint holds then the *entropies* would be equal. The *abstract-non-formal entropy substrate ideal independent-formal transform inducer*, $I'_{z,\mathrm{e,fx,j}} \in \mathrm{inducers}(z)$, maximises the *entropy* difference between the *abstract* and the *formal independent*. The *inducer* is defined

$$\begin{aligned}&I'^{*}_{z,\mathrm{e,fx,j}}(A) = \\ &\quad \{(T, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{entropy}((A * T)^{\mathrm{X}}) - \mathrm{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}}))) : \\ &\qquad\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}\end{aligned}$$

Derived alignment approximates to the *sized entropy* difference between the *abstract histogram* and the *derived histogram*,

$$\mathrm{algn}(A * T) \quad\approx\quad z \times (\mathrm{entropy}((A * T)^{\mathrm{X}}) - \mathrm{entropy}(A * T))$$

so the *abstract-non-formal entropy inducer*, $I'_{z,\mathrm{e,fx,j}}$, weakly maximises the *derived alignment*.

The *abstract-non-formal entropy inducer* is defined in terms of the *entropies* of *histograms* in the *derived variables*, $\mathrm{entropy}((A * T)^{\mathrm{X}})$ and $\mathrm{entropy}((A^{\mathrm{X}} * T)^{\mathrm{X}})$, and so ignores the *entropies* of the *underlying components*, $\{(C, \mathrm{entropy}(A * C)) : (\cdot, C) \in T^{-1}\} \in \mathrm{P}(V_A^{\mathrm{CS}}) \to \ln\mathbf{Q}_{>0}$. The discussion considers the *actualisations*, which alter the relative *independence* of the *derived* and *underlying*, and then proposes an *inducer* that maximises the *midisation pseudo-alignment*, $\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A})$. However, the *ideality* constraint restricts the *midisation pseudo-alignment* to be equal to the negative *surrealisation alignment*, so the *ideality* constraint is dropped. The *midisation pseudo-alignment substrate independent-formal transform inducer* $I_{z,\mathrm{m,fx}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, is defined

$$I^{*}_{z,\mathrm{m,fx}}(A) = \{(T, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A) - \mathrm{algn}(A * T * T^{\dagger A}) - \mathrm{algn}((A * T)^{\mathrm{X}} * T^{\odot A}))) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}$$

The computation of *midisation* is intractable, so a further approximation is required. Maximisation of *midisation* tends to move *component alignments* from *off-diagonal states* to *on-diagonal states*, balancing the high *derived alignment* of longer *diagonals* with the high *on-diagonal component alignments* of shorter *diagonals*. Thus the *midisation pseudo-alignment* varies with the *derived alignment valency density*. The *derived alignment valency-density substrate independent-formal transform inducer* $I'_{z,\mathrm{ad,fx}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, is defined

$$I'^{*}_{z,\mathrm{ad,fx}}(A) = \{(T, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * T)/w^{1/m})) : T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}\}$$

If the *independent formal* constraint is replaced by constraining the *transforms* to be *non-overlapping*, and the *inducer model* set of *substrate transforms*, $\mathcal{T}_{U_A,V_A}$, is replaced by the *limited-models infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{q}}$, then the tractable *limited-models derived alignment valency-density substrate non-overlapping infinite-layer fud inducer* $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}} \in \mathrm{inducers}(z)$, given *substrate histogram* $A \in \mathcal{A}_z$, can be defined as

$$I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}(A) = \{(F, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * F^{\mathrm{T}})/w^{1/m})) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}\}$$

This *derived alignment valency-density fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$, addresses the *formal-abstract equality* constraint, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, but ignores the *ideal transform* constraint, $A = A * T * T^{\dagger A}$. As described above, intractable *derived volume* can be addressed by expanding the *inducer models* to the *substrate fud decompositions*, $\mathcal{D}_{\mathrm{F},U_A,V_A}$, and summing the *derived alignments* of the *contingent fuds*, $\sum \mathrm{algn}(A * C * F^{\mathrm{T}}) : (C,F) \in \mathrm{cont}(D)$. Using a similar method, sections 'Decomposition alignment' and 'Tractable decomposition inducers' show how maximising the sum of the *contingent alignment valency-densities*, $\sum \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} : (C,F) \in \mathrm{cont}(D)$, of *limited-models non-overlapping infinite-layer fud decompositions*, $\mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}))$, removes *alignments* along the *decomposition* path and tends to *independent leaf components*. When *fully decomposed* the *nullable transform* of the *decomposition* is *ideal*, $A * D^{\mathrm{T}} * D^{\mathrm{T}\dagger A} = A$. The tractable *limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer*, given *non-independent substrate histogram* $A \in \mathcal{A}_z \setminus \{A^{\mathrm{X}}\}$, is defined

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A) = \\
&\quad \{(D, I^{*}_{\approx\mathbf{R}}(\sum \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} : (C,F) \in \mathrm{cont}(D))) : \\
&\qquad\qquad D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})), \\
&\qquad\qquad\qquad \forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A * C * F^{\mathrm{T}}) > 0)\}
\end{aligned}$$

where $W_F = \mathrm{der}(F)$, $w_F = |W_F^{\mathrm{C}}|$ and $m_F = |W_F|$. The *summed alignment valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is conjectured to have positive *induction correlation*. That is, it is positively correlated with the *alignment-bounded iso-transform space ideal transform maximum function*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}) \geq 0)$$

In section 'Tractable decomposition inducers', above, it is shown that, although the maximisation of the *midisation alignment* tends to minimise the *mid component size cardinality relative entropy*, the subsequent maximisation of the *idealisation alignment* tends to restore the *relative entropy* so that the maximal *relative entropy* of the *tractable limited-models summed alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is (a) greater than that of the corresponding *model* in the *tractable derived alignment valency-density fud inducer*, $I'_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}$,

$$\mathrm{entropyRelative}(A * D^{\mathrm{T}}, V_A^{\mathrm{C}} * D^{\mathrm{T}}) \quad > \quad \mathrm{entropyRelative}(A * F_{\mathrm{ad}}^{\mathrm{T}}, V_A^{\mathrm{C}} * F_{\mathrm{ad}}^{\mathrm{T}})$$

where $D \in \mathrm{maxd}(I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A))$ and $F_{\mathrm{ad}} \in \mathrm{maxd}(I'^{*}_{z,\mathrm{ad,F},\infty,\mathrm{n,q}}(A))$, and (b) comparable to that of the corresponding *model* in the *tractable derived alignment fud inducer*, $I'_{z,\mathrm{a,F},\infty,\mathrm{n,q}}$,

$$\mathrm{entropyRelative}(A * D^{\mathrm{T}}, V_A^{\mathrm{C}} * D^{\mathrm{T}}) \quad \approx \quad \mathrm{entropyRelative}(A * F_{\mathrm{a}}^{\mathrm{T}}, V_A^{\mathrm{C}} * F_{\mathrm{a}}^{\mathrm{T}})$$

where $F_{\mathrm{a}} \in \mathrm{maxd}(I'^{*}_{z,\mathrm{a,F},\infty,\mathrm{n,q}}(A))$.

In order to investigate the constraints necessary to make *tractable inducers* practicable, section 'Practicable alignment-bounding', above, considers how the *limited-models non-overlapping infinite-layer substrate fuds*, $\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}$, may be constructed. Section 'Optimisation' goes on to consider the explicit definitions of the (i) *limited-models* constraints, and (ii) *layer*-ordered *limited-underlying limited-breadth infinite-layer substrate fuds trees*, in order to define a finite search for a *practicable inducer*.

The *practicable highest-layer shuffle content alignment valency-density fud inducer*, $I'_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{d}}$, is defined,

$$\begin{aligned}
&I'^{*}_{z,\mathrm{csd},\mathrm{F},\infty,\mathrm{q},P,\mathrm{d}}(A) = \\
&\quad \{(G, I^{*}_{\mathrm{csd}}((A, A_R, G))) : \\
&\qquad\qquad |V_A| > 1,\ \{F_{\mathrm{L}}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,\mathrm{L},\mathrm{d}})), \\
&\qquad\qquad K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A,A_R,F_{\mathrm{L}},\mathrm{D},\mathrm{d}})),\ G = \mathrm{depends}(F_{\mathrm{L}}, K)\} \cup \\
&\quad \{(\emptyset, 0) : |V_A| \leq 1\}
\end{aligned}$$

where (i) the *shuffle content alignment valency-density computer* is

$$I^{*}_{\mathrm{csd}}((A, A_R, F)) = (I^{*}_{\mathrm{a}}(A * F^{\mathrm{T}}) - I^{*}_{\mathrm{a}}(A_R * F^{\mathrm{T}}))/I^{*}_{\mathrm{cvl}}(F)$$

(ii) the *valency capacity computer* is

$$I^{*}_{\mathrm{cvl}}(F) := (I^{*}_{\approx\mathrm{pow}}((w, 1/m)) : W = \mathrm{der}(F),\ w = |W^{\mathrm{C}}|,\ m = |W|)$$

(iii) the *highest-layer limited-layer limited-underlying limited-breadth fud tree searcher* is

$$Z_{P,A,A_R,\mathrm{L},\mathrm{d}} = \mathrm{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}, P_{P,A,A_R,\mathrm{L},\mathrm{d}}, \{\emptyset\})$$

(iv) the *highest-layer limited-layer limited-underlying limited-breadth fud tree searcher* neighbourhood function is

$$\begin{aligned}
&P_{P,A,A_R,\mathrm{L},\mathrm{d}}(F) = \{G : \\
&\quad G \in P_{P,A,A_R,\mathrm{L}}(F), \\
&\quad (F \neq \emptyset \implies \mathrm{maxr}(\mathrm{el}(Z_{P,A,A_R,F,\mathrm{D},\mathrm{d}})) < \mathrm{maxr}(\mathrm{el}(Z_{P,A,A_R,G,\mathrm{D},\mathrm{d}})))\}
\end{aligned}$$

(v) the *limited-layer limited-underlying limited-breadth fud tree searcher* neighbourhood function is

$$\begin{aligned}
&P_{P,A,A_R,\mathrm{L}}(F) = \{G : \\
&\quad G = F \cup \{T : K \in \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{B}})), \\
&\qquad\qquad H \in \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n},-,K})), \\
&\qquad\qquad w \in \mathrm{der}(H),\ I = \mathrm{depends}(\mathrm{explode}(H), \{w\}),\ T = I^{\mathrm{TPT}}\}, \\
&\quad \mathrm{layer}(G, \mathrm{der}(G)) \leq \mathrm{lmax}\}
\end{aligned}$$

(vi) the *limited-underlying tuple set list maximiser* is

$$Z_{P,A,A_R,F,\mathrm{B}} = \mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{B}}, P_{P,A,A_R,F,\mathrm{B}}, \mathrm{top}(\mathrm{omax}), R_{P,A,A_R,F,\mathrm{B}})$$

(vii) the *limited-underlying tuple set list maximiser* optimiser function is

$$\begin{aligned}X_{P,A,A_R,F,\mathrm{B}} = \\ &\{(K, I_{\mathrm{a}}^*(\mathrm{apply}(V_A, K, \mathrm{his}(F), A)) - I_{\mathrm{a}}^*(\mathrm{apply}(V_A, K, \mathrm{his}(F), A_R))) : \\ &\qquad K \in \mathrm{tuples}(V_A, F)\}\end{aligned}$$

(viii) the *limited-underlying tuple set list maximiser* neighbourhood function is

$$\begin{aligned}P_{P,A,A_R,F,\mathrm{B}}(B) = \{(J, X_{P,A,A_R,F,\mathrm{B}}(J)) : \\ (K,\cdot) \in B,\ w \in \mathrm{vars}(F) \cup V_A \setminus K,\ J = K \cup \{w\},\ |J^{\mathrm{C}}| \leq \mathrm{xmax}\}\end{aligned}$$

(ix) the *limited-underlying tuple set list maximiser* initial subset is

$$\begin{aligned}R_{P,A,A_R,\emptyset,\mathrm{B}} &= \{(\{w,u\}, X_{P,A,A_R,\emptyset,\mathrm{B}}(\{w,u\})) : \\ &\qquad w,u \in V_A,\ u \neq w,\ |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}\} \\ R_{P,A,A_R,F,\mathrm{B}} &= \{(\{w,u\}, X_{P,A,A_R,F,\mathrm{B}}(\{w,u\})) : \\ &\qquad w \in \mathrm{der}(F),\ u \in \mathrm{vars}(F) \cup V_A,\ u \neq w,\ |\{w,u\}^{\mathrm{C}}| \leq \mathrm{xmax}\}\end{aligned}$$

(x) the *contracted decrementing linear non-overlapping fuds list maximiser* is

$$\begin{aligned}Z_{P,A,A_R,F,\mathrm{n},-,K} = \\ &\mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{n},-,K}, N_{P,A,A_R,F,\mathrm{n},-,K}, \mathrm{top}(\mathrm{pmax}), R_{P,A,A_R,F,\mathrm{n},-,K})\end{aligned}$$

(xi) the *contracted decrementing linear non-overlapping fuds list maximiser* optimiser function is

$$\begin{aligned}X_{P,A,A_R,F,\mathrm{n},-,K} = \{(H, I_{\mathrm{csd}}^*((A, A_R, G))) : \\ H \in \mathcal{F}_{U_A,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}},\ G = \mathrm{depends}(F \cup H, \mathrm{der}(H))\}\end{aligned}$$

(xii) the *contracted decrementing linear non-overlapping fuds list maximiser* initial subset is

$$\begin{aligned}R_{P,A,A_R,F,\mathrm{n},-,K} = \{(\{M^{\mathrm{T}}\}, X_{P,A,A_R,F,\mathrm{n},-,K}(\{M^{\mathrm{T}}\})) : \\ Y \in \mathrm{B}(K),\ 2 \leq |Y| \leq \mathrm{mmax},\ M = \{J^{\mathrm{CS}\{\}} : J \in Y\}\}\end{aligned}$$

(xiii) the *contracted decrementing linear non-overlapping fuds list maximiser* neighbourhood function is

$$\begin{aligned}N_{P,A,A_R,F,\mathrm{n},-,K}(C) = \{(H \cup \{N^{\mathrm{T}}\}, X_{P,A,A_R,F,\mathrm{n},-,K}(H \cup \{N^{\mathrm{T}}\})) :\\ (H,\cdot) \in C,\ M = \mathrm{der}(H),\\ w \in M,\ |\{w\}^{\mathrm{C}}| > 2,\ Q \in \mathrm{decs}(\{w\}^{\mathrm{CS}\{\}}),\\ N = \{Q\} \cup \{\{u\}^{\mathrm{CS}\{\}} : u \in M,\ u \neq w\}\}\end{aligned}$$

(xiv) the *highest-layer limited-derived derived variables set list maximiser* is

$$Z_{P,A,A_R,F,\mathrm{D},\mathrm{d}} = \mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{D}}, P_{P,A,A_R,F,\mathrm{D}}, \mathrm{top}(\mathrm{omax}), R_{P,A,A_R,F,\mathrm{D},\mathrm{d}})$$

(xv) the *highest-layer limited-derived derived variables set list maximiser* initial subset is

$$\begin{aligned}R_{P,A,A_R,F,\mathrm{D},\mathrm{d}} \ = \ \{(J, X_{P,A,A_R,F,\mathrm{D}}(J)) :\\ w \in \mathrm{der}(F),\ u \in \mathrm{vars}(F) \setminus V_A \setminus \mathrm{vars}(\mathrm{depends}(F,\{w\})),\\ J = \{w,u\},\ |J^{\mathrm{C}}| \leq \mathrm{wmax}\}\end{aligned}$$

(xvi) the *limited-derived derived variables set list maximiser* optimiser function is

$$\begin{aligned}X_{P,A,A_R,F,\mathrm{D}} = \{(K, I^*_{\mathrm{csd}}((A, A_R, G))) :\\ K \subseteq \mathrm{vars}(F),\ K \neq \emptyset,\ G = \mathrm{depends}(F,K)\}\end{aligned}$$

(xvi) the *limited-derived derived variables set list maximiser* neighbourhood function is

$$\begin{aligned}P_{P,A,A_R,F,\mathrm{D}}(D) = \{(J, X_{P,A,A_R,F,\mathrm{D}}(J)) :\\ (K,\cdot) \in D,\ w \in \mathrm{vars}(F) \setminus V_A \setminus K,\\ J = K \cup \{w\},\ |J^{\mathrm{C}}| \leq \mathrm{wmax},\ \mathrm{der}(\mathrm{depends}(F,J)) = J\}\end{aligned}$$

where the *alignmenter* is such that $I^*_{\mathrm{a}}(A) \approx \mathrm{algn}(A)$, the *partition decrements* are

$$\mathrm{decs}(Q) := \{P : P \in \mathrm{parents}(Q),\ |P| = |Q| - 1\}$$

the *tuples* are defined

$$\mathrm{tuples}(V,F) := \{K : K \subseteq \mathrm{vars}(F) \cup V,\ (\mathrm{der}(F) \neq \emptyset \implies K \cap \mathrm{der}(F) \neq \emptyset)\}$$

el = elements, his = histograms $\in \mathcal{F} \to \mathrm{P}(\mathcal{A})$, and apply $\in \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{V}) \times \mathrm{P}(\mathcal{A}) \times \mathcal{A} \to \mathcal{A}$.

The *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer* is implemented

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd},\mathrm{D},\mathrm{F},\infty,\mathrm{q},P,\mathrm{d}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_{\emptyset}, 0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{d}})),\ \{D\} = Q
\end{aligned}$$

where (i) $D_{\emptyset} = \{(((\emptyset,\emptyset),\emptyset)\}$,(ii) the *summed shuffle content alignment valency-density computer* is

$$\begin{aligned}
&I^{*}_{\mathrm{Scsd}}((A,D)) = \\
&\quad \sum (I^{*}_{\mathrm{a}}(A * C * F^{\mathrm{T}}) - I^{*}_{\mathrm{a}}((A * C)_{R(A*C)} * F^{\mathrm{T}}))/I^{*}_{\mathrm{cvl}}(F) : (C,F) \in \mathrm{cont}(D)
\end{aligned}$$

(iii) the *highest-layer limited-models infinite-layer substrate fud decompositions tree searcher* is

$$Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{d}} = \mathrm{searchTreer}(\mathcal{D}_{\mathrm{F},\infty,U,V} \cap \mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\mathrm{q}}), P_{P,A,\mathrm{D},\mathrm{F},\mathrm{d}}, R_{P,A,\mathrm{D},\mathrm{F},\mathrm{d}})$$

(iv) the *highest-layer limited-models infinite-layer substrate fud decompositions tree searcher* neighbourhood function is

$$\begin{aligned}
&P_{P,A,\mathrm{D},\mathrm{F},\mathrm{d}}(D) = \{E : \\
&\quad (\cdot, S, G, L) \in \mathrm{maxd}(\mathrm{order}(D_{\mathbf{Q}\times\mathrm{S}\times\mathcal{X}^2}, \{(\mathrm{size}(B), S, G, L) : \\
&\qquad (L,Y) \in \mathrm{places}(D), \\
&\qquad R_L = \bigcup \mathrm{dom}(\mathrm{set}(L)),\ H_L = \bigcup \mathrm{ran}(\mathrm{set}(L)), \\
&\qquad (\cdot, F) = L_{|L|},\ W = \mathrm{der}(F), \\
&\qquad S \in W^{\mathrm{CS}} \setminus \mathrm{dom}(\mathrm{dom}(Y)), \\
&\qquad B = \mathrm{apply}(V_A, V_A, \mathrm{his}(H_L) \cup \{\{R_L \cup S\}^{\mathrm{U}}\}, A),\ \mathrm{size}(B) > 0, \\
&\qquad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,B,B_{R(B)},\mathrm{L},\mathrm{d}})), \\
&\qquad (K,a) \in \max(\mathrm{elements}(Z_{P,B,B_{R(B)},F_{\mathrm{L}},\mathrm{D},\mathrm{d}})),\ a > 0, \\
&\qquad G = \mathrm{depends}(F_{\mathrm{L}}, K)\})), \\
&\quad M = L \cup \{(|L|+1, (S,G))\}, \\
&\quad E = \mathrm{tree}(\mathrm{paths}(D) \setminus \{L\} \cup \{M\})\}
\end{aligned}$$

and (v) the *highest-layer limited-models infinite-layer substrate fud decompositions tree searcher* initial subset is

$$\begin{aligned}R_{P,A,\mathrm{D,F,d}} = \{\{((\emptyset,G),\emptyset)\} :\\ &G \in \mathrm{maxd}(\mathrm{order}(D_{\mathrm{F}},\{G :\\ &\quad F_{\mathrm{L}} \in \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_{R(A)},\mathrm{L,d}})),\\ &\quad (K,a) \in \max(\mathrm{elements}(Z_{P,A,A_{R(A)},F_{\mathrm{L}},\mathrm{D,d}})),\ a>0,\\ &\quad G = \mathrm{depends}(F_{\mathrm{L}},K)\})))\}\end{aligned}$$

5.1 Inducers and Compression

Now consider how *substrate structure alignment* and *inducers* relate to *derived history coders*.

The *fud decomposition minimum space specialising derived search function* for *history* $H \in \mathcal{H}_{U,X}$ is defined in terms of the *expanded specialising derived history coder*, $C_{\mathrm{G,T,H}}(T) \in \mathrm{coders}(\mathcal{H}_{U,X})$, as

$$Z_{\mathrm{D,F,P,m,G,T,H}}(H) \quad = \quad \{(D, -C_{\mathrm{G,T,H}}(D^{\mathrm{T}})^{\mathrm{s}}(H)) : D \in \mathcal{D}_{\mathrm{F},U,\mathrm{P}}\}$$

The *summed alignment valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, application also defines a *fud decomposition search function*, but restricted to the *limited-models non-overlapping fud decompositions*, $\mathcal{D}_{\mathrm{F},U,\mathrm{P}} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})) \subseteq \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$. Define the *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*

$$\begin{aligned}&Z_{\mathrm{D,F,P,n,q,Sd}}(H) =\\ &\quad\{(D, I^{*}_{\approx\mathbf{R}}(\textstyle\sum \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} : (C,F) \in \mathrm{cont}(D))) :\\ &\qquad\qquad D \in \mathcal{D}_{\mathrm{F},U,\mathrm{P}} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})),\ \mathrm{und}(D) \subseteq V,\\ &\qquad\qquad\qquad \forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A * C * F^{\mathrm{T}}) > 0)\} \ \cup\\ &\quad\{(D_{\mathrm{u}},0)\}\end{aligned}$$

where $V = \mathrm{vars}(H)$, $A = \mathrm{histogram}(H)$, the *unary fud decomposition* $D_{\mathrm{u}} = \{((\emptyset,\{T_{\mathrm{u}}\}),\emptyset)\}$, and the *unary transform* $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$. The addition of the *unary fud decomposition* ensures that the search is not empty, as it would otherwise be in the case, say, of *independent history*, $A = A^{\mathrm{X}}$. The domain of the *inducer search function* is a subset of the *minimum space search function*,

$$\mathrm{dom}(Z_{\mathrm{D,F,P,n,q,Sd}}(H)) \subseteq \mathrm{dom}(Z_{\mathrm{D,F,P,m,G,T,H}}(H)) = \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$$

The definition of the subset depends on the instance of the class of *limited-models*, $\mathcal{F}_{\mathrm{q}}$.

Although the *specialising derived substrate history coder*, $C_{G,V,T,H}$, is defined completely separately of the notions of *alignment* and *independence*, the properties of the minimum *coder space* are similar in many ways to the properties of the maximum *summed alignment valency-density* of the *tractable midising/idealising fud decomposition inducer*, $I'_{z,Sd,D,F,\infty,n,q}$, as is discussed below. Conjecture that in some cases, the maximum *decompositions* intersect,

$$|\mathrm{maxd}(Z_{D,F,P,n,q,Sd}(H)) \cap \mathrm{maxd}(Z_{D,F,P,m,G,T,H}(H))| \geq 0$$

More formally, conjecture that for all finite *systems* and finite *event identifier sets* there exists a class of *limited-models fuds* such that the *search functions* are positively correlated for uniform *history probability function*,

$$\begin{aligned}
&\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ (|\mathcal{H}_{U,X}| < \infty \implies \\
&\quad \exists \mathcal{F}_q \subset \mathcal{F}\ (\mathrm{covariance}(\mathcal{H}_{U,X} \times \{1/|\mathcal{H}_{U,X}|\}) \\
&\qquad\qquad (\mathrm{maxr} \circ Z_{D,F,P,m,G,T,H}, \mathrm{maxr} \circ Z_{D,F,P,n,q,Sd}) \geq 0))
\end{aligned}$$

The *fud decomposition minimum space specialising derived search function* for *history* $H \in \mathcal{H}_{U,X}$ is

$$Z_{D,F,P,m,G,T,H}(H) \quad = \quad \{(D, -C_{G,T,H}(D^{T})^{s}(H)) : D \in \mathcal{D}_{F,U,P}\}$$

It is maximised by finding the *fud decomposition* $D \in \mathcal{D}_{F,U,P}$ which minimises the *specialising derived substrate history coder space*, $C_{G,V,T,H}(D^{PVT})^{s}(H)$ where $V = \mathrm{vars}(H)$.

The minimisation of the *specialising derived substrate history coder space*, $C_{G,V,T,H}(D^{PVT})^{s}(H)$, occurs where (i) the *derived entropy* is low, (ii) the *possible derived volume* is small, (iii) the *underlying components* have high *entropy* and (iv) high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*. The minimisation of the *specialising derived substrate history coder space*, $C_{G,V,T,H}(D^{PVT})^{s}(H)$, also minimises the *specialising-canonical space difference*, $2C_{G,V,T,H}(D^{PVT})^{s}(H) - C^{s}_{H,V}(H) - C^{s}_{G,V}(H)$. *History probability functions* that have high *specialising degree of structure*, $\mathrm{structure}(U,X)(P, C_{G,T,H}(D^{T}))$, are expected to have encodings with these properties because the *degree of structure* is defined relative to the *canonical coders*.

The *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{D,F,P,n,q,Sd}(H)$, is maximised by searching for the *fud decomposition* $D \in \mathrm{maxd}(Z_{\mathrm{D,F,P,n,q,Sd}}(H)) \subset \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$, which maximises *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) = \sum \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} : (C, F) \in \mathrm{cont}(D)$, where $A = \mathrm{histogram}(H)$ and $()^{\mathrm{D}} \in \mathcal{D}_{\mathrm{F}} \to \mathcal{D}$.

In order to compare the properties of the *minimum space search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, and the *summed alignment valency-density search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, first consider the correlation between *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, and *derived entropy*, $\mathrm{entropy}(A * D^{\mathrm{T}})$. Clearly the *summed alignment valency-density* is correlated with its numerator, *summed alignment*, $\mathrm{algnSum}(U)(A, D^{\mathrm{D}}) = \sum \mathrm{algn}(A*C*F^{\mathrm{T}}) : (C, F) \in \mathrm{cont}(D)$,

$$\sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} \sim \sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})$$

Within the degree to which Stirling's approximation holds, the *contingent derived alignment* is approximately equal to the *sized* difference between the *contingent abstract entropy* and the *contingent derived entropy*,

$$\mathrm{algn}(A * C * F^{\mathrm{T}}) \approx z \times \mathrm{entropy}((A * C * F^{\mathrm{T}})^{\mathrm{X}}) - z \times \mathrm{entropy}(A * C * F^{\mathrm{T}})$$

So the *summed alignment* varies against the *summed derived entropy*

$$\sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}}) \sim - \sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{entropy}(A * C * F^{\mathrm{T}})$$

So *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, maximisation in the *summed alignment valency-density search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, tends to minimise the *summed derived entropy*,

$$\sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} \sim - \sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{entropy}(A * C * F^{\mathrm{T}})$$

As the cardinality of the *decomposition* tree increases the *summed derived entropy* decreases, because the *slices* are *aligned*, $\forall (C, F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A * C * F^{\mathrm{T}}) > 0)$.

The *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, also varies against the *derived entropy* of the *nullable transform*, $\mathrm{entropy}(A * D^{\mathrm{T}})$. If it so happens that the *decomposition* is also a *summation aligned decomposition*, $D^{\mathrm{D}} \in \mathcal{D}_{\Sigma}(A)$, then the *decomposition* is *contingently diagonalised*,

$\forall (C,T) \in \mathrm{cont}(D^{\mathrm{D}})$ $(\mathrm{diagonal}(A * C * T))$, and *contingently formal-abstract equivalent*, $\forall (C,T) \in \mathrm{cont}(D^{\mathrm{D}})$ $(A^{\mathrm{X}} * C * T = (A * C * T)^{\mathrm{X}})$. In section 'Summation aligned decomposition inducers', above, it is conjectured that the *content alignment* of a *summation aligned decomposition*, D^{D}, equals the *summation alignment*, $\mathrm{algn}(A * D^{\mathrm{T}}) - \mathrm{algn}(A^{\mathrm{X}} * D^{\mathrm{T}}) = \mathrm{algnSum}(U)(A, D^{\mathrm{D}})$. So, in the case of a *summation aligned decomposition*, the *summation alignment* varies with the *nullable transform derived alignment* and against the *nullable transform derived entropy*,

$$\begin{aligned} \sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}}) &\sim \mathrm{algn}(A * D^{\mathrm{T}}) \\ &\sim -\ \mathrm{entropy}(A * D^{\mathrm{T}}) \end{aligned}$$

Hence for a *summation aligned decomposition*, $D^{\mathrm{D}} \in \mathcal{D}_{\Sigma}(A)$,

$$\sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} \sim -\ \mathrm{entropy}(A * D^{\mathrm{T}})$$

The *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}(H)$, constrains the *fuds* to be *non-overlapping* and its maximisation tends to increase *slice midisation*. This is consistent with the *contingently formal-abstract equivalence* constraint of the *summation aligned decompositions*, $\mathcal{D}_{\Sigma}(A)$. The maximisation of the *slice alignment valency-density* tends to *contingent diagonalisation*, which is also consistent with *summation aligned decompositions*. Therefore conjecture that, even in the cases where the *decomposition* is not a *summation aligned decomposition*, $D^{\mathrm{D}} \notin \mathcal{D}_{\Sigma}(A)$, the maximisation of *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, tends to minimise the *derived entropy* of the *nullable transform*, $\mathrm{entropy}(A * D^{\mathrm{T}})$,

$$\sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} \sim -\ \mathrm{entropy}(A * D^{\mathrm{T}})$$

However, this anti-correlation between the *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, and the *derived entropy*, $\mathrm{entropy}(A * D^{\mathrm{T}})$, is not perfect. One of the reasons is that maximising the *contingent derived alignment*, $\mathrm{algn}(A * C * F^{\mathrm{T}})$, also tends to maximise the *sized contingent abstract entropy*, $z \times \mathrm{entropy}((A * C * F^{\mathrm{T}})^{\mathrm{X}})$. As discussed in section 'Maximum alignment', above, *maximum alignment*, $\mathrm{alignmentMaximum}(U)(W, z_{A*C})$ where $W = \mathrm{der}(F)$, is obtained when the *histogram* is *uniformly diagonalised*. In this case *maximum alignment* occurs when the *contingent derived histogram* is *diagonalised*, $\mathrm{diagonal}(A * C * F^{\mathrm{T}})$, and *uniform*, $|\mathrm{ran}(\mathrm{trim}(A * C * F^{\mathrm{T}}))| = 1$.

In the case of *regular derived variables* of *dimension* $m = |W|$ and *valency* $\{d\} = \{|U_w| : w \in W\}$, the *maximum alignment* approximates to

$$\text{alignmentMaximum}(U)(W, z_{A*C}) \ \approx \ z_{A*C}(m-1)\ln d$$

That is, *maximum alignment* increases weakly with increasing *diagonal*, d. *Entropy*, on the other hand, is minimised when the *histogram* is a *singleton*, $\text{entropy}(\{(\cdot, 1)\}) = 0$, so near *maximum alignment* the *entropy* no longer decreases, but instead increases as the *diagonal* becomes more *uniform*. However, in the *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\text{D,F,P,n,q,Sd}}(H)$, it is the *contingent alignment valency-density* that is maximimised, so the *maximum alignment valency-density* for *regular derived histogram* approximates to

$$\text{alignmentMaximum}(U)(W, z_{A*C})/d \ \approx \ z_{A*C}(m-1)(\ln d)/d$$

Thus the *diagonals* tend to shorten to *bivalent*, $d = 2$. Although the *diagonals* cannot shorten to *singletons*, the *derived entropy*, $\text{entropy}(A * D^{\text{T}})$, is lower in order to minimise *valency-capacity*, d. In any case, the *space* of the *specialising derived substrate history coder*, $C_{\text{G},V,\text{T,H}}$, is sometimes minimised where there are two *effective derived states*, $|(A * D^{\text{T}})^{\text{F}}| = 2$, depending on the *partition events space*, as described in section 'Derived history space', above.

In the case where the *decomposition*, $D \in \text{maxd}(Z_{\text{D,F,P,n,q,Sd}}(H)) \subset \mathcal{D}_{\text{F},U,\text{P}}$, is also a *summation aligned decomposition*, $D^{\text{D}} \in \mathcal{D}_{\Sigma}(A)$, then the *decomposition* is *contingently diagonalised*,

$$\forall (C, T) \in \text{cont}(D^{\text{D}})\ (\text{diagonal}(A * C * T))$$

and so there exists a *skeletal reduction*,

$$\exists D_{\text{s}}^{\text{D}} \in \text{reductions}(A, D^{\text{D}})\ (\text{skeletal}(A * D_{\text{s}}^{\text{DT}}))$$

The *summed derived entropy* is unchanged,

$$\sum_{(C,\{T\}) \in \text{cont}(D_{\text{s}})} \text{entropy}(A * C * T) \ = \ \sum_{(C,F) \in \text{cont}(D)} \text{entropy}(A * C * F^{\text{T}})$$

because the *off-diagonal derived states* of the *contingently diagonalised decomposition's fuds* are *ineffective* and do not contribute to the *derived entropy*. So, with respect to *summed derived entropy*, the *skeletal reductions* of a *summation aligned decomposition* of singleton *fuds* are equally well correlated with the *minimum space search function*, $Z_{\text{D,F,P,m,G,T,H}}$, than is the case without *reduction*.

In the special case of *full functional fud decomposition* $D_{\rm f} = \{((\emptyset,\{T_{\rm f}\}),\emptyset)\}$, where $T_{\rm f} = \{\{w\}^{\rm CS\{\}T} : w \in V\}^{\rm T}$, the *derived alignment* equals the *histogram alignment*, $\mathrm{algn}(A * D_{\rm f}^{\rm T}) = \mathrm{algn}(A)$, and the *derived entropy* equals the *histogram entropy*, $\mathrm{entropy}(A * D_{\rm f}^{\rm T}) = \mathrm{entropy}(A)$. Note that the *full functional decomposition*, $D_{\rm f}$, is not necessarily a *limited-models fud decomposition*, $\mathrm{trees}(\mathcal{S} \times \mathcal{F}_{\rm q})$, depending on the definition of *limited-models fuds*, $\mathcal{F}_{\rm q}$, and so the *full functional decomposition* may not be in the domain of the *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\rm D,F,P,n,q,Sd}(H)$. At the other extreme of *unary fud decomposition* $D_{\rm u} = \{((\emptyset,\{T_{\rm u}\}),\emptyset)\}$, where $T_{\rm u} = \{V^{\rm CS}\}^{\rm T}$, the *derived alignment* is zero, $\mathrm{algn}(A * D_{\rm u}^{\rm T}) = 0$, and the *derived entropy* is zero, $\mathrm{entropy}(A * D_{\rm u}^{\rm T}) = 0$.

To continue the comparison of the properties of the *minimum space search function*, $Z_{\rm D,F,P,m,G,T,H}$, and the *summed alignment valency-density search function*, $Z_{\rm D,F,P,n,q,Sd}$, now consider the correlation between *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A,D^{\rm D})$, and *component size cardinality relative entropy*, $\mathrm{entropyRelative}(A * D^{\rm T}, V^{\rm C} * D^{\rm T})$, where $D \in \mathrm{maxd}(Z_{\rm D,F,P,n,q,Sd}(H)) \subset \mathcal{D}_{{\rm F},U,{\rm P}}$, $H \in \mathcal{H}_{U,X}$, $A = \mathrm{histogram}(H)$, $V = \mathrm{vars}(H)$, $v = |V^{\rm C}|$ and $z = \mathrm{size}(A)$.

The minimisation of the *specialising derived substrate history coder space*, $C_{{\rm G},V,{\rm T,H}}(D^{\rm PVT})^{\rm s}(H)$, in the *minimum space search function*, $Z_{\rm D,F,P,m,G,T,H}(H)$, maximises the *component size cardinality relative entropy* so that high *counts* tend to be in low cardinality *components* and high cardinality *components* tend to have low *counts*. The *component size cardinality relative entropy* can be expressed in terms of *components*,

$$\begin{aligned}&\mathrm{entropyRelative}(A * D^{\rm T}, V^{\rm C} * D^{\rm T}) \;=\\ &\qquad\qquad \sum (\mathrm{size}(A * C^{\rm U})/z) \ln \frac{\mathrm{size}(A * C^{\rm U})/z}{|C|/v} : C \in D^{\rm P},\ \mathrm{size}(A * C^{\rm U}) > 0\end{aligned}$$

The *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\rm D,F,P,n,q,Sd}(H)$, is maximised by searching for the *fud decomposition* $D \in \mathrm{maxd}(Z_{\rm D,F,P,n,q,Sd}(H)) \subset \mathcal{D}_{{\rm F},U,{\rm P}}$, which maximises *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A,D^{\rm D}) = \sum \mathrm{algn}(A * C * F^{\rm T})/w_F^{1/m_F} : (C,F) \in \mathrm{cont}(D)$. The *limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer maximum function*, $\mathrm{maxr} \circ I'^{*}_{z,{\rm Sd,D,F},\infty,{\rm n,q}}$, is correlated with the *midisation pseudo-alignment substrate independent-formal transform inducer maximum function*, $\mathrm{maxr} \circ I^{*}_{z,\mathrm{m},\mathrm{fx}}$, which maximises the *midisation pseudo-alignment*. The *alignment valency-density* of a *contingent fud* of the *decomposition*, $\mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F}$, where $(C, F) \in \mathrm{cont}(D)$, varies with the *midisation pseudo-alignment*,

$$\begin{aligned}&\mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} \sim \\ &\quad \mathrm{algn}(A * C) - \mathrm{algn}(A * C * F^{\mathrm{T}} * F^{\mathrm{T}\dagger A*C}) - \mathrm{algn}((A * C * F^{\mathrm{T}})^{\mathrm{X}} * F^{\mathrm{T}\odot A*C})\end{aligned}$$

Maximisation of *midisation* tends to move *component alignments* from *off-diagonal states* to *on-diagonal states*. That is, if not *fully decomposed*, the *on-diagonal states* have high *component alignment*, $\mathrm{algn}(A * C * C') > 0$ where $(R', C') \in (F^{\mathrm{T}})^{-1}$ and $(A * C * F^{\mathrm{T}})_{R'} = \mathrm{size}(A * C * C') > 0$, while *off-diagonal states* have (i) low *component alignment*, $\mathrm{algn}(A * C * C') \approx 0$ where $(A * C * F^{\mathrm{T}})_{R'} \approx 0$, or (ii) are *independent*, $\mathrm{algn}(A * C * C') = 0 \iff A * C * C' = (A * C * C')^{\mathrm{X}}$, or (iii) are *ineffective*, $(A * C * F^{\mathrm{T}})_{R'} = 0$. If the *contingent derived histogram* is *diagonalised*, $\mathrm{diagonal}(A * C * F^{\mathrm{T}})$, then the *off-diagonal components* are necessarily *ineffective*, $(A * C * F^{\mathrm{T}})_{R'} = \mathrm{size}(A * C * C') = 0$.

Although the maximisation of the *midisation alignment* tends to minimise the *mid component size cardinality relative entropy*, $\mathrm{entropyRelative}(A * C * F^{\mathrm{T}}, C * F^{\mathrm{T}}) \approx 0$, the subsequent maximisation of the *idealisation alignment* in the *super-decomposition* tends to restore it. The increase in *relative entropy* was conjectured, in section 'Likely histograms', above, to occur where the *idealisation* is *integral*, because the logarithm of the cardinality of *integral independent histograms* varies against the *volume*. It is shown below that, regardless of whether the *idealisation* is *integral* or not, the *relative entropy* also increases during *decomposition* because of the tendency to *diagonalise* as the *midisation alignment* of the *slice* is maximised. The maximisation of *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, tends to maximise the *relative entropy* of the *nullable transform*,

$$\sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} \ \sim\ \mathrm{entropyRelative}(A * D^{\mathrm{T}}, V^{\mathrm{C}} * D^{\mathrm{T}})$$

Choose a node of the *decomposition* $((C, F), X) \in \mathrm{contingents}(D)$. In the case of a *sub-decomposition* $E \in \mathrm{subtrees}(D)$ which is such that the *component*, C, is a *component* of the *decomposition partition*, $C^{\mathrm{S}} \in E^{\mathrm{P}}$, the contribution of C to the *relative entropy* of E is

$$(\mathrm{size}(A * C)/z) \ln \frac{\mathrm{size}(A * C)/z}{|C|/v}$$

where $\text{size}(A * C) > 0$. In the *super-decomposition*, D, however, the *component*, C, is further *decomposed* so that in the case where the node has no children, $X = \emptyset$, the contribution to the *relative entropy* is instead

$$\sum (\text{size}(A * C * C')/z) \ln \frac{\text{size}(A * C * C')/z}{|C * C'|/v} : \\ (\cdot, C') \in (F^{\mathrm{T}})^{-1},\ \text{size}(A * C * C') > 0$$

The *fud*, F, is chosen such that the *midisation* is maximised and so in some cases the *relative entropy* increases,

$$\begin{aligned} &(\text{size}(A * C)/z) \ln \frac{\text{size}(A * C)/z}{|C|/v} \\ &\leq\ \sum (\text{size}(A * C * C')/z) \ln \frac{\text{size}(A * C * C')/z}{|C * C'|/v} : \\ &\qquad (\cdot, C') \in (F^{\mathrm{T}})^{-1},\ \text{size}(A * C * C') > 0 \end{aligned}$$

In the special case where the *decomposed slice* has uniform *sub-slice sizes*, $\forall (\cdot, C') \in (F^{\mathrm{T}})^{-1}\ (\text{size}(A * C * C') = \text{size}(A * C)/|(F^{\mathrm{T}})^{-1}|)$, then the *relative entropy* must increase if the *sub-components* do not all have the same cardinality, $\exists (\cdot, C') \in (F^{\mathrm{T}})^{-1}\ (|C'| \neq |C|/|(F^{\mathrm{T}})^{-1}|)$, because then

$$\exists (\cdot, C') \in (F^{\mathrm{T}})^{-1}\ \left(\frac{\text{size}(A * C * C')/z}{|C * C'|/v} \neq \frac{\text{size}(A * C)/z}{|C|/v} \right)$$

In the case where there are *ineffective components*, $\text{size}(A * C * C') = 0$, the *relative entropy* necessarily increases because the *effective underlying volume* decreases,

$$|\bigcup \{C * C' : (\cdot, C') \in (F^{\mathrm{T}})^{-1},\ \text{size}(A * C * C') > 0\}| < |C|$$

whereas the *size* of the *slice* $A * C$ is conserved,

$$\sum \text{size}(A * C * C') : (\cdot, C') \in (F^{\mathrm{T}})^{-1}\ =\ \text{size}(A * C)$$

Viewed as an optimising process, *decomposition* consists of successive alternating maximisations of *midisation pseudo-alignment* and then *idealisation alignment*. *Alignments* are removed along the *decomposition* paths so that a *fully decomposed decomposition* is *ideal*, $\text{ideal}(A, D^{\mathrm{T}})$. In the case where the node, $((C, F), X) \in \text{contingents}(D)$, has children, $X \neq \emptyset$, a child *slice* $A * C_2 \subset A * C$, where $(C_2, F_2) \in \text{dom}(X)$, which cannot contain the parent *alignment*, $\text{algn}(A * C_2 * F^{\mathrm{T}}) = 0$, again has its *fud*, F_2, chosen to maximise *midisation* from the remaining *alignment*, tending to *diagonalise* the *derived histogram*, $A * C_2 * F_2^{\mathrm{T}}$. Again, *ineffective components* and *independent components* are removed from subsequent *slices*, so the contribution to *relative entropy*, which was

$$(\mathrm{size}(A * C_2)/z) \ln \frac{\mathrm{size}(A * C_2)/z}{|C_2|/v} = (\mathrm{size}(A * C * C')/z) \ln \frac{\mathrm{size}(A * C * C')/z}{|C * C'|/v}$$

where $(\cdot, C') \in (F^{\mathrm{T}})^{-1}$ and $C * C' = C_2$, is such that now, in some cases, the *relative entropy* again increases,

$$\begin{aligned}
&(\mathrm{size}(A * C_2)/z) \ln \frac{\mathrm{size}(A * C_2)/z}{|C_2|/v} \\
&\leq \sum (\mathrm{size}(A * C_2 * C'')/z) \ln \frac{\mathrm{size}(A * C_2 * C'')/z}{|C_2 * C''|/v} : \\
&\qquad\qquad (\cdot, C'') \in (F_2^{\mathrm{T}})^{-1},\ \mathrm{size}(A * C_2 * C'') > 0
\end{aligned}$$

The *decomposition* tends to concentrate *events* into smaller and smaller *components* along the *decomposition* path as the *idealisation* is maximised, because of the asymmetric distribution of *events* in the *contingent fuds' slice partitions* as *midisation* is maximised, thus increasing the *component size cardinality relative entropy*. However, the correlation between the *minimum space search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, and the *summed alignment valency-density search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, in this respect, is not perfect. A reason is that when any *effective component*, $\mathrm{size}(A * C * C') > 0$, is *independent*, $A * C * C' = (A * C * C')^{\mathrm{X}}$, the *decomposition* is *ideal* with respect to it, and no further *slicing* of the *component* can take place, $(C * C')^{\mathrm{S}} \in D^{\mathrm{P}}$.

On the other hand, the *alignment* of the *components* varies weakly with the *component* cardinality if the *component volume* is a proper *cartesian sub-volume*, $|C * C'| \leq |(C * C')^{\mathrm{XF}}| < |V^{\mathrm{C}}|$. In the *on-diagonal* case where the *component alignments* may be high, $\mathrm{algn}(A*C*C') > 0$, the *maximum alignment* varies approximately with the logarithm of the *component* cardinality, $\ln |(C * C')^{\mathrm{XF}}|^{(n-1)/n}$. (See the section 'Maximum alignment', above, where it is shown that $\mathrm{alignmentMaximum}(U)(V, z) \approx z \ln v^{(n-1)/n}$.) In the *off-diagonal* case where the *component alignments* are low, $\mathrm{algn}(A * C * C') \approx 0$, the *expected alignment* varies with the logarithm of the *component* cardinality, $\ln |(C * C')^{\mathrm{XF}}|$. (See the section 'Minimum alignment', above, where it is conjectured that *expected alignment* varies approximately with the logarithm of the *volume*, $\ln v$ where $z \ll v$.) In both cases if the *component* is a proper *cartesian sub-volume*, $|(C * C')^{\mathrm{XF}}| < |V^{\mathrm{C}}|$, the *alignment*, $\mathrm{algn}(A * C * C')$, varies weakly with the *component* cardinality, $|C * C'|$. Hence the further *decomposition* of these *non-independent slices decomposes* larger cardinalities, tending to increase the *relative entropy*.

However, as shown in section 'Tractable decomposition inducers', above, the fraction of *derived histograms* of a given *derived geometry* that are *diagonalised*, $|\{A : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ \mathrm{diagonal}(A)\}|/|\mathcal{A}_{U,\mathrm{i},V,z}|$, increases as the *diagonals* shorten. Maximisation of *alignment valency-density* or *midisation* tends to shorten the *diagonals*, so the probability of *off-diagonal states* being completely *ineffective* is higher than would otherwise be the case. The probability of *effective independent off-diagonal derived states* that cannot be further *decomposed* is lower. In addition, maximisation of *valency-density* does not necessarily decrease the *derived dimension* where this does not lengthen the *diagonal*, so the fraction of the *derived volume* may be low, which tends to reduce the total *volume* of *effective underlying components*.

In the section above, which considers the correlation between the *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, and the *derived entropy*, $\mathrm{entropy}(A * D^{\mathrm{T}})$, it was shown that if it is the case that the *decomposition* $D \in \mathrm{maxd}(Z_{\mathrm{D,F,P,n,q,Sd}}(H))$ is a *summation aligned decomposition*, $D^{\mathrm{D}} \in \mathcal{D}_{\Sigma}(A)$, and also such that the *fuds* are all singletons, $\forall F \in \mathrm{fuds}(D)\ (|F| = 1)$, then the *summed derived entropy* of the *skeletal reduction* $D_{\mathrm{s}}^{\mathrm{D}} \in \mathrm{reductions}(A, D^{\mathrm{D}})$ is unchanged, $\sum \mathrm{entropy}(A * C * T) : (C, \{T\}) \in \mathrm{cont}(D_{\mathrm{s}}) = \sum \mathrm{entropy}(A * C * F^{\mathrm{T}}) : (C, F) \in \mathrm{cont}(D)$. While the *summation aligned decomposition*, D^{D}, may have high *component size cardinality relative entropy* because of the *contingent diagonalisation*, in the *skeletal reduction*, $D_{\mathrm{s}}^{\mathrm{D}}$, the *fuds* are *mono-derived-variate*, $\forall F \in \mathrm{fuds}(D_{\mathrm{s}})\ (|\mathrm{der}(F)| = 1)$, and therefore all of the *components* are *effective*, $\forall C \in D_{\mathrm{s}}^{\mathrm{P}}\ (\mathrm{size}(A * C) > 0)$, and there are fewer, larger *components*, $|D_{\mathrm{s}}^{\mathrm{P}}| < |D^{\mathrm{P}}|$. The *component* cardinalities, $|C|$, tend to be more correlated with the *component sizes*, $\mathrm{size}(A * C)$, and hence the *component size cardinality relative entropy* is lower in the *contingently reduced decomposition*, D_{s}. Thus, the *history space*, $C_{\mathrm{G},V,\mathrm{T,H}}(D_{\mathrm{s}}^{\mathrm{P}V\mathrm{T}})^{\mathrm{s}}(H)$, may be larger than that of the *decomposition*, $C_{\mathrm{G},V,\mathrm{T,H}}(D^{\mathrm{P}V\mathrm{T}})^{\mathrm{s}}(H)$. So the *skeletal reduction*, D_{s}, is less likely to be a maximum of the *minimum space search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$. The *skeletal reduction*, D_{s}, cannot be a maximum of the *summed alignment valency-density search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, because it has zero *summed alignment*, $\mathrm{algnValDensSum}(U)(A, D_{\mathrm{s}}^{\mathrm{D}}) = 0$.

In the special case of *full functional fud decomposition*, D_{f}, the *component size cardinality relative entropy* equals the *log volume* less the *histogram en-*

tropy

$$\text{entropyRelative}(A * D_{\mathrm{f}}^{\mathrm{T}}, V^{\mathrm{C}} * D_{\mathrm{f}}^{\mathrm{T}}) \;=\; \ln v - \text{entropy}(A)$$

At the other extreme of *unary fud decomposition*, D_{u}, the *component size cardinality relative entropy* is zero,

$$\text{entropyRelative}(A * D_{\mathrm{u}}^{\mathrm{T}}, V^{\mathrm{C}} * D_{\mathrm{u}}^{\mathrm{T}}) \;=\; 0$$

To continue the comparison of the properties of the *minimum space search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, and the *summed alignment valency-density search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, consider the correlation between *summed alignment valency-density*, $\text{algnValDensSum}(U)(A, D^{\mathrm{D}})$, and *expected component entropy*, $\text{entropyComponent}(A, D^{\mathrm{T}})$, where $D \in \text{maxd}(Z_{\mathrm{D,F,P,n,q,Sd}}(H)) \subset \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$, $H \in \mathcal{H}_{U,X}$, $A = \text{histogram}(H)$ and $z = \text{size}(A)$.

The minimisation of *specialising derived coder space*, $C_{\mathrm{G},V,\mathrm{T,H}}(D^{\mathrm{P}V\mathrm{T}})^{\mathrm{s}}(H)$, also minimises the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{T,H}}(D^{\mathrm{P}V\mathrm{T}})^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, which tends to maximise the *entropy* of the *underlying components*.

The *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}(H)$, is maximised when the *fud decomposition* $D \in \text{maxd}(Z_{\mathrm{D,F,P,n,q,Sd}}(H)) \subset \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$, is *ideal* with respect to the *histogram*, $\text{ideal}(A, D^{\mathrm{T}})$, because *contingent alignments* are successively removed along the *decomposition* paths. Thus the *components* of the *fully decomposed model* are *independent*,

$$\text{ideal}(A, D^{\mathrm{T}}) \implies \forall (\cdot, C) \in (D^{\mathrm{T}})^{-1}\ (A * C = (A * C)^{\mathrm{X}})$$

and so

$$\begin{aligned}
&\text{entropyComponent}(A, D^{\mathrm{T}}) \\
=\;& \text{expected}(\hat{A} * D^{\mathrm{T}})(\{(R, \text{entropy}(A * C)) : (R, C) \in (D^{\mathrm{T}})^{-1}\}) \\
=\;& \text{expected}(\hat{A} * D^{\mathrm{T}})(\{(R, \text{entropy}((A * C)^{\mathrm{X}})) : (R, C) \in (D^{\mathrm{T}})^{-1}\})
\end{aligned}$$

In section 'Minimum alignment', above, it is shown that the *entropy* of the *independent* of a *histogram* tends to be greater than the *entropy* of the *histogram*,

$$\text{entropy}(A * C) \leq \text{entropy}((A * C)^{\mathrm{X}})$$

So the *entropies* are expected to be higher than would be the case if the *decomposition* were not *ideal*. The maximisation of *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, tends to maximise the *expected component entropy* of the *nullable transform*,

$$\sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} \;\sim\; \mathrm{entropyComponent}(A, D^{\mathrm{T}})$$

The correlation between the *minimum space search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, and the *summed alignment valency-density search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, in this respect however, is not perfect. Maximum *entropy* is obtained when a *component* is *uniform*,

$$\mathrm{entropy}((A * C)^{\mathrm{X}}) \leq \mathrm{entropy}((A * C)^{\mathrm{XF}})$$

but an *independent component* that is not *uniform* cannot be further *decomposed* into smaller, more *uniform components* by the *fully decomposed decomposition*, $\forall C_1, C_2 \in \mathrm{dom}(\mathrm{cont}(D))\ ((A * C_1 = (A * C_1)^{\mathrm{X}}) \wedge (C_2 \neq C_1) \implies C_2 \not\subseteq C_1)$.

In the special case of *full functional fud decomposition*, D_{f}, the *expected component entropy* is zero,

$$\mathrm{entropyComponent}(A, D_{\mathrm{f}}^{\mathrm{T}}) \;=\; 0$$

At the other extreme of *unary fud decomposition*, D_{u}, the *expected component entropy* equals the *histogram entropy*,

$$\mathrm{entropyComponent}(A, D_{\mathrm{u}}^{\mathrm{T}}) \;=\; \mathrm{entropy}(A)$$

To continue the comparison of the properties of the *minimum space search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, and the *summed alignment valency-density search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, consider the correlation between *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, and *possible derived volume*, $w' = |(D^{\mathrm{T}})^{-1}|$, where $D \in \mathrm{maxd}(Z_{\mathrm{D,F,P,n,q,Sd}}(H)) \subset \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$ and $H \in \mathcal{H}_{U,X}$. As shown in section 'Decompositions', above, the *possible derived volume*, w', is bounded by the *possible derived volumes* of the individual *fuds*,

$$\begin{aligned} w' \;&\leq\; \sum_{F \in G} \left(|(F^{\mathrm{T}})^{-1}|\right) + 1 - |G| \\ &=\; \sum_{F \in G} w'_F + 1 - |G| \end{aligned}$$

where $G = \text{fuds}(D)$ and $w'_F = |(F^{\text{T}})^{-1}|$. In this case all of the *fuds* are *non-overlapping*, so

$$\begin{aligned} w' &\leq \sum_{F \in G} \left(|W_F^{\text{C}}|\right) + 1 - |G| \\ &= \sum_{F \in G} w_F + 1 - |G| \end{aligned}$$

where $W_F = \text{der}(F)$ and $w_F = |W_F^{\text{C}}|$.

The *fud decompositions* of the *alignment search*, $Z_{\text{D,F,P,n,q,Sd}}(H)$, are *limited-models fud decompositions*, $\mathcal{D}_{\text{F},U,\text{P}} \cap \text{trees}(\mathcal{S} \times (\mathcal{F}_{\text{n}} \cap \mathcal{F}_{\text{q}}))$ and hence the *fuds* are *limited-derived*, $\mathcal{F}_{\text{d}} \subset \mathcal{F}_{\text{q}} \subset \mathcal{F}$. In the case where there is a *maximum derived volume* limit $\text{wmax} \in \mathbf{N}_{\geq 4}$, then then *possible derived volume* is also explicitly limited,

$$w' \leq |G| \times \text{wmax} + 1 - |G|$$

The maximisation of the *midisation*, or its tractable counterpart, the *alignment valency-density*, of a *decomposition fud*, $(C, F) \in \text{cont}(D)$, tends to decrease the *valency-capacity*, w_F^{1/m_F}, shortening the *diagonal*. Therefore the *derived volume*, $w_F = w'_F$, also tends to decrease. The *possible derived volume* of the *decomposition*, w', is bounded by the sum of the *fuds' derived volumes*, $\sum_{F \in G} w_F$, and so it also tends to decrease.

The *possible derived volume*, w', equals the cardinality of the leaf *components* of the *decomposition*, $w' = |(D^{\text{T}})^{-1}| = |D^{\text{P}}|$, and so the *possible derived volume* does not depend on the cardinality of the *fuds*, $|G|$. For example, (i) a *multi-fud decomposition* D_1 such that $|\text{fuds}(D_1)| > 1$ and (ii) a singleton *decomposition* $D_2 = \{((\emptyset, \cdot), \emptyset)\}$, which are such that the *partitions* are equal, $D_1^{\text{P}} = D_2^{\text{P}}$, have the same *possible derived volume*, $w' = |D_1^{\text{P}}| = |D_2^{\text{P}}|$. The *possible derived volume* of a *multi-fud decomposition* does not necessarily increase with *fud* cardinality, $|G_1|$ where $G_1 = \text{fuds}(D_1)$, or *decomposition* path length, $|L|$ where $L \in \text{paths}(D_1)$.

So the *possible derived volume*, w', can be minimised by choosing a *decomposition* tree, $D \in \text{trees}(\mathcal{S} \times \mathcal{F})$, that minimises the cardinality of the *decomposition partition*, $|D^{\text{P}}|$. This is achieved in the *summed alignment valency-density search function*, $Z_{\text{D,F,P,n,q,Sd}}$, by removing *alignments* as quickly as possible along the *decomposition* paths, resulting in a few large cardinality *independent components*, $A * C = (A * C)^{\text{X}}$ where $(C, F) \in \text{cont}(D)$

and $C^{\mathrm{S}} \in D^{\mathrm{P}}$. The rate of *derived alignment* removal per additional *component* along the *decomposition* path is the *derived alignment* per *effective non-independent slice*,

$$\frac{\mathrm{algn}(A * C * F^{\mathrm{T}})}{|\{C' : (\cdot, C') \in (F^{\mathrm{T}})^{-1},\ \mathrm{algn}(A * C * C') > 0\}|} \geq \frac{\mathrm{algn}(A * C * F^{\mathrm{T}})}{|(A * C * F^{\mathrm{T}})^{\mathrm{F}}|}$$

where $(C, F) \in \mathrm{cont}(D)$. The *alignment effective-density* is less than or equal to the *alignment valency-density*,

$$\frac{\mathrm{algn}(A * C * F^{\mathrm{T}})}{|(A * C * F^{\mathrm{T}})^{\mathrm{F}}|} \leq \frac{\mathrm{algn}(A * C * F^{\mathrm{T}})}{w_F^{1/m_F}}$$

In the case of a *regular diagonalised derived histogram*, $\mathrm{diagonal}(A * C * F^{\mathrm{T}})$, of *valency* $\{d_F\} = \{|U_w| : w \in W_F\}$, which is such that $d_F = w_F^{1/m_F}$, the *alignment effective-density* equals the *alignment valency-density*,

$$\frac{\mathrm{algn}(A * C * F^{\mathrm{T}})}{|(A * C * F^{\mathrm{T}})^{\mathrm{F}}|} = \frac{\mathrm{algn}(A * C * F^{\mathrm{T}})}{d_F}$$

Thus the maximisation of the *fud's alignment valency-density* tends to maximise the rate of *derived alignment* removal per additional *component*. Along the *decomposition* path the parent *fud's alignment*, $\mathrm{algn}(A * C * F^{\mathrm{T}})$, is removed from the children *slices*, reducing the cardinality of *effective* children *slices* that are not *independent*, $|\{C' : (\cdot, C') \in (F^{\mathrm{T}})^{-1},\ \mathrm{algn}(A * C * C') > 0\}|$. By balancing the removal of *alignment* in the numerator with the creation of new *non-independent components* in the denominator, the maximisation tends to deep, narrow *decompositions* that minimise the *possible derived volume*, w'. The maximisation of *summed alignment valency-density*, $\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}})$, tends to minimise the *derived volume* of the *nullable transform*,

$$\sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} \sim 1/w'$$

The effect on the *possible derived volume*, w', of maximising the *alignment valency-density*, $\mathrm{algn}(A*C*F^{\mathrm{T}})/w_F^{1/m_F}$, is similar to the effect on the *derived entropy*, $\mathrm{entropy}(A*C*F^{\mathrm{T}})$. However, the tendency to large cardinality leaf *components*, $|C|$ where $C^{\mathrm{S}} \in D^{\mathrm{P}}$, which also have large *size*, $\mathrm{size}(A*C)$, may reduce (a) the *component size cardinality relative entropy* and (b) the *underlying component entropies*. To optimise these properties both (i) maximisation of *midisation*, and (ii) maximisation of *idealisation* by *decomposition*, are required.

While viewing the *decomposition* as constructed from a sequence of *sub-decompositions* in a computational process is a useful way of describing the removal of *alignments* along the *decomposition* paths, the search merely optimises the sum of the *contingent alignment valency-densities* for the whole *decomposition*. Therefore in some cases children *fuds* may have higher *alignment valency-densities* than their parents, $\mathrm{algn}(A*C_1*F_1^{\mathrm{T}})/w_{F_1}^{1/m_{F_1}} < \mathrm{algn}(A* C_2 * F_2^{\mathrm{T}})/w_{F_2}^{1/m_{F_2}}$, where $(C_1, F_1), (C_2, F_2) \in \mathrm{cont}(D)$ and $C_2 \subset C_1$, although the *sizes* necessarily decrease, $\mathrm{size}(A * C_1) > \mathrm{size}(A * C_2)$. However, the *maximum alignment*, approximately $z \ln v^{(n-1)/n}$, varies with *size*, and so it is often the case that the highest *alignments* are near the root of the *decomposition* tree.

In the special case of *full functional fud decomposition*, D_{f}, the *possible derived volume* is the *substrate volume*, $w' = v$, where $v = |V^{\mathrm{C}}|$. At the other extreme of *unary fud decomposition*, D_{u}, the *possible derived volume* is one, $w' = 1$.

Also, note that each non-root *fud* adds no more than $w_F - 1$ to the *possible derived volume*,

$$\begin{aligned} w' &\leq \sum_{F \in G} w_F + 1 - |G| \\ &= \sum_{F \in G} (w_F - 1) + 1 \end{aligned}$$

so there is a case for optimising the *alignment decremented-valency-density*,

$$\frac{\mathrm{algn}(A * C * F^{\mathrm{T}})}{(w_F^{1/m_F} - 1)}$$

The *alignment decremented-valency-density* has a slightly weaker *capacity* than the *alignment valency-density*, and so its maximisation would tend to longer *diagonals* and shallower, wider *decompositions*.

The comparisons above between the properties of the *minimum space search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, and the *summed alignment valency-density search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, provide evidence for the conjecture that the *search functions* are positively correlated for uniform *history probability func-*

tion,

$$\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ (|\mathcal{H}_{U,X}| < \infty \implies \exists \mathcal{F}_{\mathrm{q}} \subset \mathcal{F}\ (\mathrm{covariance}(P_{U,X})(\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,T,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,n,q,Sd}}) \geq 0))$$

where $P_{U,X} = \mathcal{H}_{U,X} \times \{1/|\mathcal{H}_{U,X}|\}$.

Although the *minimum space search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, searches for a *fud decomposition*, the *history* itself is encoded in a *specialising derived substrate history coder* parameterised only with the *transform* of the *fud decomposition*, $C_{\mathrm{G},V,\mathrm{T,H}}(D^{\mathrm{PVT}}) \in \mathrm{coders}(\mathcal{H}_{U,V,X})$.

In some cases, however, a *history* may be encoded in less *space* by means of a *specialising fud substrate history coder* if $C_{\mathrm{G},V,\mathrm{F,H}}(F)^{\mathrm{s}}(H) < C_{\mathrm{G},V,\mathrm{T,H}}(D^{\mathrm{T}})^{\mathrm{s}}(H)$, where $C_{\mathrm{G},V,\mathrm{F,H}}(F) \in \mathrm{coders}(\mathcal{H}_{U,V,X})$ and *fud* F is such that its *transform* equals that of the *fud decomposition*, $F^{\mathrm{T}} = D^{\mathrm{T}}$. It is shown in 'Derived history space', above, that in the *law-like* case where the *fud* has a *top transform*, $\exists T \in F\ (W_T = \mathrm{der}(F))$, the *space* difference is just the difference in *partitioned events space*,

$$\begin{aligned} C_{\mathrm{G},V,\mathrm{F,H}}(F)^{\mathrm{s}}(H) - C_{\mathrm{G},V,\mathrm{T,H}}(F^{\mathrm{T}})^{\mathrm{s}}(H) = \\ \sum_{T \in F} \mathrm{spaceEventsPartition}(A * \mathrm{depends}(F, V_T)^{\mathrm{T}}, T) \\ - \mathrm{spaceEventsPartition}(A, F^{\mathrm{T}}) \end{aligned}$$

which is the *size* scaled difference in *component size cardinality cross entropies*,

$$\begin{aligned} C_{\mathrm{G},V,\mathrm{F,H}}(F)^{\mathrm{s}}(H) - C_{\mathrm{G},V,\mathrm{T,H}}(F^{\mathrm{T}})^{\mathrm{s}}(H) = \\ z \times \mathrm{entropyCross}(A * F^{\mathrm{T}}, V^{\mathrm{C}} * F^{\mathrm{T}}) \\ - z \times \sum_{T \in F} \mathrm{entropyCross}(A * \mathrm{depends}(F, W_T)^{\mathrm{T}}, V_T^{\mathrm{C}} * T) \end{aligned}$$

where $V_T = \mathrm{und}(T)$ and $W_T = \mathrm{der}(T)$.

It was also conjectured that when the *specialising fud space*, $C_{\mathrm{G},V,\mathrm{F,H}}(F)^{\mathrm{s}}(H)$, is minimised, (i) the *derived entropy* decreases up the *layers*, (ii) the *possible derived volume* decreases up the *layers*, (iii) the *expected component entropy* increases up the *layers*, and (iv) the *component size cardinality cross entropy* increases up the *layers*. The optimisation of a *fud* without a *layer* limit may be made computable by building the *fud layer* by *layer*, minimising the *specialising space* at each step, until the addition of a *layer* fails to reduce the *specialising space.*

In some cases a *history* may be encoded in yet smaller *space* by means of a *specialising fud decomposition substrate history coder* if $C_{\mathrm{G},V,\mathrm{D,F,H}}(D)^{\mathrm{s}}(H) < C_{\mathrm{G},V,\mathrm{T,H}}(D^{\mathrm{T}})^{\mathrm{s}}(H)$, where $C_{\mathrm{G},V,\mathrm{D,F,H}}(D) \in \mathrm{coders}(\mathcal{H}_{U,V,X})$. This is because (i) a *specialising fud decomposition substrate history coder* allows different *slices* to have different *fuds* and (ii) complete coverage of the *substrate* is only required for whole paths, $\forall L \in \mathrm{paths}(D^*)\ (\bigcup_{(\cdot,(F,\cdot)) \in L} V_F = V)$. Therefore consider the *fud decomposition minimum space specialising fud decomposition search function* which is defined in terms of the *expanded specialising fud decomposition history coder* $C_{\mathrm{G,D,F,H}}(D) \in \mathrm{coders}(\mathcal{H}_{U,X})$,

$$Z_{\mathrm{D,F,P,m,G,D,F,H}}(H) \ = \ \{(D, -C_{\mathrm{G,D,F,H}}(D)^{\mathrm{s}}(H)) : D \in \mathcal{D}_{\mathrm{F},U,\mathrm{P}}\}$$

The *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, which is derived from the *summed alignment valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is conjectured to be positively correlated with the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D,F,P,m,G,D,F,H}}$, for uniform *history probability function*,

$$\begin{aligned}
&\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ (|\mathcal{H}_{U,X}| < \infty \implies \\
&\quad \exists \mathcal{F}_{\mathrm{q}} \subset \mathcal{F}\ (\mathrm{covariance}(P_{U,X}) \\
&\qquad\qquad (\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,D,F,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,n,q,Sd}}) \geq 0))
\end{aligned}$$

where $P_{U,X} = \mathcal{H}_{U,X} \times \{1/|\mathcal{H}_{U,X}|\}$. However, the correlation is less than or equal to the correlation with the *fud decomposition minimum space specialising derived search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$,

$$\begin{aligned}
&\mathrm{covariance}(P_{U,X})(\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,D,F,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,n,q,Sd}}) \\
\leq\ &\mathrm{covariance}(P_{U,X})(\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,T,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,n,q,Sd}})
\end{aligned}$$

The reason for this is that the *alignment search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, does not depend directly on the *transforms* of the *fuds* of the *decomposition*. The optimisation of the *alignment search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, does not necessarily minimise the *space* difference, $C_{\mathrm{G},V_F,\mathrm{F,H}}(F)^{\mathrm{s}}(H\%V_F) - C_{\mathrm{G},V_F,\mathrm{T,H}}(F^{\mathrm{T}})^{\mathrm{s}}(H\%V_F)$, whereas the optimisation of the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D,F,P,m,G,D,F,H}}$, tends to do so. For example, the *alignment search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, is independent of the *component cardinalities* of the *transforms*, so the optimisation is neutral with respect to the *component size cardinality cross entropies*,

$$\mathrm{entropyCross}(A * \mathrm{depends}(F, W_T)^{\mathrm{T}}, V_T^{\mathrm{C}} * T)$$

where $T \in F$ and $F \in \text{fuds}(D)$. Neither does the *alignment search function*, $Z_{\text{D,F,P,n,q,Sd}}$, depend directly on the *component sizes* of the *transforms* except in the case of a *top transform* $A * \text{depends}(F, W_T)^{\text{T}} = A * F^{\text{T}}$, where $W_T = W_F$. Even in this case, there is no dependency on the *component cardinalities* of the *top transform*, $\{(R, |C|) : (R, C) \in T\}$. The *alignment search function*, $Z_{\text{D,F,P,n,q,Sd}}$, does tend to maximise the *component size cardinality cross entropy*,

$$\text{entropyCross}(A * D^{\text{T}}, V^{\text{C}} * D^{\text{T}})$$

by *decomposing* high *on-diagonal component sizes*, but this is the *component size cardinality cross entropy* of the *transform* of the *decomposition*, D^{T}, rather than the *cross entropies* of the *transforms* of the *fuds* of the *decomposition*. There is no constraint that the *derived entropy* and the *possible derived volume* decreases up the *layers*, nor any constraint that the *expected component entropy* increases up the *layers*. There is no sense that the *fuds* are built *layer* by *layer* in sequence.

The *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\text{Scsd,D,F},\infty,\text{q},P,\text{d}}$, however, does impose a sequence on the search and other constraints that do not apply to the *tractable summed alignment valency-density decomposition inducer*, $I'_{z,\text{Sd,D,F},\infty,\text{n,q}}$, corresponding to the *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\text{D,F,P,n,q,Sd}}$. The *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer* is implemented in section 'Optimisation', above, as

$$\begin{aligned}
&I'^{*}_{z,\text{Scsd,D,F},\infty,\text{q},P,\text{d}}(A) = \\
&\quad \text{if}(Q \neq \emptyset, \{(D, I^{*}_{\text{Scsd}}((A, D)))\}, \{(D_{\emptyset}, 0)\}) : \\
&\qquad\qquad Q = \text{leaves}(\text{tree}(Z_{P,A,\text{D,F,d}})),\ \{D\} = Q
\end{aligned}$$

Define the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*

$$\begin{aligned}
&Z_{\text{D,F,P,q,d},P,\text{Scsd}}(H) = \\
&\quad \{(D, I^{*}_{\text{Scsd}}((A_H, D))) : Q = \text{leaves}(\text{tree}(Z_{P,A_H,\text{D,F,d}})),\ Q \neq \emptyset,\ \{D\} = Q\} \ \cup \\
&\quad \{(D_{\text{u}}, 0)\}
\end{aligned}$$

Corresponding to the conjecture that the *tractable limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\text{D,F,P,n,q,Sd}}$, and the *fud decomposition minimum space specialising* *derived search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, are positively correlated,

$$\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ (|\mathcal{H}_{U,X}| < \infty \implies \exists \mathcal{F}_{\mathrm{q}} \subset \mathcal{F}\ (\mathrm{covariance}(P_{U,X}) (\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,T,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,n,q,Sd}}) \geq 0))$$

conjecture that for all finite *systems* and finite *event identifier sets* there exists a tuple of parameters such that the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, is positively correlated with the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D,F,P,m,G,D,F,H}}$, for uniform *history probability function*,

$$\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ (|\mathcal{H}_{U,X}| < \infty \implies \exists P \in \mathcal{L}(\mathcal{X})\ (\mathrm{covariance}(P_{U,X}) (\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,D,F,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}) \geq 0))$$

Depending on the parameters, P, which imply a set of *limited-models*, $\mathcal{F}_{\mathrm{q}} \subset \mathcal{F}$, the domain of the application of the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer* is a subset of domain of the application of the *tractable summed alignment valency-density decomposition inducer*, $\mathrm{dom}(I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}(A)) \subseteq \mathrm{dom}(I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A))$, so, in some cases, the maximum *decompositions* of the *practicable search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, intersect with the maximum *decompositions* of the *tractable search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$,

$$|\mathrm{maxd}(Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}(H)) \cap \mathrm{maxd}(Z_{\mathrm{D,F,P,n,q,Sd}}(H))| \geq 0$$

and, in general, there is a high correlation

$$\mathrm{covariance}(P_{U,X})(\mathrm{maxr} \circ Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,n,q,Sd}})$$

So the relationships between the properties of the *fud decomposition minimum space specialising derived search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, and the properties of the *tractable search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, in the discussion above, also tend to hold for the relationships between the properties of the *minimum space search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, and the properties of the *practicable search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$.

In the case of the *practicable search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, however, the *fuds* of the *decomposition* are built *layer* by *layer*,

$$\forall (i, G) \in L\ (\mathrm{layer}(G, \mathrm{der}(G)) = i)$$

where $\{L\} = \text{paths}(\text{tree}(Z_{P,A,A_R,\mathrm{L},\mathrm{d}}))$ and the *highest-layer limited-layer limited-underlying limited-breadth fud tree searcher* is

$$Z_{P,A,A_R,\mathrm{L},\mathrm{d}} = \text{searchTreer}(\mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}, P_{P,A,A_R,\mathrm{L},\mathrm{d}}, \{\emptyset\})$$

So the properties of the *fuds* of the *decomposition* also depend on *layer*. In particular the *highest-layer fud tree searcher*, $Z_{P,A,A_R,\mathrm{L},\mathrm{d}}$, is constrained such that the *shuffle content alignment valency-density* of the *derived variables set* increases in each *layer*. Let the cumulative *fud* $F_{\{1\ldots i\}} = \bigcup_{j \in \{1\ldots i\}} F_j$, where $F_i = \{T : T \in F,\ \text{layer}(F, \text{der}(T)) = i\}$. Then consecutive *fuds*, $F_{\{1\ldots i\}}$ and $F_{\{1\ldots i+1\}}$ are in the path,

$$\{(i, F_{\{1\ldots i\}}), (i+1, F_{\{1\ldots i+1\}})\} \subseteq L$$

where $F = L_l$ and $l = |L|$. The *highest-layer limited-layer limited-underlying limited-breadth fud tree searcher* neighbourhood function is defined

$$\begin{aligned}
&P_{P,A,A_R,\mathrm{L},\mathrm{d}}(F) = \{G : \\
&\quad G \in P_{P,A,A_R,\mathrm{L}}(F), \\
&\quad (F \neq \emptyset \implies \text{maxr}(\text{el}(Z_{P,A,A_R,F,\mathrm{D},\mathrm{d}})) < \text{maxr}(\text{el}(Z_{P,A,A_R,G,\mathrm{D},\mathrm{d}})))\}
\end{aligned}$$

where the *highest-layer limited-derived derived variables set list maximiser* is

$$Z_{P,A,A_R,F,\mathrm{D},\mathrm{d}} = \text{maximiseLister}(X_{P,A,A_R,F,\mathrm{D}}, P_{P,A,A_R,F,\mathrm{D}}, \text{top}(\text{omax}), R_{P,A,A_R,F,\mathrm{D},\mathrm{d}})$$

so the neighbouring *layers* are such that

$$\text{maxr}(\text{el}(Z_{P,A,A_R,F_{\{1\ldots i\}},\mathrm{D},\mathrm{d}})) < \text{maxr}(\text{el}(Z_{P,A,A_R,F_{\{1\ldots i+1\}},\mathrm{D},\mathrm{d}}))$$

Let the *layer derived variables set* J_i found by the *highest-layer derived variables set list maximiser* be

$$J_i \in \text{maxd}(\text{el}(Z_{P,A,A_R,F_{\{1\ldots i\}},\mathrm{D},\mathrm{d}}))$$

The *layer derived variables set* is such that $J_i \subseteq \text{vars}(F_{\{1\ldots i\}}) \setminus V$. The *layer derived variables set*, J_i, is not necessarily equal to the *derived variables* of the whole *layer*, $\text{der}(F_{\{1\ldots i\}})$, but there must be an intersection, $J_i \cap \text{der}(F_{\{1\ldots i\}}) \neq \emptyset$. The *shuffle content alignment valency-density* of the *derived variables set* of a *layer* is $X_{P,A,A_R,F_{\{1\ldots i\}},\mathrm{D}}(J_i) = \text{maxr}(\text{el}(Z_{P,A,A_R,F_{\{1\ldots i\}},\mathrm{D},\mathrm{d}}))$ where the *limited-derived derived variables set list maximiser* optimiser function is

$$\begin{aligned}
X_{P,A,A_R,F,\mathrm{D}} = \{(K, I^*_{\mathrm{csd}}((A, A_R, G))) :& \\
&K \subseteq \text{vars}(F),\ K \neq \emptyset,\ G = \text{depends}(F, K)\}
\end{aligned}$$

the *shuffle content alignment valency-density computer* is

$$I^*_{\mathrm{csd}}((A, A_R, F)) = (I^*_{\mathrm{a}}(A * F^{\mathrm{T}}) - I^*_{\mathrm{a}}(A_R * F^{\mathrm{T}}))/I^*_{\mathrm{cvl}}(F)$$

and the *valency capacity computer* is

$$I^*_{\mathrm{cvl}}(F) := (I^*_{\approx\mathrm{pow}}((w, 1/m)) : W = \mathrm{der}(F),\ w = |W^{\mathrm{C}}|,\ m = |W|)$$

The *layer derived variables set*, J_i, intersects with the *derived variables* of the whole *layer*, $\mathrm{der}(F_{\{1\ldots i\}})$, so the *shuffle content alignment valency-density* varies with the *derived entropy*

$$\begin{aligned} X_{P,A,A_R,F_{\{1\ldots i\}},\mathrm{D}}(J_i) &\sim \mathrm{algn}(A * \mathrm{depends}(F_{\{1\ldots i\}}, J_i)^{\mathrm{T}}) \\ &\sim -z \times \mathrm{entropy}(A * \mathrm{depends}(F_{\{1\ldots i\}}, J_i)^{\mathrm{T}}) \\ &\sim -z \times \mathrm{entropy}(A * F^{\mathrm{T}}_{\{1\ldots i\}}) \end{aligned}$$

The *shuffle content alignment valency-density* of the *derived variables set* increases in each *layer*,

$$X_{P,A,A_R,F_{\{1\ldots i\}},\mathrm{D}}(J_i) < X_{P,A,A_R,F_{\{1\ldots i+1\}},\mathrm{D}}(J_{i+1})$$

so, in general, the *derived entropy* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (\mathrm{entropy}(A * F^{\mathrm{T}}_{\{1\ldots i\}}) < \mathrm{entropy}(A * F^{\mathrm{T}}_{\{1\ldots i-1\}}))$$

which is a property of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)$, and hence a property of the *specialising fud decomposition substrate history coder*, $C_{\mathrm{G},V,\mathrm{D},\mathrm{F},\mathrm{H}}(D^V)$.

Note that it is the *shuffle content alignment valency-density* that is maximised, rather than the *shuffle content alignment*. The *shuffle content alignment valency-density* varies against the *derived volume*,

$$\begin{aligned} X_{P,A,A_R,F_{\{1\ldots i\}},\mathrm{D}}(J_i) &\sim 1/|J_i^{\mathrm{C}}|^{1/|J_i|} \\ &\sim 1/|J_i^{\mathrm{C}}| \end{aligned}$$

and so, in general, the *derived volume* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (|W_i^{\mathrm{C}}| < |W_{i-1}^{\mathrm{C}}|)$$

where $W_i = \mathrm{der}(G_i)$, $G = \mathrm{depends}(F_{\mathrm{L}}, K)$, $K \in \mathrm{maxd}(\mathrm{elements}(Z_{P,A,A_R,F_{\mathrm{L}},\mathrm{D},\mathrm{d}}))$ and $\{F_{\mathrm{L}}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,A_R,\mathrm{L},\mathrm{d}}))$.

That is, the *derived entropy* and the *derived volume* tend to decrease up the *layers* of the *fuds* of the *decompositions* in both the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, and the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D,F,P,m,G,D,F,H}}$.

Another correlation that is a consequence of the *layer* by *layer* search in the *highest-layer limited-layer limited-underlying limited-breadth fud tree searcher*, $Z_{P,A,A_R,\mathrm{L,d}}$, arises in the *contracted decrementing linear non-overlapping fuds list maximiser*,

$$\begin{aligned}&Z_{P,A,A_R,F,\mathrm{n},-,K} = \\ &\quad \mathrm{maximiseLister}(X_{P,A,A_R,F,\mathrm{n},-,K}, N_{P,A,A_R,F,\mathrm{n},-,K}, \mathrm{top}(\mathrm{pmax}), R_{P,A,A_R,F,\mathrm{n},-,K})\end{aligned}$$

which *value rolls* a *tuple*, K, from the *limited-underlying tuple set list maximiser*, $Z_{P,A,A_R,F,\mathrm{B}}$, in the *limited-layer limited-underlying limited-breadth fud tree searcher* neighbourhood function,

$$\begin{aligned}P_{P,A,A_R,\mathrm{L}}(F) = \{G :\ &\\ G = F \cup \{T :\ & K \in \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{B}})),\\ & H \in \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,A,A_R,F,\mathrm{n},-,K})),\\ & w \in \mathrm{der}(H),\ I = \mathrm{depends}(\mathrm{explode}(H),\{w\}),\ T = I^{\mathrm{TPT}}\},\\ \mathrm{layer}(G,\mathrm{der}(G)) \leq \mathrm{lmax}\}&\end{aligned}$$

The *contracted decrementing linear non-overlapping fuds list maximiser* optimiser function is

$$\begin{aligned}X_{P,A,A_R,F,\mathrm{n},-,K} = \{(H, I^*_{\mathrm{csd}}((A,A_R,G))) :&\\ H \in \mathcal{F}_{U_A,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}},\ & G = \mathrm{depends}(F \cup H, \mathrm{der}(H))\}\end{aligned}$$

The *contracted decrementing linear non-overlapping fuds list maximiser* initial subset,

$$\begin{aligned}R_{P,A,A_R,F,\mathrm{n},-,K} = \{(\{M^{\mathrm{T}}\}, X_{P,A,A_R,F,\mathrm{n},-,K}(\{M^{\mathrm{T}}\})) :&\\ Y \in \mathrm{B}(K),\ 2 \leq |Y| \leq \mathrm{mmax},\ & M = \{J^{\mathrm{CS}\{\}} : J \in Y\}\}\end{aligned}$$

partitions the *tuple*, $Y \in \mathrm{B}(K)$, maximising the *shuffle content alignment valency-density*, $X_{P,A,A_R,F,\mathrm{n},-,K}(G_Y)$, of the *fud* of *self transforms* on the components of the *tuple* partition, $G_Y = \{J^{\mathrm{CS}\{\}\mathrm{T}} : J \in Y\}$. The maximisation of the *derived alignment* between the *derived variables* of the components, $\mathrm{algn}(A_F * G_Y^{\mathrm{T}})$, where $A_F = A * \prod_{(X,\cdot) \in F} X$, tends to minimise the *underlying* *alignments* within the components, $\sum_{J \in Y} \text{algn}(A_F\%J)$. So in some cases the intra-component *alignments* are less than the inter-component *alignments*, $\text{algn}(A_F\%\{v_1, v_2\}) < \text{algn}(A_F\%\{v_1, v_3\})$, where $v_1, v_2 \in J$ and $v_3 \in K \setminus J$.

The *fud*, H, resulting from the *contracted decrementing linear non-overlapping fuds list maximiser*, $Z_{P,A,A_R,F,\text{n},-,K}$, is *value rolled* from the *fud*, G_Y, of *self transforms*, *partitioning* the *underlying histogram*, $A_F\%K$. Conjecture that *alignments* of the *components* of this *partition* vary against the *derived alignment*,

$$\sum_{(\cdot,C) \in (H^{\text{T}})^{-1}} \text{algn}(A_F\%K * C) \;\;\sim\;\; -\text{algn}(A_F * H^{\text{T}})$$

Then the *fud*, H, is *exploded* into a *fud* of *transforms* each corresponding to a component of the *tuple* partition. Let $w \in \text{der}(H)$ correspond to the component J. Then $\text{der}(T) = \{w\}$ and $\text{und}(T) = J$, where $T = \text{depends}(\text{explode}(H), \{w\})^{\text{TPT}}$. The *layer*, F_i, is the set of these *transforms*, $F_{\{1\ldots i\}} = P_{P,A,A_R,\text{L}}(F_{\{1\ldots i-1\}})$. The differential between the intra-component *alignments* and the inter-component *alignments* in each *layer* implies that the *layer derived entropy* varies against the *layer expected component entropy*,

$$\begin{aligned}
&z \times \text{entropy}(A * F^{\text{T}}_{\{1\ldots i\}}) \\
&\quad\sim\;\; -\text{algn}(A * F^{\text{T}}_{\{1\ldots i\}}) \\
&\quad\sim\;\; \sum_{T \in F_i} \sum_{(\cdot,C) \in T^{-1}} \text{algn}(A * F^{\text{T}}_{\{1\ldots i-1\}}\%V_T * C) \\
&\quad\sim\;\; -\sum_{T \in F_i} \sum_{(R,C) \in T^{-1}} (A * F^{\text{T}}_{\{1\ldots i-1\}} * T)_R \times \text{entropy}(A * F^{\text{T}}_{\{1\ldots i-1\}}\%V_T * C) \\
&\quad\sim\;\; -\sum_{(R,C) \in (F_i^{\text{T}})^{-1}} (A * F^{\text{T}}_{\{1\ldots i\}})_R \times \text{entropy}(A * F^{\text{T}}_{\{1\ldots i-1\}} * C) \\
&\quad=\;\; -\text{entropyComponent}(A * F^{\text{T}}_{\{1\ldots i-1\}}, F_i^{\text{T}})
\end{aligned}$$

It has already been shown that, in general, the *derived entropy* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (\text{entropy}(A * F^{\text{T}}_{\{1\ldots i\}}) < \text{entropy}(A * F^{\text{T}}_{\{1\ldots i-1\}}))$$

so, in general, the *expected component entropy* increases up the *layers*,

$$\begin{aligned}
&\forall i \in \{2 \ldots l\} \\
&\quad(\text{entropyComponent}(A, F^{\text{T}}_{\{1\ldots i\}}) \;>\; \text{entropyComponent}(A, F^{\text{T}}_{\{1\ldots i-1\}}))
\end{aligned}$$

because the *layer derived entropy* varies against the *layer expected component entropy*. Again, this is a property of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)$, and hence a property of the *specialising fud decomposition substrate history coder*, $C_{\mathrm{G},V,\mathrm{D},\mathrm{F},\mathrm{H}}(D^V)$. That is, the *expected component entropy* tends to increase up the *layers* of the *fuds* of the *decompositions* in both the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*, $Z_{\mathrm{D},\mathrm{F},\mathrm{P},\mathrm{q},\mathrm{d},P,\mathrm{Scsd}}$, and the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D},\mathrm{F},\mathrm{P},\mathrm{m},\mathrm{G},\mathrm{D},\mathrm{F},\mathrm{H}}$.

Although the *alignments* within the *tuple* components tend to be less than the *alignments* between components, $\mathrm{algn}(A_F\%\{v_1,v_2\}) < \mathrm{algn}(A_F\%\{v_1,v_3\})$, the *tuple* is found in the *limited-underlying tuple set list maximiser*, $Z_{P,A,A_R,F,\mathrm{B}}$, by maximising the *shuffle content alignment* of the whole *tuple*, $X_{P,A,A_R,F,\mathrm{B}}(K) \sim \mathrm{algn}(A_F\%K)$. So the intra-component *alignments* are only small relative to the inter-component *alignments* and are not necessarily small absolutely, $\mathrm{algn}(A_F\%J) \geq 0$.

The last property of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)$, to be considered is the increase of the *component size cardinality cross entropy* up the *layers*. This also arises in the *contracted decrementing linear non-overlapping fuds list maximiser*, $Z_{P,A,A_R,F,\mathrm{n},-,K}$. If the component *histogram*, $A_F\%J$, where $J \in Y$ and $Y \in \mathrm{B}(K)$, is not *uniform*, $|\mathrm{ran}(A_F\%J)| \neq 1$, which is the case if component *histogram* is not *independent*, $\mathrm{algn}(A_F\%J) > 0$, then the initial *self transform*, $J^{\mathrm{CS}\{\}\mathrm{T}}$, of the component *histogram* has non-zero *component size cardinality cross entropy*,

$$\begin{aligned}
\mathrm{entropyCross}(A_F * J^{\mathrm{CS}\{\}\mathrm{T}}, J^{\mathrm{C}} * J^{\mathrm{CS}\{\}\mathrm{T}}) &= -\sum_{S \in (A_F\%J)^{\mathrm{FS}}} (\hat{A}_F\%J)_s \ln \frac{1}{|J^{\mathrm{C}}|} \\
&> 0
\end{aligned}$$

This is because the *component cardinalities* within the *layer*, $\{|C| : (\cdot,C) \in (J^{\mathrm{CS}\{\}\mathrm{T}})^{-1}\} = \{1\}$, are uniform, but the *component sizes* are not, $|\{\mathrm{size}(A_F * C) : (\cdot,C) \in (J^{\mathrm{CS}\{\}\mathrm{T}})^{-1}\}| > 1$.

The *decrementing fuds maximiser*, $Z_{P,A,A_R,F,\mathrm{n},-,K}$, *value rolls* one *value* in each step, so the largest *counts* of the component *histogram*, $A_F\%J$, tend to *roll* together at the beginning of the *derived diagonal*. The *diagonal* is constructed approximately in a sequence tending to minimise *component cardinalities* at the beginning of the *diagonal* and maximise *component cardinalities* at the end of the *diagonal* and *off-diagonal*. Maximisation of the *derived* *alignment* tends to *uniform counts* along the *diagonal*. Thus the *component size cardinality cross entropy* increases as the *diagonal* shortens below the *volume* of a component *histogram*, $|W^{\mathrm{C}}|^{1/|W|} < |J^{\mathrm{C}}|$, where $W = \mathrm{der}(H)$. Conjecture that *component size cardinality cross entropy* varies with the *derived alignment valency-density*,

$$\begin{aligned} z \times \mathrm{entropyCross}(A_F * H^{\mathrm{T}}, K^{\mathrm{C}} * H^{\mathrm{T}}) &= -\sum_{(R,C) \in (H^{\mathrm{T}})^{-1}} (A_F * H^{\mathrm{T}})_R \ln \frac{|C|}{|K^{\mathrm{C}}|} \\ &\sim \mathrm{algn}(A_F * H^{\mathrm{T}})/|W^{\mathrm{C}}|^{1/|W|} \end{aligned}$$

So the *layer derived entropy* varies against the *layer component size cardinality cross entropy*,

$$\begin{aligned} & z \times \mathrm{entropy}(A * F^{\mathrm{T}}_{\{1 \ldots i\}}) \\ & \quad \sim \ -\mathrm{algn}(A * F^{\mathrm{T}}_{\{1 \ldots i\}}) \\ & \quad \sim \ -\left(-\frac{1}{|F_i|} \sum_{T \in F_i} \sum_{(R,C) \in T^{-1}} (A * F^{\mathrm{T}}_{\{1 \ldots i-1\}} * T)_R \ln \frac{|C|}{|V^{\mathrm{C}}_T|} \right) \\ & \quad \sim \ -\left(-\sum_{(R,C) \in (F^{\mathrm{T}}_i)^{-1}} (A * F^{\mathrm{T}}_{\{1 \ldots i\}})_R \ln \frac{|C|}{|V^{\mathrm{C}}_i|} \right) \\ & \quad \sim \ -z \times \mathrm{entropyCross}(A * F^{\mathrm{T}}_{\{1 \ldots i\}}, V^{\mathrm{C}}_i * F^{\mathrm{T}}_i) \end{aligned}$$

It has already been shown that, in general, the *derived entropy* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (\mathrm{entropy}(A * F^{\mathrm{T}}_{\{1 \ldots i\}}) < \mathrm{entropy}(A * F^{\mathrm{T}}_{\{1 \ldots i-1\}}))$$

so, in general, the *component size cardinality cross entropy* increases up the *layers*,

$$\begin{aligned} & \forall i \in \{2 \ldots l\} \\ & \quad (\mathrm{entropyCross}(A * F^{\mathrm{T}}_{\{1 \ldots i\}}, V^{\mathrm{C}} * F^{\mathrm{T}}_{\{1 \ldots i\}}) > \\ & \quad\quad \mathrm{entropyCross}(A * F^{\mathrm{T}}_{\{1 \ldots i-1\}}, V^{\mathrm{C}} * F^{\mathrm{T}}_{\{1 \ldots i-1\}})) \end{aligned}$$

because the *layer derived entropy* varies against the *layer component size cardinality cross entropy*. Again, this is a property of the *specialising fud substrate history coder*, $C_{\mathrm{G,V,F,H}}(F)$, and hence a property of the *specialising fud decomposition substrate history coder*, $C_{\mathrm{G,V,D,F,H}}(D^V)$. That is, the *component size cardinality cross entropy* tends to increase up the *layers* of the *fuds* of the *decompositions* in both the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, and the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D,F,P,m,G,D,F,H}}$.

As well as retaining much of the correlation with the *fud decomposition minimum space specialising derived search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, via the *tractable search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, is additionally correlated with the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D,F,P,m,G,D,F,H}}$. This is the case even though the additional constraints implemented in the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}$, were imposed purely for practicable reasons.

The comparisons above between the properties of the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D,F,P,m,G,D,F,H}}$, and the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, provide evidence for the conjecture that the *search functions* are positively correlated for uniform *history probability function*,

$$\begin{aligned}
&\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ (|\mathcal{H}_{U,X}| < \infty \implies \\
&\quad \exists P \in \mathcal{L}(\mathcal{X})\ (\mathrm{covariance}(P_{U,X}) \\
&\qquad\qquad (\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,D,F,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}) \geq 0))
\end{aligned}$$

5.2 Artificial neural networks and Compression

The discussion above compares (i) the properties of the *tractable* and *practicable alignment inducers* to (ii) the properties of the *specialising derived history coder* and the *specialising fud decomposition history coder*. Now consider how artificial neural networks relate to the *specialising fud history coder*.

The *fud minimum space specialising fud search function* for *history* $H \in \mathcal{H}_{U,X}$ is defined in terms of the *expanded specialising fud history coder*, $C_{\mathrm{G,F,H}}(F) \in \mathrm{coders}(\mathcal{H}_{U,X})$, as

$$Z_{\mathrm{F,P,m,G,F,H}}(H) \quad = \quad \{(F, -C_{\mathrm{G,F,H}}(F)^{\mathrm{s}}(H)) : F \in \mathcal{F}_{U,\mathrm{P}}\}$$

To construct a search function for a neural network, first consider how a neuron may be represented in a *transform*. Section 'Transforms', above, has an example of a *transform* defined by a real valued function that represents a perceptron, $T = (V, w, f_\sigma(Q))$, where the *dimension* is $n = |V|$ and the function $f_\sigma(Q) \in \mathbf{R}^n :\to \mathbf{R}$ is parameterised by (i) some differentiable function $\sigma \in \mathbf{R} :\to \mathbf{R}$, called the activation function, and (ii) a vector of weights, $Q \in \mathbf{R}^{n+1}$, and is defined

$$f_\sigma(Q)(S) := \sigma(\sum_{i \in \{1 \ldots n\}} Q_i S_i \ + \ Q_{n+1})$$

Usually the activation function is such that it has a positive gradient everywhere, $\forall x \in \mathbf{R}\ (\mathrm{d}(\sigma)(x) \geq 0)$, where $\mathrm{d} \in (\mathbf{R} \to \mathbf{R}) \to (\mathbf{R} \to \mathbf{R})$ is defined $\mathrm{d}(F) := \{(x, dF(x)/dx) : x \in \mathrm{dom}(F)\}$.

The function composition of artificial neural networks may be represented by *fuds* of these *transforms*. Define nets as a subset of the set of lists of tuples of the graph and real weights,

$$\mathrm{nets} \ := \ \{G : G \in \mathcal{L}(\mathrm{P}(\mathcal{V}) \times \mathcal{V} \times \mathcal{L}(\mathbf{R})),\ \forall (\cdot, (V, \cdot, Q)) \in G\ (|Q| = |V| + 1)\}$$

Define the graph, $\mathrm{graph} \in \mathrm{nets} \to \mathcal{L}(\mathrm{P}(\mathcal{V}) \times \mathcal{V})$ as

$$\mathrm{graph}(G) \ := \ \{(i, (V, w)) : (i, (V, w, \cdot)) \in G\}$$

Define the real weights, $\mathrm{weights} \in \mathrm{nets} \to \mathcal{L}(\mathbf{R})$ as

$$\mathrm{weights}(G) \ := \ \mathrm{concat}(\{(i, Q) : (i, (\cdot, \cdot, Q)) \in G\})$$

Define the set of *transforms*, $\mathrm{fud}(\sigma) \in \mathrm{nets} \to \mathrm{P}(\mathcal{T}_{\mathrm{f}})$ as

$$\begin{aligned} &\mathrm{fud}(\sigma)(G) \ := \\ &\quad \{(\{S^V \cup \{(w, \sigma(\sum_{i \in \{1 \ldots n\}} Q_i S_i \ + \ Q_{n+1}))\} : S \in \mathbf{R}^n\} \times \{1\}, \{w\}) : \\ &\qquad\qquad\qquad\qquad\qquad\qquad (\cdot, (V, w, Q)) \in G,\ n = |V|\} \end{aligned}$$

The construction of a coordinate from a *state* is defined $()^{[]} \in \mathcal{S} \to \mathcal{L}(\mathcal{W})$ as

$$S^{[]} \ := \ \{(i, u) : ((v, u), i) \in \mathrm{order}(D_{\mathcal{V} \times \mathcal{W}}, S)\}$$

where $D_{\mathcal{V} \times \mathcal{W}}$ is an *order* on the *variables* and *values*. The converse function to construct a *state* from a coordinate $()^V \in \mathcal{L}(\mathcal{W}) \to \mathcal{S}$ is

$$S^V \ := \ \{(v, S_i) : (v, i) \in \mathrm{order}(D_{\mathcal{V}}, V)\}$$

Let the *neural net substrate fud set* $\mathcal{F}_{\infty,U,V,\sigma}$ be a subset of the *infinite-layer substrate fud set*, $\mathcal{F}_{\infty,U,V,\sigma} = \mathcal{F}_{\infty,U,V} \cap (\mathrm{fud}(\sigma) \circ \mathrm{nets})$.

An example of a *neural net substrate fud* $F \in \mathcal{F}_{\infty,U,V,\sigma}$ has $l = \text{layer}(F, \text{der}(F))$ *layers* of fixed *breadth* equal to the *underlying dimension*, $\forall i \in \{1 \ldots l\}\ (|F_i| = n)$ where $n = |V|$ and $F_i = \{T : T \in F,\ \text{layer}(F, \text{der}(T)) = i\}$, such that the *underlying* of each *transform* is the *derived* of the *layer* below, $\forall T \in F_1\ (\text{und}(T) = V)$ and $\forall i \in \{2 \ldots l\}\ \forall T \in F_i\ (\text{und}(T) = \text{der}(F_{i-1}))$.

The optimisation of artificial neural networks can be divided into unsupervised and supervised types. In the supervised case there is additional *knowledge*. First, there exists an *unknown distribution histogram* E from which the *known sample histogram*, A, is *drawn*, $A < E$. Secondly, the *substrate* can be partitioned into query *variables* $K \subset V$ and label *variables*, $V \setminus K$, such that the *distribution histogram*, E, is *causal* between the query *variables* and the label *variables*,

$$\text{split}(K, E^{\text{FS}}) \ \ \in \ \ K^{\text{CS}} \to (V \setminus K)^{\text{CS}}$$

and so the *sample histogram*, A, is also *causal*,

$$\text{split}(K, A^{\text{FS}}) \ \ \in \ \ K^{\text{CS}} \to (V \setminus K)^{\text{CS}}$$

That is, in the supervised case, there is a functional relation such that there is exactly one label *state* for every *effective* query *state*. In an optimisation, a *fud* $F \in \mathcal{F}_{\infty,U,K,\sigma}$ has its *underlying variables* restricted to the query *variables*, $\text{und}(F) \subseteq K$. The optimisation maximises the *causality* between the *derived variables* and the label *variables* by minimising some cost or loss function. At the optimum there is no error and the relation is functional,

$$\text{split}(W_F, (A * X_F \ \% \ (W_F \cup V \setminus K))^{\text{FS}}) \ \ \in \ \ W_F^{\text{CS}} \to (V \setminus K)^{\text{CS}}$$

where $X_F = \text{histogram}(F^{\text{T}})$ and $W_F = \text{der}(F)$. In some cases the choice of optimisation parameters, such as the graph or the definition of the loss function, is such that, when optimal, (i) the *model* is *causal* from the *derived variables* to the label *variables*, and (ii) a query application via *model* is equal to a query application on the *unknown distribution histogram*,

$$\forall Q \in (E\%K)^{\text{F}\{\}}\ ((Q * F^{\text{T}} * X_F * A)^{\wedge} \ \% \ (V \setminus K) = Q * E^{\text{F}} \ \% \ (V \setminus K))$$

That is, even in the case where the *sample* is *ineffective* with respect to the query, $Q^{\text{F}} \not\leq A^{\text{F}}$, if the *model* is not *over-fitted*, $(Q*F^{\text{T}})^{\text{F}} \leq (A*F^{\text{T}})^{\text{F}}$, then an estimate of the query application, $Q*E^{\text{F}} \ \% \ (V \setminus K)$, may sometimes be made.

Note that there are some cases where no set of optimisation parameters can avoid an *over-fitted model*. For example, if the *sample reduced* to the query *variables* is *independent*, $A\%K = (A\%K)^{\mathrm{X}}$, but the *distribution histogram* is not, $E\%K \neq (E\%K)^{\mathrm{X}}$, then the *self-transform* obtained from the optimisation will be *over-fitted*, $Q^{\mathrm{F}} \nleq A^{\mathrm{F}} \implies (Q * K^{\mathrm{CS}\{\}\mathrm{T}})^{\mathrm{F}} \nleq (A * K^{\mathrm{CS}\{\}\mathrm{T}})^{\mathrm{F}}$.

There are various candidates for the loss function. Given a *sample histogram* A, a *functional definition set* F, and a set of query *variables* K, the label *entropy* loss function $\mathrm{lent} \in \mathcal{A} \times \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathbf{R}$ is

$$\mathrm{lent}(A, F, K) \;\; := \;\; \sum_{(R,C) \in (F^{\mathrm{T}})^{-1}} (A * F^{\mathrm{T}})_R \times \mathrm{entropy}(A * C \ \% \ (V \setminus K))$$

where $V = \mathrm{vars}(A)$. It is not obvious, however, if the derivative of the label *entropy* function with respect to an underlying neural net weight has an analytic solution. Given $D \subset \mathbf{R}$, a numeric approximation to a *discretised fud transform*, $(V, w, \mathrm{discrete}(D, n)(f_\sigma(Q)))$ where $(V, w, f_\sigma(Q)) = F^{\mathrm{T}}$, can be defined, but its computation may be intractable.

In the case where the *derived variables* of the *fud* is a *literal frame* of the label *variables*, $W_F :\leftrightarrow: (V \setminus K)$ and $\forall v \in (V \setminus K)\ (U_v \subseteq \mathbf{R})$, the least squares loss function $\mathrm{lsq} \in \mathcal{A} \times \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathbf{R}$ is

$$\mathrm{lsq}(A, F, K) \;\; := \;\; \sum_{(S,c) \in A * X_F} \Big(c \times \sum_{i \in \{1 \ldots m\}} ((S\%W_F)^{[]}_i - (S\%(V \setminus K))^{[]}_i)^2\Big)$$

where $m = |W_F| = |(V \setminus K)|$. Let $\mathrm{lsq}(\sigma) \in \mathcal{A} \times \mathrm{nets} \times \mathrm{P}(\mathcal{V}) \to \mathbf{R}$ be

$$\mathrm{lsq}(\sigma)(A, G, K) \;\; := \;\; \mathrm{lsq}(A, \mathrm{fud}(\sigma)(G), K)$$

and its derivative with respect to the i-th weight $\mathrm{dlsq}(\sigma)(i) \in \mathcal{A} \times \mathrm{nets} \times \mathrm{P}(\mathcal{V}) \to (\mathcal{L}(\mathbf{R}) \to \mathbf{R})$ be

$$\begin{aligned}&\mathrm{dlsq}(\sigma)(i)(A, G, K) \;\; := \\ &\quad \partial_i(\{(\mathrm{weights}(G'), \mathrm{lsq}(\sigma)(A, G', K)) : G' \in \mathrm{nets},\ \mathrm{graph}(G') = \mathrm{graph}(G)\})\end{aligned}$$

where $\partial_j \in (\mathcal{L}(\mathbf{R}) \to \mathbf{R}) \to (\mathcal{L}(\mathbf{R}) \to \mathbf{R})$ is defined $\partial_j(F) := \{(Z, \partial F(Z)/\partial Z_j) : Z \in \mathrm{dom}(F)\}$.

Typically the label *variables* form a *bivalent crown*, $\mathrm{crown}(A \ \% \ (V \setminus K))$ where $\forall v \in (V \setminus K)\ (U_v = \{0, 1\})$, with each label *variable* corresponding to a label value. When the loss is zero, a query via the *model* need not compute the *component*,

$$\forall Q \in (E\%K)^{\mathrm{F}\{\}}\ (\mathrm{reframe}(Y, Q * F^{\mathrm{T}}) = Q * E^{\mathrm{F}} \ \% \ (V \setminus K))$$

where the *frame* mapping is $Y = \{(L_{W_F}(i), L_{V \setminus K}(i)) : i \in \{1 \ldots m\}\} \in W_F :\leftrightarrow: (V \setminus K)$ and $L_W = \text{flip}(\text{order}(D_{\mathcal{V}}, W))$.

Define the *substrate net set* as $\text{nets}(U, V, \sigma) = \{G : G \in \text{nets},\ \text{fud}(\sigma)(G) \in \mathcal{F}_{\infty,U,V}\}$, which is such that $\mathcal{F}_{\infty,U,V,\sigma} = \text{fud}(\sigma) \circ \text{nets}(U, V, \sigma)$.

Given a loss function, a search function for a neural network can be defined using the method of gradient descent. Let $P \in \mathcal{L}(\mathcal{X})$ be a list of search parameters. Let the activation function, *system* and query *variables* be defined in the search parameters, $\sigma, U, K \in \text{set}(P)$. Let initial *substrate net* $G_R \in \text{nets}(U, K, \sigma)$ have (a) graph $\text{graph}(G_R) \in \text{set}(P)$ and (b) arbitrary weights $R \in \mathcal{L}(\mathbf{R})$ where $R = \text{weights}(G_R) \in \text{set}(P)$. Given a *histogram* A of *variables* V in *system* U such that (i) the *variables* are real *valued*, $\forall v \in V\ (U_v \subseteq \mathbf{R})$, (ii) A is *causal*, $\text{split}(K, A^{\text{FS}}) \in K^{\text{CS}} \to (V \setminus K)^{\text{CS}}$, and (iii) the set of *derived variables*, $\text{der}(\text{fud}(\sigma)(G_R))$, is a *literal frame* of the set of label *variables*, $V \setminus K$, define the *least squares gradient descent substrate net tree searcher* as

$$Z_{P,A,\text{gr},\text{lsq}} = \text{searchTreer}(\text{nets}(U, K, \sigma), P_{P,A,\text{gr},\text{lsq}}, \{G_R\})$$

where the neighbourhood function is

$$\begin{aligned} P_{P,A,\text{gr},\text{lsq}}(G) = \{G' :\ & \text{lsq}(\sigma)(A, G, K) > t, \\ & G' \in \text{nets}(U, K, \sigma),\ \text{graph}(G') = \text{graph}(G), \\ & Q = \text{weights}(G),\ Q' = \text{weights}(G'), \\ & Q' = \{(i,\ Q_i - r \times \text{dlsq}(\sigma)(i)(A, G, K)(Q)) : i \in \{1 \ldots |Q|\}\}\} \end{aligned}$$

and loss threshold $t \in \text{set}(P)$ and rate of descent $r \in \text{set}(P)$.

Note that a practicable implementation of the *net searcher* would usually (i) perform an optimise step for each *event* rather than the whole *history*, and (ii) compute the deltas to be applied to the net weights, Q, one *layer* at a time in sequence from the top to the bottom (which is called backpropagation).

Let *history* $H \in \mathcal{H}_{U,X}$ be such that its *histogram* $A = \text{histogram}(H)$ satisfies the constraints, of (i) real *valued variables*, (ii) *causal histogram*, and (iii) a *literal frame*, imposed by the search parameters P of the *least squares gradient descent substrate net tree searcher*, $Z_{P,A,\text{gr},\text{lsq}}$. Define the *least squares gradient descent fud search function* as

$$\begin{aligned} & Z_{\text{F},\text{P},P,\text{gr},\text{lsq}}(H) = \\ & \quad \{(\text{fud}(\sigma)(G), -\text{lsq}(\sigma)(A, G, K)) : Q = \text{leaves}(\text{tree}(Z_{P,A,\text{gr},\text{lsq}})),\ \{G\} = Q\} \end{aligned}$$

It is conjectured above that for all finite *systems* and finite *event identifier sets* there exists a tuple of parameters such that the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, is positively correlated with the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D,F,P,m,G,D,F,H}}$, for uniform *history probability function*,

$$\begin{aligned}
&\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ (|\mathcal{H}_{U,X}| < \infty \implies \\
&\quad \exists P \in \mathcal{L}(\mathcal{X})\ (\mathrm{covariance}(P_{U,X}) \\
&\qquad\qquad (\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,D,F,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}) \geq 0))
\end{aligned}$$

A similar generalisation of a correlation between the *least squares gradient descent fud search function*, $Z_{\mathrm{F,P},P,\mathrm{gr,lsq}}$, and the *fud minimum space specialising fud search function*, $Z_{\mathrm{F,P,m,G,F,H}}$, cannot be made because the *history*, H, is not independent of the search parameters, P. That is, least squares gradient descent supervised neural net optimisation requires specific configuration. Conjecture, however, that in some cases the properties of the *net search function* and the *minimum space search function* are similar.

First consider the simpler relation to the *minimum space search function* for the *specialising derived history coder*. The *fud minimum space specialising derived search function* for *history* $H \in \mathcal{H}_{U,X}$ is defined in terms of the *expanded specialising derived history coder*, $C_{\mathrm{G,T,H}}(F^{\mathrm{T}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$, as

$$Z_{\mathrm{F,P,m,G,T,H}}(H) \quad = \quad \{(F, -C_{\mathrm{G,T,H}}(F^{\mathrm{T}})^{\mathrm{s}}(H)) : F \in \mathcal{F}_{U,\mathrm{P}}\}$$

The minimisation of the *specialising derived substrate history coder space*, $C_{\mathrm{G},V,\mathrm{T,H}}(F^{\mathrm{T}V})^{\mathrm{s}}(H)$, occurs where (i) the *derived entropy* is low, (ii) the *possible derived volume* is small, (iii) the *underlying components* have high *entropy* and (iv) high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*.

Note that, in some cases, particularly where there is a residual loss, the *derived histogram* may be *unit*, $A * F^{\mathrm{T}} = (A * F^{\mathrm{T}})^{\mathrm{F}}$, because the *derived values* are continuous. The infinite *derived volume* of the real *valued derived variables*, $|W^{\mathrm{C}}| = |\mathbf{R}^m| = \infty$ where $m = |W| = |V \setminus K|$, may be made finite by *discretising* with the *values* of the label *variables*, $\{(i, \mathrm{nearest}(D, r)) : (i, r) \in R^{[]}\} \in D^m$ where $D = \cup\{U_v : v \in (V \setminus K)\} \subset \mathbf{R}$ and $R \in W^{\mathrm{CS}}$. If the label *variables* form a *bivalent crown*, $\mathrm{crown}(A\ \%\ (V \setminus K))$ where $\forall v \in (V \setminus K)\ (U_v = \{0, 1\})$, then the *discretised derived volume* reduces to a finite $|W^{\mathrm{C}}_{\{0,1\}}| = |(V \setminus K)^{\mathrm{C}}| = 2^m$. In the computations of *alignment* and *entropy* that follow, the *derived variables* are *discretised* to the *values* of the label *variables*.

The initial *substrate net*, G_R, has arbitrary weights, $R = \text{weights}(G_R) \in \mathcal{L}(\mathbf{R})$, and so the corresponding initial *fud*, $F_R = \text{fud}(\sigma)(G_R)$, is likely to have a high least squares loss. That is, far from the *derived variables* and the label *variables* being *causally* related, $W_D^{\text{CS}} \to (V \setminus K)^{\text{CS}}$, they are likely to be *independent*,

$$A * X_{F_R} * \{W_D^{\text{CS}\{\}\text{T}}, (V \setminus K)^{\text{CS}\{\}\text{T}}\}^{\text{T}} \approx (A * X_{F_R} * \{W_D^{\text{CS}\{\}\text{T}}, (V \setminus K)^{\text{CS}\{\}\text{T}}\}^{\text{T}})^{\text{X}}$$

or

$$\text{algn}(A * X_{F_R} * \{W_D^{\text{CS}\{\}\text{T}}, (V \setminus K)^{\text{CS}\{\}\text{T}}\}^{\text{T}}) \approx 0$$

where $\{W_D^{\text{CS}\{\}\text{T}}, (V \setminus K)^{\text{CS}\{\}\text{T}}\}$ is the *fud* of the *self transforms* of the (i) *discretised derived variables* and (ii) label *variables*.

As the optimisation proceeds from the initial *fud*, F_R, to the optimal *fud*, $F \in \text{maxd}(Z_{\text{F},\text{P},P,\text{gr},\text{lsq}}(H))$, the loss decreases and the relation between the top *layer* and the label becomes more *causal*,

$$\text{algn}(A * X_F * \{W_D^{\text{CS}\{\}\text{T}}, (V \setminus K)^{\text{CS}\{\}\text{T}}\}^{\text{T}}) > 0$$

If the loss is zero, after *discretising*, then the relation between the *derived variables* and the label *variables* is not only *causal* but bijective, $W_D^{\text{CS}} \leftrightarrow (V \setminus K)^{\text{CS}}$. So the *self partition transforms* are highly *aligned* because *diagonalised*,

$$\text{diagonal}(A * X_F * \{W_D^{\text{CS}\{\}\text{T}}, (V \setminus K)^{\text{CS}\{\}\text{T}}\}^{\text{T}})$$

The negative least squares loss (i) varies with the *alignment* of the *self partition transforms*, (ii) varies with the *alignment* of the *reduction* to the union of the *derived variables* and label *variables*, (iii) varies against the *size scaled entropy* of the *reduction* to the union of the *derived variables* and label *variables*, and so (iv) varies against the *derived entropy* of the *fud transform*,

$$\begin{aligned}
-\text{lsq}(A, F_D, K) &\sim \text{algn}(A * X_F * \{W_D^{\text{CS}\{\}\text{T}}, (V \setminus K)^{\text{CS}\{\}\text{T}}\}^{\text{T}}) \\
&\sim \text{algn}(A * X_F \ \% \ (W_D \cup V \setminus K)) \\
&\sim -z \times \text{entropy}(A * X_F \ \% \ (W_D \cup V \setminus K)) \\
&\sim -z \times \text{entropy}(A * X_F \ \% \ W_D) \\
&= -z \times \text{entropy}(A * F_D^{\text{T}})
\end{aligned}$$

That is, as the loss, $\text{lsq}(A, F_D, K)$, is minimised the *derived entropy*, $\text{entropy}(A * F_D^{\mathrm{T}})$, tends to be minimised, which is a property of the *specialising coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(F_D^{\mathrm{T}V})$.

The *discretised derived volume* is fixed, $|W_D^{\mathrm{C}}| = |D|^m$, because the graph is fixed in the parameters, $\text{graph}(G) \in \text{set}(P)$ where $F = \text{fud}(\sigma)(G)$ and $F \in \text{maxd}(Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}(H))$. So the *derived volume* is not minimised during the optimisation. The optimisation does not share the property of low *derived volume* with the *specialising coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(F_D^{\mathrm{T}V})$. However, as the *derived alignment*, $\text{algn}(A * F_D^{\mathrm{T}})$, increases during least squares optimisation, the *causal* relation between the *discretised derived variables* and the label *variables* tends to bijective, $W_D^{\mathrm{CS}} \leftrightarrow (V \setminus K)^{\mathrm{CS}}$. So if the label is *diagonalised* then the *derived* tends to be *diagonalised*, $\text{diagonal}(A \% (V \setminus K)) \implies \text{diagonal}(A * F_D^{\mathrm{T}})$, and if the label is a *crown* then the *derived* tends to be a *crown*, $\text{crown}(A \% (V \setminus K)) \implies \text{crown}(A * F_D^{\mathrm{T}})$. In both cases the *effective derived volume* is less than the *derived volume*, $|(A * F_D^{\mathrm{T}})^{\mathrm{F}}| < |W_D^{\mathrm{C}}|$, if the label is *multi-variate*, $m \geq 2$, and *multi-valent*, $|D| \geq 2$.

The minimisation of the least squares loss function, $\text{lsq}(A, F_D, K)$, tends to minimise the label *entropy* loss function, $\text{lent}(A, F_D, K)$, as the relation between the *discretised derived variables* and the label *variables* tends to functional or *causal*, $W_D^{\mathrm{CS}} \to (V \setminus K)^{\mathrm{CS}}$. The corollary of the label *entropy* loss function,

$$\text{lent}(A, F, K) \;:=\; \sum_{(R,C) \in (F^{\mathrm{T}})^{-1}} (A * F^{\mathrm{T}})_R \times \text{entropy}(A * C \% (V \setminus K))$$

is the query *entropy* loss function $\text{qent} \in \mathcal{A} \times \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathbf{R}$,

$$\text{qent}(A, F, K) \;:=\; \sum_{(R,C) \in (F^{\mathrm{T}})^{-1}} (A * F^{\mathrm{T}})_R \times \text{entropy}(A * C \% K)$$

The *fud*, $F \in \mathcal{F}_{\infty,U,K}$, is in *substrate* K, so the query *entropy* is just the *size* scaled *expected component entropy* of the *reduced histogram*,

$$\text{qent}(A, F, K) \;=\; z \times \text{entropyComponent}(A\%K, F^{\mathrm{T}})$$

The *histogram entropy*, $\text{entropy}(A)$, is a constant, so the query *entropy*, $\text{qent}(A, F, K)$, varies against the label *entropy*, $\text{lent}(A, F, K)$. The negative least squares loss (i) varies with the negative label *entropy* loss, (ii) varies with the query *entropy*, (iii) varies with the *size scaled expected component* *entropy* of the *reduction* to the query *variables*, and so (iv) varies with the *size scaled expected component entropy*,

$$\begin{aligned}
-\ \mathrm{lsq}(A, F_D, K) &\sim -\ \mathrm{lent}(A, F_D, K) \\
&\sim \mathrm{qent}(A, F_D, K) \\
&\sim z \times \mathrm{entropyComponent}(A\%K, F_D^{\mathrm{T}}) \\
&\sim z \times \mathrm{entropyComponent}(A, F_D^{\mathrm{T}})
\end{aligned}$$

That is, as the loss, $\mathrm{lsq}(A, F_D, K)$, is minimised the *expected component entropy*, $\mathrm{entropyComponent}(A, F_D^{\mathrm{T}})$, tends to be maximised, which is a property of the *specialising coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(F_D^{\mathrm{T}V})$.

Consider the case of a *multi-variate* set of real *valued* query *variables* K, where $k = |K| \geq 2$ and $\forall x \in K\ (U_x \subseteq \mathbf{R})$, and a *neural net fud* $F \in \mathcal{F}_{\infty,U,K,\sigma}$ consisting of two *transforms*, $F = \{T_1, T_2\}$, each having the query *variables* as the *underlying*, $\mathrm{und}(T_1) = \mathrm{und}(T_2) = K$. Given a coordinate $S \in \mathbf{R}^k$ the weights of the *transforms* form a pair of hyperplanes,

$$\sum_{i \in \{1 \ldots k\}} Q_{1,i} S_i \ +\ Q_{1,k+1} \ =\ 0$$

and

$$\sum_{i \in \{1 \ldots k\}} Q_{2,i} S_i \ +\ Q_{2,k+1} \ =\ 0$$

where $Q_1, Q_2 \in \mathbf{R}^{k+1}$ are the weights correspondng to T_1, T_2. If the hyperplanes of the arbitrarily weighted initial *fud*, F_R, intersect, the acute angle between them is expected to be 45°. That is, given an activation function, σ, which is a step function, or a binary set of *discrete values*, $D = \{0, 1\}$, the probability distribution of the *component cardinalities* of the initial *fud* is bi-modal. If $(\cdot, C_1), (\cdot, C_2) \in (F_{R,\{0,1\}}^{\mathrm{T}})^{-1}$ are such that $|C_1| < |C_2|$, then it is expected that $3|C_1| = |C_2|$. So the *component cardinality entropy* of the initial *fud* is expected to be less than maximal,

$$\mathrm{entropy}(K^{\mathrm{C}} * F_{R,D}^{\mathrm{T}}) \ <\ \mathrm{entropy}(W_D^{\mathrm{C}})$$

The *derived entropy* of the initial *fud* is expected to be approximately equal to the *component cardinality entropy*,

$$\mathrm{entropy}(A * F_{R,D}^{\mathrm{T}}) \ \approx\ \mathrm{entropy}(K^{\mathrm{C}} * F_{R,D}^{\mathrm{T}})$$

and so the *component size cardinality relative entropy* of the initial *fud* is expected to be small,

$$\text{entropyRelative}(A * F^{\mathrm{T}}_{R,D}, K^{\mathrm{C}} * F^{\mathrm{T}}_{R,D}) \;\approx\; 0$$

If the *histogram*, A, is approximately uniformly distributed over the *volume*, then the *component size cardinality relative entropy* remains small during the optimisation,

$$\text{entropyRelative}(A * F^{\mathrm{T}}_{D}, K^{\mathrm{C}} * F^{\mathrm{T}}_{D}) \;\approx\; 0$$

In contrast, consider the case where the *histogram*, A, is not uniformly distributed, but clustered by label *state*. Let $Y_L \subset K^{\mathrm{CS}}$ be the set of the centres of the clusters for *effective* label *state* $L \in (A\%(V \setminus K))^{\mathrm{FS}}$. The maximum radius $r_L \in \mathbf{R}_{>0}$ is such that

$$\forall S \in A^{\mathrm{FS}} \;\Diamond L = S\%(V \setminus K) \;\exists Q \in Y_L \;(\sum_{i \in \{1 \ldots k\}} (Q^{[]}_i - S^{[]}_i)^2 \;\leq\; r_L^2)$$

Let r_C be the radius of *component* C. In the case where the *histogram* is clustered such that the cluster radius of a label *state* is much smaller than the least initial *component* radius, $\forall (\cdot, C) \in (F^{\mathrm{T}}_{R,\{0,1\}})^{-1}\ (r_L \ll r_C)$, then optimised rotations of the hyperplanes, that sweep up nearby clusters in the same label *state*, tend to be such that the magnitude of the change in the fractional *component size*, $|(A * F^{\mathrm{T}}_{2,D})(R) - (A * F^{\mathrm{T}}_{1,D})(R)|/z$, is greater than magnitude of the change in the fractional *component cardinality*, $|(K^{\mathrm{C}} * F^{\mathrm{T}}_{2,D})(R) - (K^{\mathrm{C}} * F^{\mathrm{T}}_{1,D})(R)|/|K^{\mathrm{C}}|$. So, in the clustered case, as the optimisation decreases the *derived entropy*, $\text{entropy}(A * F^{\mathrm{T}}_D)$, the *component sizes* and *component cardinalities* become less synchronised and the *component size cardinality relative entropy* increases,

$$\begin{aligned}
-\ \text{lsq}(A, F_D, K) &\sim -z \times \text{entropy}(A * F^{\mathrm{T}}_D) \\
&\sim z \times \text{entropyRelative}(A * F^{\mathrm{T}}_D, K^{\mathrm{C}} * F^{\mathrm{T}}_D) \\
&= z \times \text{entropyRelative}(A * F^{\mathrm{T}}_D, V^{\mathrm{C}} * F^{\mathrm{T}}_D)
\end{aligned}$$

The same reasoning applies to *fuds* consisting of more than two *transforms*, $|F| > 2$, but note that at higher *fud* cardinalities the initial *component cardinality entropy*, $\text{entropy}(K^{\mathrm{C}} * F^{\mathrm{T}}_{R,D})$, tends to be multi-modal and so approximates more closely to the *uniform cartesian derived entropy*, $\text{entropy}(W^{\mathrm{C}}_D)$. So there is less freedom for the *relative entropy* of the *fud* to increase during optimisation. In the case of *multi-layer fuds*, however, the *breadth* can be constrained and so the *relative entropy* of taller, narrrower *fuds* may be higher than in shorter, wider *fuds* of the same cardinality.

In general, in the clustered case, the optimised *fud* is such that high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*, which, again, is a property of the *specialising coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(F_D^{\mathrm{T}V})$.

Overall, the comparisons above suggest that, given search parameters P, there sometimes exists a subset of *histories* $\mathcal{H}_{U,X,P} \subset \mathcal{H}_{U,X}$ satisfying the constraints of (i) real *valued variables*, (ii) *causal histogram*, (iii) a *literal frame*, and (iv) clustered *histogram* such that there is a positive correlation between the *least squares gradient descent fud search function*, $Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}$, and the *fud minimum space specialising derived search function*, $Z_{\mathrm{F},\mathrm{P},\mathrm{m},\mathrm{G},\mathrm{T},\mathrm{H}}$,

$$\mathrm{covariance}(P_{U,X,P})(\mathrm{maxr} \circ Z_{\mathrm{F},\mathrm{P},\mathrm{m},\mathrm{G},\mathrm{T},\mathrm{H}}, \mathrm{maxr} \circ Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}) \quad \geq \quad 0$$

where $P_{U,X,P} = \mathcal{H}_{U,X,P} \times \{1/|\mathcal{H}_{U,X,P}|\}$.

Now consider the relation to the *minimum space search function* for the *specialising fud history coder*. It is conjectured that when the *specialising fud space*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H)$, is minimised in the *fud minimum space specialising fud search function*, $Z_{\mathrm{F},\mathrm{P},\mathrm{m},\mathrm{G},\mathrm{F},\mathrm{H}}$, (i) the *derived entropy* decreases up the *layers*, (ii) the *possible derived volume* decreases up the *layers*, (iii) the *expected component entropy* increases up the *layers*, and (iv) the *component size cardinality cross entropy* increases up the *layers*. The optimisation of a *fud* without a *layer* limit may be made computable by building the *fud layer* by *layer*, minimising the *specialising space* at each step, until the addition of a *layer* fails to reduce the *specialising space*.

In the case of the *net search function*, $Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}$, the *substrate nets* are not built *layer* by *layer* during the optimisation because the graph, $\mathrm{graph}(G)$ where $F = \mathrm{fud}(\sigma)(G)$ and $F \in \mathrm{maxd}(Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}(H))$, is fixed in the parameters, $\mathrm{graph}(G) = \mathrm{graph}(G_R) \in \mathrm{set}(P)$. The properties of the *nets* do vary *layer* by *layer*, however, because the optimisation of the least squares loss function minimises the square of the distance between the top *layer*, $W = \mathrm{der}(F)$, and the label *variables*, $V \setminus K$. So the top *layer* is more closely *aligned* to the label *variables* than the other *layers*.

The loss function with respect to the neuron weights is composed of *layers*. The second order sensitivity of the loss function generally increases with the *layer*,

$$\mathrm{ddlsq}(\sigma)(i)(A, G, K)(Q) \quad < \quad \mathrm{ddlsq}(\sigma)(j)(A, G, K)(Q)$$

where $Q = \text{weights}(G)$, weights $i, j \in \{1 \dots |Q|\}$ parameterise corresponding *transforms* $T_i, T_j \in F$ such that $\text{layer}(F, \text{der}(T_i)) < \text{layer}(F, \text{der}(T_j))$, and the second order derivative of the least squares loss function is defined

$$\begin{aligned}&\text{ddlsq}(\sigma)(i)(A, G, K) \ := \\ &\quad \partial_i^2(\{(\text{weights}(G'), \text{lsq}(\sigma)(A, G', K)) : G' \in \text{nets},\ \text{graph}(G') = \text{graph}(G)\})\end{aligned}$$

So gradient descent optimisation resolves more quickly for higher *layers* than lower *layers*.

Let F_i be the i-th *layer* of the *fud*, F, where $i \in \{1 \dots l\}$ and $l = \text{layer}(F, W)$. The second order sensitivity of the loss function generally increases with the *layer* because, although the relation between a lower *layer* and a higher *layer* is always *causal*, $W_{i-1,D}^{\text{CS}} \to W_{i,D}^{\text{CS}}$, it is not usually bijective, $W_{i-1,D}^{\text{CS}} \leftrightarrow W_{i,D}^{\text{CS}}$. For example, the relation between *layers* is sometimes only partially bijective, $J_{i-1,D}^{\text{CS}} \leftrightarrow J_{i,D}^{\text{CS}}$ where $J_{i-1} \subset W_{i-1}$ and $J_i \subset W_i$. The relation is always functional so the degree of multijectivity is the query *entropy*, $\text{qent}(A * F_{\{1 \dots i-1\},D}^{\text{T}}, F_{i,D}, W_{i-1,D})$. When the query *entropy* is zero, the relation is *effectively* bijective, corresponding to a *self transform*, $W_{i-1,D}^{\text{CS}\{\}\text{T}}$. When the query *entropy* is maximised, the relation is multijective, corresponding to a *unary transform*, $\{W_{i-1,D}^{\text{CS}}\}^{\text{T}}$.

For the same reason it is less likely for the lower *layer* to be *causal* to the label *variables*, $W_{i-1,D}^{\text{CS}} \to (V \setminus K)^{\text{CS}}$, than for the higher *layer* to be *causal* to the label *variables*, $W_{i,D}^{\text{CS}} \to (V \setminus K)^{\text{CS}}$. So, in general, the *alignment* between the *layer variables* and the label *variables* increases up the *layers*. For $i \in \{2 \dots l\}$,

$$\begin{aligned}&\text{algn}(A * X * \{W_{i,D}^{\text{CS}\{\}\text{T}}, (V \setminus K)^{\text{CS}\{\}\text{T}}\}^{\text{T}}) \ > \\ &\qquad\qquad \text{algn}(A * X * \{W_{i-1,D}^{\text{CS}\{\}\text{T}}, (V \setminus K)^{\text{CS}\{\}\text{T}}\}^{\text{T}})\end{aligned}$$

and

$$\text{algn}(A * X \ \% \ (W_{i,D} \cup V \setminus K)) \ > \ \text{algn}(A * X \ \% \ (W_{i-1,D} \cup V \setminus K))$$

where $X = \prod_{T \in F} \text{his}(T)$. So the *entropy* between the *layer variables* and the label *variables* tends to decrease up the *layers*,

$$\text{entropy}(A * X \ \% \ (W_{i,D} \cup V \setminus K)) \ < \ \text{entropy}(A * X \ \% \ (W_{i-1,D} \cup V \setminus K))$$

Therefore conjecture that, in general, the *derived entropy* also decreases up the *layers*, regardless of the label *variables*,

$$\forall i \in \{2 \dots l\} \ (\text{entropy}(A * F_{\{1 \dots i\},D}^{\text{T}}) < \text{entropy}(A * F_{\{1 \dots i-1\},D}^{\text{T}}))$$

which is a property of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F_D)$.

In the comparison above between the *least squares gradient descent fud search function*, $Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}$, and the *fud minimum space specialising derived search function*, $Z_{\mathrm{F},\mathrm{P},\mathrm{m},\mathrm{G},\mathrm{T},\mathrm{H}}$ it is shown that the *derived entropy*, $\mathrm{entropy}(A * F_D^{\mathrm{T}})$, (i) varies with the *effective derived volume*, $|(A * F_D^{\mathrm{T}})^{\mathrm{F}}|$, (ii) varies against the *expected component entropy*, $\mathrm{entropyComponent}(A, F_D^{\mathrm{T}})$, and (iii) varies against the *component size cardinality relative entropy*, $\mathrm{entropyRelative}(A * F_D^{\mathrm{T}}, V^{\mathrm{C}} * F_D^{\mathrm{T}})$. So conjecture that, in general, (i) the *effective derived volume* decreases up the *layers*,

$$\forall i \in \{2 \dots l\}\ (|(A * F_{\{1 \dots i\},D}^{\mathrm{T}})^{\mathrm{F}}| < |(A * F_{\{1 \dots i-1\},D}^{\mathrm{T}})^{\mathrm{F}}|)$$

(ii) the *expected component entropy* increases up the *layers*,

$$\forall i \in \{2 \dots l\}\ (\mathrm{entropyComponent}(A, F_{\{1 \dots i\},D}^{\mathrm{T}})\ >\ \mathrm{entropyComponent}(A, F_{\{1 \dots i-1\},D}^{\mathrm{T}}))$$

and (iii) the *component size cardinality relative entropy* increases up the *layers*,

$$\forall i \in \{2 \dots l\}\ (\mathrm{entropyRelative}(A * F_{\{1 \dots i\},D}^{\mathrm{T}}, V_D^{\mathrm{C}} * F_{\{1 \dots i\},D}^{\mathrm{T}})\ >\ \mathrm{entropyRelative}(A * F_{\{1 \dots i-1\},D}^{\mathrm{T}}, V_D^{\mathrm{C}} * F_{\{1 \dots i-1\},D}^{\mathrm{T}}))$$

Again, these properties are also properties of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F_D)$.

To conclude, the comparisons above suggest that, given search parameters P, there sometimes exists a subset of *histories* $\mathcal{H}_{U,X,P} \subset \mathcal{H}_{U,X}$ satisfying the constraints of (i) real *valued variables*, (ii) *causal histogram*, (iii) a *literal frame*, and (iv) clustered *histogram* such that there is a positive correlation between the *least squares gradient descent fud search function*, $Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}$, and the *fud minimum space specialising fud search function*, $Z_{\mathrm{F},\mathrm{P},\mathrm{m},\mathrm{G},\mathrm{F},\mathrm{H}}$,

$$\mathrm{covariance}(P_{U,X,P})(\mathrm{maxr} \circ Z_{\mathrm{F},\mathrm{P},\mathrm{m},\mathrm{G},\mathrm{F},\mathrm{H}}, \mathrm{maxr} \circ Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}})\ \geq\ 0$$

where $P_{U,X,P} = \mathcal{H}_{U,X,P} \times \{1/|\mathcal{H}_{U,X,P}|\}$.

An example of additional supervised *knowledge* is where it is *known* that the *substrate* exhibits some *symmetry*. The supervised *models* can be constrained to exhibit these *symmetries* by copying common *submodels* amongst them. For example, in the case where the *substrate* represents a visual or auditory field with translational symmetry, the common *submodel* can consist of a relative *model* on a *frame* subset of the *substrate* which is copied across the whole *substrate* by adding translation offsets to the *frame variables*.

If the optimisation of artificial neural networks is of the unsupervised type, there is no *knowledge* of a *causal* label. An example of an unsupervised optimisation is the auto-encoder [2]. Here the method of least squares gradient descent is used but the label is simply the *substrate* itself. Let $Y \in V :\leftrightarrow: V_Y$ be a mapping from the sample *variables*, V, to a disjoint *reframed* set, V_Y, such that the *reframe* is *literal*, $\forall(v,w) \in Y\ (U_w = U_v)$. The *histogram* may be extended by dotting with the *reframe*,

$$A_Y = \{(S \cup \text{reframe}(Y,S), c) : (S,c) \in A\}$$

Then the query *variables* are the *substrate*, $K = V \in \text{set}(P)$, the label *variables* are the *reframed substrate*, V_Y, and the search is performed on the dotted *histogram*, $Z_{P,A_Y,\text{gr},\text{lsq}}$.

If the auto-encoder's graph is such that all *layers* have the same *breadth*, $\forall i \in \{1 \ldots l\}\ (|F_i| = n)$, then the likely *model* is the *over-fitted effective self transform*, $F^{\text{T}} = (A^{\text{FS}\{\}} \cup \{V^{\text{CS}} \setminus A^{\text{FS}}\})^{\text{T}}$ or the *full functional transform*, $F^{\text{T}} = \{\{w\}^{\text{CS}\{\}\text{T}} : w \in V\}^{\text{T}}$. However, if the *fud* has an hourglass shape such that there is an intermediate *layer* F_a which has a *breadth* less than all other *layers*, $\forall i \in \{1 \ldots l\}\ (i \neq a \implies |F_i| > |F_a|)$, then it may be expected that (i) the *derived entropy* decreases up to this *layer*,

$$\forall i \in \{2 \ldots a\}\ (\text{entropy}(A * F^{\text{T}}_{\{1 \ldots i\},D}) < \text{entropy}(A * F^{\text{T}}_{\{1 \ldots i-1\},D}))$$

(ii) the *effective derived volume* decreases up to this *layer*,

$$\forall i \in \{2 \ldots a\}\ (|(A * F^{\text{T}}_{\{1 \ldots i\},D})^{\text{F}}| < |(A * F^{\text{T}}_{\{1 \ldots i-1\},D})^{\text{F}}|)$$

(iii) the *expected component entropy* increases up to this *layer*,

$$\forall i \in \{2 \ldots a\}$$
$$(\text{entropyComponent}(A, F^{\text{T}}_{\{1 \ldots i\},D})\ >\ \text{entropyComponent}(A, F^{\text{T}}_{\{1 \ldots i-1\},D}))$$

and (iv) the *component size cardinality relative entropy* increases up to this *layer*,

$$\forall i \in \{2 \ldots a\}$$
$$(\text{entropyRelative}(A * F^{\text{T}}_{\{1 \ldots i\},D}, V_D^{\text{C}} * F^{\text{T}}_{\{1 \ldots i\},D})\ >$$
$$\text{entropyRelative}(A * F^{\text{T}}_{\{1 \ldots i-1\},D}, V_D^{\text{C}} * F^{\text{T}}_{\{1 \ldots i-1\},D}))$$

Above the intermediate *layer*, a, it is expected that there is little change in these properties. For example, $\text{entropy}(A * F^{\text{T}}_{\{1 \ldots l\},D}) \approx \text{entropy}(A * F^{\text{T}}_{\{1 \ldots a\},D})$.

5.3 Classical induction

The following sections consider how *induction* is related to *likelihood* and *sensitivity*. First consider *classical induction*.

As defined at the beginning of this section $P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ is an *unknown history probability function* in the non-empty *histories* $\mathcal{H}_{U,X}$, where U is a non-empty finite *system* and $X \subset \mathcal{X}$ is a non-empty *unknown* finite set of *event identifiers*.

Now, similarly to section 'Derived history space', above, let $H_{\text{h}} \in \mathcal{H}_{U,X}$ be a *distribution history* with *substrate* V_{h} equal to the *system variables*, $V_{\text{h}} = \text{vars}(H_{\text{h}}) = \text{vars}(U)$. Its *volume* is $v_{\text{h}} = |V_{\text{h}}^{\text{C}}|$. Its domain is the entire set of *event identifiers*, $\text{ids}(H_{\text{h}}) = X$, so that the *size* $z_{\text{h}} = |H_{\text{h}}|$ equals the cardinality of the *event identifiers*, $z_{\text{h}} = |X|$. Thus the *distribution history* is a left total *state*-valued function of the *event identifiers*, $H_{\text{h}} \in X :\to V_{\text{h}}^{\text{CS}}$. The *historically distributed history probability function* $P_{U,X,H_{\text{h}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ is defined

$$\begin{aligned}
P_{U,X,H_{\text{h}}} \;&:=\; \Big(\bigcup \{\{(H,1) : H \subseteq H_{\text{h}}\%V_H,\ |H| = z_H\}^{\wedge} : \\
&\qquad\qquad\qquad V_H \subseteq V_{\text{h}},\ z_H \in \{1 \ldots z_{\text{h}}\}\}\Big)^{\wedge} \cup \\
&\qquad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \not\subseteq H_{\text{h}}\%V_H\} \cup \{(\emptyset,0)\} \\
&=\; \{(H, 1/(z_{\text{h}} 2^{v_{\text{h}}} \tbinom{z_{\text{h}}}{z_H})) : H \in \mathcal{H}_{U,X},\ H \subseteq H_{\text{h}}\%V_H,\ H \neq \emptyset\} \cup \\
&\qquad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \not\subseteq H_{\text{h}}\%V_H\} \cup \{(\emptyset,0)\}
\end{aligned}$$

where $V_H = \text{vars}(H)$, $z_H = |H|$, $\hat{X} = \text{normalise}(X)$ and $\text{normalise}(\emptyset) = \emptyset$.

The *historically distributed history probability*, $P_{U,X,H_{\text{h}}}(H)$, is the *probability* of *drawing* the *history* $H \subseteq H_{\text{h}}\%V_H$ of arbitrary *variables* $V_H \subseteq V_{\text{h}}$ and *size* $z_H \in \{1 \ldots z_{\text{h}}\}$ from *distribution history* $H_{\text{h}} \in \mathcal{H}_{U,X}$. All subsets of the *distribution history* for a given set of *variables* and *size* are defined as equally probable,

$$\forall V \subseteq V_{\text{h}}\ \forall H, G \subseteq H_{\text{h}}\%V\ (|G| = |H| \implies P_{U,X,H_{\text{h}}}(G) = P_{U,X,H_{\text{h}}}(H))$$

Note that this definition does not assume that the subsets of the *distribution history*, $\text{P}(H_{\text{h}}\%V_H)$, are equally probable, $P_{U,X,H_{\text{h}}} \neq \{(H,1) : V_H \subseteq$

$V_{\mathrm{h}},\ H \subseteq H_{\mathrm{h}}\%V_H\}^{\wedge}$. Equi-probable subsets would imply that there is a modal *sample size* at $z_{\mathrm{h}}/2$. Here it is assumed that there is no constraint on the *sample size* other than it is non-zero and less than or equal to the *distribution size*, $1 \leq z_H \leq z_{\mathrm{h}}$. So *sizes* are defined as equi-probable, $\forall z \in \{1 \ldots z_{\mathrm{h}}\}\ (\sum(P_{U,X,H_{\mathrm{h}}}(H) : H \in \mathcal{H}_{U,X},\ |H| = z) = 1/z_{\mathrm{h}})$.

Now for arbitrary non-empty *drawn history* $H \subseteq H_{\mathrm{h}}\%V_H$, the *historical probability* of *drawing without replacement* its *histogram* $A_H = \mathrm{histogram}(H)$ from the *distribution histogram* $E_{\mathrm{h}} = \mathrm{histogram}(H_{\mathrm{h}})$, is the expected *historically distributed history probability* of the *histogram*, A_H, times the normalising factor,

$$\hat{Q}_{\mathrm{h}}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = z_{\mathrm{h}} 2^{v_{\mathrm{h}}} \sum(P_{U,X,H_{\mathrm{h}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *historical distribution* is

$$Q_{\mathrm{h}}(E, z)(A) \;=\; \prod_{S \in A^{\mathrm{S}}} \binom{E_S}{A_S} = \prod_{S \in A^{\mathrm{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!} \in \mathbf{N}_{>0}$$

and the *historical probability distribution* is normalised,

$$\hat{Q}_{\mathrm{h}}(E, z)(A) \;=\; Q_{\mathrm{h}}(E, z)(A)/\tbinom{z_E}{z}$$

The *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},U}$, can equally well be expressed in terms of the *historically distributed history probability function*, $P_{U,X,H_{\mathrm{h}}}$,

$$\hat{Q}_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = z_{\mathrm{h}} 2^{v_{\mathrm{h}}} \sum(P_{U,X,H_{\mathrm{h}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *distribution histogram* is *complete*, $E_{\mathrm{h}} = \mathrm{histogram}(H_{\mathrm{h}}) + V_{\mathrm{h}}^{\mathrm{CZ}} \in \mathcal{A}_{U,\mathrm{i},V_{\mathrm{h}},z_{\mathrm{h}}}$, the *histograms* are *complete*, $A_H = \mathrm{histogram}(H) + V_H^{\mathrm{CZ}} \in \mathcal{A}_{U,\mathrm{i},V_H,z_H}$, the *stuffed historical distribution*, $Q_{\mathrm{h},U}$, is defined

$$\begin{aligned} Q_{\mathrm{h},U}(E, z) \;=\; & \{(A + A^{\mathrm{CZ}}, f) : (A, f) \in Q_{\mathrm{h}}(E, z)\} \cup \\ & (\mathcal{A}_{U,\mathrm{i},V,z} \setminus \{A + A^{\mathrm{CZ}} : A \in \mathrm{dom}(Q_{\mathrm{h}}(E, z))\}) \times \{0\} \end{aligned}$$

and the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},U}$, is defined $\hat{Q}_{\mathrm{h},U}(E, z) := \mathrm{normalise}(Q_{\mathrm{h},U}(E, z))$.

In *classical induction* it is assumed that the *history probability function*, P, is *historically distributed*,

$$P = P_{U,X,H_{\mathrm{h}}}$$

where the *distribution history*, H_{h}, is *unknown*, but there exists a non-empty *observation* or *sample history* $H_{\mathrm{o}} \subseteq H_{\mathrm{h}}\%V_{\mathrm{o}}$ of *known size* $z_{\mathrm{o}} = |H_{\mathrm{o}}| > 0$ in *known variables* $V_{\mathrm{o}} \subseteq V_{\mathrm{h}}$ that has a *known complete histogram* $A_{\mathrm{o}} = \mathrm{histogram}(H_{\mathrm{o}}) + V_{\mathrm{o}}^{\mathrm{CZ}}$. The *system*, U, is *known* at least for the *observation variables*, V_{o}. The *distribution history*, H_{h}, is *unknown*, so the *historically distributed history probability* of the *sample history*, $P_{U,X,H_{\mathrm{h}}}(H_{\mathrm{o}})$, is also *unknown*, except that it is non-zero,

$$P_{U,X,H_{\mathrm{h}}}(H_{\mathrm{o}}) > 0$$

because the *sample history* exists. The *complete distribution histogram*, $E_{\mathrm{h}} = \mathrm{histogram}(H_{\mathrm{h}}) + V_{\mathrm{h}}^{\mathrm{CZ}}$, is *unknown*, so the *stuffed historical probability* of the *sample histogram*, $\hat{Q}_{\mathrm{h},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$, is also *unknown*, except that is non-zero, $\hat{Q}_{\mathrm{h},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}}) > 0$. In order to estimate the *distribution histogram*, E_{h}, and hence the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})$, and the *historically distributed history probability function*, $P_{U,X,H_{\mathrm{h}}}$, the *likelihood function* for the *probability distribution* must be defined. See appendix 'Likelihood functions and Fisher information', below.

First make the further *induction* assumption that the *distribution history size*, z_{h}, is large with respect to the *sample size* $z_{\mathrm{o}} = \mathrm{size}(A_{\mathrm{o}})$, so that, in the limit, the *stuffed historical probability*, $\hat{Q}_{\mathrm{h},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$, approximates to the *generalised multinomial probability*, $\hat{Q}_{\mathrm{m},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$. That is, if $z_{\mathrm{o}} \ll z_{\mathrm{h}}$ then

$$\hat{Q}_{\mathrm{h},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \approx \hat{Q}_{\mathrm{m},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

where $E_{\mathrm{o}} = E_{\mathrm{h}}\%V_{\mathrm{o}}$ and the *generalised multinomial probability* is

$$\hat{Q}_{\mathrm{m},U}(E, z)(A) \;=\; \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} \hat{E}_S^{A_S}$$

where *integral substrate histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$ is *drawn with replacement* from $E \in \mathcal{A}_{U,V,z_E}$.

The *maximum likelihood estimate* and the *Fisher information* of the *generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m},U}$, are well defined, but may also be considered by noting that this *distribution* approximates to the *generalised multiple binomial probability distribution*, $\hat{Q}_{\mathrm{b},U}$,

$$\hat{Q}_{\mathrm{m},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \approx \hat{Q}_{\mathrm{b},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

where the *generalised multiple binomial probability distribution* is defined

$$\hat{Q}_{\mathrm{b},U}(E, z)(A) \;=\; \prod_{S \in V^{\mathrm{CS}}} \binom{z}{A_S} \hat{E}_S^{A_S} (1 - \hat{E}_S)^{z - A_S}$$

where *multiple support histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,\{0\ldots z\}}$ is *drawn with replacement* from $E \in \mathcal{A}_{U,V,z_E}$. The approximation is best when the *entropy* of the *distribution histogram*, $\mathrm{entropy}(E)$, is high.

Likelihood functions are parameterised by a real tuple or coordinate, $\mathbf{R}^n$. In order to construct a coordinate from a *histogram* define $()^{\square} \in \mathcal{A} \to \mathcal{L}(\mathbf{Q}_{\geq 0})$ as

$$A^{\square} := \{(i,c) : ((S,c),i) \in \mathrm{order}(D_{\mathcal{S}\times\mathbf{Q}}, A)\}$$

where $D_{\mathcal{S}}$ is an *order* on the *states*. If A is *complete*, $A^{\mathrm{U}} = A^{\mathrm{C}}$, then $A^{\square} \in \mathbf{R}^v$, where $v = |A^{\mathrm{C}}|$.

The *multiple binomial parameterised probability density function* $\mathrm{mbppdf}(z) \in \mathrm{ppdfs}(v,v)$, where $v = |V^{\mathrm{C}}|$, is defined

$$\begin{aligned}
&\mathrm{mbppdf}(z)(E) := \\
&\quad \{(A, \prod_{i \in \{1\ldots v\}} \frac{z!}{\Gamma_!(A_i)\ \Gamma_!(z-A_i)} E_i^{A_i} (1-E_i)^{z-A_i}) : A \in \mathbf{R}^v_{[0,z]}\} \cup \\
&\quad (\mathbf{R}^v \setminus \mathbf{R}^v_{[0,z]}) \times \{0\}
\end{aligned}$$

where $z \in \mathbf{N}_{>0}$, $E \in \mathbf{R}^v_{(0,1)}$ and $\Gamma_!$ is the unit-translated gamma function. The *multiple binomial likelihood function* $\mathrm{mblf}(z) \in \mathrm{lfs}(v,v)$ is defined

$$\mathrm{mblf}(z)(A) \;:=\; \{(E, \mathrm{mbppdf}(z)(E)(A)) : E \in \mathbf{R}^v_{(0,1)}\}$$

where $A \in \mathbf{R}^v$.

These definitions only require that each parameter is in the open set between zero and one, $E \in \mathbf{R}^v_{(0,1)} = \{r : r \in \mathbf{R},\ 0 < r < 1\}^v$, so E is not necessarily a *probability function*. That is, in some cases $E \neq \hat{E} \in \mathcal{P}$. This is to allow well defined partial derivatives in free parameters. So $\partial_i(\mathrm{mblf}(z)(A))(E)$ is the sensitivity of the *likelihood* to the i-th parameter at E, where $\partial_j \in (\mathcal{L}(\mathbf{R}) \to \mathbf{R}) \to (\mathcal{L}(\mathbf{R}) \to \mathbf{R})$ is defined $\partial_j(F) := \{(Z, \partial F(Z)/\partial Z_j) : Z \in \mathrm{dom}(F)\}$ and F is a continuous function.

In the case where the *volume* is at least two, $v > 1$, and the *distribution histogram* is *completely effective*, $E^{\mathrm{F}} = V^{\mathrm{C}} \implies \hat{E}^{\square} \in \mathbf{R}^v_{(0,1)}$, the *multiple binomial parameterised probability density* and the *multiple binomial likelihood* equals the *generalised multiple binomial probability*,

$$\mathrm{mbppdf}(z)(\hat{E}^{\square})(A^{\square}) = \mathrm{mblf}(z)(A^{\square})(\hat{E}^{\square}) = \hat{Q}_{\mathrm{b},U}(E,z)(A)$$

As shown in the appendix, the *binomial parameterised probability density function*, $\mathrm{bppdf}(n)(p)$, is defined

$$\mathrm{bppdf}(n)(p)(k) \quad := \quad \frac{n!}{\Gamma_! k \ \Gamma_!(n-k)} p^k (1-p)^{n-k} \in \mathbf{R}_{(0,1)}$$

and the corresponding *likelihood function* is $\mathrm{blf}(n)(k)(p) := \mathrm{bppdf}(n)(p)(k)$, where $n > 0$ and $0 < p < 1$. Given observation coordinate $k_{\mathrm{o}} \in \mathbf{R}_{(0,n)}$ the *maximum likelihood estimate* for the parameter of the *probability density function* is the modal *likelihood*, $\{\tilde{p}\} = \mathrm{maxd}(\mathrm{blf}(n)(k_{\mathrm{o}}))$, which is $\tilde{p} = k_{\mathrm{o}}/n$. Here the gradient of the *likelihood function* is zero, $\mathrm{d}(\mathrm{blf}(n)(k_{\mathrm{o}}))(\tilde{p}) = \mathrm{d}(\mathrm{blf}(n)(k_{\mathrm{o}}))(k_{\mathrm{o}}/n) = 0$, where $\mathrm{d} \in (\mathbf{R} \to \mathbf{R}) \to (\mathbf{R} \to \mathbf{R})$ is defined $\mathrm{d}(F) := \{(x, dF(x)/dx) : x \in \mathrm{dom}(F)\}$.

The *multiple binomial parameterised probability density function*, $\mathrm{mbppdf}(z)$, is the product of a set of independent *binomial parameterised probability density functions*, $\mathrm{bppdf}(z)$,

$$\mathrm{mbppdf}(z)(E)(A) = \prod_{i \in \{1 \ldots v\}} \mathrm{bppdf}(z)(E_i)(A_i)$$

and so, given non-singleton *volume*, $v_{\mathrm{o}} = |V_{\mathrm{o}}^{\mathrm{C}}| > 1$, and a *completely effective sample histogram*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}} \implies \hat{A}_{\mathrm{o}}^{\square} \in \mathbf{R}_{(0,1)}^{v_{\mathrm{o}}}$, the *maximum likelihood estimate* is $\tilde{E}_{\mathrm{o}}^{\square} = \hat{A}_{\mathrm{o}}^{\square}$, where $\{\tilde{E}_{\mathrm{o}}^{\square}\} = \mathrm{maxd}(\mathrm{mblf}(z_{\mathrm{o}})(A_{\mathrm{o}}^{\square}))$. Thus, in *classical induction*, in the case of *completely effective sample histogram*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}} \implies E_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the *maximum likelihood estimate* $\tilde{E}_{\mathrm{o}} \in \mathcal{A}_{U,V_{\mathrm{o}},1}$ of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *generalised multiple binomial probability distribution*, $\hat{Q}_{\mathrm{b},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is

$$\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

The *maximum likelihood estimate* in this case is a rational-valued function, $\tilde{E}_{\mathrm{o}}^{\square} = \hat{A}_{\mathrm{o}}^{\square} \in \mathbf{N} \to \mathbf{Q}_{\geq 0}$, so the *maximum likelihood estimate* can also be written as the maximisation of the *complete congruent histograms* of unit *size*,

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{b},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

The *maximum likelihood estimate* is not computable as a maximisation. Although the *substrate histograms* are countably infinite, $\mathcal{A}_{U,V_{\mathrm{o}},1} \leftrightarrow \mathbf{N}$, the maximisation never terminates. An approximation to the continuous case may be made by using a scaling factor. The *scaled complete integral congruent histograms* equals the *complete congruent histograms* in the limit

$$\lim_{k \to \infty} \{A/Z_k : A \in \mathcal{A}_{U,\mathrm{i},V,kz}\} = \mathcal{A}_{U,V,z}$$

where $k \in \mathbf{N}_{>0}$ and $Z_k = \text{scalar}(k)$. The finite approximation to the *maximum likelihood estimate* is

$$\{\tilde{E}_{\text{o}}\} \approx \text{maxd}(\{(E/Z_k, \hat{Q}_{\text{b},U}(E/Z_k, z_{\text{o}})(A_{\text{o}})) : E \in \mathcal{A}_{U,\text{i},V_{\text{o}},k}\})$$

The approximation improves as the scaling factor, k, increases.

The *normalised mean* of the *generalised multiple binomial probability distribution* at the *maximum likelihood estimate* equals the *maximum likelihood estimate* of the *distribution histogram*,

$$\text{normalise}(\text{mean}(\hat{Q}_{\text{b},U}(\tilde{E}_{\text{o}}, z_{\text{o}}))) = \tilde{E}_{\text{o}} = \hat{A}_{\text{o}}$$

The *multinomial parameterised probability density function* $\text{mppdf}(z) \in \text{ppdfs}(v, v)$, where $v \in \mathbf{N}_{>0}$, is defined

$$\begin{aligned}
&\text{mppdf}(z)(E) := \\
&\quad \{(A, \frac{n!}{\prod_{i \in \{1 \ldots v\}} \Gamma_! A_i} \prod_{i \in \{1 \ldots v\}} E_i^{A_i}) : A \in \mathbf{R}^v_{[0,z]},\ \sum_{i \in \{1 \ldots v\}} A_i = z\} \cup \\
&\quad \{(A, 0) : A \in \mathbf{R}^v_{[0,z]},\ \sum_{i \in \{1 \ldots v\}} A_i \neq z\} \cup \\
&\quad (\mathbf{R}^v \setminus \mathbf{R}^v_{[0,z]}) \times \{0\}
\end{aligned}$$

where $z \in \mathbf{N}_{>0}$, $E \in \mathbf{R}^v_{(0,1)}$ and $\sum_{i \in \{1 \ldots v\}} E_i = 1$, otherwise $\text{mppdf}(z)(E)$ is undefined.

The *multinomial likelihood function* $\text{mlf}(z) \in \text{lfs}(v, v)$ is defined

$$\text{mlf}(n)(A) \ := \ \{(E, \text{mppdf}(z)(E)(A)) : E \in \mathbf{R}^v_{(0,1)}\}$$

where $A \in \mathbf{R}^v$. Note that the *multinomial likelihood function* only requires that each parameter is in the open set between zero and one, $E \in \mathbf{R}^v_{(0,1)} = \{r : r \in \mathbf{R},\ 0 < r < 1\}^v$, so E is not necessarily a *probability function*. That is, in some cases $E \neq \hat{E} \notin \mathcal{P}$.

In the case where the *volume* is at least two, $v > 1$, and the *distribution histogram* is *completely effective*, $E^{\text{F}} = V^{\text{C}} \implies \hat{E}^{[]} \in \mathbf{R}^v_{(0,1)}$, the *multinomial parameterised probability density* and the *multinomial likelihood* equals the *generalised multinomial probability*,

$$\text{mppdf}(z)(\hat{E}^{[]})(A^{[]}) = \text{mlf}(z)(A^{[]})(\hat{E}^{[]}) = \hat{Q}_{\text{m},U}(E, z)(A)$$

The *maximum likelihood estimate* for the parameter of the *multinomial parameterised probability density function*, $\{\tilde{E}_{\mathrm{o}}^{\square}\} = \mathrm{maxd}(\mathrm{mlf}(z_{\mathrm{o}})(A_{\mathrm{o}}^{\square}))$, is equal to the *maximum likelihood estimate* for the parameter of the *multiple binomial parameterised probability density function*,

$$\{\tilde{E}_{\mathrm{o}}^{\square}\} = \mathrm{maxd}(\mathrm{mlf}(z_{\mathrm{o}})(A_{\mathrm{o}}^{\square})) = \mathrm{maxd}(\mathrm{mblf}(z_{\mathrm{o}})(A_{\mathrm{o}}^{\square}))$$

That is, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$.

Again, the *maximum likelihood estimate* can also be written as the maximisation of the *complete congruent histograms* of unit *size*,

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

and the *normalised mean* of the *generalised multinomial probability distribution* at the *maximum likelihood estimate* equals the *maximum likelihood estimate* of the *distribution histogram*,

$$\mathrm{normalise}(\mathrm{mean}(\hat{Q}_{\mathrm{m},U}(\tilde{E}_{\mathrm{o}}, z_{\mathrm{o}}))) = \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

In the *multinomial distribution* one of the *states* is not free, because $\mathrm{sum}(\hat{A}_{\mathrm{o}}) = z_{\mathrm{o}}$, but the *maximum likelihood estimate* remains constrained to *completely effective sample histogram*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$. This would be the case even if the *distribution history size*, z_{h}, were *known*.

Finally, the *maximum likelihood estimate* for the parameter of the *historical parameterised probability density function* corresponding to the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},U}$, is conjectured to be the same as the *maximum likelihood estimate* for the parameter of the *multinomial parameterised probability density function*, $\mathrm{maxd}(\mathrm{mlf}(z_{\mathrm{o}})(A_{\mathrm{o}}^{\square}))$. That is, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$.

To conclude, in *classical induction* where (i) the *history probability function* is *historically distributed*, $P = P_{U,X,H_{\mathrm{h}}}$, (ii) the *volume* is non-singleton, $v_{\mathrm{o}} > 1$, and (iii) the *sample histogram* is *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, is simply estimated to be equal to the *sample probability histogram*, $\hat{A}_{\mathrm{o}}$,

$$\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

Consider the *maximum likelihood estimate* of the *generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m},U}(E,z)$, in the case where the *distribution* is the *maximum likelihood estimate*, $\tilde{E} = \hat{A}$. The logarithm of the *generalised multinomial probability* is

$$\ln \hat{Q}_{\mathrm{m},U}(A,z)(A) = \ln z! - z \ln z - \sum_{S \in A^{\mathrm{S}}} \ln A_S! + \sum_{S \in A^{\mathrm{FS}}} A_S \ln A_S$$

Applying Stirling's approximation, $\ln n! = n \ln n - n + \mathrm{O}(\ln n)$, the *log likelihood* varies against the sum of the logarithm of the *histogram*

$$\begin{aligned} \ln \hat{Q}_{\mathrm{m},U}(A,z)(A) &\sim -\sum_{S \in A^{\mathrm{FS}}} \ln A_S \\ &= -\mathrm{sum}(\ln(A)) \end{aligned}$$

where $\ln \in (\mathcal{X} \to \mathbf{Q}) \to (\mathcal{X} \to \ln \mathbf{Q}_{>0})$ is defined as $\ln(X) := \{(x, \ln q) : (x,q) \in X,\ q > 0\}$. This is to say that the *likelihood* varies against the product of the *counts* of the *histogram*,

$$\hat{Q}_{\mathrm{m},U}(A,z)(A) \sim 1/\prod_{S \in A^{\mathrm{FS}}} A_S$$

The sum of the logarithm of the *histogram* varies with the *entropy* of the *histogram*,

$$\mathrm{sum}(\ln(A)) \sim \mathrm{entropy}(A)$$

The so the *log-likelihood* varies against the *histogram entropy*,

$$\ln \hat{Q}_{\mathrm{m},U}(A,z)(A) \sim -\mathrm{entropy}(A)$$

Note that the *entropy* is not scaled by the *size*.

In *classical induction* where (i) the *history probability function* is *historically distributed*, $P = P_{U,X,H_{\mathrm{h}}}$, (ii) the *volume* is non-singleton, $v_{\mathrm{o}} > 1$, and (iii) the *sample histogram* is *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iv) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (v) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$ where $A_{\mathrm{o},z_{\mathrm{h}}} = \mathrm{scalar}(z_{\mathrm{h}}) * \hat{A}_{\mathrm{o}}$, then the *log likelihood* of the *stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the sum of the logarithm of the *sample histogram*

$$\ln \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim -\mathrm{sum}(\ln(A_{\mathrm{o}}))$$

and (b) varies against the *sample entropy*

$$\ln \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim -\ \mathrm{entropy}(A_{\mathrm{o}})$$

The *Fisher information* of the parameter p of the *binomial parameterised probability density function*, $\mathrm{bppdf}(n)(p)$, is the second moment of the *log-likelihood* sensitivity,

$$\begin{aligned} I_{\mathrm{bppdf}(n)}(p) &:= \int_0^n (\mathrm{d}(\ln \circ \mathrm{blf}(n)(k))(p))^2 \times \mathrm{bppdf}(n)(p)(k)\ dk \\ &= \frac{n}{p(1-p)} \end{aligned}$$

where $n > 0$ and $0 < p < 1$. The *Fisher information* of the parameter, $I_{\mathrm{bppdf}(n)}(p)$, is minimised where $p = 0.5$. In this case the *Fisher information* is $I_{\mathrm{bppdf}(n)}(0.5) = 4n$. If an *observation coordinate* is $k_{\mathrm{o}} = n/2$, then the *maximum likelihood estimate*, $\tilde{p} = k_{\mathrm{o}}/n = 0.5$, minimises the *Fisher information*. The *Fisher information* is maximised at the extremes of the parameter. As the parameter, p, tends to zero or one, the *Fisher information* tends to infinity. The *Fisher information* is proportional to the size, n.

The *multiple binomial parameterised probability density function*, $\mathrm{mbppdf}(z)$, is the product of a set of independent *binomial parameterised probability density functions*, $\mathrm{bppdf}(z)$, so the *Fisher information* of the *multiple binomial parameterised probability density function* is the sum,

$$\begin{aligned} I_{\mathrm{mbppdf}(z)}(E) &= \sum_{i \in \{1 \ldots v\}} I_{\mathrm{bppdf}(z)}(E_i) \\ &= \sum_{i \in \{1 \ldots v\}} \frac{z}{E_i(1-E_i)} \end{aligned}$$

where $z \in \mathbf{N}_{>0}$ and $E \in \mathbf{R}^v_{(0,1)}$.

The *sum sensitivity* of the *generalised multiple binomial probability distribution*, $\hat{Q}_{\mathrm{b},U}(E,z)$, to the *distribution histogram*, E, is defined as the *Fisher information* of the *multiple binomial parameterised probability density function*, $\mathrm{mbppdf}(z)$. Define the *sensitivity* of a *state* for a *complete distribution* as $\mathrm{sensitivity}(U) \in \mathcal{Q}_U \to (\mathcal{S}_U \to \mathbf{R}_{\geq 0})$, and for the *generalised multiple binomial probability distribution* as

$$\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{b},U}(E,z))(S) := \frac{z}{\hat{E}_S(1-\hat{E}_S)}$$

where $\hat{E}_S \notin \{0, 1\}$. The *sum sensitivity* is

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{b},U}(E, z))) := \sum_{S \in E^{\text{FS}}} \frac{z}{\hat{E}_S(1 - \hat{E}_S)}$$

where $|E^{\text{F}}| > 1$.

In the case of non-singleton *volume*, $v > 1$, and *completely effective distribution histogram*, $E^{\text{F}} = V^{\text{C}}$, the *sensitivity* is equal to the *Fisher information*,

$$\text{sensitivity}(U)(\hat{Q}_{\text{b},U}(E, z))(S) := I_{\text{bppdf}(z)}(\hat{E}_i^{\square}) = \frac{z}{\hat{E}_S(1 - \hat{E}_S)}$$

where $S \in V^{\text{CS}}$ and $i \in \{1 \dots v\}$ corresponds to S. The *sum sensitivity* is

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{b},U}(E, z))) := I_{\text{mbppdf}(z)}(\hat{E}^{\square}) = \sum_{S \in V^{\text{CS}}} \frac{z}{\hat{E}_S(1 - \hat{E}_S)}$$

The *sensitivity* of the *generalised multinomial probability distribution*, $\hat{Q}_{\text{m},U}(E, z)$, is conjectured to be equal to *sensitivity* of the *generalised multiple binomial probability distribution*, $\hat{Q}_{\text{b},U}(E, z)$,

$$\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(E, z))(S) = \text{sensitivity}(U)(\hat{Q}_{\text{b},U}(E, z))(S) = \frac{z}{\hat{E}_S(1 - \hat{E}_S)}$$

and hence the *sum sensitivities* are equal

$$\begin{aligned} \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(E, z))) &= \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{b},U}(E, z))) \\ &= \sum_{S \in V^{\text{CS}}} \frac{z}{\hat{E}_S(1 - \hat{E}_S)} \end{aligned}$$

The *sum sensitivity* varies with *size*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(E, z))) \sim z$$

The *variance* of *state* S in the *generalised multinomial probability distribution* is

$$\text{var}(U)(\hat{Q}_{\text{m},U}(E, z))(S) = z\hat{E}_S(1 - \hat{E}_S)$$

The *sum variance* is shown in 'Multinomial distributions', above, to vary with the *scaled entropy*,

$$\text{sum}(\text{var}(U)(\hat{Q}_{\text{m},U}(E, z))) \sim z \times \text{entropy}(E)$$

The *sensitivity* varies against the *variance*,

$$\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E,z))(S) \sim -\mathrm{var}(U)(\hat{Q}_{\mathrm{m},U}(E,z))(S)$$

so the *sum sensitivity* varies against the *scaled entropy*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E,z))) \quad \sim \quad -z \times \mathrm{entropy}(E)$$

The *entropy* is maximised, and the *sum sensitivity* minimised, when the *distribution histogram* is *uniform*, $E = V^{\mathrm{C}}$,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(V^{\mathrm{C}},z))) = \frac{v^2 z}{(v-1)}$$

where $v = |V^{\mathrm{C}}| > 1$. For large *volume*, $v \gg 1$, the *uniform sum sensitivity* is asymptotically proportionate to the *volume*, v. That is, *sample histograms* that have large *volumes* and *sizes* tend to be more *sensitive* to the *distribution histogram*, E, than smaller *sample histograms*.

The *sensitivity* of the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},U}(E,z)$, is conjectured to vary with the *sensitivity* of the *generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m},U}(E,z)$, and hence the *sum sensitivities* also vary together

$$\begin{aligned}\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(E,z))) &\sim \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E,z))) \\ &= \sum_{S \in V^{\mathrm{CS}}} \frac{z}{\hat{E}_S(1-\hat{E}_S)}\end{aligned}$$

As the *distribution history size* exceeds the *sample size*, $z_E \gg z$, in the limit the *sum sensitivity* of the *stuffed historical probability distribution* tends to equal the *sum sensitivity* of the *generalised multinomial probability distribution*,

$$\lim_{z_E \to \infty} \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(E,z))) = \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E,z)))$$

In *classical induction* where (i) the *history probability function* is *historically distributed*, $P = P_{U,X,H_{\mathrm{h}}}$, (ii) the *volume* is non-singleton, $v_{\mathrm{o}} > 1$, and (iii) the *sample histogram* is *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},U}(E_{\mathrm{o}},z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iv) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (v) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then (a) the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate* is approximately

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \approx \quad \sum_{S \in V_{\mathrm{o}}^{\mathrm{CS}}} \frac{z_{\mathrm{o}}}{\hat{A}_{\mathrm{o}}(S)\ (1 - \hat{A}_{\mathrm{o}}(S))}$$

(b) the *sum sensitivity* varies against the *size* scaled *entropy*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad -z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}})$$

The *sum sensitivity* varies against the *size* scaled *entropy*, and the *log-likelihood* also varies against the *entropy*, albeit not *size* scaled, so conjecture that the *sum sensitivity* varies weakly with the *log-likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad \ln \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

This is consistent with the discussion in appendix 'Likelihood functions and Fisher information' where it is conjectured that, in some cases, the sensitivity of the *probability density function* to parameter at the *maximum likelihood estimate* varies with the *log-likelihood*.

In the discussion above of the *maximum likelihood estimate* and *sum sensitivity* in *classical induction*, the *sample histogram* is constrained to be *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$. This allows the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, to be made by deriving the *likelihood function* of the *historical parameterised probability density function* from the *binomial likelihood function*,

$$\mathrm{blf}(n)(k)(p) := \frac{n!}{\Gamma_! k\ \Gamma_!(n-k)} p^k (1-p)^{n-k}$$

which is only defined for non-zero, non-unit parameter, $0 < p < 1$.

Similarly, the *sum sensitivity* is derived from the *Fisher information* of the *binomial likelihood function*. The *Fisher information* tends to infinity in the limit,

$$\lim_{p \to 0} I_{\mathrm{bppdf}(n)}(p) = \lim_{p \to 0} \frac{n}{p(1-p)} = \infty$$

If it is the case that the *sample histogram* is neither *singleton* nor *completely effective*, $1 < |A_{\mathrm{o}}^{\mathrm{F}}| < v_{\mathrm{o}}$, then the coordinate has smaller, but not unit, dimension, $(\hat{A}_{\mathrm{o}} * A_{\mathrm{o}}^{\mathrm{F}})^{\square} \in \mathcal{R}_{(0,1)}^{|A_{\mathrm{o}}^{\mathrm{F}}|} \notin \{\{1\}, \mathcal{R}_{(0,1)}^{v_{\mathrm{o}}}\}$, and so the *maximum likelihood estimate* and *sum sensitivity* must be restricted to a subset of the *cartesian* *states*, $A_{\rm o}^{\rm FS} \subset V_{\rm o}^{\rm CS}$. The *maximum likelihood estimate* for an *incompletely effective non-singleton sample histogram* is then

$$(\tilde{E}_{\rm o} * A_{\rm o}^{\rm F})^\wedge = \hat{A}_{\rm o}$$

The *maximum likelihood estimate* for the *ineffective states*, $\tilde{E}_{\rm o} \setminus (\tilde{E}_{\rm o} * A_{\rm o}^{\rm F})$, remains *unknown*. In addition, the *effective* normalising factor, $1/\text{size}(\tilde{E}_{\rm o} * A_{\rm o}^{\rm F})$, is *unknown*.

Similarly, the approximation of the *sum sensitivity* is restricted to the *effective states*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{{\rm h},U}(A_{{\rm o},z_{\rm h}} * A_{\rm o}^{\rm F}, z_{\rm o}))) \;\approx\; \sum_{S \in A_{\rm o}^{\rm FS}} \frac{z_{\rm o}}{\hat{A}_{\rm o}(S)\ (1 - \hat{A}_{\rm o}(S))}$$

This may be an underestimate, however, because of the *unknown effective* normalisation. The *sum sensitivity* of the *ineffective states* is *unknown* because there is no *draw* from $\tilde{E}_{\rm o} \setminus (\tilde{E}_{\rm o} * A_{\rm o}^{\rm F})$.

The *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*, $\text{sum}(\text{sensitivity}(U)(\hat{Q}_{{\rm h},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})))$, can be related to queries on the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\rm o} = \hat{A}_{\rm o}$. Let *non-zero histogram* $Q \in \mathcal{A}_U$, be a query *histogram* in the *variables* $K = \text{vars}(Q)$ that are a subset of the *sample variables*, $K \subseteq V_{\rm o}$. The normalisation of the query *histogram* is a *probability histogram*, $\hat{Q} \in \mathcal{A} \cap \mathcal{P}$. The *estimated transform induced* from the *maximum likelihood estimate*, $\hat{A}_{\rm o}$, for the query *variables*, K, is $T_{\hat{A}_{\rm o},K} = (\hat{A}_{\rm o}, (V_{\rm o} \setminus K)) \in \mathcal{T}$. The *estimated transformed product* is $\hat{Q} * T_{\hat{A}_{\rm o},K} = \hat{Q} * (\hat{A}_{\rm o}, (V_{\rm o} \setminus K)) \in \mathcal{A} \cap \mathcal{P}'$. The *estimated conditional transform induced* from $\hat{A}_{\rm o}$ and K is $T'_{\hat{A}_{\rm o},K} = (\hat{A}_{\rm o}/(\hat{A}_{\rm o}\%K), (V_{\rm o} \setminus K)) \in \mathcal{T}$. In the case where the *sample histogram* is *completely effective*, $A_{\rm o}^{\rm F} = V_{\rm o}^{\rm C} \implies Q^{\rm F} \leq (A_{\rm o}\%K)^{\rm F}$, the *estimated transformed conditional product* is a *probability histogram*,

$$\hat{Q} * T'_{\hat{A}_{\rm o},K} = \hat{Q} * (\frac{\hat{A}_{\rm o}}{\hat{A}_{\rm o}\%K}, (V_{\rm o} \setminus K)) \in \mathcal{A} \cap \mathcal{P}$$

The *sum sensitivity* is a property of the *distribution*, so, in the case where the query *histogram* consists of one *effective state*, $|Q^{\rm F}| = 1 \implies Q + K^{\rm CZ} \in \mathcal{A}_{U,{\rm i},K,1}$, (i) *expand* the query *histogram* to the *sample variables*, $V_{\rm o}$, and (ii) *scale* the *expanded* query *histogram* to the *sample size*, $z_{\rm o}$. Now the *estimated* *transformed conditional product* can be rewritten in terms of a *draw* of the *sample size*, z_{o}, from the *distribution histogram*

$$\begin{aligned}\hat{Q} * T'_{\hat{A}_{\mathrm{o}},K} &= \{(N, \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, 1)(\hat{Q} * \{N\}^{\mathrm{U}} + V_{\mathrm{o}}^{\mathrm{CZ}})) : N \in (V_{\mathrm{o}} \setminus K))^{\mathrm{CS}}\}^{\wedge} \\ &= \{(N, (\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(Z_{\mathrm{o}} * \hat{Q} * \{N\}^{\mathrm{U}} + V_{\mathrm{o}}^{\mathrm{CZ}}))^{1/z_{\mathrm{o}}}) \\ &\qquad\qquad : N \in (V_{\mathrm{o}} \setminus K))^{\mathrm{CS}}\}^{\wedge}\end{aligned}$$

where $Z_{\mathrm{o}} = \mathrm{scalar}(z_{\mathrm{o}})$.

The *application* of the *model induced* from the *maximum likelihood estimate* to the query *histogram*, $\hat{Q}*T'_{\hat{A}_{\mathrm{o}},K}$, can be viewed as a *probability function* of label, $T'(\hat{A}_{\mathrm{o}}, \hat{Q}) \in ((V_{\mathrm{o}} \setminus K)^{\mathrm{CS}} \to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$, parameterised by (i) the *maximum likelihood estimate*, $\hat{A}_{\mathrm{o}}$, and (ii) the query *histogram*, Q. The *model* application is relatively independent of the query *state*, S_Q where $\{S_Q\} = Q^{\mathrm{FS}}$, and query *variables*, K. The *model* application depends on the geometric scaling of the *historical distribution*, $\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$, so the *query sensitivity* to the *distribution histogram* varies with the *sum sensitivity* of the *historical distribution* at the *maximum likelihood estimate* divided by the *sample size*,

$$\begin{aligned}\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}} &\approx \sum_{S \in V_{\mathrm{o}}^{\mathrm{CS}}} \frac{1}{\hat{A}_{\mathrm{o}}(S)\ (1 - \hat{A}_{\mathrm{o}}(S))} \\ &\sim -\ \mathrm{entropy}(A_{\mathrm{o}})\end{aligned}$$

That is, as the *sample entropy* increases, the *sum sensitivity* of the query to the *model* implied by the *sample* decreases.

If it is *known* that the label *variables* are a function of the query *variables*,

$$\mathrm{split}(K, E_{\mathrm{o}}^{\mathrm{FS}}) \in K^{\mathrm{CS}} \to (V_{\mathrm{o}} \setminus K)^{\mathrm{CS}}$$

then the *distribution histogram*, E_{o}, is *known* to be *ineffective*, $E_{\mathrm{o}}^{\mathrm{F}} < V_{\mathrm{o}}^{\mathrm{C}}$, and so the *sample histogram* cannot be *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} \neq V_{\mathrm{o}}^{\mathrm{C}}$. However, the *effectiveness* of the *distribution histogram* in the query *variables* is not *known*, $(E_{\mathrm{o}}\%K)^{\mathrm{F}} \leq K^{\mathrm{C}}$, unless the *sample histogram* is *completely effective* in the query *variables*, $(A_{\mathrm{o}}\%K)^{\mathrm{F}} = K^{\mathrm{C}} \implies (E_{\mathrm{o}}\%K)^{\mathrm{F}} = K^{\mathrm{C}}$. So if the *sample histogram* is *ineffective* in the query *variables*, $(A_{\mathrm{o}}\%K)^{\mathrm{F}} < K^{\mathrm{C}}$, then there still exists an *unknown* normalising factor,

$$1/\mathrm{size}(\tilde{E}_{\mathrm{o}} * A_{\mathrm{o}}^{\mathrm{F}}\ \%\ K) = 1/\mathrm{size}(\tilde{E}_{\mathrm{o}} * A_{\mathrm{o}}^{\mathrm{F}})$$

In this case the *maximum likelihood estimate* remains restricted,

$$\begin{aligned}(\tilde{E}_{\mathrm{o}} * A_{\mathrm{o}}^{\mathrm{F}}\ \%\ K)^{\wedge} &= \hat{A}_{\mathrm{o}}\ \%\ K \\ (\tilde{E}_{\mathrm{o}} * A_{\mathrm{o}}^{\mathrm{F}})^{\wedge} &= \hat{A}_{\mathrm{o}}\end{aligned}$$

and the *sum sensitivity* may be an underestimate.

5.4 Classical independent induction

In *classical induction* it is assumed that the *history probability function* is *historically distributed*, $P = P_{U,X,H_{\mathrm{h}}}$. Consider the related case of *classical independent induction* where all *drawn histories* are *known* to be *independent*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{x}}$, where the *independent historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{x}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$, is

$$\begin{aligned} P_{U,X,H_{\mathrm{h}},\mathrm{x}} \;:=\; & \Big(\bigcup\{\{(H,1) : H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H,\ A_H = A_H^{\mathrm{X}}\}^{\wedge} : \\ & \qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\ & \{(H,0) : H \in \mathcal{H}_{U,X},\ A_H \neq A_H^{\mathrm{X}}\} \cup \\ & \{(H,0) : H \in \mathcal{H}_{U,X},\ H \nsubseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset,0)\} \end{aligned}$$

That is, *drawn histories* are *necessarily independent*, $\forall H \in \mathcal{H}_{U,X}\ (P_{U,X,H_{\mathrm{h}},\mathrm{x}}(H) > 0 \implies A_H = A_H^{\mathrm{X}})$.

Now, since *drawing* the *distribution history*, H_{h}, itself is always possible, $P_{U,X,H_{\mathrm{h}},\mathrm{x}}(H_{\mathrm{h}}) > 0$, the *distribution histogram* is *known* to be *independent*, $E_{\mathrm{h}} = E_{\mathrm{h}}^{\mathrm{X}}$, as well as the *sample histogram*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{X}}$. Given a *drawn history* $H \in \mathcal{H}_{U,X}$, where $P_{U,X,H_{\mathrm{h}},\mathrm{x}}(H) > 0$, the *independent historical probability* of *histogram* $A_H = \mathrm{histogram}(H) + V_H^{\mathrm{CZ}} \in \mathcal{A}_{U,\mathrm{i},V_H,z_H}$ is now conditional,

$$\hat{Q}_{\mathrm{h,x},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) \;\propto\; \sum(P_{U,X,H_{\mathrm{h}},\mathrm{x}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *independent conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,x},U}$, is defined

$$\begin{aligned} \hat{Q}_{\mathrm{h,x},U}(E,z) \;=\; & \{(A, \frac{\hat{Q}_{\mathrm{h},U}(E,z)(A)}{\sum(\hat{Q}_{\mathrm{h},U}(E,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B = B^{\mathrm{X}})}) \\ & \qquad : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A = A^{\mathrm{X}}\} \cup \\ & \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \neq A^{\mathrm{X}}\} \end{aligned}$$

The denominator, $\sum(\hat{Q}_{\mathrm{h},U}(E,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B = B^{\mathrm{X}})$, is a constant and so the *independent conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,x},U}$, is just the normalisation of the *stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},U}$, where the *histogram* is *independent*,

$$\begin{aligned} \hat{Q}_{\mathrm{h,x},U}(E,z) \;=\; & \{(A, \hat{Q}_{\mathrm{h},U}(E,z)(A)) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A = A^{\mathrm{X}}\}^{\wedge} \cup \\ & \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \neq A^{\mathrm{X}}\} \end{aligned}$$

The *maximum likelihood estimate* now corresponds to

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E^{\mathrm{X}}, \hat{Q}_{\mathrm{h,x},U}(E^{\mathrm{X}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

Conjecture that, though the *sample histogram*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{X}}$, is in the denominator, the *maximum likelihood estimate* is as before, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, because the *distribution histogram*, E^{X}, is also *independent*. The *sum sensitivity*, however, is lower now. In section 'Alignment and independent histograms', above, it is shown that the *multinomial probability density* of an *independent histogram* A^{X} of *size* z and *variables* V *drawn* from an *independent distribution* E^{X} is approximately equal to the product of the *multinomial probability densities* of the *reduced independent histogram* $A^{\mathrm{X}}\%\{w\}$, where $w \in V$, *drawn* from the *reduced independent distribution* $E^{\mathrm{X}}\%\{w\}$

$$\mathrm{mpdf}(U)(E^{\mathrm{X}}, z)(A^{\mathrm{X}}) \approx \prod_{w \in V} \mathrm{mpdf}(U)(E^{\mathrm{X}}\%\{w\}, z)(A^{\mathrm{X}}\%\{w\})$$

where the *multinomial probability density function* is defined

$$\mathrm{mpdf}(U)(E, z) := \{(A, \frac{\Gamma_! z}{\prod_{S \in A^{\mathrm{S}}} \Gamma_! A_S} \prod_{S \in A^{\mathrm{S}}} \hat{E}_S^{A_S}) : A \in \mathcal{A}_{U,V,z}\}$$

In this case where the *sample histogram* is *integral*, $A = A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, the *generalised multinomial probability distribution* approximates

$$\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)(A^{\mathrm{X}}) \approx \prod_{w \in V} \hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}\%\{w\}, z)(A^{\mathrm{X}}\%\{w\})$$

and therefore the *sum sensitivity* of the numerator is the sum of the *sum sensitivities* of the *perimeter*, which is less than the *sum sensitivity* of the *volume*,

$$\begin{aligned}
&\sum_{w \in V} \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}\%\{w\}, z))) \\
&\qquad = \sum_{w \in V} \sum_{S \in \{w\}^{\mathrm{CS}}} \frac{z}{\hat{E}^{\mathrm{X}}\%\{w\}(S)\ (1 - \hat{E}^{\mathrm{X}}\%\{w\}(S))} \\
&\qquad \leq \sum_{S \in V^{\mathrm{CS}}} \frac{z}{\hat{E}_S(1 - \hat{E}_S)} \\
&\qquad = \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E^{\mathrm{X}}, z)))
\end{aligned}$$

The *sum sensitivity* of *independent conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,x},U}$, is conjectured to be less than the *sum sensitivity* of the numerator because the denominator has *sum sensitivity*

$$\begin{aligned}&\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,x},U}(E^{\text{X}}, z)))\\&\qquad\leq \sum_{w \in V} \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h},U}(E^{\text{X}}\%\{w\}, z)))\end{aligned}$$

The *perimeter sum sensitivities* are minimised when the *distribution histogram* is *uniform*, $E = V^{\text{C}}$, and so the *perimeters* are *uniform*, $\forall w \in V\ ((E\%\{w\})^{\wedge} = (\{w\}^{\text{C}})^{\wedge})$. In the case of *regular distribution histogram* of *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_w| : w \in V\}$, the minimum *sum sensitivity* is

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,x},U}(V^{\text{C}}, z))) \leq \frac{ndz}{(d-1)} \leq d^n z \leq \frac{d^{2n} z}{(d^n - 1)} = \frac{v^2 z}{(v-1)}$$

where $v = |V^{\text{C}}| > 1$. That is, in the *independent* case, *sum sensitivity* varies with *dimension*, n, rather than *volume*, v.

In *classical independent induction* where (i) the *history probability function* is *independent historically distributed*, $P = P_{U,X,H_{\text{h}},\text{x}}$, (ii) the *volume* is non-singleton, $v_{\text{o}} > 1$, and (iii) the *sample histogram* is *completely effective*, $A^{\text{F}}_{\text{o}} = V^{\text{C}}_{\text{o}}$, the *maximum likelihood estimate*, $\tilde{E}_{\text{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\text{o}}$, in the *independent conditional stuffed historical probability distribution*, $\hat{Q}_{\text{h,x},U}(E_{\text{o}}, z_{\text{o}})$, is $\tilde{E}_{\text{o}} = \hat{A}_{\text{o}}$, so, if it is also the case that (iv) the *distribution history size* is large with respect to the *sample size*, $z_{\text{h}} \gg z_{\text{o}}$, and such that (v) the *scaled probability sample histogram* is *integral*, $A_{\text{o},z_{\text{h}}} \in \mathcal{A}_{\text{i}}$, then the *sum sensitivity* of the *independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies as the sum of the *perimeter sum sensitivities*,

$$\begin{aligned}&\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,x},U}(A_{\text{o},z_{\text{h}}}, z_{\text{o}}))) \sim\\&\qquad\sum_{w \in V_{\text{o}}} \sum_{S \in \{w\}^{\text{CS}}} \frac{z_{\text{o}}}{\hat{A}^{\text{X}}_{\text{o}}\%\{w\}(S)\ (1 - \hat{A}^{\text{X}}_{\text{o}}\%\{w\}(S))}\end{aligned}$$

Given a *mono-effective* query *histogram* $Q = \{S_Q\}^{\text{U}}$, where $S_Q \in K^{\text{CS}}$ and $K \subset V_{\text{o}}$, the *estimated transformed conditional product* is $\hat{Q} * T'_{\hat{A}_{\text{o}},K}$, where the *estimated conditional transform induced* from the *sample histogram*, $\hat{A}_{\text{o}}$, and the query *variables*, K, is $T'_{\hat{A}_{\text{o}},K} = (\hat{A}_{\text{o}}/(\hat{A}_{\text{o}}\%K), (V_{\text{o}} \setminus K))$. The *query sensitivity* varies as the *sum sensitivity* of the *independent conditional historical distribution* at the *maximum likelihood estimate* divided by the *sample* *size*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,x},U}(A_{\text{o},z_\text{h}}, z_\text{o})))/z_\text{o} \sim \sum_{w \in V_\text{o}} \sum_{S \in \{w\}^\text{CS}} \frac{1}{\hat{A}_\text{o}^\text{X}\%\{w\}(S)\ (1 - \hat{A}_\text{o}^\text{X}\%\{w\}(S))}$$

Of course, in the case of *independent sample histogram*, $\hat{A}_\text{o} = \hat{A}_\text{o}^\text{X}$, the label *variables*, $V_\text{o} \setminus K$, are *independent* of the query *variables*, K, and so the *estimated transformed conditional product* is trivial,

$$\hat{Q} * T'_{\hat{A}_\text{o}^\text{X},K} = \hat{A}_\text{o}^\text{X}\ \%\ (V_\text{o} \setminus K)$$

That is, in spite of lower *query sensitivity* to the *estimate* of the *unknown distribution histogram*, E_o^X, there is no functional or *causal* relation between the query *variables* and the label *variables*,

$$\text{split}(K, \hat{A}_\text{o}^\text{XFS}) \notin K^\text{CS} \to (V_\text{o} \setminus K)^\text{CS}$$

This is true for any query in the contrived case of *independent historically distributed history probability function*, $P = P_{U,X,H_\text{h},\text{x}}$.

5.5 Classical modelled induction

Having considered (i) the case of *classical induction*, where the *history probability function* is *historically distributed*, $P = P_{U,X,H_\text{h}}$, and (ii) the special case of *classical independent induction* where the *history probability function* is *independent historically distributed*, $P = P_{U,X,H_\text{h},\text{x}}$, now consider (iii) the special case of *classical modelled induction*.

5.5.1 Necessary derived

Given some *known substrate transform* in the *sample variables* $T_\text{o} \in \mathcal{T}_{U,V_\text{o}}$, the *derived histogram* of the *distribution probability histogram* is $\hat{E}_\text{h} * T_\text{o}$. In *classical modelled induction*, while the *distribution probability histogram*, $\hat{E}_\text{h}$, remains *unknown*, the *derived distribution probability histogram*, $\hat{E}_\text{h} * T_\text{o}$, is *known* and *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have a *derived probability histogram* equal to the *known derived distribution probability histogram*, $\hat{A}_H * T_\text{o} = \hat{E}_\text{h} * T_\text{o}$. Define the *iso-derived historically distributed* *history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$,

$$\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}} := \\
&\quad \Big(\bigcup \{\{(H,1) : H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H,\ \hat{A}_H * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}\}^{\wedge} : \\
&\qquad\qquad\qquad\qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ \hat{A}_H * T_{\mathrm{o}} \neq \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}\} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \nsubseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset,0)\}
\end{aligned}$$

For *drawn histories* the *derived probability histogram* is *necessary*, $\forall H \in \mathcal{H}_{U,X}\ (P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H) > 0 \implies \hat{A}_H * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}})$. Not all *sizes* and sets of *variables* are necessarily *drawable*. That is, in some cases, $\exists z \in \{1 \ldots z_{\mathrm{h}}\}\ \exists V \subseteq V_{\mathrm{h}}\ \forall H \in \mathcal{H}_{U,X}\ ((z_H = z) \wedge (V_H = V) \implies P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H) = 0)$. A *size* $z \in \{1 \ldots z_{\mathrm{h}}\}$ can be *drawn* if (i) the *variables* $V \subseteq V_{\mathrm{h}}$ are a superset of the *transform underlying*, $\mathrm{und}(T_{\mathrm{o}})$, and (ii) the *scaled derived distribution histogram*, $\mathrm{scalar}(z) * \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}$, is *integral*,

$$\begin{aligned}
&(\mathrm{und}(T_{\mathrm{o}}) \subseteq V)\ \wedge\ (\mathrm{scalar}(z) * \hat{E}_{\mathrm{h}} * T_{\mathrm{o}} \in \mathcal{A}_{\mathrm{i}}) \implies \\
&\qquad \exists H \in \mathcal{H}_{U,X}\ ((z_H = z) \wedge (V_H = V) \wedge (P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H) > 0))
\end{aligned}$$

The *distribution history* can always be *drawn*, so the *probability function* is not a *weak probability function*, $\sum_{H \in \mathcal{H}_{U,X}} P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H) = 1$.

All *iso-derived* subsets of the *distribution history* for a given set of *variables* and *size* are defined as equally probable,

$$\begin{aligned}
&\forall V \subseteq V_{\mathrm{h}}\ \forall H, G \subseteq H_{\mathrm{h}}\%V \\
&\qquad\qquad (A_G * T_{\mathrm{o}} = A_H * T_{\mathrm{o}} \implies P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(G) = P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H))
\end{aligned}$$

In *classical modelled induction* the *history probability function* is *iso-derived historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}$.

Given a *drawn history* $H \in \mathcal{H}_{U,X}$, where $P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H) > 0$, the *iso-derived historical probability* of *histogram* $A_H = \mathrm{histogram}(H) + V_H^{\mathrm{CZ}} \in \mathcal{A}_{U,\mathrm{i},V_H,z_H}$ is now conditional,

$$\begin{aligned}
&\frac{Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T_{\mathrm{o}},z_H}(A_H * T_{\mathrm{o}})} Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(B)} = \\
&\qquad\qquad \frac{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}
\end{aligned}$$

The *iso-derived historical probability* may be expressed in terms of a *histogram distribution* which is not explicitly conditional on the *necessary derived*, $\hat{E}_\mathrm{h} * T_\mathrm{o}$,

$$\hat{Q}_{\mathrm{h},\mathrm{d},T_\mathrm{o},U}(E_\mathrm{h}\%V_H, z_H)(A_H) \;\;\propto\;\; \sum(P_{U,X,H_\mathrm{h},\mathrm{d},T_\mathrm{o}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *iso-derived conditional stuffed historical probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{d},T,U}(E,z) \\
&\quad := \;\; \{(A, \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\
&\qquad\;\; \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \not\leq E\}
\end{aligned}$$

which is defined if $z \leq \mathrm{size}(E)$. The *derived histogram* valued *integral histogram* function $D_{U,\mathrm{i},T,z}$ is defined

$$D_{U,\mathrm{i},T,z} = \{(A, A*T) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The finite set of *iso-deriveds* of *derived histogram* $A*T$ is

$$D^{-1}_{U,\mathrm{i},T,z}(A*T) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B*T = A*T\}$$

which is such that the *lifted iso-deriveds* is a singleton, $\{B*T : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B*T = A*T\} = \{A*T\}$.

In the case where all the *derived* are possible,

$$\forall A' \in \mathrm{ran}(D_{U,\mathrm{i},T,z})\ \exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ ((A*T = A')\ \wedge\ (A \leq E))$$

the normalisation of the *iso-derived conditional stuffed historical probability distribution* is a fraction $1/|\mathrm{ran}(D_{U,\mathrm{i},T,z})|$,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{d},T,U}(E,z) \\
&\quad = \;\; \{(A, \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},T,z})|}\frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}
\end{aligned}$$

The case of possible *derived* is equivalent to possible *iso-derived*, $\forall A' \in \mathrm{ran}(D_{U,\mathrm{i},T,z})\ (\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A')} Q_{\mathrm{h},U}(E,z)(B) > 0)$. All *derived* are possible in the case where the least *count* of the *distribution histogram* is greater than or equal to the *sample size*, $z \leq \mathrm{mind}(E_\mathrm{h})$.

In the case of possible *derived* the *iso-derived historical probability* is

$$\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = \\ \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},T_{\mathrm{o}},z_H})|}\frac{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}$$

In the case of a *full functional transform*, $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}}$, the *iso-derived* is a singleton of the *sample histogram*, $D^{-1}_{U,\mathrm{i},T_{\mathrm{s}},z}(A*T_{\mathrm{s}}) = \{A\}$, and so the denominator equals the numerator, $\sum_{B \in D^{-1}_{U,\mathrm{i},T_{\mathrm{s}},z}(A*T_{\mathrm{s}})} Q_{\mathrm{h},U}(E,z)(B) = Q_{\mathrm{h},U}(E,z)(A)$. Thus the scaled *iso-derived historically distributed history probability* is certain, $|\mathcal{A}_{U,\mathrm{i},V,z}| \times \hat{Q}_{\mathrm{h,d},T_{\mathrm{s}},U}(E,z)(Z * \hat{E}) = 1$, where $Z = \mathrm{scalar}(z)$. In this case, the *distribution probability histogram*, $\hat{E}$, is *known*, because $\hat{E} * T_{\mathrm{s}}$ is *known*, and so everything is *known*.

At the other extreme of a *unary transform*, $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, the *iso-derived* includes all *substrate histograms*, $D^{-1}_{U,\mathrm{i},T_{\mathrm{u}},z}(A*T_{\mathrm{u}}) = \mathcal{A}_{U,\mathrm{i},V,z}$, and the normalised denominator is one, $\sum_{B \in D^{-1}_{U,\mathrm{i},T_{\mathrm{u}},z}(A*T_{\mathrm{u}})} \hat{Q}_{\mathrm{h},U}(E,z)(B) = 1$. Thus the *iso-derived conditional stuffed historical probability distribution* equals the *underlying stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{u}},U}(E,z) = \hat{Q}_{\mathrm{h},U}(E,z)$. In this case, nothing is *known*, because $\hat{E} * T_{\mathrm{u}} = \{(\{(\{V^{\mathrm{CS}}\}, V^{\mathrm{CS}})\}, 1)\}$ is trivially *known*. In this case *classical modelled induction* reduces to *classical induction*.

The *iso-derived conditional generalised multinomial probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\mathrm{m,d},T,U}(E,z) \\
&:= \ \{(A, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \leq E^{\mathrm{F}}\}^{\wedge} \cup \\
&\qquad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \nleq E^{\mathrm{F}}\}
\end{aligned}$$

which is defined if $\mathrm{size}(E) > 0$.

The case where all the *derived* are possible is weaker than for *historical*,

$$\forall A' \in \mathrm{ran}(D_{U,\mathrm{i},T,z})\ \exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ ((A * T = A')\ \wedge\ (A^{\mathrm{F}} \leq E^{\mathrm{F}}))$$

In this case the *iso-derived conditional generalised multinomial probability distribution* is

$$\begin{aligned}&\hat{Q}_{\mathrm{m,d},T,U}(E,z)\\ &= \{(A, \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},T,z})|}\frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}\end{aligned}$$

Let $A_{\mathrm{o}} \in \mathcal{A}_{U,\mathrm{i},V_{\mathrm{o}},z_{\mathrm{o}}}$ be a *known sample integral histogram* of *size* z_{o} in the *underlying variables* of the *transform* $V_{\mathrm{o}} = \mathrm{und}(T_{\mathrm{o}})$. It is assumed that the *distribution history size*, z_{h}, is large with respect to the *sample size* $z_{\mathrm{o}} = \mathrm{size}(A_{\mathrm{o}})$, so that, in the limit, the *iso-derived historical probability*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$, approximates to the *iso-derived multinomial probability*, $\hat{Q}_{\mathrm{m,d},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$. That is, if $z_{\mathrm{o}} \ll z_{\mathrm{h}}$ then

$$\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(E_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}}) \approx \hat{Q}_{\mathrm{m,d},T_{\mathrm{o}},U}(E_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$$

where $E_{\mathrm{o}} = E_{\mathrm{h}}\%V_{\mathrm{o}}$.

The *iso-derived conditional generalised multinomial parameterised probability density function*, $\mathrm{mdtppdf}(T,z) \in \mathrm{ppdfs}(v,v)$, and *iso-derived conditional generalised multinomial likelihood function*, $\mathrm{mdtlf}(T,z) \in \mathrm{lfs}(v,v)$, corresponding to the *iso-derived conditional generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m,d},T,U}$, are not given explicitly here, but are such that

$$\mathrm{mdtppdf}(T,z)(\hat{E}^{\square})(A^{\square}) = \mathrm{mdtlf}(T,z)(A^{\square})(\hat{E}^{\square}) = \hat{Q}_{\mathrm{m,d},T,U}(E,z)(A)$$

Now in the case of *classical modelled induction* where the *transform*, T_{o}, is *known*, the *real maximum likelihood estimate* $\tilde{E}'_{\mathrm{o}} \in \mathbf{R}^{v_{\mathrm{o}}}_{(0,1)}$ for the parameter of the *iso-derived multinomial parameterised probability density function* is

$$\{\tilde{E}'_{\mathrm{o}}\} = \mathrm{maxd}(\mathrm{mdtlf}(T_{\mathrm{o}},z_{\mathrm{o}})(A^{\square}_{\mathrm{o}}))$$

which is such that $\forall i \in \{1 \ldots v_{\mathrm{o}}\}\ (\partial_i(\mathrm{mdtlf}(T_{\mathrm{o}},z_{\mathrm{o}})(A^{\square}_{\mathrm{o}}))(\tilde{E}'_{\mathrm{o}}) = 0)$. The *maximum likelihood estimate* $\tilde{E}'_{\mathrm{o}}$ is only defined in the case where the *sample histogram* is *completely effective*, $A^{\mathrm{F}}_{\mathrm{o}} = V^{\mathrm{C}}_{\mathrm{o}} \implies \hat{A}^{\square}_{\mathrm{o}} \in \mathbf{R}^{v_{\mathrm{o}}}_{(0,1)}$, because the *binomial likelihood function* is only defined for the open set. That is, $\mathrm{d}(\mathrm{blf}(z_{\mathrm{o}})(0))$ is undefined and so the derivative of the *iso-derived multinomial parameterised probability density function* is undefined where there are *ineffective states*.

In the case of *completely effective sample histogram*, $A^{\mathrm{F}}_{\mathrm{o}} = V^{\mathrm{C}}_{\mathrm{o}}$, the maximisation for *known transform*, T_{o}, of the *iso-derived conditional generalised* *multinomial probability* parameterised by the *complete congruent histograms* of unit *size* is a singleton of the *rational maximum likelihood estimate*

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m,d},T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

The *real maximum likelihood estimate*, $\tilde{E}'_{\mathrm{o}}$, is not necessarily a rational coordinate, $\mathbf{R}^{v_{\mathrm{o}}}_{(0,1)} \supset \mathbf{Q}^{v_{\mathrm{o}}}_{(0,1)}$, and so the *rational maximum likelihood estimate* is not necessarily equal to the *real maximum likelihood estimate*. However, it is conjectured that the maximisation of the *distribution* approximates to the maximisation of the *likelihood function*,

$$\tilde{E}_{\mathrm{o}}^{\square} \approx \tilde{E}'_{\mathrm{o}}$$

In the case where the *sample histogram* is not *completely effective*, $A^{\mathrm{F}}_{\mathrm{o}} < V^{\mathrm{C}}_{\mathrm{o}}$, the maximisation of the *iso-derived conditional generalised multinomial probability distribution* for *known transform* is well defined, unlike the *parameterised probability density function*, but is not necessarily a singleton

$$|\mathrm{max}(\{(E, \hat{Q}_{\mathrm{m,d},T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})| \geq 1$$

In the case where the maximisation of the *iso-derived conditional generalised multinomial probability distribution* is a singleton, it is equal to the *normalised derived-dependent*, $\tilde{E}_{\mathrm{o}} = \hat{A}^{\mathrm{D}(T_{\mathrm{o}})}_{\mathrm{o}}$, where the *derived-dependent* $A^{\mathrm{D}(T)} \in \mathcal{A}_{U,V,z}$ is defined in 'Likely histograms', above, as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-derived*,

$$\{A^{\mathrm{D}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *derived-dependent*, $A^{\mathrm{D}(T)}$, is sometimes not computable. The finite approximation to the *derived-dependent* is

$$\{A^{\mathrm{D}(T)}_k\} = \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})$$

The approximation, $A^{\mathrm{D}(T)}_k \approx A^{\mathrm{D}(T)}$, improves as the scaling factor, k, increases.

Unlike in *classical non-modelled induction* where the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, is equal to the *sample probability histogram*, $\hat{A}_{\mathrm{o}}$, in *classical* *modelled induction* the *maximum likelihood estimate* is not necessarily equal to the *sample probability histogram*. It is only in the case where the *sample histogram* is *natural* that the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger} \implies A_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

Otherwise, the overall *maximum likelihood estimate*, which is the *derived-dependent*, is near the *histogram*, $\tilde{E}_{\mathrm{o}} \sim \hat{A}_{\mathrm{o}}$, only in as much as it is far from the *naturalisation*, $\tilde{E}_{\mathrm{o}} \nsim \hat{A}_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$.

The requirement that the *distribution history* itself be *drawable*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H_{\mathrm{h}}) > 0$, has been ignored so far. This requirement modifies the maximisation to add the constraint that the *maximum likelihood estimate* be *iso-derived*, $\tilde{E}_{\mathrm{o}} * T_{\mathrm{o}} = \hat{A}_{\mathrm{o}} * T_{\mathrm{o}}$,

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m,d},T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ E * T_{\mathrm{o}} = \hat{A}_{\mathrm{o}} * T_{\mathrm{o}}\})$$

So, strictly speaking, the *maximum likelihood estimate* is only approximately equal to the *normalised derived-dependent*, $\tilde{E}_{\mathrm{o}} \approx \hat{A}_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})}$, if the *derived-dependent* is not *iso-derived*, $A_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})} * T_{\mathrm{o}} \neq A_{\mathrm{o}} * T_{\mathrm{o}}$. In the special case, however, where the *sample histogram* is *natural*, the *maximum likelihood estimate* is exactly equal to the *normalised derived-dependent*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})} = \hat{A}_{\mathrm{o}}$, because $A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger} * T_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}}$.

In *classical modelled induction*, where (i) the *history probability function* is *iso-derived historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample histogram* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-derived conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is

$$\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

In section 'Iso-sets', above, the degree to which an *integral iso-set* $I \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$, where $A \in I$, is said to be *law-like*, or the *iso-derivedence*, is defined as

$$\frac{|I\ \cap\ D_{U,\mathrm{i},T,z}^{-1}(A * T)|}{|I\ \cup\ D_{U,\mathrm{i},T,z}^{-1}(A * T)|}$$

In the case of *classical modelled induction* the *integral iso-set* is the *integral iso-derived*, $I = D_{U,\mathrm{i},T,z}^{-1}(A * T)$, and so *classical modelled induction* is maximally *law-like*.

The *iso-abstractence* of the *iso-deriveds* equals the *iso-derivedence* of the *iso-abstracts*,

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A*T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})|}$$

So *classical modelled induction* is not maximally *entity-like* if the *iso-deriveds* is a proper subset of the *iso-abstracts*, $D^{-1}_{U,T,z}(A*T) \subset Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})$. This is the case if the *derived* is not *independent*, $A*T \neq (A*T)^{\mathrm{X}}$.

Given the *known substrate transform*, T_{o}, consider the *log likelihood* of the *iso-derived conditional generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m,d},T_{\mathrm{o}},U}$, at the *maximum likelihood estimate*.

In section 'Likely histograms', above, the logarithm of the *maximum conditional probability* with respect to the *dependent-analogue* is conjectured to vary with the *relative space* with respect to the *independent-analogue*. In the case of *iso-derived conditional*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{D}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{D}(T)},z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \quad \sim \quad \mathrm{spaceRelative}(A*T*T^{\dagger})(A)$$

where the *distribution-relative multinomial space* is defined, in section 'Likely histograms', above, as

$$\mathrm{spaceRelative}(E)(A) \quad := \quad -\ln \frac{\mathrm{mpdf}(U)(E,z)(A)}{\mathrm{mpdf}(U)(E,z)(E)}$$

In section 'Transform alignment', above, because the set of *iso-deriveds* is *law-like*, it is shown that, in the case where the *dependent analogue* is in the *iso-set*, the difference in *relative space* between the *histogram* and the *dependent* must be in the differences between the *relative spaces* of the *components*,

$$\begin{aligned}
&A^{\mathrm{D}(T)} \in D^{-1}_{U,T,z}(A*T) \implies \\
&\quad \sum_{(\cdot,C)\in T^{-1}} \mathrm{spaceRelative}(A*T*T^{\dagger}*C)(A*C) \\
&\leq \sum_{(\cdot,C)\in T^{-1}} \mathrm{spaceRelative}(A*T*T^{\dagger}*C)(A^{\mathrm{D}(T)}*C)
\end{aligned}$$

So, in the case of the *derived-dependent*, the difference in *relative space* between the *histogram* and the *dependent* must be in the difference between the *non-independent* terms of the *alignments*,

$$A^{\mathrm{D}(T)} \in D_{U,T,z}^{-1}(A * T) \quad \implies \quad \sum_{S \in V^{\mathrm{CS}}} \ln \Gamma_! A_S \leq \sum_{S \in V^{\mathrm{CS}}} \ln \Gamma_! A_S^{\mathrm{D}(T)}$$

The sum of the *relative spaces* of the *components* approximates to

$$\sum_{(\cdot,C) \in T^{-1}} \mathrm{spaceRelative}(A * T * T^{\dagger} * C)(A * C) \ \approx$$
$$z + \sum_{S \in V^{\mathrm{CS}}} \ln \Gamma_! A_S - \sum_{(R,C) \in T^{-1}} (A * T)_R \ln((A * T)_R / |C|)$$

So the sum of the *relative spaces* of the *components* varies with the *non-independent* term of the *histogram alignment*,

$$\sum_{(\cdot,C) \in T^{-1}} \mathrm{spaceRelative}(A * T * T^{\dagger} * C)(A * C) \ \sim \ \sum_{S \in V^{\mathrm{CS}}} \ln \Gamma_! A_S$$

which is independent of the *transform*, T. The sum of the *relative spaces* of the *components* varies against the *size* scaled *component size cardinality relative entropy*,

$$\sum_{(\cdot,C) \in T^{-1}} \mathrm{spaceRelative}(A * T * T^{\dagger} * C)(A * C) \ \sim$$
$$-z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

The *derived-dependent* varies with the *histogram*, $\tilde{E}_{\mathrm{o}} \sim \hat{A}_{\mathrm{o}}$, so conjecture that in the case where the *sample* is not *natural*, $A \neq A * T * T^{\dagger} \implies \mathrm{spaceRelative}(A * T * T^{\dagger})(A) > 0$, the *log-likelihood* varies with the *non-independent* term of the *histogram alignment*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{D}(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{D}(T)}, z)(B) : B \in D_{U,\mathrm{i},T,z}^{-1}(A * T)} \ \sim \ \sum_{S \in V^{\mathrm{CS}}} \ln \Gamma_! A_S$$

and varies against the *size* scaled *component size cardinality relative entropy*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{D}(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{D}(T)}, z)(B) : B \in D_{U,\mathrm{i},T,z}^{-1}(A * T)} \ \sim$$
$$-z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

In *classical modelled induction*, where (i) the *history probability function* is *iso-derived historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}$, given some *substrate* *transform* in the *sample variables* $T_o \in \mathcal{T}_{U,V_o}$, if it is the case that (ii) the *sample histogram* is not *natural*, $A_o \neq A_o * T_o * T_o^{\dagger}$, (iii) the *distribution history size* is large with respect to the *sample size*, $z_h \gg z_o$, and such that (iv) the *scaled estimate distribution histogram* is *integral*, $\tilde{E}_{o,z_h} \in \mathcal{A}_i$ where $\tilde{E}_{o,z_h} = \text{scalar}(z_h) * \tilde{E}_o$, then the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *relative space* of the *sample* with respect to the *naturalisation*,

$$\ln \hat{Q}_{h,d,T_o,U}(\tilde{E}_{o,z_h}, z_o)(A_o) \sim \text{spaceRelative}(A_o * T_o * T_o^{\dagger})(A_o)$$

varies with the *non-independent* term of the *sample alignment*,

$$\ln \hat{Q}_{h,d,T_o,U}(\tilde{E}_{o,z_h}, z_o)(A_o) \sim \sum_{S \in V_o^{CS}} \ln \Gamma_! A_o(S)$$

and varies against the *size* scaled *component size cardinality relative entropy*,

$$\ln \hat{Q}_{h,d,T_o,U}(\tilde{E}_{o,z_h}, z_o)(A_o) \sim -z_o \times \text{entropyRelative}(A_o * T_o, V_o^{C} * T_o)$$

Given the *known substrate transform*, T_o, consider the *log likelihood* of the *iso-derived conditional generalised multinomial probability distribution*, $\hat{Q}_{m,d,T_o,U}$, at the *maximum likelihood estimate*, in the special case where the *histogram* is *natural*, $A_o = A_o * T_o * T_o^{\dagger} \implies \tilde{E}_o = \hat{A}_o^{D(T_o)} = \hat{A}_o$.

First consider the cardinality of the *integral iso-deriveds*. Let a pair of *substrate transforms* $T_1, T_2 \in \mathcal{T}_{U,V}$ be such that (i) both are *natural*, $A*T_1*T_1^{\dagger} = A * T_2 * T_2^{\dagger} = A$, (ii) the first *derived* is *independent*, $A * T_1 = (A * T_1)^X$, (iii) the second *derived* is not *independent*, $A*T_2 \neq (A*T_2)^X$, but (iv) the second *abstract* equals the first *derived*, $(A * T_2)^X = A * T_1 = (A * T_1)^X$. In section 'Iso-sets', above, it is shown that if and only if the *derived* is *independent* then the *iso-deriveds* equals the *iso-abstracts*,

$$A * T = (A * T)^X \iff D_{U,T,z}^{-1}(A * T) = Y_{U,T,W,z}^{-1}((A * T)^X)$$

So the second *integral iso-deriveds* is a proper subset of the first *integral iso-deriveds*,

$$D_{U,i,T_2,z}^{-1}(A * T_2) \subset Y_{U,i,T_1,W,z}^{-1}((A * T_1)^X) = D_{U,i,T_1,z}^{-1}(A * T_1)$$

and the denominator of the second *iso-derived conditional multinomial probability* is necessarily less than the denominator of the first *iso-derived conditional multinomial probability*,

$$\sum_{B \in D_{U,i,T_2,z}^{-1}(A*T_2)} \hat{Q}_{m,U}(A, z)(B) < \sum_{B \in D_{U,i,T_1,z}^{-1}(A*T_1)} \hat{Q}_{m,U}(A, z)(B)$$

So the second *iso-derived conditional multinomial probability* at the *maximum likelihood estimate* is necessarily greater than the first *iso-derived conditional multinomial probability* at the *maximum likelihood estimate*,

$$\frac{\hat{Q}_{\mathrm{m},U}(A,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T_2,z}(A*T_2)} \hat{Q}_{\mathrm{m},U}(A,z)(B)} \quad > \quad \frac{\hat{Q}_{\mathrm{m},U}(A,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T_1,z}(A*T_1)} \hat{Q}_{\mathrm{m},U}(A,z)(B)}$$

or

$$\hat{Q}_{\mathrm{m,d},T_2,U}(A,z)(A) \quad > \quad \hat{Q}_{\mathrm{m,d},T_1,U}(A,z)(A)$$

That is, in this case the second *transform* is more *likely* than the first *transform*. As shown in 'Minimum alignment', above, the *independent entropy* is always at least the *histogram entropy*, $\forall A \in \mathcal{A}$ $(\mathrm{entropy}(A^{\mathrm{X}}) \geq \mathrm{entropy}(A))$. The second *derived* is not *independent* so its *entropy* is necessarily less than the *entropy* of the first *derived*, $\mathrm{entropy}(A * T_2) < \mathrm{entropy}(A * T_1)$. It is conjectured in 'Transform alignment', above, that the *log iso-abstractence* of the second *transform* varies with the *size* scaled *derived entropy* and against the *size* scaled *independent derived entropy*,

$$\begin{aligned}
\ln \frac{|D^{-1}_{U,\mathrm{i},T,z}(A*T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})|} \quad &\sim \quad z \times \mathrm{entropy}(A*T) - z \times \mathrm{entropy}((A*T)^{\mathrm{X}}) \\
&\approx \quad -\ \mathrm{algn}(A*T)
\end{aligned}$$

hence conjecture that the *log likelihood* at the *maximum likelihood estimate* varies against the *derived entropy*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \quad \sim \quad -z \times \mathrm{entropy}(A*T)$$

This can be refined by considering the cardinality of the set of *integral iso-deriveds* which may be stated explicitly as the product of the weak compositions of the *components*,

$$|D^{-1}_{U,\mathrm{i},T,z}(A*T)| \quad = \quad \prod_{(R,C) \in T^{-1}} \frac{((A*T)_R + |C| - 1)!}{(A*T)_R!\ (|C|-1)!}$$

It is shown in 'Integral iso-sets and entropy', above, that the *integral iso-deriveds log-cardinality* varies against the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}
&\ln |D^{-1}_{U,\mathrm{i},T,z}(A*T)| \ \sim \\
&\quad -\ \big((z+v) \times \mathrm{entropy}(A*T + V^{\mathrm{C}}*T) \\
&\qquad\qquad -z \times \mathrm{entropy}(A*T)\ - v \times \mathrm{entropy}(V^{\mathrm{C}}*T)\big)
\end{aligned}$$

In the domain where the *size* is greater than the *volume*, $z > v$, the *integral iso-deriveds log-cardinality* varies against the *volume* scaled *component cardinality size relative entropy*,

$$\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \quad \sim \quad -v \times \mathrm{entropyRelative}(V^{\mathrm{C}} * T, A * T)$$

In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *integral iso-deriveds log-cardinality* varies against the *size* scaled *component size cardinality relative entropy*,

$$\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \quad \sim \quad -z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

In both domains the *integral iso-deriveds log-cardinality* varies against the *relative entropy*. That is, *integral iso-deriveds log-cardinality* is minimised when (a) the *cross entropy* is maximised and (b) the *component entropy* is minimised. The *cross entropy* is maximised when high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

The *log likelihood* varies against the *iso-derived log-cardinality*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) &\propto \ln \frac{Q_{\mathrm{m},U}(A,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} Q_{\mathrm{m},U}(A,z)(B)} \\ &\sim -\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)|\end{aligned}$$

So the *log likelihood* varies with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \sim \\ &\quad (z+v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\ &\qquad\qquad -z \times \mathrm{entropy}(A * T) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)\end{aligned}$$

In the domain where the *size* is greater than the *volume*, $z > v$, the *log likelihood* varies with the *volume* scaled *component cardinality size relative entropy*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \quad \sim \quad v \times \mathrm{entropyRelative}(V^{\mathrm{C}} * T, A * T)$$

The *volume* scaled *component cardinality size relative entropy* approximates to the negative logarithm of the *cartesian derived multinomial probability* with respect to the *derived*, so in this domain the *log likelihood* varies against the *cartesian derived multinomial probability*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \quad \sim \quad -\ln \hat{Q}_{\mathrm{m},U}(A * T, v)(V^{\mathrm{C}} * T)$$

The *sample histogram* is *natural*, $A_{\rm o} = A_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger}$, so even if the *sample* is not *completely effective*, $A_{\rm o}^{\rm F} < V_{\rm o}^{\rm C}$, the *component size* of *effective components* is always at least equal to the *component cardinality*, $\forall (R,C) \in T_{\rm o}^{-1}\ ((A_{\rm o} * T_{\rm o})_R > 0 \implies (A_{\rm o} * T_{\rm o})_R \geq |C|)$, and this domain, where the *log likelihood* varies with the *component cardinality size relative entropy*, applies.

In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *log likelihood* varies with the *size* scaled *component size cardinality relative entropy*,

$$\ln \hat{Q}_{{\rm m,d},T,U}(A,z)(A) \quad \sim \quad z \times \mathrm{entropyRelative}(A * T, V^{\rm C} * T)$$

The *size* scaled *component size cardinality relative entropy* approximates to the negative logarithm of the *derived multinomial probability* with respect to the *cartesian derived*, so in this domain the *log likelihood* varies against the *derived multinomial probability*,

$$\ln \hat{Q}_{{\rm m,d},T,U}(A,z)(A) \quad \sim \quad -\ln \hat{Q}_{{\rm m},U}(V^{\rm C} * T, z)(A * T)$$

If the *size* is less than the *volume*, $z_{\rm o} < v_{\rm o}$, the *effective volume* is necessarily less than *cartesian*, $A_{\rm o}^{\rm F} < V_{\rm o}^{\rm C}$, so sometimes the *sample* merely approximates to the *naturalisation*, $A_{\rm o} \approx A_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger}$.

In *classical modelled induction*, where (i) the *history probability function* is *iso-derived historically distributed*, $P = P_{U,X,H_{\rm h},{\rm d},T_{\rm o}}$, given some *substrate transform* in the *sample variables* $T_{\rm o} \in \mathcal{T}_{U,V_{\rm o}}$, if it is the case that (ii) the *sample histogram* is *natural*, $A_{\rm o} = A_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger}$, then the *maximum likelihood estimate*, $\tilde{E}_{\rm o}$, of the *unknown distribution probability histogram*, $\hat{E}_{\rm o}$, in the *iso-derived conditional stuffed historical probability distribution*, $\hat{Q}_{{\rm h,d},T_{\rm o},U}(E_{\rm o}, z_{\rm o})$, is $\tilde{E}_{\rm o} = \hat{A}_{\rm o}$, so, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\rm h} \gg z_{\rm o}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{{\rm o},z_{\rm h}} \in \mathcal{A}_{\rm i}$, then the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}
&\ln \hat{Q}_{{\rm h,d},T_{\rm o},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o}) \ \sim \\
&\qquad (z_{\rm o} + v_{\rm o}) \times \mathrm{entropy}(A_{\rm o} * T_{\rm o} + V_{\rm o}^{\rm C} * T_{\rm o}) \\
&\qquad\qquad -z_{\rm o} \times \mathrm{entropy}(A_{\rm o} * T_{\rm o}) \ - v_{\rm o} \times \mathrm{entropy}(V_{\rm o}^{\rm C} * T_{\rm o})
\end{aligned}$$

In the case where the *size* is greater than the *volume*, $z_{\rm o} > v_{\rm o}$, the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *volume* scaled *component cardinality size relative entropy*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) &\sim v_{\mathrm{o}} \times \mathrm{entropyRelative}(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}, A_{\mathrm{o}} * T_{\mathrm{o}}) \\ &\sim -\ln \hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}} * T_{\mathrm{o}}, v_{\mathrm{o}})(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})\end{aligned}$$

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) &\sim z_{\mathrm{o}} \times \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) \\ &\sim -\ln \hat{Q}_{\mathrm{m},U}(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}} * T_{\mathrm{o}})\end{aligned}$$

In other words, the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *size* scaled *component size cardinality cross entropy*,

$$\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim z_{\mathrm{o}} \times \mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})$$

and against the *size* scaled *derived entropy*

$$\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim -z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

So, in this case, the *log likelihood* is maximised when (a) the *derived entropy* is minimised, and (b) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

This case, where the *sample* is approximately *natural*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, and the *maximum likelihood estimate* is approximately equal to the *naturalisation*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})} \approx A_{\mathrm{o}}$, may be compared to the case where the *sample* is not *natural*, $A_{\mathrm{o}} \neq A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, and the *maximum likelihood estimate* is not equal to the *naturalisation*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})} \neq A_{\mathrm{o}}$. In the *natural* case the *log likelihood* varies with the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned}&A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger} \implies \\ &\quad \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim z_{\mathrm{o}} \times \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})\end{aligned}$$

whereas in the non-*natural* case the *log likelihood* varies against the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned}&A_{\mathrm{o}} \neq A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger} \implies \\ &\quad \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim -\ z_{\mathrm{o}} \times \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})\end{aligned}$$

Now consider the *multinomial probability* terms that appear in the numerator and denominator of the *iso-derived conditional generalised multinomial probability distribution*. The logarithm of the *iso-derived conditional multinomial probability* at the *maximum likelihood estimate*, where the *sample* is *natural*, $A = A * T * T^{\dagger}$, is in proportion to the logarithm of the *sample multinomial probability* in the numerator divided by the sum of the *iso-derived multinomial probabilities* in the denominator,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \quad \propto \quad \ln \frac{\hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B)}$$

Given *integral iso-derived histogram* $B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)$ the logarithm of the *generalised multinomial probability*, where the *distribution histogram* is *natural*, $\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B)$, can be re-written in terms of *components*,

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B) \\
&\quad = \ln z! - z \ln z \; - \sum_{S \in B^{\mathrm{FS}}} \ln B_S! \; + \sum_{S \in B^{\mathrm{FS}}} B_S \ln(A * T * T^{\dagger})_S \\
&\quad = \ln z! - z \ln z \; - \sum_{(\cdot,C) \in T^{-1}} \sum_{S \in C^{\mathrm{S}}} \ln B_S! + \sum_{(R,\cdot) \in T^{-1}} (A * T)_R \ln \frac{(A * T)_R}{(V^{\mathrm{C}} * T)_R}
\end{aligned}$$

In the case of *natural distribution histogram*, $A = A * T * T^{\dagger}$, the *permutorial* term, $\sum_{(R,\cdot) \in T^{-1}} (A * T)_R \ln(A * T)_R/(V^{\mathrm{C}} * T)_R$, does not depend on the *iso-derived*, B, only on the *distribution histogram*, A. That is, the *permutorial* term is constant for all *iso-derived*. The *permutorial* term is proportional to the *size* scaled *component size cardinality relative entropy*,

$$\sum_{(R,\cdot) \in T^{-1}} (A * T)_R \ln \frac{(A * T)_R}{(V^{\mathrm{C}} * T)_R} \quad \propto \quad z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

The logarithm of the *multinomial probability* of the *iso-derived*, $\ln \hat{Q}_{\mathrm{m},U}(A*T*T^{\dagger}, z)(B)$, may be compared to the logarithm of the *multinomial probability* of the *derived*, $\ln \hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)$,

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{m},U}(A * T, z)(A * T) \\
&\quad = \ln z! - z \ln z \; - \sum_{(R,\cdot) \in T^{-1}} \ln(A * T)_R! \; + \sum_{(R,\cdot) \in T^{-1}} (A * T)_R \ln(A * T)_R
\end{aligned}$$

The *permutorial* term of the *derived* is proportional to the negative *size* scaled *derived entropy*,

$$\begin{aligned}
\sum_{(R,\cdot) \in T^{-1}} (A * T)_R \ln(A * T)_R \quad &\propto \quad z \times \mathrm{expected}(\hat{A} * T)(\hat{A} * T) \\
&= \quad -z \times \mathrm{entropy}(A * T)
\end{aligned}$$

The numerator is the *sample multinomial probability*. In the case where the *iso-derived* equals the *sample histogram*, $B = A$, the *multinomial* term, $\sum_{(\cdot,C)\in T^{-1}} \sum_{S \in C^{\mathrm{S}}} \ln B_S!$, simplifies to $\sum_{(R,\cdot)\in T^{-1}} (V^{\mathrm{C}} * T)_R \ln((A*T)_R/(V^{\mathrm{C}} * T)_R)!$, and the *multinomial probability* is maximised,

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger}) \\
&\quad = \ln z! - z \ln z \; - \sum_{(R,\cdot)\in T^{-1}} (V^{\mathrm{C}} * T)_R \ln \frac{(A*T)_R}{(V^{\mathrm{C}} * T)_R}! \\
&\qquad\qquad + \sum_{(R,\cdot)\in T^{-1}} (A*T)_R \ln \frac{(A*T)_R}{(V^{\mathrm{C}} * T)_R}
\end{aligned}$$

The difference between *multinomial probabilities* is the difference in *multinomial* terms, and varies with the difference in *size* scaled *entropies*,

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger}) \; - \; \ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B) \\
&\quad = \sum_{(\cdot,C)\in T^{-1}} \sum_{S \in C^{\mathrm{S}}} \ln B_S! \; - \sum_{(R,\cdot)\in T^{-1}} (V^{\mathrm{C}} * T)_R \ln \frac{(A*T)_R}{(V^{\mathrm{C}} * T)_R}! \\
&\quad \sim \; z \times \mathrm{entropy}(A * T * T^{\dagger}) - z \times \mathrm{entropy}(B)
\end{aligned}$$

The *sample multinomial probability* can be re-written,

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger}) \\
&\quad = \ln z! - z \ln z \; - \sum_{S \in A^{\mathrm{FS}}} \ln A_S! \; + \sum_{S \in A^{\mathrm{FS}}} A_S \ln A_S \\
&\quad = \ln z! - z \ln z \; - \sum_{(R,\cdot)\in T^{-1}} (V^{\mathrm{C}} * T)_R \ln \frac{(A*T)_R}{(V^{\mathrm{C}} * T)_R}! \\
&\qquad\qquad + \sum_{(R,\cdot)\in T^{-1}} (V^{\mathrm{C}} * T)_R \frac{(A*T)_R}{(V^{\mathrm{C}} * T)_R} \ln \frac{(A*T)_R}{(V^{\mathrm{C}} * T)_R}
\end{aligned}$$

In the case where the *cartesian derived* is *uniform*, $\forall (R,\cdot) \in T^{-1}\ ((V^{\mathrm{C}} * T)_R = v/w')$ where $w' = |T^{-1}|$, the *sample multinomial probability* can be written in terms of the *derived multinomial probability*,

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger}) \\
&\quad = \ln z! - z \ln z \; - \frac{v}{w'} \ln \frac{zw'}{v}! + z \ln \frac{zw'}{v} + \\
&\qquad\qquad \frac{v}{w'} \ln \hat{Q}_{\mathrm{m},U}(A * T, \frac{zw'}{v})(1/Z_{v/w'} * A * T) \\
&\quad \approx \ln z! - z \ln z \; - \frac{v}{w'} \ln \frac{zw'}{v}! + z \ln \frac{zw'}{v} \; + \ln \hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)
\end{aligned}$$

where $Z_{v/w'} = \text{scalar}(v/w')$.

The *component size cardinality relative entropy* is the *component size cardinality cross entropy* minus the *component size entropy* or *derived entropy*,

$$\begin{aligned}&\text{entropyRelative}(A * T, V^{\mathrm{C}} * T) \ = \\ &\qquad\qquad \text{entropyCross}(A * T, V^{\mathrm{C}} * T) \ - \ \text{entropy}(A * T)\end{aligned}$$

When the *cartesian derived* is *uniform*, $\text{ran}(\hat{V}^{\mathrm{C}} * T) = \{1/w'\}$, the *component size cardinality cross entropy* is a constant, $\ln w'$, and the *component size cardinality relative entropy* is in proportion to the negative *derived entropy*.

$$\text{expected}(\hat{A} * T)\left(\ln \frac{\hat{A} * T}{Z_{1/w'}}\right) \ = \ \ln w' \ - \ \text{entropy}(A * T)$$

In the case where the *cartesian derived* is *uniform*, the *derived multinomial probability* varies most closely to the *sample multinomial probability* and so, although the *sample multinomial probability* also appears in the denominator, the *iso-derived conditional multinomial probability* varies with the *derived multinomial probability*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A) \ \sim \ \ln \hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)$$

In the case where the *cartesian derived* is not *uniform* and the *cross entropy* is greater than the logarithm of the *possible derived volume*, $\text{entropyCross}(A * T, V^{\mathrm{C}} * T) > \ln w'$, the *relative entropy* exceeds the *uniform cartesian derived relative entropy*,

$$\text{entropyRelative}(A * T, V^{\mathrm{C}} * T) \ > \ \text{expected}(\hat{A} * T)\left(\ln \frac{\hat{A} * T}{Z_{1/w'}}\right)$$

and the *permutorial* term of the *sample multinomial probability* exceeds that of the scaled *derived multinomial probability*,

$$\sum_{(R,\cdot) \in T^{-1}} (A * T)_R \ln \frac{(A * T)_R}{(V^{\mathrm{C}} * T)_R} \ > \ \sum_{(R,\cdot) \in T^{-1}} (A * T)_R \ln \frac{(A * T)_R}{v/w'}$$

In this case the scaled logarithm of the *derived multinomial probability*, plus constants, is necessarily less than the logarithm of the *sample multinomial probability*

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger}) \\ &\quad > \ \ln z! - z \ln z \ - \frac{v}{w'} \ln \frac{zw'}{v}! + z \ln \frac{zw'}{v} + \\ &\qquad\qquad\qquad\qquad \frac{v}{w'} \ln \hat{Q}_{\mathrm{m},U}(A * T, \frac{zw'}{v})(1/Z_{v/w'} * A * T)\end{aligned}$$

and so the scaled logarithm of the *derived multinomial probability* approximates more closely to the lower *entropy iso-derived* that are not equal to the *sample*, $B \neq A$,

$$\begin{aligned}
&\exists B \in D^{-1}_{U,\mathrm{i},T,z}(A * T) \\
&\quad (\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A) - \frac{v}{w'} \ln \hat{Q}_{\mathrm{m},U}(A * T, \frac{zw'}{v})(1/Z_{v/w'} * A * T) \\
&\quad > \ \ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B) - \frac{v}{w'} \ln \hat{Q}_{\mathrm{m},U}(A * T, \frac{zw'}{v})(1/Z_{v/w'} * A * T))
\end{aligned}$$

In this case of high *relative entropy*, the *derived multinomial probability* still varies with the *sample multinomial probability*, but varies more closely to the other *iso-derived* that appear in the denominator, so now the *iso-derived conditional multinomial probability* varies against the *derived multinomial probability*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A) \ \sim \ - \ \ln \hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)$$

The degree of anti-correlation varies with the *relative entropy*.

In the third case where the *cartesian derived* is not *uniform* but the *cross entropy* is less than the logarithm of the *possible derived volume*, $\mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) < \ln w'$, the *relative entropy* is less than the *uniform cartesian derived relative entropy*,

$$\mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T) \ < \ \mathrm{expected}(\hat{A} * T) \left(\ln \frac{\hat{A} * T}{Z_{1/w'}} \right)$$

Now the scaled logarithm of the *derived multinomial probability*, plus constants, is necessarily greater than the logarithm of the *sample multinomial probability*

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger}) \\
&\quad < \ \ln z! - z \ln z \ - \frac{v}{w'} \ln \frac{zw'}{v}! + z \ln \frac{zw'}{v} + \\
&\qquad\qquad \frac{v}{w'} \ln \hat{Q}_{\mathrm{m},U}(A * T, \frac{zw'}{v})(1/Z_{v/w'} * A * T)
\end{aligned}$$

and so the *sample* is the closest of the *iso-derived*. In this case of low *relative entropy*, the *iso-derived conditional multinomial probability* varies with the *derived multinomial probability*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A) \ \sim \ \ln \hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)$$

As the *cross entropy* tends to the *derived entropy*, the *relative entropy* tends to zero, so the *derived* and *cartesian derived* become perfectly synchronised and the *sample* tends to *uniform*, $\hat{A} = \hat{V}^{\mathrm{C}}$.

In *classical modelled induction*, where (i) the *history probability function* is *iso-derived historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample histogram* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-derived conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iii) the *component size cardinality relative entropy* is high, $\mathrm{entropyCross}(A_{\mathrm{o}}*T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}}*T_{\mathrm{o}}) > \ln w'_{\mathrm{o}}$, (iv) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (v) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *derived multinomial probability*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad - \ \ln \hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}} * T_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}} * T_{\mathrm{o}})$$

Consider further the *multinomial probability* terms that appear in the numerator and denominator of the *iso-derived conditional generalised multinomial probability distribution* where the *distribution histogram* is *natural*, $A*T*T^{\dagger}$. It is shown above that that the logarithm of the *generalised multinomial probability*, re-written in terms of *components*, is,

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B) \\
&\quad = \ \ln z! - z \ln z \ - \sum_{(\cdot,C) \in T^{-1}} \sum_{S \in C^{\mathrm{S}}} \ln B_S! + \sum_{(R,\cdot) \in T^{-1}} (A * T)_R \ln \frac{(A * T)_R}{(V^{\mathrm{C}} * T)_R}
\end{aligned}$$

The *permutorial* term is constant for all *iso-derived*, so the *iso-derived conditional multinomial probability* simplifies to

$$\frac{\hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B)} \ = \ 1/ \sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \prod_{S \in V^{\mathrm{CS}}} \frac{A_S!}{B_S!}$$

The minimum terms in the denominator are such that the *iso-derived histogram* $B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)$ is *singleton* in all of its *components*, $\forall (\cdot, C) \in T^{-1}(|(B * C)^{\mathrm{F}}| = 1)$. The logarithm of a minimum term approximates to the *size* scaled *component size cardinality cross entropy*,

$$\begin{aligned}
\ln \prod_{S \in V^{\mathrm{CS}}} \frac{A_S!}{B_S!} &= \sum_{(R,\cdot) \in T^{-1}} (V^{\mathrm{C}} * T)_R \ln \frac{(A * T)_R}{(V^{\mathrm{C}} * T)_R}! - \sum_{(R,\cdot) \in T^{-1}} \ln (A * T)_R! \\
&\approx -\sum_{(R,\cdot) \in T^{-1}} (A * T)_R \ln (V^{\mathrm{C}} * T)_R \\
&\propto z \times \mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T)
\end{aligned}$$

So, in the case where the *size* is greater than the *volume*, $z > v$, and the *sample* is *natural*, $A = A * T * T^{\dagger}$, the logarithm of the *iso-derived conditional multinomial probability* varies against the *size* scaled *component size cardinality cross entropy*,

$$\ln \frac{\hat{Q}_{\mathrm{m},U}(A,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \hat{Q}_{\mathrm{m},U}(A,z)(B)} \sim -z \times \mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T)$$

This is somewhat contrary to relationship with respect to the *iso-derived* cardinality where it was shown that the logarithm of the *iso-derived conditional multinomial probability* varies with the *volume* scaled *component cardinality size cross entropy*,

$$\ln \frac{\hat{Q}_{\mathrm{m},U}(A,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \hat{Q}_{\mathrm{m},U}(A,z)(B)} \sim v \times \mathrm{entropyCross}(V^{\mathrm{C}} * T, A * T)$$

However, note that the logarithm of the cardinality of the *singleton-component iso-derived* is only $\sum_{(R,\cdot) \in T^{-1}} \ln (V^{\mathrm{C}} * T)_R$, which is much smaller than the logarithm of the cardinality of the *iso-derived*, approximately $\sum_{(R,\cdot) \in T^{-1}} (V^{\mathrm{C}} * T)_R \ln ((A * T)_R / (V^{\mathrm{C}} * T)_R)$, especially in the case where the *component cardinality size relative entropy* is low. So the negative correlation between the *iso-derived conditional multinomial probability* and the *size* scaled *component size cardinality cross entropy* is weak.

Also, in the near-*natural* case where the *trimmed sample* is *unit*, $A * A^{\mathrm{F}} = A^{\mathrm{F}}$, the logarithm of a minimum term approximates to the *size* scaled *derived entropy*,

$$\begin{aligned}
\ln \prod_{S \in V^{\mathrm{CS}}} \frac{A_S!}{B_S!} &= -\sum_{(R,\cdot) \in T^{-1}} \ln (A * T)_R! \\
&\approx z \times \mathrm{entropy}(A * T)
\end{aligned}$$

So, in this case where the *size* is less than the *volume*, $z < v$, and the *sample* is near-*natural*, $A \approx A * T * T^{\dagger}$, the logarithm of the *iso-derived conditional multinomial probability* varies against the *size* scaled *derived entropy*,

$$\ln \frac{\hat{Q}_{\mathrm{m},U}(A,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \hat{Q}_{\mathrm{m},U}(A,z)(B)} \quad \sim \quad -z \times \mathrm{entropy}(A * T)$$

which agrees with the relationship with respect to the *iso-derived* cardinality where it was shown that the logarithm of the *iso-derived conditional multinomial probability* varies with the *size* scaled *component size cardinality relative entropy*,

$$\ln \frac{\hat{Q}_{\mathrm{m},U}(A,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \hat{Q}_{\mathrm{m},U}(A,z)(B)} \quad \sim \quad z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

In section 'Derived history space', above, the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)$, is constructed,

$$\begin{aligned} &C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T) = \\ &\quad \mathrm{coderHistorySubstrateDerivedSpecialising}(U, X, T, D_{\mathrm{S}}, D_{\mathrm{X}}) \\ &\qquad\qquad \in \ \mathrm{coders}(\mathcal{H}_{U,V,X}) \end{aligned}$$

where $\mathcal{H}_{U,V,X} = \{H : H \in \mathcal{H}_{U,X},\ \mathrm{vars}(H) = V\}$. The *space* of the *specialising coder* is

$$\begin{aligned} \mathrm{space}(C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T))(H) &= \mathrm{spaceIds}(|X|, |H|) + \\ &\quad \mathrm{spaceCountsDerived}(U)(A, T) + \\ &\quad \mathrm{spaceClassification}(A * T) + \\ &\quad \mathrm{spaceEventsPartition}(A, T) \\ &= \mathrm{spaceIds}(|X|, |H|) + \\ &\quad \ln \frac{(z + w' - 1)!}{z!\ (w' - 1)!} + \\ &\quad \ln z! - \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A * T)_R! + \\ &\quad \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C| \end{aligned}$$

The *space* of the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)$, varies (i) with the *possible derived volume*, w', where the *possible derived* *volume* is less than the *size*, $w' < z$, otherwise with the *size* scaled *log possible derived volume*, $z \ln w'$, and (ii) against the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned} C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) \ \sim \\ & (w' \ :\ w' < z) \ +\ (z \ln w' \ :\ w' \geq z) \\ & -\ z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T) \end{aligned}$$

So the *space* of the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)$, is minimised when (a) the *possible derived volume* is minimised, (b) the *derived entropy* or *component size entropy* is minimised, and (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *integral iso-deriveds log-cardinality* varies against the *size* scaled *component size cardinality relative entropy*. The sum of the *derived classification space* and the *partitioned events space* varies against the *size* scaled *component size cardinality relative entropy*. So the *integral iso-deriveds log-cardinality* varies with the sum of the *derived classification space* and the *partitioned events space*,

$$\begin{aligned} \ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \ &\sim\ -z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T) \\ &\sim\ \mathrm{spaceClassification}(A * T) + \\ &\quad\ \mathrm{spaceEventsPartition}(A, T) \end{aligned}$$

So in this domain the *log likelihood* varies against the *specialising space*.

Conjecture that in the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising derived substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim\ -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

The *iso-derived conditional stuffed historical probability distribution log likelihood* is maximised and the *specialising derived substrate history coder space* is minimised by varying the *transform* such that (i) the *derived entropy* is low, and (ii) high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*.

In section 'Derived history space', above, the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, is shown to be characterised by certain properties. The *specialising-canonical space difference* varies (i) with twice the *possible derived volume*, $2w'$, where $w' < z$, otherwise with twice the *size* scaled *log possible derived volume*, $2z \ln w'$, (ii) with the *size* scaled *derived entropy*, (iii) against twice the *size* scaled *component size cardinality cross entropy* and (iv) against the *size* scaled *size expected component entropy*,

$$\begin{aligned}
2C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) \ \sim \\
&2\big((w' \ : \ w' < z) \ + \ (z \ln w' \ : \ w' \geq z)\big) \\
&+ z \times \mathrm{entropy}(A * T) \\
&- 2z \times \mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) \\
&- z \times \mathrm{entropyComponent}(A, T)
\end{aligned}$$

So the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, is minimised when (a) the *possible derived volume* is minimised, (b) the *derived entropy* is minimised, (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*, and (d) the *expected component entropy* is maximised.

The *canonical* term, $C^{\mathrm{s}}_{\mathrm{H},V}(H) + C^{\mathrm{s}}_{\mathrm{G},V}(H)$, is independent of the *model*, T, so properties of the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, are also properties of the *specialising space*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H)$. So conjecture that in *classical modelled induction* where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising-canonical space difference*,

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \\
&-(2C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}})^{\mathrm{s}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{H},V_{\mathrm{o}}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{G},V_{\mathrm{o}}}(H_{\mathrm{o}}))
\end{aligned}$$

In the special case where the *histogram* is *natural*, $A = A * T * T^{\dagger} \implies \tilde{E} = \hat{A}^{\mathrm{D}(T)} = \hat{A}$, and the *component size cardinality cross entropy* is greater than the logarithm of the *possible derived volume*, $\mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) > \ln w'$, so the *relative entropy* is high, conjecture that the *iso-derived conditional multinomial probability* at the *maximum likelihood estimate* varies with the *underlying-derived relative multinomial probability*,

$$\frac{\hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B)} \sim \frac{\hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})}{\hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)}$$

This may be generalised to cases where the *histogram* is not *natural*, $A \neq A*T*T^{\dagger}$, but only approximated to the *naturalisation*, $A \approx A*T*T^{\dagger}$, such that the *relative space* with respect to the *naturalisation*, $\mathrm{spaceRelative}(A * T * T^{\dagger})(A)$, is small,

$$\frac{\hat{Q}_{\mathrm{m},U}(E, z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \hat{Q}_{\mathrm{m},U}(E, z)(B)} \sim \frac{\hat{Q}_{\mathrm{m},U}(E, z)(A)}{\hat{Q}_{\mathrm{m},U}(E * T, z)(A * T)}$$

In the case of a *full functional transform*, $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}}$, this correlation is exact, $\sum_{B \in D^{-1}_{U,\mathrm{i},T_{\mathrm{s}},z}(A*T_{\mathrm{s}})} \hat{Q}_{\mathrm{m},U}(E, z)(B) = \hat{Q}_{\mathrm{m},U}(E*T_{\mathrm{s}}, z)(A*T_{\mathrm{s}})$, because $D^{-1}_{U,\mathrm{i},T_{\mathrm{s}},z}(A * T_{\mathrm{s}}) = \{A\}$. At the other extreme of a *unary transform*, $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, the correlation is also exact, $\sum_{B \in D^{-1}_{U,\mathrm{i},T_{\mathrm{u}},z}(A*T_{\mathrm{u}})} \hat{Q}_{\mathrm{m},U}(E, z)(B) = \hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{u}}, z)(A * T_{\mathrm{u}}) = 1$, because $D^{-1}_{U,\mathrm{i},T_{\mathrm{u}},z}(A * T_{\mathrm{u}}) = \mathcal{A}_{U,\mathrm{i},V,z}$.

The numerator of the *underlying-derived relative multinomial probability* corresponds to the *unknown underlying*, while the denominator corresponds to the *known derived*. Thus, in the case of high *component size cardinality relative entropy*, the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},T,U}$, is conjectured to vary with the *unknown-known multinomial probability distribution sum sensitivity difference*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T,U}(E, z))) \sim \\ &\quad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E, z))) - \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E * T, z)))\end{aligned}$$

and so the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* is sometimes less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T,U}(E, z))) \\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(E, z)))\end{aligned}$$

In the case of a *full functional transform*, T_{s}, the *iso-derived historically distributed history probability* is a constant, so the *iso-derived conditional stuffed historical probability distribution sum sensitivity* is zero,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{s}},U}(E, z))) = 0$$

Here the *underlying* is *known* and so the *unknown-known multinomial probability distribution sum sensitivity difference* is also zero,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E,z))) = \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E * T_{\mathrm{s}},z)))$$

In the case of a *unary transform*, T_{u}, the *iso-derived conditional stuffed historical probability distribution sum sensitivity* equals the *stuffed historical probability distribution sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{u}},U}(E,z))) = \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(E,z)))$$

Here the *derived multinomial probability distribution sum sensitivity* is undefined, because the *derived volume* is singular, $w = 1$, and so the *unknown-known multinomial probability distribution sum sensitivity difference* is undefined.

Given a *transform* $T \in \mathcal{T}_{U,V}$ the *multinomial probability* can be written in terms of its *component multinomial probabilities*,

$$\begin{aligned}
&\hat{Q}_{\mathrm{m},U}(E,z)(A) \\
&\quad= \frac{z!}{\prod_{S \in V^{\mathrm{CS}}} A_S!} \prod_{S \in V^{\mathrm{CS}}} \hat{E}_S^{A_S} \\
&\quad= \frac{z!}{\prod_{(R,\cdot) \in T^{-1}} (A * T)_R!} \prod_{(R,C) \in T^{-1}} \frac{(A * T)_R!}{\prod_{S \in C^{\mathrm{S}}} (A * C)_S!} \prod_{S \in C^{\mathrm{S}}} (\hat{E} * C)_S^{(A*C)_S} \\
&\quad= \frac{z!}{\prod_{(R,\cdot) \in T^{-1}} (A * T)_R!} \prod_{(R,C) \in T^{-1}} \hat{Q}_{\mathrm{m},U}(E * C, (A * T)_R)(A * C)
\end{aligned}$$

where the *distribution histogram* is *completely effective*, $E^{\mathrm{F}} = V^{\mathrm{C}}$. The *derived multinomial probability* is

$$\begin{aligned}
\hat{Q}_{\mathrm{m},U}(E * T,z)(A * T) &= \frac{z!}{\prod_{R \in (V^{\mathrm{C}} * T)^{\mathrm{FS}}} (A * T)_R!} \prod_{R \in (V^{\mathrm{C}} * T)^{\mathrm{FS}}} (\hat{E} * T)_R^{(A*T)_R} \\
&= \frac{z!}{\prod_{(R,\cdot) \in T^{-1}} (A * T)_R!} \prod_{(R,\cdot) \in T^{-1}} (\hat{E} * T)_R^{(A*T)_R}
\end{aligned}$$

So the *underlying-derived relative multinomial probability* can be rewritten in terms of *components*,

$$\frac{\hat{Q}_{\mathrm{m},U}(E,z)(A)}{\hat{Q}_{\mathrm{m},U}(E * T,z)(A * T)} = \frac{\prod_{(R,C) \in T^{-1}} \hat{Q}_{\mathrm{m},U}(E * C, (A * T)_R)(A * C)}{\prod_{(R,\cdot) \in T^{-1}} (\hat{E} * T)_R^{(A*T)_R}}$$

The *unknown-known multinomial probability distribution sum sensitivity difference* has several properties corresponding to the numerators and denominators of each side of this equation.

First, the *unknown-known sum sensitivity difference* varies with the *sum sensitivity* of the numerator of the left hand side, which is the *multinomial probability distribution sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E,z))) \quad = \quad \sum_{S \in V^{\mathrm{CS}}} \frac{z}{\hat{E}_S(1-\hat{E}_S)}$$

As shown above, the *multinomial probability distribution sum sensitivity* varies against the *scaled entropy*, so the *iso-derived conditional stuffed historical probability distribution sum sensitivity* varies against the *underlying entropy*

$$\begin{aligned}\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T,U}(E,z))) \quad &\sim \quad \sum_{S \in V^{\mathrm{CS}}} \frac{z}{\hat{E}_S(1-\hat{E}_S)} \\ &\sim \quad -z \times \mathrm{entropy}(E)\end{aligned}$$

The *underlying entropy*, $\mathrm{entropy}(E)$, is independent of the *transform*, T, and so remains constant during any optimisation of the *sum sensitivity* by varying the *transform*.

Second, the *unknown-known sum sensitivity difference* varies against the *sum sensitivity* of the denominator of the left hand side, which is the *derived multinomial probability distribution sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E*T,z))) \quad = \quad \sum_{(R,\cdot) \in T^{-1}} \frac{z}{(\hat{E}*T)_R \ (1-(\hat{E}*T)_R)}$$

so the *iso-derived conditional stuffed historical probability distribution sum sensitivity* varies with the *derived entropy*,

$$\begin{aligned}\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T,U}(E,z))) \quad &\sim \quad -\sum_{(R,\cdot) \in T^{-1}} \frac{z}{(\hat{E}*T)_R \ (1-(\hat{E}*T)_R)} \\ &\sim \quad z \times \mathrm{entropy}(E*T)\end{aligned}$$

Third, the *unknown-known sum sensitivity difference* varies with the *sum sensitivity* of the numerator of the right hand side, which is the sum of the *unknown multinomial probability distribution sum sensitivities*,

$$\sum_{(R,C) \in T^{-1}} \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(E * C, (A * T)_R))) = \sum_{(R,C) \in T^{-1}} \sum_{S \in C^{\text{S}}} \frac{(A * T)_R}{\hat{E}_S(1 - \hat{E}_S)}$$

so the *iso-derived conditional stuffed historical probability distribution sum sensitivity* varies against the *unknown size* scaled *expected component entropy*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,d},T,U}(E, z))) \quad \sim \quad -z \times \text{entropyComponent}(E, T)$$

Fourth, the denominator of the right hand side, $\prod_{(R,\cdot) \in T^{-1}} (\hat{E} * T)_R^{(A*T)_R}$, is the *permutorial* part of the *derived multinomial probability*, $\hat{Q}_{\text{m},U}(E * T, z)(A * T)$. The other part is the *multinomial coefficient*, $z!/\prod_{R \in (A*T)^{\text{FS}}} (A * T)_R!$, which does not depend on the *distribution histogram*, E. So the *sum sensitivity* of the denominator of the right hand side varies with the *sum sensitivity* of the denominator of the left hand side, which is the *derived multinomial probability distribution sum sensitivity*, $\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(E * T, z)))$. Again, the *iso-derived conditional stuffed historical probability distribution sum sensitivity* varies with the *derived entropy*, $\text{entropy}(E * T)$.

In *classical modelled induction*, where (i) the *history probability function* is *iso-derived historically distributed*, $P = P_{U,X,H_{\text{h}},\text{d},T_{\text{o}}}$, given some *substrate transform* in the *sample variables* $T_{\text{o}} \in \mathcal{T}_{U,V_{\text{o}}}$, if it is the case that (ii) the *sample histogram* is *natural*, $A_{\text{o}} = A_{\text{o}} * T_{\text{o}} * T_{\text{o}}^{\dagger}$, then the *maximum likelihood estimate*, $\tilde{E}_{\text{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\text{o}}$, in the *iso-derived conditional stuffed historical probability distribution*, $\hat{Q}_{\text{h,d},T_{\text{o}},U}(E_{\text{o}}, z_{\text{o}})$, is $\tilde{E}_{\text{o}} = \hat{A}_{\text{o}}$, so, if it is also the case that (iii) the *component size cardinality relative entropy* is high, $\text{entropyCross}(A_{\text{o}} * T_{\text{o}}, V_{\text{o}}^{\text{C}} * T_{\text{o}}) > \ln w'_{\text{o}}$, (iv) the *distribution history size* is large with respect to the *sample size*, $z_{\text{h}} \gg z_{\text{o}}$, and such that (v) the *scaled probability sample histogram* is *integral*, $A_{\text{o},z_{\text{h}}} \in \mathcal{A}_{\text{i}}$, then the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,d},T_{\text{o}},U}(A_{\text{o},z_{\text{h}}}, z_{\text{o}}))) \leq \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h},U}(A_{\text{o},z_{\text{h}}}, z_{\text{o}})))$$

(b) varies with the *derived entropy*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

and (c) varies against the *unknown size* scaled *expected component entropy*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim -z_{\mathrm{o}} \times \mathrm{entropyComponent}(A_{\mathrm{o}}, T_{\mathrm{o}})$$

The *derived entropy* varies with the *derived classification space*, and so varies with the *specialising derived substrate history coder space*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H)$. Conjecture that in the case where the *sample* equals the *naturalisation*, $A_{\mathrm{o}} = A_{\mathrm{o}}*T_{\mathrm{o}}*T_{\mathrm{o}}^{\dagger}$, the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *specialising derived substrate history coder space*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

Both the *iso-derived conditional stuffed historical probability distribution sum sensitivity* and the *specialising derived substrate history coder space* are minimised by varying the *transform* such that the *derived entropy* is low.

Also, the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H)-C^{\mathrm{s}}_{\mathrm{H},V}(H)-C^{\mathrm{s}}_{\mathrm{G},V}(H)$, varies with *derived entropy* and against the *size* scaled *expected component entropy*, so conjecture that in the case where the *sample* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *specialising-canonical space difference*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim 2C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}})^{\mathrm{s}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{H},V_{\mathrm{o}}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{G},V_{\mathrm{o}}}(H_{\mathrm{o}})$$

Both the *iso-derived conditional stuffed historical probability distribution sum sensitivity* and the *specialising-canonical space difference* are minimised by varying the *transform* such that (a) the *derived entropy* is low and (b) the *underlying components* have high *entropy*.

Altogether, in *classical modelled induction* where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, and the *relative entropy* is high, the *sum sensitivity* has similar properties as the *log-likelihood* but with the correlations reversed. Conjecture that in this case the *sum sensitivity* varies against the *log-likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim -\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

That is, in the high *relative entropy natural* case, the maximisation of the *log-likelihood* also tends to minimise the *sum sensitivity* to the *maximum likelihood estimate*. This is opposite to the relationship between the *sum sensitivity* and the *log-likelihood* in *classical non-modelled induction*, which was found to be weakly positively correlated.

In the case where there are *ineffective possible derived states*, $|(A_\mathrm{o} * T_\mathrm{o})^\mathrm{F}| < w'_\mathrm{o}$ where $w'_\mathrm{o} = |T_\mathrm{o}^{-1}|$, the *distribution histogram* is *known* to be *incompletely effective*, $E_\mathrm{o}^\mathrm{F} < V_\mathrm{o}^\mathrm{C}$. The *states known* to be *ineffective* are in the *ineffective components*,

$$\forall (R,C) \in T_\mathrm{o}^{-1}\ ((A_\mathrm{o} * T_\mathrm{o})_R = 0 \implies E_\mathrm{o} * C = C^\mathrm{Z})$$

Now the *classical modelled induction* assumption of a *completely effective sample histogram*, $A_\mathrm{o}^\mathrm{F} = V_\mathrm{o}^\mathrm{C}$, can be weakened to requiring only that (i) there are at least two *effective states*, $|A_\mathrm{o}^\mathrm{F}| > 1$, and (ii) the *components* of *effective derived states* are *effective*, $\forall (R,C) \in T_\mathrm{o}^{-1}\ ((A_\mathrm{o} * T_\mathrm{o})_R > 0 \implies (A_\mathrm{o} * C)^\mathrm{F} = C)$. The *maximum likelihood estimate* is unchanged, $\tilde{E}_\mathrm{o} = \hat{A}_\mathrm{o}$. Although the coordinate has smaller dimension, $(\hat{A}_\mathrm{o} * A_\mathrm{o}^\mathrm{F})^{\square} \in \mathcal{R}_{(0,1)}^{|A_\mathrm{o}^\mathrm{F}|} \neq \mathcal{R}_{(0,1)}^{v_\mathrm{o}}$, there is no *effective* normalising factor, $1/\mathrm{size}(\tilde{E}_\mathrm{o} * A_\mathrm{o}^\mathrm{F}) = 1$, and both the *log likelihood* and the *sum sensitivity* are the same as for the case of *completely effective derived histogram*.

In the case where the requirement of *completely effective effective components* does not hold, $\exists (R,C) \in T_\mathrm{o}^{-1}\ ((A_\mathrm{o} * T_\mathrm{o})_R > 0\ \wedge\ (A_\mathrm{o} * C)^\mathrm{F} < C)$, then there is an *unknown effective* normalising factor for each of the *incompletely effective components*, $\mathrm{size}(\tilde{E}_\mathrm{o} * C)/\mathrm{size}(\tilde{E}_\mathrm{o} * C * A_\mathrm{o}^\mathrm{F})$.

Note, however, that the maximisation of the *log likelihood* or the minimisation of the *sum sensitivity* tend to maximise the *component size cardinality relative entropy* and so *components* with larger *sizes* tend to have smaller *volumes* and the cardinality of *ineffective component states*, $C \setminus (A_\mathrm{o} * C)^\mathrm{F}$, tends to be minimised.

In the case where the *histogram* is *naturalised*, $A_\mathrm{o} = A_\mathrm{o} * T_\mathrm{o} * T_\mathrm{o}^\dagger$, the *effective components* are *completely effective*, $\forall (R,C) \in T_\mathrm{o}^{-1}\ ((A_\mathrm{o} * T_\mathrm{o})_R > 0 \implies (A_\mathrm{o} * C)^\mathrm{F} = C)$, and so there are no *unknown effective* normalising factors.

In the case where there are *ineffective possible derived states*, but there are at least two *effective derived states*, $1 < |(A_\mathrm{o} * T_\mathrm{o})^\mathrm{F}| < w'_\mathrm{o}$, then the *derived* coordinate has smaller, but not unit, dimension, $(\hat{A}_{\mathrm{o}} * T_{\mathrm{o}} * (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{F}})^{\square} \notin \{\{1\}, \mathcal{R}_{(0,1)}^{w'_{\mathrm{o}}}\}$, and the *derived multinomial probability distribution sum sensitivity* is,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A_{\mathrm{o},z_{\mathrm{h}}} * T_{\mathrm{o}} * (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{F}}, z_{\mathrm{o}}))) = \\ &\qquad\qquad \sum_{R \in (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{FS}}} \frac{z_{\mathrm{o}}}{(\hat{A}_{\mathrm{o}} * T_{\mathrm{o}})_R \ (1 - (\hat{A}_{\mathrm{o}} * T_{\mathrm{o}})_R)}\end{aligned}$$

In the case where the *derived histogram* is *known*, the *derived effective* normalising factor is *known* to be one, $1/\mathrm{size}(\tilde{E}_{\mathrm{o}} * T_{\mathrm{o}} * (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{F}}) = 1$. If, however, the *knowledge* of the *derived histogram* is less than certain and there is some doubt about the *effectiveness* of weakly *effective states*, it may be noted that as the *effectiveness* of the *derived states* decreases, the *derived entropy* decreases and the *derived sensitivity* increases, tending to infinity in the limit,

$$\lim_{(A_{\mathrm{o}} * T_{\mathrm{o}})_R \to 0} \frac{z_{\mathrm{o}}}{(A_{\mathrm{o}} * T_{\mathrm{o}})_R \ (1 - (A_{\mathrm{o}} * T_{\mathrm{o}})_R)} = \infty$$

In this domain of low *derived entropy* the variation of the *derived sensitivity* remains against the *derived entropy* as *ineffectiveness* increases. That is, even in the case of uncertain *derived histogram*, $\hat{A}_{\mathrm{o}} * T_{\mathrm{o}} \approx \hat{E}_{\mathrm{o}} * T_{\mathrm{o}}$, the *iso-derived conditional stuffed historical probability distribution sum sensitivity* continues to vary with the *derived entropy*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad z \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

Just as for *non-modelled classical induction* the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate*, $\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))$, can be related to queries on the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, in the special case where (i) the *sample histogram* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, and (ii) the *component size cardinality relative entropy* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln w'_{\mathrm{o}}$. In the case of *classical modelled induction*, the given *substrate transform* must be such that its *contraction* has *underlying variables* that are a subset of the query *variables*, $\mathrm{und}(T_{\mathrm{o}}^{\%}) \subseteq K$. In the case where the query *histogram* consists of one *effective state*, $Q = \{(S_Q, 1)\}$, there exists an *effective derived state* R_Q, where $\{R_Q\} = (Q * T_{\mathrm{o}}^{\%})^{\mathrm{FS}}$. The corresponding *underlying component* is $C_Q = T_{\mathrm{o}}^{-1}(R_Q)$. In this case the application of the query via the *model* equals the application via the *component* directly, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K) = (A_{\mathrm{o}} * C_Q)^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K)$. If the *possible derived volume* is non singular, $w'_{\mathrm{o}} > 1$, the query *histogram* itself cannot be *drawn* from the *distribution history*,

$\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o}},1)(Q * \{N\}^{\mathrm{U}}) = 0$, where $N \in (V_{\mathrm{o}} \setminus K)^{\mathrm{CS}}$, because the query *derived probability histogram* is not equal to the *known derived distribution probability histogram*, $\hat{Q} * \{N\}^{\mathrm{U}} * T_{\mathrm{o}} \neq \hat{A}_{\mathrm{o}} * T_{\mathrm{o}}$. The application of the query must be in terms of a modified *sample histogram*,

$$\begin{aligned}
&(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K) = \\
&\quad \{(N, (\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}))^{1/z_{\mathrm{o}}}) : N \in (V_{\mathrm{o}} \setminus K)^{\mathrm{CS}}, \\
&\qquad\qquad A_{Q,N} = A_{\mathrm{o}} - (A_{\mathrm{o}} * C_Q) + ((A_{\mathrm{o}} * C_Q) \ \% \ K * \{N\}^{\mathrm{U}})\}^{\wedge}
\end{aligned}$$

where his = histogram. If the *sample histogram* is *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the modified *sample histogram*, $A_{Q,N}$, can be *drawn* from the *distribution*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}) > 0$, because its *derived* is equal to the *known derived*, $A_{Q,N} * T_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}}$. The modified *sample histogram* is in the *iso-deriveds*, $A_{Q,N} \in D^{-1}_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}}(A_{\mathrm{o}} * T_{\mathrm{o}})$, and so only the numerator of the *iso-derived conditional stuffed historical probability* has changed,

$$\frac{\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N})}{\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} = \frac{\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N})}{\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}$$

In the case of (i) a *completely effective histogram*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}} \implies Q^{\mathrm{F}} \leq (A_{\mathrm{o}}\%K)^{\mathrm{F}}$, and (ii) a *self transform* with respect to query *variables*, $T_{\mathrm{s}} = K^{\mathrm{CS}\{\}V_{\mathrm{o}}\mathrm{T}}$, the query application via the *model* equals the *estimated transformed conditional product*,

$$(Q * T_{\mathrm{s}}^{\%} * \mathrm{his}(T_{\mathrm{s}}^{\%}) * A_{\mathrm{o}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K) = \hat{Q} * T'_{\hat{A}_{\mathrm{o}},K} \in \mathcal{A} \cap \mathcal{P}$$

As for *non-modelled classical induction*, the *model* application depends on the geometric scaling of the *historical distribution*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$, so the *query sensitivity* to the *distribution histogram* varies with the *sum sensitivity* of the *historical distribution* at the *maximum likelihood estimate* divided by the *sample size*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}$$

Although the query application via the *model* is sometimes not equal to the *estimated transformed conditional product*, the *query sensitivity* to the *distribution histogram* is sometimes lower,

$$\begin{aligned}
&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}} \\
&\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}
\end{aligned}$$

Similarly, where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the *model likelihood* of the *distribution histogram* is sometimes higher,

$$\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{Q,N}) \ \geq \ \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{Q,N})$$

In other words, querying via the *known derived* of the *model* sometimes reduces the *sensitivity* to the *unknown* and increases the *likelihood* at the cost of modifying the query.

If it is *known* that the *sample* is not *natural*, $A_o \neq A_o * T_o * T_o^\dagger$, for example, if it is *known* that the label *variables* are a function of the query *variables*

$$\mathrm{split}(K, E_o^{FS}) \in K^{CS} \to (V_o \setminus K)^{CS}$$

then in some cases the *states* of the modified query, $A_{Q,N}$, may be *ineffective* in the *effective sample component*,

$$(A_{Q,N} * C_Q)^F \not\leq (A_o * C_Q)^F$$

In these cases there is an *unknown* normalising factor for the *component*,

$$\mathrm{size}(\tilde{E}_o * C_Q)/\mathrm{size}(\tilde{E}_o * C_Q * A_o^F)$$

but there is not necessarily an *unknown* normalising factor in the query *variables*

$$\mathrm{size}(\tilde{E}_o * C_Q \ \% \ K)/\mathrm{size}(\tilde{E}_o * C_Q \ \% \ K * A_o^F)$$

That is, even if the *non-modelled* query is *ineffective*, $Q^F \cap A_o^F\%K = \emptyset$, if the *derived* is *effective*, $(Q * T_o)^F \leq (A_o * T_o)^F$, then it is *necessarily* the case that the query, Q, may be carried out via the *model* by modifying it to $A_{Q,N}$, which for some N there exists a *drawn history*, $\exists N \in (V_o \setminus K)^{CS}\ (\hat{Q}_{h,d,T_o,U}(A_{o,z_h}, z_o)(A_{Q,N}) > 0)$, subject to the *unknown* normalising factor if $((A_o * C_Q)\%K)^F < (C_Q\%K)^F$.

5.5.2 Necessary derived functional definition set

So far the discussion of *classical modelled induction* has considered the case where the *known model* is a *transform*. Consider extending the *model* first to *functional definition sets* and then to *fud decompositions*.

Given some *known substrate fud*, $F_o \in \mathcal{F}_{U,V_o}$, such that there exists a *top transform*, $\exists T \in F_o\ (\mathrm{der}(T) = \mathrm{der}(F_o))$, the *derived histogram set* of the *distribution probability histogram* is $\{\hat{E}_h * T_{F_o} : T \in F_o\}$, where $T_F := \mathrm{depends}(F, \mathrm{der}(T))^T$. In *classical functional definition set induction*, while the *distribution probability histogram*, $\hat{E}_h$, remains *unknown*, the *derived distribution probability histogram set*, $\{\hat{E}_h * T_{F_o} : T \in F_o\}$, is *known* and *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have a *derived probability histogram* equal to the *known derived distribution probability histogram* for each of the *transforms* of the *fud*, $\forall T \in F_{\mathrm{o}}\ (\hat{A}_H * T_{F_{\mathrm{o}}} = \hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}})$. Define the *iso-fud historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$,

$$
\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}} := \\
&\quad \big(\bigcup \{\{(H,1) : H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H, \\
&\qquad\qquad\qquad \forall T \in F_{\mathrm{o}}\ (\hat{A}_H * T_{F_{\mathrm{o}}} = \hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}})\}^{\wedge} : \\
&\qquad\qquad\qquad\qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\big)^{\wedge} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ \exists T \in F_{\mathrm{o}}\ (\hat{A}_H * T_{F_{\mathrm{o}}} \neq \hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}})\} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \not\subseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset,0)\}
\end{aligned}
$$

In *classical functional definition set induction* the *history probability function* is *iso-fud historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}$.

If the *fud* is a singleton, $F_{\mathrm{o}} = \{T_{\mathrm{o}}\}$, *classical functional definition set induction* reduces to *classical transform induction*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},\{T_{\mathrm{o}}\}} = P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}$.

The *iso-fud historical probability* may be expressed in terms of a *histogram distribution*,

$$
\hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) \ \ \propto \ \ \sum(P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)
$$

where the *iso-fud conditional stuffed historical probability distribution* is defined

$$
\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{d},F,U}(E,z) \\
&\quad := \ \ \{(A, \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},F,z}(\{A*T_F : T \in F\})} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\
&\qquad\quad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \not\leq E\}
\end{aligned}
$$

which is defined if $z \leq \mathrm{size}(E)$. The *derived histogram set* valued *integral histogram* function $D_{U,\mathrm{i},F,z}$ is defined

$$
D_{U,\mathrm{i},F,z} = \{(A, \{A * T_F : T \in F\}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}
$$

The finite set of *iso-fuds* of *derived histogram set* $\{A * T_F : T \in F\}$ is

$$
D^{-1}_{U,\mathrm{i},F,z}(\{A * T_F : T \in F\}) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ \forall T \in F\ (B * T_F = A * T_F)\}
$$

In this case the *top transform* exists, $\exists T \in F~(\mathrm{der}(T) = \mathrm{der}(F))$, so the set of *iso-fuds* is a *law-like* subset of the *iso-deriveds*,

$$D_{U,F,z}^{-1}(\{A * T_F : T \in F\}) \subseteq D_{U,F^{\mathrm{T}},z}^{-1}(A * F^{\mathrm{T}})$$

and therefore *necessary derived fud* is stricter than *necessary derived*. That is, a *history* can only be *drawn* in *classical functional definition set induction* if it can be *drawn* in *classical transform induction* for the *transform* of the *fud*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}(H) > 0 \implies P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}^{\mathrm{T}}}(H) > 0$.

The *iso-fud conditional generalised multinomial probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\mathrm{m,d},F,U}(E,z) \\
&\quad := \quad \{(A, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in D_{U,\mathrm{i},F,z}^{-1}(\{A * T_F : T \in F\})} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},~A^{\mathrm{F}} \leq E^{\mathrm{F}}\}^{\wedge} \cup \\
&\qquad\quad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},~A^{\mathrm{F}} \not\leq E^{\mathrm{F}}\}
\end{aligned}$$

which is defined if $\mathrm{size}(E) > 0$.

It is assumed that the *distribution history size*, z_{h}, is large with respect to the *sample size* $z_{\mathrm{o}} = \mathrm{size}(A_{\mathrm{o}})$, so that, in the limit, the *iso-fud historical probability*, $\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$, approximates to the *iso-fud multinomial probability*, $\hat{Q}_{\mathrm{m,d},F_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$. That is, if $z_{\mathrm{o}} \ll z_{\mathrm{h}}$ then

$$\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(E_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}}) \approx \hat{Q}_{\mathrm{m,d},F_{\mathrm{o}},U}(E_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$$

where $E_{\mathrm{o}} = E_{\mathrm{h}}\%V_{\mathrm{o}}$.

In the case of *completely effective sample histogram*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the maximisation for *known fud*, F_{o}, of the *iso-fud conditional generalised multinomial probability* parameterised by the *complete congruent histograms* of unit *size* is a singleton of the *rational maximum likelihood estimate*

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m,d},F_{\mathrm{o}},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

In the case where the maximisation of the *iso-fud conditional generalised multinomial probability distribution* is a singleton, it is equal to the *normalised fud-dependent*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{D_F}(F_{\mathrm{o}})}$, where the *fud-dependent* $A^{\mathrm{D_F}(F)} \in \mathcal{A}_{U,V,z}$ is defined in 'Likely histograms', above, as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-fud*,

$$\{A^{\mathrm{D_F}(F)}\} = \\ \mathrm{maxd}(\{(E, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum Q_{\mathrm{m},U}(E,z)(B) : B \in D_{U,\mathrm{i},F,z}^{-1}(D_{U,F,z}(A))}) : E \in \mathcal{A}_{U,V,z}\})$$

The *independent analogue* is the *fud-independent*, $A^{\mathrm{E_F}(F)} \in \mathcal{A}_{U,V,z}$, defined,

$$\{A^{\mathrm{E_F}(F)}\} = \\ \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D_{U,\mathrm{i},F,z}^{-1}(D_{U,F,z}(A)))) : E \in \mathcal{A}_{U,V,z}\})$$

The *fud-independent* approximates to the arithmetic *average* of the *naturalisations*,

$$A^{\mathrm{E_F}(F)} \approx Z_{1/|F|} * \sum_{T \in F} A * T_F * T_F^{\dagger}$$

It is only in the case where the *histogram* equals the *fud-independent* that the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})} \implies A_{\mathrm{o}}^{\mathrm{D_F}(F_{\mathrm{o}})} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

Otherwise, the overall *maximum likelihood estimate*, which is the *fud-dependent*, is near the *histogram*, $\tilde{E}_{\mathrm{o}} \sim \hat{A}_{\mathrm{o}}$, only in as much as it is far the *fud-independent*, $\tilde{E}_{\mathrm{o}} \nsim \hat{A}_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$.

In *classical functional definition set induction*, where (i) the *history probability function* is *iso-fud historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}$, given some *substrate fud* in the *sample variables* $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, such that there exists a *top transform*, $\exists T \in F_{\mathrm{o}}\ (\mathrm{der}(T) = \mathrm{der}(F_{\mathrm{o}}))$, if it is the case that (ii) the *sample histogram* equals the *fud-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-fud conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is

$$\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

Given the *known substrate fud*, F_{o}, consider the *log likelihood* of the *iso-fud conditional generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m,d},F_{\mathrm{o}},U}$, at the *maximum likelihood estimate*, in the special case where *sample histogram* equals the *fud-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{D_F}(F_{\mathrm{o}})} = \hat{A}_{\mathrm{o}}$.

The set of *iso-fuds* is the intersection of the *iso-deriveds* of each *transform*

$$D^{-1}_{U,F,z}(\{A * T_F : T \in F\}) \quad = \quad \bigcap_{T \in F} D^{-1}_{U,T_F,z}(A * T_F)$$

So the cardinality of the set of *integral iso-fuds* is less than or equal to the product of the weak compositions of the *components* for any *transform*,

$$\forall T \in F \left(|D^{-1}_{U,\mathrm{i},F,z}(D_{U,F,z}(A))| \leq \prod_{(R,C) \in T_F^{-1}} \frac{((A * T_F)_R + |C| - 1)!}{(A * T_F)_R!\ (|C| - 1)!} \right)$$

and the logarithm of the *integral iso-fuds* cardinality is less or equal to the *integral iso-deriveds log-cardinality* for any *transform*,

$$\forall T \in F \ \left(\ln |D^{-1}_{U,\mathrm{i},F,z}(D_{U,F,z}(A))| \leq \ln |D^{-1}_{U,\mathrm{i},T_F,z}(A * T_F)|\right)$$

In the case where the *volume* is much greater than one, $v \gg 1$, the *integral iso-deriveds log-cardinality* approximates to the negative *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \ &\approx \\ &-\big((z + v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\ &\qquad -z \times \mathrm{entropy}(A * T) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)\big)\end{aligned}$$

In the domain where the *size* is greater than the *volume*, $z > v$, the *integral iso-deriveds log-cardinality* varies against the *volume* scaled *component cardinality size relative entropy*,

$$\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \quad \sim \quad -v \times \mathrm{entropyRelative}(V^{\mathrm{C}} * T, A * T)$$

In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *integral iso-deriveds log-cardinality* varies against the *size* scaled *component size cardinality relative entropy*,

$$\ln |D^{-1}_{U,\mathrm{i},T,z}(A * T)| \quad \sim \quad -z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)$$

The *log likelihood* varies against the *log iso-fud cardinality*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{m,d},F,U}(A, z)(A) \quad &\propto \quad \ln \frac{Q_{\mathrm{m},U}(A, z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},F,z}(D_{U,F,z}(A))} Q_{\mathrm{m},U}(A, z)(B)} \\ &\sim \quad -\ln |D^{-1}_{U,\mathrm{i},F,z}(D_{U,F,z}(A))|\end{aligned}$$

and so varies against the *integral iso-deriveds log-cardinalities* for all of the *transforms*

$$\forall T \in F \ \left(\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \sim -\ \ln |D^{-1}_{U,\mathrm{i},T_F,z}(A * T_F)|\right)$$

So the *log likelihood* varies with the sum of the *size-volume* scaled *component size cardinality sum relative entropies*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \ \sim \\ \sum_{T \in F} \big((z+v) \times \mathrm{entropy}(A * T_F + V^{\mathrm{C}} * T_F) \\ -z \times \mathrm{entropy}(A * T_F)\ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T_F)\big)\end{aligned}$$

In the domain where the *size* is greater than the *volume*, $z > v$, the *log likelihood* varies with the sum of the *volume* scaled *substrate component cardinality size relative entropies*,

$$\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \ \sim \ v \times \sum_{T \in F} \mathrm{entropyRelative}(V^{\mathrm{C}} * T_F, A * T_F)$$

and, in the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *log likelihood* varies with the sum of the *size* scaled *substrate component size cardinality relative entropies*,

$$\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \ \sim \ z \times \sum_{T \in F} \mathrm{entropyRelative}(A * T_F, V^{\mathrm{C}} * T_F)$$

In *classical functional definition set induction*, where (i) the *history probability function* is *iso-fud historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}$, given some *substrate fud* in the *sample variables* $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, such that there exists a *top transform*, $\exists T \in F_{\mathrm{o}}\ (\mathrm{der}(T) = \mathrm{der}(F_{\mathrm{o}}))$, if it is the case that (ii) the *sample histogram* equals the *fud-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-fud conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the sum of *size-volume* scaled *component size cardinality sum relative entropies* of all *transforms*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ &\sim \\ &\sum_{T \in F_{\mathrm{o}}} \big((z_{\mathrm{o}} + v_{\mathrm{o}}) \times \mathrm{entropy}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}} + V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}}) \\ &\qquad -z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}) \ - v_{\mathrm{o}} \times \mathrm{entropy}(V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}})\big)\end{aligned}$$

In the case where the *size* is greater than the *volume*, $z_{\mathrm{o}} > v_{\mathrm{o}}$, the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the sum of *volume* scaled *component cardinality size relative entropies* of all *transforms*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ v_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropyRelative}(V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}}, A_{\mathrm{o}} * T_{F_{\mathrm{o}}})$$

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample histogram* approximates to the *fud-independent histogram*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, or $\mathrm{spaceRelative}(A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})})(A_{\mathrm{o}}) \approx 0$, the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with sum of the *size* scaled *component size cardinality relative entropies* for all *transforms*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ z_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}})$$

In other words, the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the sum of the *size* scaled *component size cardinality cross entropies* of all *transforms*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ z_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropyCross}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}})$$

and against the sum of the *size* scaled *derived entropies* for all *transforms*

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ -z_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropy}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}})$$

So, in this case, the *log likelihood* is maximised when (a) the sum of the *derived entropies* of all *transforms* is minimised, and (b) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for all *transforms*.

It is shown in section 'Necessary derived', above, that in the case where (i) the *sample* is *natural*, $A = A * T * T^{\dagger}$, and (ii) the *component size cardinality cross entropy* is greater than the logarithm of the *possible derived volume*, $\mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) > \ln w'$, so that the *relative entropy* is high, then the logarithm of the *iso-derived conditional multinomial probability* varies against the logarithm of the *derived multinomial probability*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \quad \sim \quad - \ln \hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)$$

Extending the *model* from *transform*, T, to *functional definition set*, F, conjecture that in the case where (i) the *sample histogram* equals the *fud-independent*, $A = A^{\mathrm{E_F}(F)}$, and (ii) the *cross entropies* are sufficient, $\forall T \in F\ (\mathrm{entropyCross}(A * T_F, V^{\mathrm{C}} * T_F) > \ln |T_F^{-1}|)$, the logarithm of the *iso-fud conditional multinomial probability* varies against the sum of the logarithms of the *derived multinomial probabilities*,

$$\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \quad \sim \quad -\sum_{T \in F} \ln \hat{Q}_{\mathrm{m},U}(A * T_F, z)(A * T_F)$$

In *classical functional definition set induction*, where (i) the *history probability function* is *iso-fud historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}$, given some *substrate fud* in the *sample variables* $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, such that there exists a *top transform*, $\exists T \in F_{\mathrm{o}}\ (\mathrm{der}(T) = \mathrm{der}(F_{\mathrm{o}}))$, if it is the case that (ii) the *sample histogram* equals the *fud-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-fud conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iii) the *component size cardinality relative entropies* are high, $\forall T \in F_{\mathrm{o}}\ (\mathrm{entropyCross}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}}) > \ln |T_{F_{\mathrm{o}}}^{-1}|)$, (iv) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (v) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the sum of logarithms of the *derived multinomial probabilities* of the *transforms*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\sum_{T \in F_{\mathrm{o}}} \ln \hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, z_{\mathrm{o}})(A_{\mathrm{o}} * T_{F_{\mathrm{o}}})$$

In section 'Derived history space', above, the *specialising fud substrate history coder* is constructed

$$C_{\mathrm{G},V,\mathrm{F,H}}(F) = \mathrm{coderHistorySubstrateFudSpecialising}(U, X, F, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,V,X})$$

In the *law-like* case where the *fud* has a *top transform*, $\exists T \in F\ (W_T = \mathrm{der}(F))$, the *space* is

$$\begin{aligned}
&\mathrm{space}(C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F))(H) = \\
&\qquad\qquad \mathrm{spaceIds}(|X|,|H|) + \\
&\qquad\qquad \mathrm{spaceCountsDerived}(U)(A,F^{\mathrm{T}}) + \\
&\qquad\qquad \mathrm{spaceClassification}(A * F^{\mathrm{T}}) + \\
&\qquad\qquad \sum_{T \in F} \mathrm{spaceEventsPartition}(A * \mathrm{depends}(F,V_T)^{\mathrm{T}},T)
\end{aligned}$$

Let w' be the *possible derived volume* of the *transform* of the *fud*, $w' = |(F^{\mathrm{T}})^{-1}|$. The *space* of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)$, varies (i) with the *possible fud derived volume*, w', where the *possible fud derived volume* is less than the *size*, $w' < z$, otherwise with the *size* scaled *log possible fud derived volume*, $z \ln w'$, (ii) with the *size* scaled *transform fud derived entropy* and (iii) against the sum of the *size* scaled *component size cardinality cross entropies* of the *transforms* of the *fud*,

$$\begin{aligned}
C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) \ \sim& \\
&\quad (w' \ :\ w' < z)\ +\ (z \ln w' \ :\ w' \geq z) \\
&+ z \times \mathrm{entropy}(A * F^{\mathrm{T}}) \\
&- z \times \sum_{T \in F} \mathrm{entropyCross}(A * T_F, V_T^{\mathrm{C}} * T)
\end{aligned}$$

So the *space* of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)$, is minimised when (a) the *possible fud derived volume* is minimised, (b) the *derived entropy* or *component size entropy* of the *fud transform* is minimised, and (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for each of the *fud transforms*.

In the domain where the *size* is less than or equal to the *volume*, $z \leq v$, the *log likelihood* varies with the sum of the *size* scaled *substrate component size cardinality cross entropies*,

$$\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \ \ \sim\ \ z \times \sum_{T \in F} \mathrm{entropyCross}(A * T_F, V^{\mathrm{C}} * T_F)$$

so conjecture that the *log likelihood* varies with the sum of the *size* scaled *layer component size cardinality cross entropies*,

$$\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \ \ \sim\ \ z \times \sum_{T \in F} \mathrm{entropyCross}(A * T_F, V_T^{\mathrm{C}} * T)$$

In addition the *log likelihood* varies against *size* scaled *fud transform derived entropy*,

$$\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \quad \sim \quad -z \times \mathrm{entropy}(A * F^{\mathrm{F}})$$

because the *log likelihood* varies against all of the *transform derived entropies* including that of the *top transform*. Together, conjecture that in this domain the *log likelihood* varies against the *specialising space*.

Conjecture that in the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-independent*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising fud substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F,H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}}))(H_{\mathrm{o}})$$

where F^V is the *expansion* that adds a *unary transform* in the remaining *underlying variables*, $F \cup \{\{(V \setminus \mathrm{und}(F))^{\mathrm{CS}}\}^{\mathrm{T}}\}$. The *iso-fud conditional stuffed historical probability distribution log likelihood* is maximised and the *specialising fud substrate history coder space* is minimised by varying the *fud* such that (i) the *fud transform derived entropy* is low, and (ii) high *counts* are in low cardinality *components* and high cardinality *components* have low *counts* for all *transforms*.

In section 'Derived history space', above, the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{F,H}}(F)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, is shown to be characterised by certain properties. The *specialising-canonical space difference* varies (i) with twice the total *possible derived volume* of the *transforms*, where the *possible derived volumes* are less than the *size*, otherwise with twice the total *size* scaled *log possible derived volume*, (ii) with the sum of the *size* scaled *derived entropies*, (iii) against twice the sum of the *size* scaled *component size cardinality cross entropies* and (iv) against the sum of the *size* scaled *size expected component entropies*,

$$\begin{aligned}
2C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) \ \sim \\
\sum_{T \in F} 2\big((w'_T \ : \ w'_T < z) \ + \ (z \ln w'_T \ : \ w'_T \geq z)\big) \\
+ \sum_{T \in F} z \times \mathrm{entropy}(A * T_F) \\
- \sum_{T \in F} 2z \times \mathrm{entropyCross}(A * T_F, V_T^{\mathrm{C}} * T) \\
- \sum_{T \in F} z \times \mathrm{entropyComponent}(A * \mathrm{dep}(F, V_T)^{\mathrm{T}}, T)
\end{aligned}$$

where $w'_T = |T^{-1}|$ and $T_F = \mathrm{dep}(F, W_T)^{\mathrm{T}}$. So the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, is minimised when (a) the total *possible derived volume* is minimised, (b) the total *derived entropy* is minimised, (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for each *transform*, and (d) the total *expected component entropy* is maximised. It was also conjectured that when the *specialising-canonical space difference* is minimised, (i) the *derived entropy* decreases up the *layers*, (ii) the *possible derived volume* decreases up the *layers*, (iii) the *expected component entropy* increases up the *layers*, and (iv) the *component size cardinality cross entropy* increases up the *layers*. The *canonical* terms, $C^{\mathrm{s}}_{\mathrm{H},V}(H)$ and $C^{\mathrm{s}}_{\mathrm{G},V}(H)$, are independent of the *model*, so these properties are also the properties of the *specialising derived substrate history coder space*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H)$.

So conjecture that in *classical functional definition set induction* where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-independent*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E}_{\mathrm{F}}(F_{\mathrm{o}})}$, the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising-canonical space difference*,

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \\
-(2C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F},\mathrm{H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{H},V_{\mathrm{o}}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{G},V_{\mathrm{o}}}(H_{\mathrm{o}}))
\end{aligned}$$

In the special case where (i) the *sample histogram* equals the *fud-independent*, $A = A^{\mathrm{E}_{\mathrm{F}}(F)} \implies \tilde{E} = \hat{A}^{\mathrm{D}_{\mathrm{F}}(F)} = \hat{A}$, and (ii) the *cross entropies* are sufficient, $\forall T \in F\ (\mathrm{entropyCross}(A * T_F, V^{\mathrm{C}} * T_F) > \ln |T_F^{-1}|)$, so that the *relative entropies* are high, conjecture that the *iso-fud conditional multinomial probability* at the *maximum likelihood estimate* varies with the product of the *underlying-derived relative multinomial probabilities*,

$$\frac{\hat{Q}_{\mathrm{m},U}(A,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},F,z}(\{A*T_F : T \in F\})} \hat{Q}_{\mathrm{m},U}(A,z)(B)} \sim \frac{\hat{Q}_{\mathrm{m},U}(A,z)(A)}{\prod_{T \in F} \hat{Q}_{\mathrm{m},U}(A*T_F,z)(A*T_F)}$$

Following the reasoning in *classical transform induction*, above, the *sum sensitivity* of the *iso-fud conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},T,U}$, is conjectured to vary with the *unknown-known multinomial probability distribution sum sensitivity difference*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F,U}(E,z))) \sim \\ &\quad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E,z))) - \sum_{T \in F} \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(E*T_F,z)))\end{aligned}$$

The *iso-fud conditional stuffed historical probability distribution sum sensitivity* varies against the *underlying entropy*

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F,U}(E,z))) \quad \sim \quad -z \times \mathrm{entropy}(E)$$

The *iso-fud conditional stuffed historical probability distribution sum sensitivity* varies with the sum of the *derived entropies*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F,U}(E,z))) \quad \sim \quad \sum_{T \in F} z \times \mathrm{entropy}(E*T_F)$$

The *iso-fud conditional stuffed historical probability distribution sum sensitivity* varies against the sum of the *size* scaled *expected component entropies*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F,U}(E,z))) \sim \\ &\quad -\sum_{T \in F} z \times \mathrm{entropyComponent}(E*\mathrm{dep}(F,V_T)^{\mathrm{T}},T)\end{aligned}$$

where dep = depends. In *classical functional definition set induction*, where (i) the *history probability function* is *iso-fud historically distributed*, $P = P_{U,X,H_\mathrm{h},\mathrm{d},F_\mathrm{o}}$, given some *substrate fud* in the *sample variables* $F_\mathrm{o} \in \mathcal{F}_{U,V_\mathrm{o}}$, such that there exists a *top transform*, $\exists T \in F_\mathrm{o}\ (\mathrm{der}(T) = \mathrm{der}(F_\mathrm{o}))$, if it is the case that (ii) the *sample histogram* equals the *fud-independent*, $A_\mathrm{o} = A_\mathrm{o}^{\mathrm{E_F}(F_\mathrm{o})}$, then the *maximum likelihood estimate*, $\tilde{E}_\mathrm{o}$, then the *maximum likelihood estimate*, $\tilde{E}_\mathrm{o}$, of the *unknown distribution probability histogram*, $\hat{E}_\mathrm{o}$, in the *iso-fud conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},F_\mathrm{o},U}(E_\mathrm{o},z_\mathrm{o})$, is $\tilde{E}_\mathrm{o} = \hat{A}_\mathrm{o}$, so, if it is also the case that (iii) the *component size cardinality relative entropies* are high, $\forall T \in F_\mathrm{o}\ (\mathrm{entropyCross}(A_\mathrm{o}*T_{F_\mathrm{o}}, V_\mathrm{o}^\mathrm{C}*T_{F_\mathrm{o}}) > \ln|T_{F_\mathrm{o}}^{-1}|)$,

(iv) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (v) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *sum sensitivity* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\end{aligned}$$

(b) varies with the total *derived entropy*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim \sum_{T \in F_{\mathrm{o}}} z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}})$$

and (c) varies against the total *size* scaled *expected component entropy*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim\\ &\qquad -\sum_{T \in F_{\mathrm{o}}} z_{\mathrm{o}} \times \mathrm{entropyComponent}(A_{\mathrm{o}} * \mathrm{dep}(F_{\mathrm{o}}, V_T)^{\mathrm{T}}, T)\end{aligned}$$

Conjecture that in the *natural* case the *sum sensitivity* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *specialising fud substrate history coder space*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F},\mathrm{H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})$$

Both the *iso-fud conditional stuffed historical probability distribution sum sensitivity* and the *specialising fud substrate history coder space* are minimised by varying the *fud* such that the *derived entropy* is low.

Also, conjecture that the *sum sensitivity* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *specialising-canonical space difference*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim\\ &\qquad\qquad 2C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F},\mathrm{H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{H},V_{\mathrm{o}}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{G},V_{\mathrm{o}}}(H_{\mathrm{o}})\end{aligned}$$

Both the *iso-fud conditional stuffed historical probability distribution sum sensitivity* and the *specialising-canonical space difference* are minimised by varying the *fud* such that (a) the *derived entropy* is low and (b) the *underlying components* have high *entropy*.

Altogether, in *classical functional definition set induction* where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-independent*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, and the *relative entropies* are high, the *sum sensitivity* has similar properties as the *log-likelihood* but with the correlations reversed. Conjecture that in this case the *sum sensitivity* varies against the *log-likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad -\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

That is, in the *natural* case, the maximisation of the *log-likelihood* also tends to minimise the *sum sensitivity* to the *maximum likelihood estimate*.

The *sum sensitivity* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))$$

can be related to queries on the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, in the special case where (i) the *sample histogram* equals the *fud-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, and (ii) the *component size cardinality relative entropies* are high, $\forall T \in F_{\mathrm{o}}\ (\mathrm{entropyCross}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}}) > \ln |T_{F_{\mathrm{o}}}^{-1}|)$. In the case of *classical functional definition set induction*, the given *substrate fud* must be such that its set of *underlying variables* is a subset of the query *variables*, $\mathrm{und}(F_{\mathrm{o}}) \subseteq K$. In the case where the query *histogram* consists of one *effective state*, $Q = \{(S_Q, 1)\}$, there exists an *effective derived state* for each of the *transforms*, $\{R_Q : T \in F_{\mathrm{o}},\ \{R_Q\} = (Q * T_{F_{\mathrm{o}}})^{\mathrm{FS}}\}$. The corresponding *underlying component* is the intersection $C_Q = \bigcap\{T_{F_{\mathrm{o}}}^{-1}(R_Q) : T \in F_{\mathrm{o}},\ \{R_Q\} = (Q * T_{F_{\mathrm{o}}})^{\mathrm{FS}}\}$. This *component* is a subset of that for *transform induction*, $C_Q \subseteq (F_{\mathrm{o}}^{\mathrm{T}})^{-1}(R_Q)$ where $\{R_Q\} = (Q * F_{\mathrm{o}}^{\mathrm{T}})^{\mathrm{FS}}$. In this case the application of the query via the *model* equals the application via the *component* directly, $\bigcap\{Q * T_{F_{\mathrm{o}}} * \mathrm{his}(T_{F_{\mathrm{o}}}) * A_{\mathrm{o}}\ \%\ V_{\mathrm{o}} : T \in F_{\mathrm{o}}\}^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K) = (A_{\mathrm{o}} * C_Q)^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$. If any *possible derived volume* is non singular, $|T_{F_{\mathrm{o}}}^{-1}| > 1$ where $T \in F_{\mathrm{o}}$, the query *histogram* itself cannot be *drawn* from the *distribution history*, $\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o}}, 1)(Q * \{N\}^{\mathrm{U}}) = 0$, where $N \in (V_{\mathrm{o}} \setminus K)^{\mathrm{CS}}$, because at least one query *derived probability histogram* is not equal to the corresponding *known derived distribution probability histogram*, $\exists T \in F_{\mathrm{o}}\ (\hat{Q} * \{N\}^{\mathrm{U}} * T_{F_{\mathrm{o}}} \neq \hat{A}_{\mathrm{o}} * T_{F_{\mathrm{o}}})$. The application of the query must be in terms of a modified *sample histogram*,

$$\begin{aligned}
&\bigcap\{Q * T_{F_{\mathrm{o}}} * \mathrm{his}(T_{F_{\mathrm{o}}}) * A_{\mathrm{o}}\ \%\ V_{\mathrm{o}} : T \in F_{\mathrm{o}}\}^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K) = \\
&\quad \{(N, (\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}))^{1/z_{\mathrm{o}}}) : N \in (V_{\mathrm{o}} \setminus K)^{\mathrm{CS}}, \\
&\qquad A_{Q,N} = A_{\mathrm{o}} - (A_{\mathrm{o}} * C_Q) + ((A_{\mathrm{o}} * C_Q)\ \%\ K * \{N\}^{\mathrm{U}})\}^{\wedge}
\end{aligned}$$

where his = histogram. If the *sample histogram* is *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the modified *sample histogram*, $A_{Q,N}$, can be *drawn* from the *distribution*, $\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}) > 0$, because the modified *sample histogram* is an *iso-fud*, $\forall T \in F_{\mathrm{o}}\ (A_{Q,N} * T_{F_{\mathrm{o}}} = A_{\mathrm{o}} * T_{F_{\mathrm{o}}})$. The *model* application depends on the geometric scaling of the *historical distribution*, $\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$, so the *query sensitivity* to the *distribution histogram* varies with the *sum sensitivity* of the *historical distribution* at the *maximum likelihood estimate* divided by the *sample size*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}$$

Although the query application via the *model* is sometimes not equal to the *estimated transformed conditional product*, the *query sensitivity* to the *distribution histogram* is sometimes lower,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}\\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}\end{aligned}$$

Similarly, where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the *model likelihood* of the *distribution histogram* is sometimes higher,

$$\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{Q,N}) \quad \geq \quad \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{Q,N})$$

In other words, querying via the *known derived* of the *model* sometimes reduces the *sensitivity* to the *unknown* and increases the *likelihood* at the cost of modifying the query.

5.5.3 Necessary derived functional definition set decomposition

The last section extended the *model* from *transforms* to *functional definition sets*. Now extend further to *functional definition set decompositions*. This discussion is very similar to that of the previous section, except that now the *fuds* are *contingent* on the *slice*.

Given some non-empty *known substrate fud decomposition*, $D_{\mathrm{o}} \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\}$, such that there exists a *top transform* for all of the *fuds*, $\forall F \in \mathrm{fuds}(D_{\mathrm{o}})\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$, the *component derived set* of the *distribution probability histogram* is $\{(C, \{\hat{E}_{\mathrm{h}} * C * T_F : T \in F\}) : (C, F) \in \mathrm{cont}(D_{\mathrm{o}})\}$, where $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$ and $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$. In *classical functional definition set decomposition induction*, while the *distribution probability histogram*, $\hat{E}_{\mathrm{h}}$, remains *unknown*, the *component derived distribution probability set*, $\{(C, \{\hat{E}_{\mathrm{h}} * C * T_F : T \in F\}) : (C, F) \in$

$\mathrm{cont}(D_{\mathrm{o}})\}$, is *known* and *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have a *derived probability histogram* equal to the *known derived distribution probability histogram* for each of the *transforms* of the *fud* for each *slice*, $\forall (C,F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (\hat{A}_H * C * T_F = \hat{E}_{\mathrm{h}} * C * T_F)$. Define the *iso-fud-decomposition historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$,

$$
\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}} := \\
&\quad \Big(\bigcup \{\{(H,1) : H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H, \\
&\qquad\qquad \forall (C,F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (\hat{A}_H * C * T_F = \hat{E}_{\mathrm{h}} * C * T_F)\}^{\wedge} : \\
&\qquad\qquad\qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \dots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X}, \\
&\qquad\qquad \exists (C,F) \in \mathrm{cont}(D_{\mathrm{o}})\ \exists T \in F\ (\hat{A}_H * C * T_F \neq \hat{E}_{\mathrm{h}} * C * T_F)\} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \not\subseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset,0)\}
\end{aligned}
$$

In *classical functional definition set decomposition induction* the *history probability function* is *iso-fud-decomposition historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}}$.

If the *decomposition* only has a root node, $D_{\mathrm{o}} = \{((\emptyset,F_{\mathrm{o}}),\emptyset)\}$, *classical functional definition set decomposition induction* reduces to *classical functional definition set induction*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},\{((\emptyset,F_{\mathrm{o}}),\emptyset)\}} = P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}$. If the root *fud* is a singleton, $F_{\mathrm{o}} = \{T_{\mathrm{o}}\}$, *classical functional definition set decomposition induction* reduces to *classical derived induction*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},\{((\emptyset,\{T_{\mathrm{o}}\}),\emptyset)\}} = P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}$.

The *iso-fud-decomposition historical probability* may be expressed in terms of a *histogram distribution*,

$$\hat{Q}_{\mathrm{h},\mathrm{d},D_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) \quad \propto \quad \sum (P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *iso-fud-decomposition conditional stuffed historical probability distribution* is defined

$$
\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{d},D,U}(E,z) \\
&\quad := \ \{(A, \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A))} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\
&\qquad\quad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \not\leq E\}
\end{aligned}
$$

which is defined if $z \leq \mathrm{size}(E)$. The *component-derived-set* function valued function of the *substrate histograms* $D_{U,D,\mathrm{F},z} \in \mathcal{A}_{U,V,z} \to (\mathcal{A}_U \to \mathrm{P}(\mathcal{A}_U))$ is defined

$$D_{U,D,\mathrm{F},z} = \{(A, \{(C, \{A{*}C{*}T_F : T \in F\}) : (C,F) \in \mathrm{cont}(D)\}) : A \in \mathcal{A}_{U,V,z}\}$$

The finite set of *iso-fud-decompositions* of *component-derived-set* $D_{U,D,\mathrm{F},z}(A)$ is

$$\begin{aligned} D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A)) \ &= \\ &\{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ \forall (C,F) \in \mathrm{cont}(D)\ \forall T \in F\ (B * C * T_F = A * C * T_F)\} \end{aligned}$$

In this case the *top transform* exists for all *fuds*, $\forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$, so the set of *iso-fud-decompositions* is a *law-like* subset of the *iso-deriveds*,

$$D^{-1}_{U,D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A)) \ \subseteq \ D^{-1}_{U,D^{\mathrm{T}},z}(A * D^{\mathrm{T}})$$

and therefore *necessary derived fud decomposition* is stricter than *necessary derived*. That is, a *history* can only be *drawn* in *classical functional definition set decomposition induction* if it can be *drawn* in *classical transform induction* for the *transform* of the *fud decomposition*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}}(H) > 0 \implies P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}^{\mathrm{T}}}(H) > 0$.

The set of *iso-fud-decompositions* is also a *law-like* subset of the *iso-fuds* for the root node,

$$\begin{aligned} D^{-1}_{U,D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A)) \ &\subseteq \ D^{-1}_{U,F,z}(\{A * T_F : T \in F\}) \\ &\subseteq \ D^{-1}_{U,F^{\mathrm{T}},z}(A * F^{\mathrm{T}}) \end{aligned}$$

where $\{((\emptyset, F), \cdot)\} = D$.

The *iso-fud-decomposition conditional generalised multinomial probability distribution* is defined

$$\begin{aligned} &\hat{Q}_{\mathrm{m,d},D,U}(E,z) \\ &:= \ \{(A, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A))} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \leq E^{\mathrm{F}}\}^{\wedge} \cup \\ &\qquad \{(A, 0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \not\leq E^{\mathrm{F}}\} \end{aligned}$$

which is defined if $\mathrm{size}(E) > 0$.

It is assumed that the *distribution history size*, z_{h}, is large with respect to the *sample size* $z_{\mathrm{o}} = \mathrm{size}(A_{\mathrm{o}})$, so that, in the limit, the *iso-fud-decomposition* *historical probability*, $\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$, approximates to the *iso-fud-decomposition multinomial probability*, $\hat{Q}_{\mathrm{m,d},D_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$. That is, if $z_{\mathrm{o}} \ll z_{\mathrm{h}}$ then

$$\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \approx \hat{Q}_{\mathrm{m,d},D_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

where $E_{\mathrm{o}} = E_{\mathrm{h}}\%V_{\mathrm{o}}$.

In the case of *completely effective sample histogram*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the maximisation for *known fud decomposition*, D_{o}, of the *iso-fud-decomposition conditional generalised multinomial probability* parameterised by the *complete congruent histograms* of unit *size* is a singleton of the *rational maximum likelihood estimate*

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m,d},D_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

In the case where the maximisation of the *iso-fud-decomposition conditional generalised multinomial probability distribution* is a singleton, it is equal to the *normalised fud-decomposition-dependent*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{D_{D,F}}(D_{\mathrm{o}})}$, where the *fud-decomposition-dependent* $A^{\mathrm{D_{D,F}}(D)} \in \mathcal{A}_{U,V,z}$ is defined in 'Likely histograms', above, as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-fud-decomposition*,

$$\begin{aligned}&\{A^{\mathrm{D_{D,F}}(D)}\} = \\ &\quad \mathrm{maxd}(\{(E, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum Q_{\mathrm{m},U}(E,z)(B) : B \in D_{U,\mathrm{i},D,\mathrm{F},z}^{-1}(D_{U,D,\mathrm{F},z}(A))}) : E \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *independent analogue* is the *fud-decomposition-independent*, $A^{\mathrm{E_{D,F}}(D)} \in \mathcal{A}_{U,V,z}$, defined,

$$\begin{aligned}&\{A^{\mathrm{E_{D,F}}(D)}\} = \\ &\quad \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D_{U,\mathrm{i},D,\mathrm{F},z}^{-1}(D_{U,D,\mathrm{F},z}(A)))) : E \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *fud-decomposition-independent* approximates to the *scaled sum* of the *slice* arithmetic *average* of the *naturalisations*,

$$A^{\mathrm{E_{D,F}}(D)} \approx Z_z * \left(\sum_{(C,F)\in\mathrm{cont}(D)} \left(Z_{1/|F|} * \sum_{T \in F} A * C * T_F * T_F^{\dagger}\right)\right)^{\wedge}$$

It is only in the case where the *histogram* equals the *fud-decomposition-independent* that the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D_{\mathrm{o}})} \implies A_{\mathrm{o}}^{\mathrm{D_{D,F}}(D_{\mathrm{o}})} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

Otherwise, the overall *maximum likelihood estimate*, which is the *fud decomposition dependent*, is near the *histogram*, $\tilde{E}_{\rm o} \sim \hat{A}_{\rm o}$, only in as much as it is far the *fud-decomposition-independent*, $\tilde{E}_{\rm o} \nsim \hat{A}_{\rm o}^{\rm E_{D,F}(\mathit{D}_o)}$.

In *classical functional definition set decomposition induction*, where (i) the *history probability function* is *iso-fud-decomposition historically distributed*, $P = P_{U,X,H_{\rm h},{\rm d},D_{\rm o}}$, given some *substrate fud decomposition* in the *sample variables* $D_{\rm o} \in \mathcal{D}_{{\rm F},U,V_{\rm o}} \setminus \{\emptyset\}$, such that there exists a *top transform* for all of the *fuds*, $\forall F \in \mathrm{fuds}(D_{\rm o})\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$, if it is the case that (ii) the *sample histogram* equals the *fud-decomposition-independent*, $A_{\rm o} = A_{\rm o}^{\rm E_{D,F}(\mathit{D}_o)}$, then the *maximum likelihood estimate*, $\tilde{E}_{\rm o}$, of the *unknown distribution probability histogram*, $\hat{E}_{\rm o}$, in the *iso-fud-decomposition conditional stuffed historical probability distribution*, $\hat{Q}_{{\rm h},{\rm d},D_{\rm o},U}(E_{\rm o}, z_{\rm o})$, is

$$\tilde{E}_{\rm o} = \hat{A}_{\rm o}$$

Given the *known substrate fud decomposition*, $D_{\rm o}$, consider the *log likelihood* in the special case where *sample histogram* equals the *fud-decomposition-independent*, $A_{\rm o} = A_{\rm o}^{\rm E_{D,F}(\mathit{D}_o)} \implies \tilde{E}_{\rm o} = \hat{A}_{\rm o}^{\rm D_{D,F}(\mathit{D}_o)} = \hat{A}_{\rm o}$.

In *classical functional definition set decomposition induction*, if it is the case that (ii) the *sample histogram* equals the *fud-decomposition-independent*, $A_{\rm o} = A_{\rm o}^{\rm E_{D,F}(\mathit{D}_o)}$, then the *maximum likelihood estimate*, $\tilde{E}_{\rm o}$, of the *unknown distribution probability histogram*, $\hat{E}_{\rm o}$, in the *iso-fud-decomposition conditional stuffed historical probability distribution*, $\hat{Q}_{{\rm h},{\rm d},D_{\rm o},U}(E_{\rm o}, z_{\rm o})$, is $\tilde{E}_{\rm o} = \hat{A}_{\rm o}$, so, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\rm h} \gg z_{\rm o}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{{\rm o},z_{\rm h}} \in \mathcal{A}_{\rm i}$, then the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the sum of *size-volume* scaled *component size cardinality sum relative entropies* of all *transforms* for all *slices*,

$$\begin{aligned}
&\ln \hat{Q}_{{\rm h},{\rm d},D_{\rm o},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o}) \ \sim \\
&\qquad \sum_{(C,F) \in \mathrm{cont}(D_{\rm o})} \sum_{T \in F} \big((z_{A_{\rm o}*C} + |C|) \times \mathrm{entropy}(A_{\rm o} * C * T_F + C * T_F) \\
&\qquad\qquad - z_{A_{\rm o}*C} \times \mathrm{entropy}(A_{\rm o} * C * T_F)\ - |C| \times \mathrm{entropy}(C * T_F)\big)
\end{aligned}$$

In the case where the *size* is greater than the *volume*, $z_{\rm o} > v_{\rm o}$, the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the sum of *volume* scaled *component cardinality size relative entropies* of all *transforms* for all *slices*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim \\ &\qquad \sum_{(C,F) \in \mathrm{cont}(D_{\mathrm{o}})} \big(|C| \times \sum_{T \in F} \mathrm{entropyRelative}(C * T_F, A_{\mathrm{o}} * C * T_F)\big)\end{aligned}$$

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample histogram* approximates to the *fud-decomposition-independent histogram*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D_{\mathrm{o}})}$, or $\mathrm{spaceRelative}(A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D_{\mathrm{o}})})(A_{\mathrm{o}}) \approx 0$, the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with sum of the *size* scaled *component size cardinality relative entropies* of all *transforms* for all *slices*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim \\ &\qquad \sum_{(C,F) \in \mathrm{cont}(D_{\mathrm{o}})} \big(z_{A_{\mathrm{o}} * C} \times \sum_{T \in F} \mathrm{entropyRelative}(A_{\mathrm{o}} * C * T_F, C * T_F)\big)\end{aligned}$$

So, in this case, the *log likelihood* is maximised when (a) the sum of the *derived entropies* of all *transforms* for all *slices* is minimised, and (b) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for all *transforms* for all *slices*.

If it is also the case that the *component size cardinality relative entropies* are high, $\forall (C,F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (\mathrm{entropyCross}(A_{\mathrm{o}} * C * T_F, C * T_F) > \ln |T_F^{-1}|)$, then the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the sum of logarithms of the *derived multinomial probabilities* of the *transforms* for all *slices*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim \\ &\qquad - \sum_{(C,F) \in \mathrm{cont}(D_{\mathrm{o}})} \sum_{T \in F} \ln \hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}} * C * T_F, z_{A_{\mathrm{o}} * C})(A_{\mathrm{o}} * C * T_F)\end{aligned}$$

In section 'Derived history space', above, the *specialising fud decomposition substrate history coder* $C_{\mathrm{G},V,\mathrm{D,F,H}}(F) \in \mathrm{coders}(\mathcal{H}_{U,V,X})$ is constructed

$$\begin{aligned}&C_{\mathrm{G},V,\mathrm{D,F,H}}(F) = \\ &\qquad \mathrm{coderHistorySubstrateFudDecompSpecialising}(U, X, F, D_{\mathrm{S}}, D_{\mathrm{X}})\end{aligned}$$

Conjecture that, in the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-decomposition-independent*, $A_{\mathrm{o}} \approx$

$A_{\rm o}^{\rm E_{D,F}(\mathit{D}_o)}$, the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising fud decomposition substrate history coder space*,

$$\ln \hat{Q}_{\rm h,d,\mathit{D}_o,\mathit{U}}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o}) \quad \sim \quad -\ \mathrm{space}(C_{{\rm G},V_{\rm o},{\rm D,F,H}}(D_{\rm o}^{V_{\rm o}}))(H_{\rm o})$$

where D^V is the *expansion* that adds a *unary transform* in the remaining *underlying variables* to the leaf *fuds* in the *decomposition* tree such that the *fud* of each path of the *application* tree has complete coverage of the *substrate*,

$$\forall L \in \mathrm{paths}(D^*)\ (\bigcup_{(\cdot,(F,\cdot)) \in L} V_F = V)$$

where $V_F = \mathrm{und}(F)$.

Conjecture that, in this case, the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising-canonical space difference*,

$$\ln \hat{Q}_{\rm h,d,\mathit{D}_o,\mathit{U}}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o}) \ \sim \\ -(2C_{{\rm G},V_{\rm o},{\rm D,F,H}}(D_{\rm o}^{V_{\rm o}})^{\rm s}(H_{\rm o}) - C^{\rm s}_{{\rm H},V_{\rm o}}(H_{\rm o}) - C^{\rm s}_{{\rm G},V_{\rm o}}(H_{\rm o}))$$

In *classical functional definition set decomposition induction*, where (i) the *history probability function* is *iso-fud-decomposition historically distributed*, $P = P_{U,X,H_{\rm h},{\rm d},D_{\rm o}}$, given some *substrate fud decomposition* in the *sample variables* $D_{\rm o} \in \mathcal{D}_{{\rm F},U,V_{\rm o}} \setminus \{\emptyset\}$, such that there exists a *top transform* for all of the *fuds*, $\forall F \in \mathrm{fuds}(D_{\rm o})\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$, if it is the case that (ii) the *sample histogram* equals the *fud-decomposition-independent*, $A_{\rm o} = A_{\rm o}^{\rm E_{D,F}(\mathit{D}_o)}$, then the *maximum likelihood estimate*, $\tilde{E}_{\rm o}$, then the *maximum likelihood estimate*, $\tilde{E}_{\rm o}$, of the *unknown distribution probability histogram*, $\hat{E}_{\rm o}$, in the *iso-fud-decomposition conditional stuffed historical probability distribution*, $\hat{Q}_{\rm h,d,\mathit{D}_o,\mathit{U}}(E_{\rm o}, z_{\rm o})$, is $\tilde{E}_{\rm o} = \hat{A}_{\rm o}$, so, if it is also the case that (iii) the *component size cardinality relative entropies* are high, $\forall (C,F) \in \mathrm{cont}(D_{\rm o})\ \forall T \in F\ (\mathrm{entropyCross}(A_{\rm o} * C * T_F, C * T_F) > \ln |T_F^{-1}|)$, (iv) the *distribution history size* is large with respect to the *sample size*, $z_{\rm h} \gg z_{\rm o}$, and such that (v) the *scaled probability sample histogram* is *integral*, $A_{{\rm o},z_{\rm h}} \in \mathcal{A}_{\rm i}$, then the *sum sensitivity* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\rm h,d,\mathit{D}_o,\mathit{U}}(A_{{\rm o},z_{\rm h}}, z_{\rm o}))) \\ \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm h},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})))$$

(b) varies with the total *derived entropy*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}}))) \sim \\ &\qquad\sum_{(C,F)\in\mathrm{cont}(D_{\mathrm{o}})} \big(z_{A_{\mathrm{o}}*C} \times \sum_{T\in F} \mathrm{entropy}(A_{\mathrm{o}} * C * T_F)\big)\end{aligned}$$

and (c) varies against the total *size* scaled *expected component entropy*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}}))) \sim \\ &\quad -\sum_{(C,F)\in\mathrm{cont}(D_{\mathrm{o}})} \big(z_{A_{\mathrm{o}}*C} \times \sum_{T\in F} \mathrm{entropyComponent}(A_{\mathrm{o}} * C * \mathrm{dep}(F,V_T)^{\mathrm{T}},T)\big)\end{aligned}$$

Conjecture that in the *natural* case the *sum sensitivity* of the *iso fud decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *specialising fud decomposition substrate history coder space*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}}))) \sim C_{\mathrm{G},V_{\mathrm{o}},\mathrm{D,F,H}}(D_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})$$

Also, conjecture that the *sum sensitivity* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *specialising-canonical space difference*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}}))) \sim \\ &\qquad 2C_{\mathrm{G},V_{\mathrm{o}},\mathrm{D,F,H}}(D_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{H},V_{\mathrm{o}}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{G},V_{\mathrm{o}}}(H_{\mathrm{o}})\end{aligned}$$

So conjecture that in the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-independent*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D_{\mathrm{o}})}$, the *sum sensitivity* varies against the *log-likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}}))) \sim -\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}})$$

That is, in the *natural* case, the maximisation of the *log-likelihood* also tends to minimise the *sum sensitivity* to the *maximum likelihood estimate.*

The *sum sensitivity* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})))$$

can be related to queries on the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, in the special case where (i) the *sample histogram* equals the *fud-decomposition-independent*, $A_o = A_o^{E_{D,F}(D_o)}$, and (ii) the *component size cardinality relative entropies* are high, $\forall (C, F) \in \text{cont}(D_o)\ \forall T \in F\ (\text{entropyCross}(A_o * C * T_F, C * T_F) > \ln |T_F^{-1}|)$. In the case of *classical functional definition set decomposition induction*, the given *substrate fud decomposition* must be such that its set of *underlying variables* is a subset of the query *variables*, $\text{und}(D_o) \subseteq K$. In the case where the query *histogram* consists of one *effective state*, $Q = \{(S_Q, 1)\}$, there exists an *effective derived state* for each of the *transforms* along one of the *slice* paths, $\{(C, R_Q) : (C, F) \in \text{cont}(D_o),\ Q * C \neq \emptyset,\ T \in F,\ \{R_Q\} = (Q * T_F)^{\text{FS}}\}$. The corresponding *underlying component* is the intersection

$$\begin{aligned} C_Q = \bigcap \{ & C * T_F^{-1}(R_Q) : \\ & (C, F) \in \text{cont}(D_o),\ Q * C \neq \emptyset,\ T \in F,\ \{R_Q\} = (Q * T_F)^{\text{FS}}\} \end{aligned}$$

This *component* is a subset of that for *transform induction*, $C_Q \subseteq (D_o^{\text{T}})^{-1}(R_Q)$ where $\{R_Q\} = (Q * D_o^{\text{T}})^{\text{FS}}$. In this case the application of the query via the *model* equals the application via the *component* directly, $\bigcap \{Q * T_F * \text{his}(T_F) * (A_o * C)\ \%\ V_o : (C, F) \in \text{cont}(D_o),\ Q * C \neq \emptyset,\ T \in F\}^\wedge\ \%\ (V_o \setminus K) = (A_o * C_Q)^\wedge\ \%\ (V_o \setminus K)$. If any *possible derived volume* is non singular, $|T_F^{-1}| > 1$, where $(C, F) \in \text{cont}(D_o)$, $Q * C \neq \emptyset$ and $T \in F$, the query *histogram* itself cannot be *drawn* from the *distribution history*, $\hat{Q}_{\text{h,d},D_o,U}(A_o, 1)(Q * \{N\}^{\text{U}}) = 0$, where $N \in (V_o \setminus K)^{\text{CS}}$, because at least one query *derived probability histogram* is not equal to the corresponding *known derived distribution probability histogram*, $\exists (C, F) \in \text{cont}(D_o)\ \exists T \in F\ (\hat{Q} * C * \{N\}^{\text{U}} * T_F \neq \hat{A}_o * C * T_F)$. The application of the query must be in terms of a modified *sample histogram*,

$$\begin{aligned} & \bigcap \{Q * T_F * \text{his}(T_F) * (A_o * C)\ \%\ V_o : \\ & \qquad (C, F) \in \text{cont}(D_o),\ Q * C \neq \emptyset,\ T \in F\}^\wedge\ \%\ (V_o \setminus K) \\ = \quad & \{(N, (\hat{Q}_{\text{h,d},D_o,U}(A_{o,z_h}, z_o)(A_{Q,N}))^{1/z_o}) : N \in (V_o \setminus K)^{\text{CS}}, \\ & \qquad A_{Q,N} = A_o - (A_o * C_Q) + ((A_o * C_Q)\ \%\ K * \{N\}^{\text{U}})\}^\wedge \end{aligned}$$

where his = histogram. If the *sample histogram* is *completely effective*, $A_o^{\text{F}} = V_o^{\text{C}}$, the modified *sample histogram*, $A_{Q,N}$, can be *drawn* from the *distribution*, $\hat{Q}_{\text{h,d},D_o,U}(A_{o,z_h}, z_o)(A_{Q,N}) > 0$, because the modified *sample histogram* is an *iso-fud-decomposition*, $\forall (C, F) \in \text{cont}(D_o)\ \forall T \in F\ (A_{Q,N} * C * T_F = A_o * C * T_F)$. The *model* application depends on the geometric scaling of the *historical distribution*, $\hat{Q}_{\text{h,d},D_o,U}(A_{o,z_h}, z_o)$, so the *query sensitivity* to the *distribution histogram* varies with the *sum sensitivity* of the *historical distribution* at the *maximum likelihood estimate* divided by the *sample size*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,d},D_o,U}(A_{o,z_h}, z_o)))/z_o$$

Although the query application via the *model* is sometimes not equal to the *estimated transformed conditional product*, the *query sensitivity* to the *distribution histogram* is sometimes lower,

$$\begin{aligned}\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,d},D_{\text{o}},U}(A_{\text{o},z_{\text{h}}}, z_{\text{o}})))/z_{\text{o}} \\ \leq \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h},U}(A_{\text{o},z_{\text{h}}}, z_{\text{o}})))/z_{\text{o}}\end{aligned}$$

Similarly, where the *size* is less than the *volume*, $z_{\text{o}} < v_{\text{o}}$, the *model likelihood* of the *distribution histogram* is sometimes higher,

$$\hat{Q}_{\text{h,d},D_{\text{o}},U}(A_{\text{o},z_{\text{h}}}, z)(A_{Q,N}) \quad \geq \quad \hat{Q}_{\text{h},U}(A_{\text{o},z_{\text{h}}}, z)(A_{Q,N})$$

In other words, querying via the *known derived* of the *model* sometimes reduces the *sensitivity* to the *unknown* and increases the *likelihood* at the cost of modifying the query.

5.5.4 Unknown necessary derived

In the discussion above of *classical modelled induction*, the *iso-set conditional stuffed historical probability distribution likelihood* and *sum sensitivity* relations are correlations rather than approximations or equivalences. In the case where the *models* are *transforms*, the variation over both (i) the set of *probability distribution substrate histograms*, $\mathcal{A}_{U,V_{\text{o}},1}$, and (ii) the set of *substrate transforms*, $\mathcal{T}_{U,V_{\text{o}}}$, has been informally implicit in the correlations. In the discussion above, the *model*, $T_{\text{o}} \in \mathcal{T}_{U,V_{\text{o}}}$, is *known* and the *derived*, $\hat{E}_{\text{h}} * T_{\text{o}}$, is both *necessary* and *known*. Optimisation can be done to find the *maximum likelihood estimate* of the *distribution histogram* for *known model*,

$$\{\tilde{E}_{\text{o}}\} = \text{maxd}(\{(E, \hat{Q}_{\text{m,d},T_{\text{o}},U}(E, z_{\text{o}})(A_{\text{o}})) : E \in \mathcal{A}_{U,V_{\text{o}},1}\})$$

Now consider the case where the *derived* is still *necessary*, $\hat{A}_{\text{o}} * T_{\text{o}} = \hat{E}_{\text{h}} * T_{\text{o}}$, but the *model*, T_{o}, is *unknown* and so the *derived* is *unknown*. Here the *transform*, T, is considered to be a parameter of the *iso-derived conditional generalised multinomial probability distribution*, $\hat{Q}_{\text{m,d},T,U}(E, z)$, along with the *distribution histogram*, E. Note, however, that while the *distribution histogram*, E, can be mapped to a real vector, $\hat{E}^{\square} \in \mathbf{R}^v$, in the *iso-derived conditional generalised multinomial parameterised probability density function*, $\text{mdtppdf}(T, z)(\hat{E}^{\square}) \in (\mathbf{R}^v :\to \mathbf{R})$, and hence can be a continuous argument to the *iso-derived conditional generalised multinomial likelihood function*, $\text{mdtlf}(T, z)(A^{\square}) \in (\mathbf{R}^v \to \mathbf{R})$, there is no straightforward mapping $\hat{T}^{\square}$ for the *transform*. Another problem is that the *iso-derived conditional generalised multinomial probability*, $\hat{Q}_{\text{m,d},T,U}(E, z_{\text{o}})(A_{\text{o}})$, is not explicitly constrained so that the *derived* is *necessary*, $\hat{A}_{\text{o}} * T = \hat{E}_{\text{h}} * T$. However, the *maximum likelihood estimate* for the pair $(\tilde{E}_{\rm o},\tilde{T}_{\rm o})$ can be defined as an optimisation of the *multinomial probability* conditional on the *iso-derived* where both the *distribution histogram* and *transform* are treated as arguments to a likelihood function,

$$\begin{aligned}&(\tilde{E}_{\rm o},\tilde{T}_{\rm o})\\ &\in\ \mathrm{maxd}(\{((E,T),\frac{Q_{\rm m,U}(E,z_{\rm o})(A_{\rm o})}{\sum_{B \in D^{-1}_{U,{\rm i},T,z_{\rm o}}(A_{\rm o}*T)} Q_{\rm m,U}(E,z_{\rm o})(B)}) :\\ &\qquad\qquad E \in \mathcal{A}_{U,V_{\rm o},1},\ T \in \mathcal{T}_{U,V_{\rm o}}\})\end{aligned}$$

The *sensitivity* to parameter is now with respect to the pair, (E,T), and not just with respect to the *distribution histogram*, E. Again, there is no mapping of the *transform* to a coordinate, $\hat{T}^{\square}$, so the *sensitivity* with respect to the *distribution-transform* pair at the *maximum likelihood estimate*, $(\tilde{E}_{\rm o},\tilde{T}_{\rm o})$, may be approximated as the sum of (i) the *sum sensitivity* of the *iso-derived conditional multinomial probability distribution* at the *maximum likelihood estimate*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm m,d},\tilde{T}_{\rm o},U}(\tilde{E}_{\rm o},z_{\rm o})))$$

and (ii) the negative logarithm of the cardinality of the *maximum likelihood estimate models*,

$$-\ \ln|\mathrm{max}(\{(T,\frac{Q_{\rm m,U}(\tilde{E}_{\rm o},z_{\rm o})(A_{\rm o})}{\sum_{B \in D^{-1}_{U,{\rm i},T,z_{\rm o}}(A_{\rm o}*T)} Q_{\rm m,U}(\tilde{E}_{\rm o},z_{\rm o})(B)}) : T \in \mathcal{T}_{U,V_{\rm o}}\})|$$

Although the *maximum likelihood estimate* and the *sensitivity* with respect to the pair, $(\tilde{E}_{\rm o},\tilde{T}_{\rm o})$, can be defined, there is, however, no singular solution to the optimisation with respect to the *distribution probability histogram*, E,

$$\begin{aligned}&\mathrm{maxd}(\{((E,T),\frac{Q_{\rm m,U}(E,z_{\rm o})(A_{\rm o})}{\sum_{B \in D^{-1}_{U,{\rm i},T,z_{\rm o}}(A_{\rm o}*T)} Q_{\rm m,U}(E,z_{\rm o})(B)}) :\\ &\qquad\qquad E \in \mathcal{A}_{U,V_{\rm o},1},\ T \in \mathcal{T}_{U,V_{\rm o}}\}) \supseteq \mathcal{A}_{U,V_{\rm o},1} \times \{T_{\rm s}\}\end{aligned}$$

where $T_{\rm s}$ is a *self transform*, for example the *self partition transform*, $T_{\rm s} = V^{\rm CS\{\}T}$ or the *full functional transform*, $T_{\rm s} = \{\{w\}^{{\rm CS}\{\}V{\rm T}} : w \in V\}^{\rm T}$. When the *transform* is a *self transform* the denominator equals the numerator, $\sum_{B \in D^{-1}_{U,{\rm i},T_{\rm s},z}(A*T_{\rm s})} Q_{\rm m,U}(E,z)(B) = Q_{\rm m,U}(E,z)(A)$, and the solution is degenerate. That is, the maximisation does not yield a single *maximum likelihood estimate* for the *distribution probability histogram*, $\tilde{E}_{\rm o}$.

In the case where the *derived* is *necessary* but *unknown*, the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\rm o}$, is just the *self transform*, $\tilde{T}_{\rm o} = T_{\rm s}$, which is the trivial case where everything is *known*.

5.5.5 Uniform possible derived induction

This singular solution for *unknown transform* can be addressed by making the *transform* more like a continuous vector. That is, by avoiding discontinuities in the *history probability function*. Consider the case where it is *unknown* if the given *histogram*, A_{o}, is a *sample histogram drawn* from the *distribution histogram*, E_{h}, so, in some cases $P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H_{\mathrm{o}}) = 0$. That is, it is *known* that some *derived* is *necessary*, $\exists B \in \mathcal{A}_{U,\mathrm{i},V_{\mathrm{o}},z_{\mathrm{o}}}\ (\hat{B} * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}})$, but not whether the given *derived histogram* is *necessary*, $\hat{A}_{\mathrm{o}} * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}$.

In the *necessary* given *derived* case, a *probability function* $P_{\mathrm{d}} \in (\mathcal{A}_{\mathrm{i}} \to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ of the *derived* can be defined as

$$P_{\mathrm{d}} = \{(Z_{\mathrm{o}} * \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}, 1)\} \ \cup\ ((\mathrm{ran}(D_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}}) \setminus \{Z_{\mathrm{o}} * \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}\}) \times \{0\})$$

where $Z_{\mathrm{o}} * \hat{E}_{\mathrm{h}} * T_{\mathrm{o}} \in \mathcal{A}_{\mathrm{i}}$ and $Z_{\mathrm{o}} = \mathrm{scalar}(z_{\mathrm{o}})$. That is, the *sample derived* is certain, $P_{\mathrm{d}}(A_{\mathrm{o}}*T_{\mathrm{o}}) = P_{\mathrm{d}}(Z_{\mathrm{o}}*\hat{E}_{\mathrm{h}}*T_{\mathrm{o}}) = 1$. The *expected iso-derived probability* in this *probability function*, P_{d}, is that of the *necessary iso-derived*,

$$\begin{aligned}\mathrm{expected}(P_{\mathrm{d}}, \{(A', \sum_{B \in D^{-1}_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}}(A')} Q_{\mathrm{m},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(B)) : A' \in \mathrm{ran}(D_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}})\}) \\ = \sum_{B \in D^{-1}_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}}(Z_{\mathrm{o}}*\hat{E}_{\mathrm{h}}*T_{\mathrm{o}})} Q_{\mathrm{m},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(B)\end{aligned}$$

In the not *necessary* given *derived* case, the probability of the *sample derived* is not certain, $P_{\mathrm{d,p}}(A_{\mathrm{o}} * T_{\mathrm{o}}) \notin \{0, 1\}$. In the absence of further *knowledge* it is assumed that the given *derived*, $\hat{A}_{\mathrm{o}} * T_{\mathrm{o}}$, is at least *possible* and that the *probability function* $P_{\mathrm{d,p}} \in (\mathcal{A}_{\mathrm{i}} \to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ of the *derived* is uniform,

$$P_{\mathrm{d,p}} = \mathrm{ran}(D_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}}) \times \{1/|\mathrm{ran}(D_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}})|\}$$

Now the *expected iso-derived probability* in this *probability function*, $P_{\mathrm{d,p}}$, is

$$\begin{aligned}\mathrm{expected}(P_{\mathrm{d,p}}, \{(A', \sum_{B \in D^{-1}_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}}(A')} Q_{\mathrm{m},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(B)) : A' \in \mathrm{ran}(D_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}})\}) \\ = 1/|\mathrm{ran}(D_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}})|\end{aligned}$$

This is to assume that the choice of *derived* per se is arbitrary. This relaxation of the constraint that the *sample* be *necessarily drawn* from the *iso-derived* of the *distribution*, $P_{\mathrm{d}}(A_{\mathrm{o}} * T_{\mathrm{o}}) = 1$, to the constraint that the *sample* be *possibly drawn* from the *iso-derived* of the *distribution*, $P_{\mathrm{d,p}}(A_{\mathrm{o}} * T_{\mathrm{o}}) > 0$, is equivalent to assuming that the *sample* is *drawn* from the *uniform possible iso-derived historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$, which is defined as the solution to

$$\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}} := \\
&\quad \Big(\bigcup\{\{(H, 1/\sum(P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}(G) : \\
&\qquad\qquad G \subseteq H_{\mathrm{h}}\%V_H,\ |G| = z_H,\ A_G * T_{\mathrm{o}} = A_H * T_{\mathrm{o}})) : \\
&\qquad\qquad\quad H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H\}^{\wedge} : \\
&\qquad\qquad\qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\
&\quad \{(H, 0) : H \in \mathcal{H}_{U,X},\ H \not\subseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset, 0)\}
\end{aligned}$$

All *iso-derived* subsets of the *distribution history* for a given set of *variables* and *size* are defined as equally probable,

$$\begin{aligned}
&\forall V \subseteq V_{\mathrm{h}}\ \forall H, G \subseteq H_{\mathrm{h}}\%V \\
&\qquad (A_G * T_{\mathrm{o}} = A_H * T_{\mathrm{o}} \implies P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}(G) = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}(H))
\end{aligned}$$

The *uniform possible iso-derived historically distributed history probability function* is such that given a *drawn history* $H \in \mathcal{H}_{U,X}$

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = \\
&\qquad \frac{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}
\end{aligned}$$

The *possible history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}$, is related to the *iso-derived conditional historical probability distribution*, $\hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)$, in the same way as for the *necessary* case, $P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}$, except that the normalising fraction is restored. In the case where all *derived* are *possible* the normalising fraction is $1/|\mathrm{ran}(D_{U,\mathrm{i},T_{\mathrm{o}},z_H})|$,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = \\
&\qquad \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},T_{\mathrm{o}},z_H})|} \frac{Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T_{\mathrm{o}},z_H}(A_H * T_{\mathrm{o}})} Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(B)}
\end{aligned}$$

Any *historically drawn history* is *possible*,

$$\forall H \subseteq H_{\mathrm{h}}\%V_H\ (H \neq \emptyset \implies P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}(H) > 0)$$

but sometimes the *probability* is lower than in the *necessary* case,

$$\forall H \subseteq H_{\mathrm{h}}\%V_H\ (P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H) > 0 \iff P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}(H) \leq P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H))$$

Now it can be seen that the *history probability function* is more continuous in the sense that the *uniform possible* domain may be larger than the *necessary* domain,

$$|\{H : H \in \mathcal{H}_{U,X},\ P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}(H) > 0\}| \geq \\ |\{H : H \in \mathcal{H}_{U,X},\ P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H) > 0\}|$$

The *uniform possible log likelihood* has similar properties to the *necessary log likelihood*.

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(E,z)(A) = \ln \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} Q_{\mathrm{m},U}(E,z)(B)} - \ln |\mathrm{ran}(D_{U,\mathrm{i},T,z})|$$

The cardinality of the *derived*, $|\mathrm{ran}(D_{U,\mathrm{i},T,z})|$, is equal to the cardinality of the *possible derived substrate histograms*,

$$|\mathrm{ran}(D_{U,\mathrm{i},T,z})| = \frac{(z+w'-1)!}{z!\ (w'-1)!}$$

where $w' = |T^{-1}|$. So the additional term in the *uniform possible log likelihood*, $-\ln |\mathrm{ran}(D_{U,\mathrm{i},T,z})|$, varies against the *possible derived volume*, w', where the *possible derived volume* is less than the *size*, $w' < z$, otherwise against the *size* scaled *log possible derived volume*, $z \ln w'$,

$$\begin{aligned} -\ln |\mathrm{ran}(D_{U,\mathrm{i},T,z})| &= -\ \mathrm{spaceCountsDerived}(U)(A,T) \\ &\sim -\ ((w'\ :\ w' < z)\ +\ (z \ln w'\ :\ w' \geq z)) \end{aligned}$$

In the case where the *sample* is *natural*, $A = A * T * T^{\dagger}$, the *uniform possible log likelihood* varies (i) against the *possible derived volume*, w', where the *possible derived volume* is less than the *size*, $w' < z$, otherwise against the *size* scaled *log possible derived volume*, $z \ln w'$, and (ii) with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned} &\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A) \sim \\ &\quad -\ ((w'\ :\ w' < z)\ +\ (z \ln w'\ :\ w' \geq z)) \\ &\quad +\ (z+v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\ &\qquad\qquad -z \times \mathrm{entropy}(A * T)\ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T) \end{aligned}$$

Note that the meaning of *possible* in *possible derived volume*, w', is that the *derived state* is *effective*, whereas the meaning of *possible* in *uniform possible log likelihood*, $\ln \hat{Q}_{\mathrm{m,d},T,U}(A,z)(A)$, is that the *distribution frequency* is effective or non-zero.

In the case where the *cross entropy* is greater than the logarithm of the *possible derived volume*, $\mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) > \ln w'$, so that the *component size cardinality relative entropy* is high, the *iso-derived conditional multinomial probability* varies against the *derived multinomial probability*,

$$\ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A) \quad \sim \quad - \ \ln \hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)$$

and the *sum sensitivity* is less than or equal to the *multinomial sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,d},T,U}(A, z))) \quad \leq \quad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A, z)))$$

In the case where *size* is greater than the *volume*, $z > v$, the *uniform possible log likelihood* varies (i) against the *possible derived volume*, w', where the *possible derived volume* is less than the *size*, $w' < z$, otherwise against the *size* scaled *log possible derived volume*, $z \ln w'$, and (ii) with the *volume* scaled *component cardinality size relative entropy*,

$$\begin{aligned} \ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A) \ \sim \\ & - ((w' \ : \ w' < z) \ + \ (z \ln w' \ : \ w' \geq z)) \\ & + v \times \mathrm{entropyRelative}(V^{\mathrm{C}} * T, A * T) \end{aligned}$$

In the case where the *sample* is near *natural*, $A \approx A * T * T^{\dagger}$, and the *size* is less than or equal to the *volume*, $z \leq v$, the *uniform possible log likelihood* varies (i) against the *possible derived volume*, w', where the *possible derived volume* is less than the *size*, $w' < z$, otherwise against the *size* scaled *log possible derived volume*, $z \ln w'$, and (ii) with the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned} \ln \hat{Q}_{\mathrm{m,d},T,U}(A, z)(A) \ \sim \\ & - ((w' \ : \ w' < z) \ + \ (z \ln w' \ : \ w' \geq z)) \\ & + z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T) \end{aligned}$$

In this case the correlation properties of *uniform possible derived induction* relate to the correlation properties of the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)$, more closely than those of *necessary derived induction.* The *specialising space* varies (i) with the *possible derived volume*, w', where the *possible derived volume* is less than the *size*, $w' < z$, otherwise with the *size* scaled *log possible derived volume*, $z \ln w'$, and (ii) against the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned} C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) \ \sim \\ & (w' \ : \ w' < z) \ + \ (z \ln w' \ : \ w' \geq z) \\ & - z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T) \end{aligned}$$

In *classical uniform possible modelled induction*, where (i) the *history probability function* is *uniform possible iso-derived historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample* equals the *naturalisation*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-derived conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the *possible derived volume*, w'_{o}, where the *possible derived volume* is less than the *size*, $w'_{\mathrm{o}} < z_{\mathrm{o}}$, otherwise against the *size* scaled *log possible derived volume*, $z_{\mathrm{o}} \ln w'_{\mathrm{o}}$,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\; -\,((w'_{\mathrm{o}} \,:\, w'_{\mathrm{o}} < z_{\mathrm{o}}) \,+\, (z_{\mathrm{o}} \ln w'_{\mathrm{o}} \,:\, w' \geq z_{\mathrm{o}}))$$

and (b) with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim \\ &\quad (z_{\mathrm{o}} + v_{\mathrm{o}}) \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}} + V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) \\ &\qquad\qquad -z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}}) \;-\, v_{\mathrm{o}} \times \mathrm{entropy}(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})\end{aligned}$$

So the *uniform possible log likelihood* is maximised when (a) the *possible derived volume* is minimised, (b) the *component entropy* is minimised, and (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

If, in addition, the *component size cardinality relative entropy* is high,

$$\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln w'_{\mathrm{o}}$$

the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\end{aligned}$$

In the case where the *size* is greater than the *volume*, $z_{\mathrm{o}} > v_{\mathrm{o}}$, the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *volume* scaled *component cardinality size relative entropy*,

$$\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad v_{\mathrm{o}} \times \mathrm{entropyRelative}(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}, A_{\mathrm{o}} * T_{\mathrm{o}})$$

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the *possible derived volume*, w'_{o}, where the *possible derived volume* is less than the *size*, $w'_{\mathrm{o}} < z_{\mathrm{o}}$, otherwise against the *size* scaled *log possible derived volume*, $z_{\mathrm{o}} \ln w'_{\mathrm{o}}$,

$$\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ ((w'_{\mathrm{o}}\ :\ w'_{\mathrm{o}} < z_{\mathrm{o}})\ +\ (z_{\mathrm{o}} \ln w'_{\mathrm{o}}\ :\ w' \geq z_{\mathrm{o}}))$$

(b) varies with the *size* scaled *component size cardinality relative entropy*,

$$\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad z_{\mathrm{o}} \times \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})$$

so (c) varies against the *specialising derived substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

So the *uniform possible log likelihood* is maximised, in this case, when (a) the *possible derived volume* is minimised, (b) the *derived entropy* is minimised, (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*, and (d) the *expected component entropy* is maximised.

Conjecture that, in the case of high *component size cardinality relative entropy*, the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *specialising derived substrate history coder space*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

and so the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* varies against the *log-likelihood* of the *iso-derived conditional stuffed historical probability distribution*

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad -\ \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

As described in section 'Necessary derived', the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate*, $\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,d},T_{\text{o}},U}(A_{\text{o},z_{\text{h}}}, z_{\text{o}})))$, can be related to queries on the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\text{o}} = \hat{A}_{\text{o}}$, in the special case where (i) the *sample histogram* is *natural*, $A_{\text{o}} = A_{\text{o}} * T_{\text{o}} * T_{\text{o}}^{\dagger}$, and (ii) the *component size cardinality relative entropy* is high, $\text{entropyCross}(A_{\text{o}} * T_{\text{o}}, V_{\text{o}}^{\text{C}} * T_{\text{o}}) > \ln w'_{\text{o}}$. In the case of *classical modelled induction*, the given *substrate transform* must be such that its *contraction* has *underlying variables* that are a subset of the query *variables*, $\text{und}(T_{\text{o}}^{\%}) \subseteq K$. In the case where the query *histogram* consists of one *effective state*, $Q = \{(S_Q, 1)\}$, there exists an *effective derived state* R_Q, where $\{R_Q\} = (Q * T_{\text{o}}^{\%})^{\text{FS}}$. The corresponding *underlying component* is $C_Q = T_{\text{o}}^{-1}(R_Q)$. If the *possible derived volume* is non singular, $w'_{\text{o}} > 1$, the query *histogram* itself cannot be *drawn* from the *distribution history*, $\hat{Q}_{\text{h,d},T_{\text{o}},U}(A_{\text{o}}, 1)(Q * \{N\}^{\text{U}}) = 0$, where $N \in (V_{\text{o}} \setminus K)^{\text{CS}}$, because the query *derived probability histogram* is not equal to a *uniform possible derived distribution probability histogram*, $\hat{Q} * \{N\}^{\text{U}} * T_{\text{o}} \neq \hat{A}_{\text{o}} * T_{\text{o}}$. The application of the query must be in terms of a modified *sample histogram*,

$$\begin{aligned}
&(Q * T_{\text{o}}^{\%} * \text{his}(T_{\text{o}}^{\%}) * A_{\text{o}})^{\wedge} \ \% \ (V_{\text{o}} \setminus K) = \\
&\quad \{(N, (\hat{Q}_{\text{h,d},T_{\text{o}},U}(A_{\text{o},z_{\text{h}}}, z_{\text{o}})(A_{Q,N}))^{1/z_{\text{o}}}) : N \in (V_{\text{o}} \setminus K)^{\text{CS}}, \\
&\qquad\quad A_{Q,N} = A_{\text{o}} - (A_{\text{o}} * C_Q) + ((A_{\text{o}} * C_Q) \ \% \ K * \{N\}^{\text{U}})\}^{\wedge}
\end{aligned}$$

The *model* application depends on the geometric scaling of the *historical distribution*, $\hat{Q}_{\text{h,d},T_{\text{o}},U}(A_{\text{o},z_{\text{h}}}, z_{\text{o}})$, so the *query sensitivity* to the *distribution histogram* varies with the *sum sensitivity* of the *historical distribution* at the *maximum likelihood estimate* divided by the *sample size*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,d},T_{\text{o}},U}(A_{\text{o},z_{\text{h}}}, z_{\text{o}})))/z_{\text{o}}$$

Now consider the case where the *model*, T_{o}, is *unknown*. The *maximum likelihood estimate* for the pair $(\tilde{E}_{\text{o}}, \tilde{T}_{\text{o}})$ in the *uniform possible* case is

$$(\tilde{E}_{\text{o}}, \tilde{T}_{\text{o}}) \ \in \ \text{maxd}(\{((E,T), \hat{Q}_{\text{m,d},T,U}(E, z_{\text{o}})(A_{\text{o}})) : E \in \mathcal{A}_{U,V_{\text{o}},1}, \ T \in \mathcal{T}_{U,V_{\text{o}}}\})$$

If there is a unique maximum for the *distribution probability histogram*, $\tilde{E}_{\text{o}}$, this can be rewritten in terms of the *derived-dependent*,

$$\tilde{T}_{\text{o}} \ \in \ \text{maxd}(\{(T, \hat{Q}_{\text{m,d},T,U}(A_{\text{o}}^{\text{D}(T)}, z_{\text{o}})(A_{\text{o}})) : T \in \mathcal{T}_{U,V_{\text{o}}}\})$$

Strictly speaking, this is only the case for the subset of *substrate transforms*, $\mathcal{T}_{U,V_o}$, for which the *derived-dependent histogram*, $A_o^{D(T)}$, is defined,

$$\{T : T \in \mathcal{T}_{U,V_o},$$
$$|\max(\{(D, \frac{Q_{m,U}(D,z_o)(A_o)}{\sum Q_{m,U}(D,z_o)(B) : B \in D^{-1}_{U,i,T,z_o}(A_o * T)}) : D \in \mathcal{A}_{U,V_o,z_o}\})| = 1\}$$

Even if the *derived-dependent histogram*, $A_o^{D(T)}$ is defined, the *maximum likelihood estimate* for the *model*, $\tilde{T}_o$, is not necessarily computable because the *derived-dependent histogram*, $A_o^{D(T)}$, is sometimes not computable.

If the optimisation is restricted to *natural transforms*, $A_o = A_o * T * T^{\dagger} \implies A_o^{D(T)} = A_o$, then the optimisation is

$$\tilde{T}_o \in \text{maxd}(\{(T, \hat{Q}_{m,d,T,U}(A_o,z_o)(A_o)) : T \in \mathcal{T}_{U,V_o},\ A_o = A_o * T * T^{\dagger}\})$$

In this case, all the *derived* are *possible* because the *distribution* equals the *sample*, so the optimisation is

$$\tilde{T}_o \in \text{maxd}(\{(T, \frac{1}{|\text{ran}(D_{U,i,T,z_o})|} \frac{Q_{m,U}(A_o,z_o)(A_o)}{\sum_{B \in D^{-1}_{U,i,T,z_o}(A_o * T)} Q_{m,U}(A_o,z_o)(B)}) :$$
$$T \in \mathcal{T}_{U,V_o},\ A_o = A_o * T * T^{\dagger}\})$$

Now, the set of *maximum likelihood estimates* for the *model*, $\tilde{T}_o$, is computable.

In this case, the numerator, $Q_{m,U}(A_o,z_o)(A_o)$, is constant.

The *maximum likelihood estimate* for the *model* is not *self*, $\tilde{T}_o \neq T_s$, if

$$\frac{1}{|\text{ran}(D_{U,i,\tilde{T}_o,z_o})|} \frac{Q_{m,U}(A_o,z_o)(A_o)}{\sum_{B \in D^{-1}_{U,i,\tilde{T}_o,z_o}(A_o * \tilde{T}_o)} Q_{m,U}(A_o,z_o)(B)} > \frac{1}{|\text{ran}(\mathcal{A}_{U,i,V_o,z_o})|}$$

which is the case if the *iso-derived conditional multinomial probability* is greater than the inverted average *iso-derived* cardinality,

$$\frac{Q_{m,U}(A_o,z_o)(A_o)}{\sum_{B \in D^{-1}_{U,i,\tilde{T}_o,z_o}(A_o * \tilde{T}_o)} Q_{m,U}(A_o,z_o)(B)} > \frac{|\text{ran}(D_{U,i,\tilde{T}_o,z_o})|}{|\text{dom}(D_{U,i,\tilde{T}_o,z_o})|}$$

The *sample* is *natural*, $A_o = A_o * T_o * T_o^{\dagger}$, so the *permutorial* term is constant for all *iso-derived*, and the *iso-derived conditional multinomial probability* simplifies,

$$\frac{\hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B)} \;=\; 1/\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A*T)} \prod_{S \in V^{\mathrm{CS}}} \frac{A_S!}{B_S!}$$

The *maximum likelihood estimate* for the *model* is not *self*, $\tilde{T}_{\mathrm{o}} \neq T_{\mathrm{s}}$, if

$$\sum_{B \in D^{-1}_{U,\mathrm{i},\tilde{T}_{\mathrm{o}},z_{\mathrm{o}}}(A_{\mathrm{o}}*\tilde{T}_{\mathrm{o}})} \prod_{S \in V_{\mathrm{o}}^{\mathrm{CS}}} \frac{A_{\mathrm{o}}(S)!}{B_S!} \;<\; \frac{|\mathrm{dom}(D_{U,\mathrm{i},\tilde{T}_{\mathrm{o}},z_{\mathrm{o}}})|}{|\mathrm{ran}(D_{U,\mathrm{i},\tilde{T}_{\mathrm{o}},z_{\mathrm{o}}})|}$$

The terms of the sum are less than or equal to one, $\prod_{S \in V^{\mathrm{CS}}} A_S!/B_S! \leq 1$, so the *model* is not *self* at least in the case where the *iso-derived* cardinality is less than the average *iso-derived* cardinality,

$$|D^{-1}_{U,\mathrm{i},\tilde{T}_{\mathrm{o}},z_{\mathrm{o}}}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}})| \;<\; \frac{|\mathrm{dom}(D_{U,\mathrm{i},\tilde{T}_{\mathrm{o}},z_{\mathrm{o}}})|}{|\mathrm{ran}(D_{U,\mathrm{i},\tilde{T}_{\mathrm{o}},z_{\mathrm{o}}})|}$$

The *sample* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, so the *maximum likelihood estimate* for the *model* is not *unary*, $\tilde{T}_{\mathrm{o}} \neq T_{\mathrm{u}}$, if the *sample* is not *cartesian*, $\hat{A}_{\mathrm{o}} \neq \hat{V}_{\mathrm{o}}^{\mathrm{C}}$.

In some cases the *maximum likelihood estimate* for the *model* is neither *self* nor *unary*, $\tilde{T}_{\mathrm{o}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$.

In *classical uniform possible modelled induction*, where (i) the *history probability function* is *uniform possible iso-derived historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}$, given some *unknown substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample histogram* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, then the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-derived conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, and, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, in the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is

$$\tilde{T}_{\mathrm{o}} \;\in\; \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h},\mathrm{d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger}\})$$

and in some cases the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, is non-trivial,

$$\tilde{T}_{\mathrm{o}} \;\notin\; \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$$

In the case where the *component size cardinality relative entropy* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution-model* pair is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\end{aligned}$$

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the *sample* is sometimes not *natural*. In this case the search must be constrained, such that the *sample* is near *natural*, by a limit u to the *relative space* of the *sample* with respect to its *naturalisation*. The *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, in the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is

$$\begin{aligned}\tilde{T}_{\mathrm{o}} \in \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\\ \mathrm{spaceRelative}(A_{\mathrm{o}} * T * T^{\dagger})(A_{\mathrm{o}}) \leq u\})\end{aligned}$$

Note that the choice of the limit, u, is not necessary in the stricter case where the *trimmed sample* is *unit*, $\mathrm{trim}(A_{\mathrm{o}}) = A_{\mathrm{o}}^{\mathrm{F}}$. In this case the *sample history* is bijective, $H_{\mathrm{o}} \in X \leftrightarrow V_{\mathrm{o}}^{\mathrm{CS}}$. That is, each *event* has a unique *state*. Often, in these cases, the *size* is much less than the *volume*, $z_{\mathrm{o}} \ll v_{\mathrm{o}}$. In this case of *sparse histogram*, where $A_{\mathrm{o}} * A_{\mathrm{o}}^{\mathrm{F}} = A_{\mathrm{o}}^{\mathrm{F}}$, all *transforms* are near *natural*, so the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, simplifies to

$$\tilde{T}_{\mathrm{o}} \in \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

In this case, the *over-fitted effective self transform* is the solution to the optimisation,

$$A_{\mathrm{o}} * A_{\mathrm{o}}^{\mathrm{F}} = A_{\mathrm{o}}^{\mathrm{F}} \implies \tilde{T}_{\mathrm{o}} = (A_{\mathrm{o}}^{\mathrm{FS}\{\}} \cup \{V_{\mathrm{o}}^{\mathrm{CS}} \setminus A_{\mathrm{o}}^{\mathrm{FS}}\})^{\mathrm{T}}$$

The *effective self transform* is *self* only for *effective states*, $A_{\mathrm{o}}^{\mathrm{FS}\{\}}$. It has a remainder *component* for all of the *ineffective states*, $V_{\mathrm{o}}^{\mathrm{CS}} \setminus A_{\mathrm{o}}^{\mathrm{FS}}$. The *derived volume* is $z_{\mathrm{o}} + 1$.

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the properties of the maximisation of the *log likelihood*, $\ln \hat{Q}_{\mathrm{m,d},T_{\mathrm{o}},U}(A_{\mathrm{o}}, z)(A_{\mathrm{o}})$, are consistent with the properties of the minimisation of the *sum sensitivity*, $\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,d},T_{\mathrm{o}},U}(A_{\mathrm{o}}, z)))$. So conjecture that in *classical uniform* *possible modelled induction*, where the *size* is less than the *volume*, but the *sample* is near *natural*, and the *relative entropy* is high, the *sum sensitivity* varies against the *log likelihood*, and the optimisation tends to minimise the *sensitivity* to the *distribution*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad -\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

Similarly, the *query sensitivity* to the *distribution*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}$$

is also minimised by the optimisation of *log-likelihood*.

There is no mapping of the *transform* to a coordinate, $\hat{T}^{\square}$, so the *sensitivity* to the *model*, $\tilde{T}_{\mathrm{o}}$ cannot be calculated directly as the *Fisher information* of a centrally distributed real *likelihood function*. Instead the *sensitivity* to the *model* is defined as the negative logarithm of the cardinality of the *maximum likelihood estimate models*,

$$- \ \ln |\mathrm{max}(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger}\})|$$

Although there is an anti-correlation between the *log-likelihood* and the *sensitivity* to the *distribution*, it is not necessarily the case that there is also an anti-correlation between the *log-likelihood* and the *sensitivity* to the *model*.

5.5.6 Uniform possible derived functional definition set induction

Again, consider extending the *model* for *uniform possible derived induction* from *transforms* to *functional definition sets*.

Given some *known substrate fud*, $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, such that there exists a *top transform*, $\exists T \in F_{\mathrm{o}}\ (\mathrm{der}(T) = \mathrm{der}(F_{\mathrm{o}}))$, the *derived histogram set* of the *distribution probability histogram* is $\{\hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}} : T \in F_{\mathrm{o}}\}$, where $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$. Consider the case where it is *unknown* if the given *histogram*, A_{o}, is a *sample histogram drawn* from the *distribution histogram*, E_{h}, so, in some cases $P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}(H_{\mathrm{o}}) = 0$. That is, it is *known* that some *derived histogram set* is *necessary*, $\exists B \in \mathcal{A}_{U,\mathrm{i},V_{\mathrm{o}},z_{\mathrm{o}}}\ \forall T \in F_{\mathrm{o}}\ (\hat{B} * T_{F_{\mathrm{o}}} = \hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}})$, but not whether the given *derived histogram set* is *necessary*, $\forall T \in F_{\mathrm{o}}\ (\hat{A} * T_{F_{\mathrm{o}}} = \hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}})$. In the absence of further *knowledge* it is assumed that the given *derived histogram set*, $\{\hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}} : T \in F_{\mathrm{o}}\}$ is at least *possible* and that the *probability function* of the *derived histogram set* is uniform. This relaxation of the constraint that the *sample* be *necessarily drawn* from the *iso-fud* of the *distribution* to the constraint that the *sample* be *possibly drawn* from the *iso-fud* of the *distribution* is equivalent to assuming that the *sample* is *drawn* from the *uniform possible iso-fud historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$, which is defined as the solution to

$$\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}} := \\
&\quad \Big(\bigcup \{\{(H, 1/\sum(P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}(G) : \\
&\qquad\qquad G \subseteq H_{\mathrm{h}}\%V_H,\ |G| = z_H, \\
&\qquad\qquad\quad \forall T \in F_{\mathrm{o}}\ (A_G * T_{F_{\mathrm{o}}} = A_H * T_{F_{\mathrm{o}}}))) : \\
&\qquad\qquad\qquad H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H\}^{\wedge} : \\
&\qquad\qquad\qquad\quad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\
&\quad \{(H, 0) : H \in \mathcal{H}_{U,X},\ H \not\subseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset, 0)\}
\end{aligned}$$

All *iso-fud* subsets of the *distribution history* for a given set of *variables* and *size* are defined as equally probable,

$$\begin{aligned}
&\forall V \subseteq V_{\mathrm{h}}\ \forall H, G \subseteq H_{\mathrm{h}}\%V \\
&\quad (\forall T \in F_{\mathrm{o}}\ (A_G * T_{F_{\mathrm{o}}} = A_H * T_{F_{\mathrm{o}}}) \implies P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}(G) = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}(H))
\end{aligned}$$

The *uniform possible iso-fud historically distributed history probability function* is such that given a *drawn history* $H \in \mathcal{H}_{U,X}$

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = \\
&\qquad \frac{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}
\end{aligned}$$

The *possible history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}$, is related to the *iso-fud conditional historical probability distribution*, $\hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)$, in the same way as for the *necessary* case, $P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}$, except that the normalising fraction is restored. In the case where all *derived histogram sets* are *possible* the normalising fraction is $1/|\mathrm{ran}(D_{U,\mathrm{i},F_{\mathrm{o}},z_H})|$,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = \\
&\qquad \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},F_{\mathrm{o}},z_H})|} \frac{Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H)}{\sum_{B \in D^{-1}_{U,\mathrm{i},F_{\mathrm{o}},z_H}(D_{U,F_{\mathrm{o}},z_H}(A_H))} Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(B)}
\end{aligned}$$

Any *historically drawn history* is *possible*,

$$\forall H \subseteq H_{\mathrm{h}}\%V_H\ (H \neq \emptyset \implies P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}(H) > 0)$$

but sometimes the *probability* is lower than in the *necessary* case,

$$\forall H \subseteq H_{\mathrm{h}}\%V_H\ (P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}(H) > 0 \iff P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}(H) \leq P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}(H))$$

The *uniform possible log likelihood* has similar properties to the *necessary log likelihood*.

$$\begin{aligned}\ln \hat{Q}_{\mathrm{m,d},F,U}(E,z)(A) \ &= \\ &\ln \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},F,z}(D_{U,F,z}(A))} Q_{\mathrm{m},U}(E,z)(B)} - \ln |\mathrm{ran}(D_{U,\mathrm{i},F,z})|\end{aligned}$$

The cardinality of the *derived histogram sets*, $|\mathrm{ran}(D_{U,\mathrm{i},F,z})|$, is greater than or equal to the cardinality of the *possible derived substrate histograms* of the *fud transform*,

$$\begin{aligned}|\mathrm{ran}(D_{U,\mathrm{i},F,z})| &\geq |\mathrm{ran}(D_{U,\mathrm{i},F^{\mathrm{T}},z})| \\ &= \frac{(z + |(F^{\mathrm{T}})^{-1}| - 1)!}{z!\ (|(F^{\mathrm{T}})^{-1}| - 1)!}\end{aligned}$$

So the additional term in the *uniform possible log likelihood*, $-\ln|\mathrm{ran}(D_{U,\mathrm{i},F,z})|$, varies, for each *transform* $T \in F$, against the *possible derived volume*, $|T_F^{-1}|$, where the *possible derived volume* is less than the *size*, $|T_F^{-1}| < z$, otherwise against the *size* scaled *log possible derived volume*, $z \ln |T_F^{-1}|$,

$$\begin{aligned}-\ln |\mathrm{ran}(D_{U,\mathrm{i},F,z})| &\sim \\ &-\sum_{T \in F} \big((|T_F^{-1}| \ :\ |T_F^{-1}| < z)\ +\ (z \ln |T_F^{-1}| \ :\ |T_F^{-1}| \geq z)\big)\end{aligned}$$

The *fud-independent*, $A^{\mathrm{E_F}(F)} \in \mathcal{A}_{U,V,z}$, is defined,

$$\begin{aligned}\{A^{\mathrm{E_F}(F)}\} \ &= \\ &\mathrm{maxd}(\{(E, \textstyle\sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},F,z}(D_{U,F,z}(A)))) : E \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *fud-independent* approximates to the arithmetic *average* of the *naturalisations*,

$$A^{\mathrm{E_F}(F)} \ \approx \ Z_{1/|F|} * \sum_{T \in F} A * T_F * T_F^{\dagger}$$

In the case where the *sample* is equal to the *fud-independent*, $A = A^{\mathrm{E_F}(F)}$, the *uniform possible log likelihood* varies (i) against the sum of the *possible derived volumes* or *size* scaled *log possible derived volumes*, and (ii) with the sum of the *size-volume* scaled *component size cardinality sum relative entropies*,

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \ \sim& \\
&- \sum_{T \in F} \big((|T_F^{-1}| \ : \ |T_F^{-1}| < z) \ + \ (z \ln |T_F^{-1}| \ : \ |T_F^{-1}| \geq z) \big) \\
&+ \sum_{T \in F} \big((z+v) \times \mathrm{entropy}(A * T_F + V^{\mathrm{C}} * T_F) \\
&\qquad\qquad -z \times \mathrm{entropy}(A * T_F) \ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T_F) \big)
\end{aligned}$$

In the case where the *cross entropies* are sufficient, $\forall T \in F\ (\mathrm{entropyCross}(A * T_F, V^{\mathrm{C}} * T_F) > \ln |T_F^{-1}|)$, the logarithm of the *iso-fud conditional multinomial probability* varies against the sum of the logarithms of the *derived multinomial probabilities*,

$$\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \ \ \sim \ \ -\sum_{T \in F} \ln \hat{Q}_{\mathrm{m},U}(A * T_F, z)(A * T_F)$$

and the *sum sensitivity* is less than or equal to the *multinomial sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,d},F,U}(A,z))) \ \ \leq \ \ \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A,z)))$$

In the case where *size* is greater than the *volume*, $z > v$, the *uniform possible log likelihood* varies (i) against the sum of the *possible derived volumes* or *size* scaled *log possible derived volumes*, and (ii) with the sum of the *volume* scaled *component cardinality size relative entropies*,

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \ \sim& \\
&- \sum_{T \in F} \big((|T_F^{-1}| \ : \ |T_F^{-1}| < z) \ + \ (z \ln |T_F^{-1}| \ : \ |T_F^{-1}| \geq z) \big) \\
&+ v \times \sum_{T \in F} \mathrm{entropyRelative}(V^{\mathrm{C}} * T_F, A * T_F)
\end{aligned}$$

In the case where the *sample* is near *natural*, $A \approx A^{\mathrm{E_F}(F)}$, and the *size* is less than or equal to the *volume*, $z \leq v$, the *uniform possible log likelihood* varies (i) against the sum of the *possible derived volumes* or *size* scaled *log possible derived volumes*, and (ii) with the sum of the *size* scaled *component size cardinality relative entropies*,

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{m,d},F,U}(A,z)(A) \ \sim& \\
&- \sum_{T \in F} \big((|T_F^{-1}| \ : \ |T_F^{-1}| < z) \ + \ (z \ln |T_F^{-1}| \ : \ |T_F^{-1}| \geq z) \big) \\
&+ z \times \sum_{T \in F} \mathrm{entropyRelative}(A * T_F, V^{\mathrm{C}} * T_F)
\end{aligned}$$

In this case the correlation properties of *uniform possible fud induction* relate to the correlation properties of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)$, more closely than those of *necessary fud induction*. The *specialising space* varies (i) with the *possible fud derived volume*, $|(F^{\mathrm{T}})^{-1}|$, or the *size* scaled *log possible fud derived volume*, $z \ln |(F^{\mathrm{T}})^{-1}|$, (ii) with the *size* scaled *fud transform derived entropy* and (iii) against the sum of the *size* scaled *component size cardinality cross entropies* of the *transforms* of the *fud*,

$$\begin{aligned}
&C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F)^{\mathrm{s}}(H) \ \sim \\
&\qquad\qquad (|(F^{\mathrm{T}})^{-1}| \ : \ |(F^{\mathrm{T}})^{-1}| < z) \ + \ (z \ln |(F^{\mathrm{T}})^{-1}| \ : \ |(F^{\mathrm{T}})^{-1}| \geq z) \\
&\qquad\quad + z \times \mathrm{entropy}(A * F^{\mathrm{T}}) \\
&\qquad\quad - z \times \sum_{T \in F} \mathrm{entropyCross}(A * T_F, V_T^{\mathrm{C}} * T)
\end{aligned}$$

In *classical uniform possible functional definition set induction*, where (i) the *history probability function* is *uniform possible iso-fud historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}$, given some *substrate fud* in the *sample variables* $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, such that there exists a *top transform*, $\exists T \in F_{\mathrm{o}}\ (\mathrm{der}(T) = \mathrm{der}(F_{\mathrm{o}}))$, if it is the case that (ii) the *sample* equals the *fud-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-fud conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the sum of the *possible derived volumes* or *size* scaled *log possible derived volumes*

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \\
&\qquad - \sum_{T \in F_{\mathrm{o}}} \big((|T_{F_{\mathrm{o}}}^{-1}| \ : \ |T_{F_{\mathrm{o}}}^{-1}| < z_{\mathrm{o}}) \ + \ (z_{\mathrm{o}} \ln |T_{F_{\mathrm{o}}}^{-1}| \ : \ |T_{F_{\mathrm{o}}}^{-1}| \geq z_{\mathrm{o}})\big)
\end{aligned}$$

and (b) with the sum of the *size-volume* scaled *component size cardinality sum relative entropies*,

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \\
&\qquad + \sum_{T \in F_{\mathrm{o}}} \big((z_{\mathrm{o}} + v_{\mathrm{o}}) \times \mathrm{entropy}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}} + V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}}) \\
&\qquad\qquad\qquad\qquad - z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}) \ - v_{\mathrm{o}} \times \mathrm{entropy}(V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}})\big)
\end{aligned}$$

So the *uniform possible log likelihood* is maximised when (a) the total *possible derived volume* is minimised, (b) the sum of the *derived entropy* of all *transforms* is minimised, and (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for all *transforms*.

If, in addition, the *component size cardinality relative entropies* are high,

$$\forall T \in F_{\mathrm{o}}\ (\mathrm{entropyCross}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}}) > \ln |T_{F_{\mathrm{o}}}^{-1}|)$$

the *sum sensitivity* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\end{aligned}$$

The *query sensitivity*, $\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}$, is also lower.

In the case where the *size* is greater than the *volume*, $z_{\mathrm{o}} > v_{\mathrm{o}}$, the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *volume* scaled *component cardinality size relative entropies*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim\ v_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropyRelative}(V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}}, A_{\mathrm{o}} * T_{F_{\mathrm{o}}})$$

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-independent*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the sum of the *possible derived volumes* or *size* scaled *log possible derived volumes*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim\\ &\quad -\sum_{T \in F_{\mathrm{o}}} \big((|T_{F_{\mathrm{o}}}^{-1}| \ :\ |T_{F_{\mathrm{o}}}^{-1}| < z_{\mathrm{o}}) \ +\ (z_{\mathrm{o}} \ln |T_{F_{\mathrm{o}}}^{-1}| \ :\ |T_{F_{\mathrm{o}}}^{-1}| \geq z_{\mathrm{o}})\big)\end{aligned}$$

(b) varies with the *size* scaled *component size cardinality relative entropies*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim\ z_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}})$$

so (c) varies against the *specialising fud substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F},\mathrm{H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})$$

and (d) varies against the *specialising-canonical space difference*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ -(2C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F},\mathrm{H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{H},V_{\mathrm{o}}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{G},V_{\mathrm{o}}}(H_{\mathrm{o}}))$$

where F^V is the *expansion* that adds a *unary transform* in the remaining *underlying variables*, $F \cup \{\{(V \setminus \mathrm{und}(F))^{\mathrm{CS}}\}^{\mathrm{T}}\}$.

So the *uniform possible log likelihood* is maximised, in this case, when (a) the total *possible derived volume* is minimised, (b) the total *derived entropy* is minimised, (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for each *transform*, and (d) the total *expected component entropy* is maximised. It is also conjectured that, (i) the *derived entropy* decreases up the *layers*, (ii) the *possible derived volume* decreases up the *layers*, (iii) the *expected component entropy* increases up the *layers*, and (iv) the *component size cardinality cross entropy* increases up the *layers*.

Conjecture that, in the case of high *component size cardinality relative entropies*, the *sum sensitivity* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *specialising fud substrate history coder space*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F},\mathrm{H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})$$

and so the *sum sensitivity* of the *iso-fud conditional stuffed historical probability distribution* varies against the *log-likelihood* of the *iso-fud conditional stuffed historical probability distribution*

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad -\ \ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

Now consider the case where the *model*, F_{o}, is *unknown*. The *maximum likelihood estimate* for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{F}_{\mathrm{o}})$ in the *uniform possible* case is

$$\begin{aligned}(\tilde{E}_{\mathrm{o}}, \tilde{F}_{\mathrm{o}}) \ \in \ & \mathrm{maxd}(\{((E,F), \hat{Q}_{\mathrm{m,d},F,U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : \\ & \qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F)\})\end{aligned}$$

If there is a unique maximum for the *distribution probability histogram*, $\tilde{E}_{\mathrm{o}}$, this can be rewritten in terms of the *fud-dependent*,

$$\begin{aligned}\tilde{F}_{\mathrm{o}} \ \in \ & \mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{m,d},F,U}(A_{\mathrm{o}}^{\mathrm{D_F}(F)}, z_{\mathrm{o}})(A_{\mathrm{o}})) : \\ & \qquad F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F)\})\end{aligned}$$

If the optimisation is restricted such that the *sample* is equal to the *fud-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F)} \implies A_{\mathrm{o}}^{\mathrm{D_F}(F)} = A_{\mathrm{o}}$, then the optimisation is

$$\begin{aligned}\tilde{F}_{\mathrm{o}} \in\ & \mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{m,d},F,U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : \\ & \qquad F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F),\ A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F)}\})\end{aligned}$$

In this case, all of the *derived histogram sets* are *possible* because the *distribution* equals the *sample*, so the optimisation is

$$\begin{aligned}\tilde{F}_{\mathrm{o}} \in\ & \mathrm{maxd}(\{(F, \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},F,z_{\mathrm{o}}})|} \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in D_{U,\mathrm{i},F,z_{\mathrm{o}}}^{-1}(D_{U,F,z_{\mathrm{o}}}(A_{\mathrm{o}}))} Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(B)}) : \\ & \qquad F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F),\ A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F)}\})\end{aligned}$$

Now, the set of *maximum likelihood estimates* for the *model*, $\tilde{F}_{\mathrm{o}}$, is computable, if an approximation is used for the *fud-independent*, $A_{\mathrm{o}}^{\mathrm{E_F}(F)}$,

$$\begin{aligned}\tilde{F}_{\mathrm{o}} \in\ & \mathrm{maxd}(\{(F, \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},F,z_{\mathrm{o}}})|} \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in D_{U,\mathrm{i},F,z_{\mathrm{o}}}^{-1}(D_{U,F,z_{\mathrm{o}}}(A_{\mathrm{o}}))} Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(B)}) : \\ & \qquad F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F),\ A_{\mathrm{o}} = Z_{1/|F|} * \sum_{T \in F} A_{\mathrm{o}} * T_F * T_F^{\dagger}\})\end{aligned}$$

In some cases the *maximum likelihood estimate* for the *model* is neither *self* nor *unary*, $\tilde{F}_{\mathrm{o}} \notin \{\{T_{\mathrm{s}}\}, \{T_{\mathrm{u}}\}\}$.

In *classical uniform possible functional definition set induction*, where (i) the *history probability function* is *uniform possible iso-fud historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}$, given some *unknown substrate fud* in the *sample variables* $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, such that there exists a *top transform*, $\exists T \in F_{\mathrm{o}}\ (\mathrm{der}(T) = \mathrm{der}(F_{\mathrm{o}}))$, if it is the case that (ii) the *sample histogram* equals the *fud-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, then the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-fud conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, and, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *maximum likelihood estimate* of the *model*, $\tilde{F}_{\mathrm{o}}$, in the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is

$$\begin{aligned}\tilde{F}_{\mathrm{o}} \in\ & \mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{h,d},F,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : \\ & \qquad F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F),\ A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F)}\})\end{aligned}$$

and in some cases the *maximum likelihood estimate* for the *model*, $\tilde{F}_{\mathrm{o}}$, is non-trivial,

$$\tilde{F}_{\mathrm{o}} \quad \notin \quad \{\{T_{\mathrm{s}}\},\{T_{\mathrm{u}}\}\}$$

In the case where the *component size cardinality relative entropies* are high,

$$\forall T \in F_{\mathrm{o}}\ (\mathrm{entropyCross}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}}) > \ln |T_{F_{\mathrm{o}}}^{-1}|)$$

the *sum sensitivity* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution-model* pair is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},\tilde{F}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\end{aligned}$$

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is approximated

$$\begin{aligned}\tilde{F}_{\mathrm{o}} \quad \in \quad &\mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{h,d},F,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : \\ &\qquad F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F),\ A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_F}(F)}\})\end{aligned}$$

In this case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the properties of the maximisation of the *log likelihood*, $\ln \hat{Q}_{\mathrm{m,d},F_{\mathrm{o}},U}(A_{\mathrm{o}}, z)(A_{\mathrm{o}})$, are consistent with the properties of the minimisation of the *sum sensitivity*, $\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,d},F_{\mathrm{o}},U}(A_{\mathrm{o}}, z)))$. So conjecture that in *classical uniform possible functional definition set induction*, where the *size* is less than the *volume*, but the *sample* approximates to the *fud-independent*, and the *relative entropies* are high, the *sum sensitivity* varies against the *log likelihood*, and the optimisation tends to minimise the *sensitivity* to the *distribution*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},\tilde{F}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad -\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

Similarly, the *query sensitivity* to the *distribution*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},\tilde{F}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}$$

is also minimised by the optimisation of *log-likelihood*.

5.5.7 Uniform possible derived functional definition set decomposition induction

The last section extended the *model* from *transforms* to *functional definition sets*. Now extend further to *functional definition set decompositions*. This discussion is very similar to that of the previous section, except that now the *fuds* are *contingent* on the *slice*.

Given some non-empty *known substrate fud decomposition*, $D_{\mathrm{o}} \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\}$, such that there exists a *top transform* for all of the *fuds*, $\forall F \in \mathrm{fuds}(D_{\mathrm{o}})\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$, the *component derived set* of the *distribution probability histogram* is $\{(C, \{\hat{E}_{\mathrm{h}} * C * T_F : T \in F\}) : (C,F) \in \mathrm{cont}(D_{\mathrm{o}})\}$, where $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$ and $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$. Consider the case where it is *unknown* if the given *histogram*, A_{o}, is a *sample histogram drawn* from the *distribution histogram*, E_{h}, so, in some cases $P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}}(H_{\mathrm{o}}) = 0$. That is, it is *known* that some *component derived set* is *necessary*, $\exists B \in \mathcal{A}_{U,\mathrm{i},V_{\mathrm{o}},z_{\mathrm{o}}}\ \forall (C,F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (\hat{B} * C * T_F = \hat{E}_{\mathrm{h}} * C * T_F)$, but not whether the given *component derived set* is *necessary*, $\forall (C,F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (\hat{A} * C * T_F = \hat{E}_{\mathrm{h}} * C * T_F)$. In the absence of further *knowledge* it is assumed that the given *component derived set*, $\{\hat{E}_{\mathrm{h}} * C * T_F : (C,F) \in \mathrm{cont}(D_{\mathrm{o}}),\ T \in F\}$ is at least *possible* and that the *probability function* of the *component derived set* is uniform. This relaxation of the constraint that the *sample* be *necessarily drawn* from the *iso-fud-decomposition* of the *distribution* to the constraint that the *sample* be *possibly drawn* from the *iso-fud-decomposition* of the *distribution* is equivalent to assuming that the *sample* is *drawn* from the *uniform possible iso-fud-decomposition historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$, which is defined as the solution to

$$
\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}} := \\
&\quad \Big(\bigcup \{\{(H, 1/\textstyle\sum(P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}}(G) : \\
&\qquad\qquad G \subseteq H_{\mathrm{h}}\%V_H,\ |G| = z_H,\ \forall (C,F) \in \mathrm{cont}(D_{\mathrm{o}}) \\
&\qquad\qquad\quad \forall T \in F\ (A_G * C * T_F = A_H * C * T_F))) : \\
&\qquad\qquad\qquad H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H\}^{\wedge} : \\
&\qquad\qquad\qquad\quad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \nsubseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset, 0)\}
\end{aligned}
$$

All *iso-fud-decomposition* subsets of the *distribution history* for a given set of *variables* and *size* are defined as equally probable,

$$\forall V \subseteq V_{\mathrm{h}}\ \forall H, G \subseteq H_{\mathrm{h}}\%V$$
$$(\forall (C,F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (A_G * C * T_F = A_H * C * T_F) \implies$$
$$P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}}(G) = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}}(H))$$

The *uniform possible iso-fud-decomposition historically distributed history probability function* is such that given a *drawn history* $H \in \mathcal{H}_{U,X}$

$$\hat{Q}_{\mathrm{h},\mathrm{d},D_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = \frac{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}$$

The *possible history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}}$, is related to the *iso-fud-decomposition conditional historical distribution*, $\hat{Q}_{\mathrm{h},\mathrm{d},D_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)$, in the same way as for the *necessary* case, $P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}}$, except that the normalising fraction is restored. In the case where all *component derived sets* are *possible* the normalising fraction is $1/|\mathrm{ran}(D_{U,\mathrm{i},D_{\mathrm{o}},\mathrm{F},z_H})|$,

$$\hat{Q}_{\mathrm{h},\mathrm{d},D_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},D_{\mathrm{o}},\mathrm{F},z_H})|} \frac{Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H)}{\sum_{B \in D^{-1}_{U,\mathrm{i},D_{\mathrm{o}},\mathrm{F},z_H}(D_{U,D_{\mathrm{o}},\mathrm{F},z_H}(A_H))} Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(B)}$$

Any *historically drawn history* is *possible*,

$$\forall H \subseteq H_{\mathrm{h}}\%V_H\ (H \neq \emptyset \implies P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}}(H) > 0)$$

but sometimes the *probability* is lower than in the *necessary* case,

$$\forall H \subseteq H_{\mathrm{h}}\%V_H\ (P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}}(H) > 0 \iff P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}}(H) \leq P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}}(H))$$

The *uniform possible log likelihood* has similar properties to the *necessary log likelihood*.

$$\ln \hat{Q}_{\mathrm{m},\mathrm{d},D,U}(E,z)(A) = \ln \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A))} Q_{\mathrm{m},U}(E,z)(B)} - \ln|\mathrm{ran}(D_{U,\mathrm{i},D,\mathrm{F},z})|$$

In *classical uniform possible functional definition set decomposition induction*, where (i) the *history probability function* is *uniform possible iso-fud-decomposition historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},D_{\mathrm{o}}}$, given some *substrate fud* in the *sample variables* $D_{\mathrm{o}} \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\}$, such that there exists a *top transform* for all of the *fuds*, $\forall F \in \mathrm{fuds}(D_{\mathrm{o}})\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$, if it is the case that (ii) the *sample* equals the *fud-decomposition-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D_{\mathrm{o}})}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-fud-decomposition conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the sum of the *possible derived volumes* or *size* scaled *log possible derived volumes* of the *slices*

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \\
&\quad -\sum_{(C,F)\in \mathrm{cont}(D_{\mathrm{o}})} \sum_{T \in F} \big((|T_F^{-1}| \ :\ |T_F^{-1}| < z_{A_{\mathrm{o}}*C}) \ + \\
&\qquad\qquad\qquad\qquad (z_{A_{\mathrm{o}}*C} \ln |T_F^{-1}| \ :\ |T_F^{-1}| \geq z_{A_{\mathrm{o}}*C}) \big)
\end{aligned}$$

and (b) with the sum of the *size-volume* scaled *component size cardinality sum relative entropies* for all *slices*,

$$\begin{aligned}
&\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \\
&\quad \sum_{(C,F)\in \mathrm{cont}(D_{\mathrm{o}})} \sum_{T \in F} \big((z_{A_{\mathrm{o}}*C} + |C|) \times \mathrm{entropy}(A_{\mathrm{o}} * C * T_F + C * T_F) \\
&\qquad\qquad -z_{A_{\mathrm{o}}*C} \times \mathrm{entropy}(A_{\mathrm{o}} * C * T_F) \ - |C| \times \mathrm{entropy}(C * T_F) \big)
\end{aligned}$$

So the *uniform possible log likelihood* is maximised when (a) the total *possible derived volume* is minimised, (b) the sum of the *derived entropy* of all *transforms* for all *slices* is minimised, and (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for all *transforms* for all *slices*.

If, in addition, the *component size cardinality relative entropies* are high,

$$\forall (C, F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (\mathrm{entropyCross}(A_{\mathrm{o}} * C * T_F, C * T_F) > \ln |T_F^{-1}|)$$

the *sum sensitivity* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}
&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \\
&\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))
\end{aligned}$$

The *query sensitivity*, $\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},D_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})))/z_\mathrm{o}$, is also lower.

In the case where the *size* is greater than the *volume*, $z_\mathrm{o} > v_\mathrm{o}$, the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *volume* scaled *component cardinality size relative entropies* for all *slices*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},D_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o}) \ &\sim \\ &\sum_{(C,F)\in\mathrm{cont}(D_\mathrm{o})} \Big(|C| \times \sum_{T \in F} \mathrm{entropyRelative}(C * T_F, A_\mathrm{o} * C * T_F)\Big)\end{aligned}$$

In the case where the *size* is less than the *volume*, $z_\mathrm{o} < v_\mathrm{o}$, but the *sample histogram* approximates to the *fud-decomposition-independent histogram*, $A_\mathrm{o} \approx A_\mathrm{o}^{\mathrm{E}_{\mathrm{D,F}}(D_\mathrm{o})}$, or $\mathrm{spaceRelative}(A_\mathrm{o}^{\mathrm{E}_{\mathrm{D,F}}(D_\mathrm{o})})(A_\mathrm{o}) \approx 0$, the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the sum of the *possible derived volumes* or *size* scaled *log possible derived volumes* of the *slices*

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},D_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o}) \ &\sim \\ -\sum_{(C,F)\in\mathrm{cont}(D_\mathrm{o})} \sum_{T \in F} &\Big((|T_F^{-1}| \ : \ |T_F^{-1}| < z_{A_\mathrm{o} * C}) \ + \\ &\qquad (z_{A_\mathrm{o} * C} \ln |T_F^{-1}| \ : \ |T_F^{-1}| \geq z_{A_\mathrm{o} * C})\Big)\end{aligned}$$

(b) varies with sum of the *size* scaled *component size cardinality relative entropies* of all *transforms* for all *slices*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},D_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o}) \ &\sim \\ &\sum_{(C,F)\in\mathrm{cont}(D_\mathrm{o})} \Big(z_{A_\mathrm{o} * C} \times \sum_{T \in F} \mathrm{entropyRelative}(A_\mathrm{o} * C * T_F, C * T_F)\Big)\end{aligned}$$

so (c) varies against the *specialising fud decomposition substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,d},D_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o}) \ \sim \ -C_{\mathrm{G},V_\mathrm{o},\mathrm{D,F,H}}(D_\mathrm{o}^{V_\mathrm{o}})^\mathrm{s}(H_\mathrm{o})$$

and (d) varies against the *specialising-canonical space difference*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},D_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o}) \ &\sim \\ &-(2C_{\mathrm{G},V_\mathrm{o},\mathrm{D,F,H}}(D_\mathrm{o}^{V_\mathrm{o}})^\mathrm{s}(H_\mathrm{o}) - C^\mathrm{s}_{\mathrm{H},V_\mathrm{o}}(H_\mathrm{o}) - C^\mathrm{s}_{\mathrm{G},V_\mathrm{o}}(H_\mathrm{o}))\end{aligned}$$

where D^V is the *expansion* that adds a *unary transform* in the remaining *underlying variables* to the leaf *fuds* in the *decomposition* tree such that the *fud* of each path of the *application* tree has complete coverage of the *substrate*,

$$\forall L \in \text{paths}(D^*)\ (\bigcup_{(\cdot,(F,\cdot)) \in L} V_F = V)$$

where $V_F = \text{und}(F)$.

So the *uniform possible log likelihood* is maximised, in this case, when (a) the total *possible derived volume* is minimised, (b) the total *derived entropy* is minimised, (c) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for all *transforms* for all *slices*, and (d) the total *expected component entropy* is maximised. It is also conjectured that, for all *fuds*, (i) the *derived entropy* decreases up the *layers*, (ii) the *possible derived volume* decreases up the *layers*, (iii) the *expected component entropy* increases up the *layers*, and (iv) the *component size cardinality cross entropy* increases up the *layers*.

Conjecture that, in the case of high *component size cardinality relative entropies*, the *sum sensitivity* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *specialising fud decomposition substrate history coder space*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,d},D_\text{o},U}(A_{\text{o},z_\text{h}}, z_\text{o}))) \quad \sim \quad C_{\text{G},V_\text{o},\text{D,F,H}}(D_\text{o}^{V_\text{o}})^{\text{s}}(H_\text{o})$$

and so the *sum sensitivity* of the *iso-fud-decomposition conditional stuffed historical probability distribution* varies against the *log-likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution*

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,d},D_\text{o},U}(A_{\text{o},z_\text{h}}, z_\text{o}))) \quad \sim \quad -\ \ln \hat{Q}_{\text{h,d},D_\text{o},U}(A_{\text{o},z_\text{h}}, z_\text{o})(A_\text{o})$$

Now consider the case where the *model*, D_o, is *unknown*. The *maximum likelihood estimate* for the pair $(\tilde{E}_\text{o}, \tilde{D}_\text{o})$ in the *uniform possible* case is

$$\begin{aligned}(\tilde{E}_\text{o}, \tilde{D}_\text{o}) \quad \in \quad & \text{maxd}(\{((E,D), \hat{Q}_{\text{m,d},D,U}(E, z_\text{o})(A_\text{o})) : \\ & \quad E \in \mathcal{A}_{U,V_\text{o},1}, \\ & \quad\quad D \in \mathcal{D}_{\text{F},U,V_\text{o}} \setminus \{\emptyset\},\ \forall F \in \text{fuds}(D)\ \exists T \in F\ (W_T = W_F)\})\end{aligned}$$

If there is a unique maximum for the *distribution probability histogram*, $\tilde{E}_\text{o}$, this can be rewritten in terms of the *fud-decomposition-dependent*,

$$\begin{aligned}\tilde{D}_\text{o} \quad \in \quad & \text{maxd}(\{(D, \hat{Q}_{\text{m,d},D,U}(A_\text{o}^{\text{D}_{\text{D,F}}(D)}, z_\text{o})(A_\text{o})) : \\ & \quad\quad D \in \mathcal{D}_{\text{F},U,V_\text{o}} \setminus \{\emptyset\},\ \forall F \in \text{fuds}(D)\ \exists T \in F\ (W_T = W_F)\})\end{aligned}$$

If the optimisation is restricted such that the *sample* is equal to the *fud-decomposition-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)} \implies A_{\mathrm{o}}^{\mathrm{D_{D,F}}(D)} = A_{\mathrm{o}}$, then the optimisation is

$$\begin{aligned}\tilde{D}_{\mathrm{o}} \in \ & \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{m,d},D,U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : \\ & \quad D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F), \\ & \quad\quad A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)}\})\end{aligned}$$

In this case, all of the *component derived sets* are *possible* because the *distribution* equals the *sample*, so the optimisation is

$$\begin{aligned}\tilde{D}_{\mathrm{o}} \in \ & \mathrm{maxd}(\{(D, \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},D,\mathrm{F},z_{\mathrm{o}}})|} \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in D_{U,\mathrm{i},D,\mathrm{F},z_{\mathrm{o}}}^{-1}(D_{U,D,\mathrm{F},z_{\mathrm{o}}}(A_{\mathrm{o}}))} Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(B)}) : \\ & \quad D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F), \\ & \quad\quad A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)}\})\end{aligned}$$

Now, the set of *maximum likelihood estimates* for the *model*, $\tilde{D}_{\mathrm{o}}$, is computable, if an approximation is used for the *fud-decomposition-independent*, $A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)}$,

$$\begin{aligned}\tilde{D}_{\mathrm{o}} \in & \\ & \mathrm{maxd}(\{(D, \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},D,\mathrm{F},z_{\mathrm{o}}})|} \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in D_{U,\mathrm{i},D,\mathrm{F},z_{\mathrm{o}}}^{-1}(D_{U,D,\mathrm{F},z_{\mathrm{o}}}(A_{\mathrm{o}}))} Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(B)}) : \\ & \quad D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F), \\ & \quad A_{\mathrm{o}} = Z_{z_{\mathrm{o}}} * \left(\sum_{(C,F) \in \mathrm{cont}(D)} \left(Z_{1/|F|} * \sum_{T \in F} A_{\mathrm{o}} * C * T_F * T_F^{\dagger} \right) \right)^{\wedge} \})\end{aligned}$$

In some cases the *maximum likelihood estimate* for the *model* is neither *self* nor *unary*, $\tilde{D}_{\mathrm{o}} \notin \{\{((\emptyset, \{T_{\mathrm{s}}\}), \emptyset)\}, \{((\emptyset, \{T_{\mathrm{u}}\}), \emptyset)\}\}$.

In *classical uniform possible functional definition set decomposition induction*, where (i) the *history probability function* is *uniform possible iso-fud-decomposition historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d,p},D_{\mathrm{o}}}$, given some *unknown substrate fud decomposition* in the *sample variables* $D_{\mathrm{o}} \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\}$, such that there exists a *top transform* for all of the *fuds*, $\forall F \in \mathrm{fuds}(D_{\mathrm{o}})\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$, if it is the case that (ii) the *sample histogram* equals the *fud-decomposition-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D_{\mathrm{o}})}$, then the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-fud-decomposition conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, and, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *maximum likelihood estimate* of the *model*, $\tilde{D}_{\mathrm{o}}$, in the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is

$$\begin{aligned}\tilde{D}_{\mathrm{o}} \ \in\ \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{h,d},D,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) :\\ D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F),\\ A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)}\})\end{aligned}$$

and in some cases the *maximum likelihood estimate* for the *model*, $\tilde{D}_{\mathrm{o}}$, is non-trivial,

$$\tilde{D}_{\mathrm{o}} \ \notin\ \{\{((\emptyset, \{T_{\mathrm{s}}\}), \emptyset)\}, \{((\emptyset, \{T_{\mathrm{u}}\}), \emptyset)\}\}$$

In the case where the *component size cardinality relative entropies* are high,

$$\forall (C, F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (\mathrm{entropyCross}(A_{\mathrm{o}} * C * T_F, C * T_F) > \ln |T_F^{-1}|)$$

the *sum sensitivity* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution-model* pair is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},\tilde{D}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\\ \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\end{aligned}$$

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is approximated

$$\begin{aligned}\tilde{D}_{\mathrm{o}} \ \in\ \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{h,d},D,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) :\\ D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F),\\ A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)}\})\end{aligned}$$

In this case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the properties of the maximisation of the *log likelihood*, $\ln \hat{Q}_{\mathrm{m,d},D_{\mathrm{o}},U}(A_{\mathrm{o}}, z)(A_{\mathrm{o}})$, are consistent with the properties of the minimisation of the *sum sensitivity*, $\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,d},D_{\mathrm{o}},U}(A_{\mathrm{o}}, z)))$. So conjecture that in *classical uniform possible functional definition set decomposition induction*, where the *size* is less than the *volume*, but the *sample* approximates to the *fud-decomposition-independent*, and the *relative entropies* are high, the *sum sensitivity* varies against the *log likelihood*, and the optimisation tends to minimise the *sensitivity* to the *distribution*, $\tilde{E}_{\rm o} = \hat{A}_{\rm o}$,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\rm h,d,\tilde{D}_o,U}(A_{{\rm o},z_{\rm h}}, z_{\rm o}))) \quad \sim \quad -\ln \hat{Q}_{\rm h,d,D_o,U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o})$$

Similarly, the *query sensitivity* to the *distribution*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\rm h,d,\tilde{D}_o,U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})))/z_{\rm o}$$

is also minimised by the optimisation of *log-likelihood.*

5.5.8 Specialising induction

Although the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\rm o}$, is computable for *uniform possible derived induction*, where the *sample histogram* is *natural*, $A_{\rm o} = A_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger}$, the computation is not tractable. The *derived* function, $D_{U,{\rm i},T,z_{\rm o}} \in \mathcal{A}_{U,{\rm i},V_{\rm o},z_{\rm o}} :\to \mathcal{A}_{U,{\rm i},W,z_{\rm o}}$, is intractable because its computation requires the computation of its domain of the *substrate histograms*, $\mathcal{A}_{U,{\rm i},V_{\rm o},z_{\rm o}}$. The *substrate histograms* have cardinality $|\mathcal{A}_{U,{\rm i},V_{\rm o},z_{\rm o}}| = (z_{\rm o} + v_{\rm o} - 1)!/(z_{\rm o}!\ (v_{\rm o} - 1)!)$, which is exponential in the *dimension*, $|V_{\rm o}|$. In addition, in the case for *uniform possible derived induction* the computation of the set of *substrate transforms*, $\mathcal{T}_{U,V_{\rm o}}$, is intractable. The cardinality of the *substrate transforms* is $|\mathcal{T}_{U,V_{\rm o}}| = 2^{\mathrm{bell}(v_{\rm o})}$, which is factorial in the *volume*, $v_{\rm o}$.

The discussion of *specialising induction*, below, considers this issue of *intractability*, firstly by constructing a somewhat artificial *history probability function* explicitly defined by *specialising space*, and then by showing how its *log likelihood* correlates to *tractable induction.* Whereas in *uniform possible derived induction* the *natural log likelihood* is merely anti-correlated to the *specialising space*, here in *specialising induction* the *log likelihood* is strictly proportional to the negative *specialising space*, whether *natural* or not. That is, the *induction* assumptions are amended by replacing the notion that *histories* are *conditionally drawn* from a *distribution history*, with a more explicit assertion of a *degree of structure* with respect to a *specialising coder* for arbitrary *sample substrate variables* and *sample size.*

Consider the *specialising history probability function* $P_{U,X,{\rm G},T_{\rm o},{\rm H}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ which is defined such that the probability of a *history* is inversely proportional to the bounding integer, for which the *space* is the logarithm, of the integer encoding of the *history* in the *specialising coder*, given a *known substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$,

$$\begin{aligned}
&P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}} := \\
&\quad \Big(\bigcup \{\{(H, \exp(-C_{\mathrm{G,T,H}}(T_{\mathrm{o}})^{\mathrm{s}}(H))) : \\
&\qquad\qquad H \in \mathcal{H}_{U,X},\ \mathrm{vars}(H) = V_H,\ |H| = z_H\}^{\wedge} : \\
&\qquad\qquad\qquad V_H \subseteq \mathrm{vars}(U),\ z_H \in \{1 \ldots |X|\}\}\Big)^{\wedge} \cup \\
&\quad \{(\emptyset, 0)\}
\end{aligned}$$

where the *specialising derived substrate history coder* is

$$C_{\mathrm{G},V,\mathrm{T,H}}(T) \ = \ \mathrm{coderHistorySubstrateDerivedSpecialising}(U, X, T, D_{\mathrm{S}}, D_{\mathrm{X}})$$

and the *expanded specialising derived history coder* $C_{\mathrm{G,T,H}}(T) \in \mathrm{coders}(\mathcal{H}_{U,X})$ *expands* the *transform* to the *history variables*, V_H, where the *history variables* is a superset of the *underlying variables*, $V = \mathrm{und}(T)$, and otherwise defaults to an *index coder*,

$$\begin{aligned}
C_{\mathrm{G,T,H}}(T)^{\mathrm{s}}(H) \ = \ & (C_{\mathrm{G},V_H,\mathrm{T,H}}(T^{\mathrm{P}V_H\mathrm{T}})^{\mathrm{s}}(H) + s_{|V_H|} \ : \ V_H \supseteq V) \ + \\
& (C^{\mathrm{s}}_{\mathrm{H}}(H) \ : \ V_H \not\supseteq V)
\end{aligned}$$

where $s_n = \mathrm{spaceVariables}(U)(n)$.

All *non-empty histories* are *possible* in *specialising induction*, $\forall H \in \mathcal{H}_{U,X} \setminus \{\emptyset\}\ (P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}(H) > 0)$.

All *histories* having the same *specialising space* for a given set of *variables* and *size* are defined as equally probable,

$$\begin{aligned}
&\forall H, G \in \mathcal{H}_{U,V,z,X} \\
&\quad (C_{\mathrm{G,T,H}}(T_{\mathrm{o}})^{\mathrm{s}}(G) = C_{\mathrm{G,T,H}}(T_{\mathrm{o}})^{\mathrm{s}}(H) \implies P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}(G) = P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}(H))
\end{aligned}$$

The *specialising history probability function*, $P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, may be compared to a *probability function* $P \in \mathcal{P}$ for which there exists an *entropy coder* $C \in \mathrm{coders}(Y)$. *Entropy coders* need no normalisation, $\forall x \in Y\ (C^{\mathrm{s}}(x) = \ln 1/P_x)$ or $\forall x \in Y\ (P_x = \exp(-C^{\mathrm{s}}(x)))$. The *expected space* of an *entropy coder* is the *entropy*, $\mathrm{expected}(P)(C^{\mathrm{s}}) = \mathrm{entropy}(P)$, so a *derived history coder* $C \in \mathrm{coders}(\mathcal{H}_{U,X})$ that is also an *entropy coder* is maximally *compressing*, $\mathrm{structure}(U, X)(P, C) = 1$. The *specialising history probability function*, $P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, may have less than maximum *degree of structure* because (i) *histories* having *variables* which are not a superset of the *underlying variables*,

$V_{\rm o}$, of the given *transform*, $T_{\rm o}$, are encoded in a *canonical coder*, and (ii) there is a normalisation by *variables* and *size*, (V_H, z_H).

In *specialising induction* there is no *distribution histogram*, so the *drawn history* is parameterised only by *substrate variables* and *size*, (V_H, z_H). Nor is any *sample* constrained to equal its *naturalisation*, $A_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger}$. If a *history* is *possible* in *uniform possible derived induction*, then it is *possible* in *specialising induction*,

$$\forall H \in \mathcal{H}_{U,X}\ (P_{U,X,H_{\rm h},{\rm d},{\rm p},T_{\rm o}}(H) > 0 \implies P_{U,X,{\rm G},T_{\rm o},{\rm H}}(H) > 0)$$

The *specialising space* is the same for all members of an *iso-derived*, $\forall B \in D^{-1}_{U,{\rm i},T,z}(A*T)\ (C_{\rm G,T,H}(T)^{\rm s}(H_B) = C_{\rm G,T,H}(T)^{\rm s}(H_A))$, so all *iso-derived* subsets of the *distribution history* for a given set of *variables* and *size* are not only equally *iso-derived probable*,

$$\begin{aligned}&\forall V \subseteq V_{\rm h}\ \forall H, G \subseteq H_{\rm h}\%V\\ &\qquad (A_G * T_{\rm o} = A_H * T_{\rm o} \implies P_{U,X,H_{\rm h},{\rm d},{\rm p},T_{\rm o}}(G) = P_{U,X,H_{\rm h},{\rm d},{\rm p},T_{\rm o}}(H))\end{aligned}$$

but also equally *specialising probable*,

$$\begin{aligned}&\forall V \subseteq V_{\rm h}\ \forall H, G \subseteq H_{\rm h}\%V\\ &\qquad (A_G * T_{\rm o} = A_H * T_{\rm o} \implies P_{U,X,{\rm G},T_{\rm o},{\rm H}}(G) = P_{U,X,{\rm G},T_{\rm o},{\rm H}}(H))\end{aligned}$$

Given a *history* $H \in \mathcal{H}_{U,X}$, such that $H \neq \emptyset$, the *specialised historical probability* of *histogram* $A_H = \text{histogram}(H) + V_H^{\rm CZ} \in \mathcal{A}_{U,{\rm i},V_H,z_H}$ is

$$\hat{Q}_{{\rm G},T_{\rm o},{\rm H},U}(z_H)(A_H) \ \propto\ \sum(P_{U,X,{\rm G},T_{\rm o},{\rm H}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *specialising probability distribution* is defined

$$\begin{aligned}&\hat{Q}_{{\rm G},T,{\rm H},U}(z)\ :=\\ &\qquad \{(A, \frac{z!}{\prod_{S \in A^{\rm FS}} A_S!} \times \exp(-C_{\rm G,T,H}(T)^{\rm s}(H_A))) : A \in \mathcal{A}_{U,{\rm i},V,z}\}^{\wedge}\end{aligned}$$

where $V = \text{und}(T)$ and $H_A = \text{history}(A)$.

The *log likelihood* is proportional to the *classification space* of the *underlying histogram* less the *specialising space* of the corresponding *history*,

$$\ln \hat{Q}_{{\rm G},T,{\rm H},U}(z)(A) \ \propto\ \text{spaceClassification}(A)\ -\ \text{space}(C_{{\rm G},V,{\rm T},{\rm H}}(T))(H_A)$$

The *space* of the *specialising coder* is

$$\begin{aligned}
\mathrm{space}(C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T))(H) &= \mathrm{spaceIds}(|X|,|H|) + \\
&\quad \mathrm{spaceCountsDerived}(U)(A,T) + \\
&\quad \mathrm{spaceClassification}(A * T) + \\
&\quad \mathrm{spaceEventsPartition}(A,T) \\
&= \mathrm{spaceIds}(|X|,|H|) + \\
&\quad \ln \frac{(z+w'-1)!}{z!\ (w'-1)!} + \\
&\quad \ln z! - \sum_{R \in (A*T)^{\mathrm{S}}} \ln (A*T)_R! + \\
&\quad \sum_{(R,C) \in T^{-1}} (A*T)_R \ln |C|
\end{aligned}$$

The *space* of the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)$, varies (i) with the *possible derived volume*, w', where the *possible derived volume* is less than the *size*, $w' < z$, otherwise with the *size* scaled *log possible derived volume*, $z \ln w'$, and (ii) against the *size* scaled *component size cardinality relative entropy*,

$$\begin{aligned}
&C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) \sim \\
&\qquad (w' \ :\ w' < z) \ +\ (z \ln w' \ :\ w' \geq z) \\
&\qquad - z \times \mathrm{entropyRelative}(A * T, V^{\mathrm{C}} * T)
\end{aligned}$$

The *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H)$, varies (i) with twice the *possible derived volume*, $2w'$, where $w' < z$, otherwise with twice the *size* scaled *log possible derived volume*, $2z \ln w'$, (ii) with the *size* scaled *derived entropy*, (iii) against twice the *size* scaled *component size cardinality cross entropy* and (iv) against the *size* scaled *size expected component entropy*,

$$\begin{aligned}
&2C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) - C^{\mathrm{s}}_{\mathrm{H},V}(H) - C^{\mathrm{s}}_{\mathrm{G},V}(H) \sim \\
&\qquad 2\big((w' \ :\ w' < z) \ +\ (z \ln w' \ :\ w' \geq z)\big) \\
&\qquad + z \times \mathrm{entropy}(A * T) \\
&\qquad - 2z \times \mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) \\
&\qquad - z \times \mathrm{entropyComponent}(A, T)
\end{aligned}$$

The *canonical* term, $C^{\mathrm{s}}_{\mathrm{H},V}(H) + C^{\mathrm{s}}_{\mathrm{G},V}(H)$, is independent of the *model*, T, so properties of the *specialising-canonical space difference*, $2C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H) -$

$C^{\rm s}_{{\rm H},V}(H) - C^{\rm s}_{{\rm G},V}(H)$, are also properties of the *specialising space*, $C_{{\rm G},V,{\rm T},{\rm H}}(T)^{\rm s}(H)$.

The *specialising log likelihood* varies (i) against twice the *possible derived volume*, $2w'$, where $w' < z$, otherwise against twice the *size* scaled *log possible derived volume*, $2z \ln w'$, (ii) with the *size* scaled *underlying entropy*, (iii) against the *size* scaled *derived entropy*, (iv) with twice the *size* scaled *component size cardinality cross entropy* and (v) with the *size* scaled *size expected component entropy*,

$$\begin{aligned}
\ln \hat{Q}_{{\rm G},T,{\rm H},U}(z)(A) \ \sim \ & \\
& -\ 2\big((w' \ : \ w' < z) \ + \ (z \ln w' \ : \ w' \geq z)\big) \\
& +\ z \times \mathrm{entropy}(A) \\
& -\ z \times \mathrm{entropy}(A * T) \\
& +\ 2z \times \mathrm{entropyCross}(A * T, V^{\rm C} * T) \\
& +\ z \times \mathrm{entropyComponent}(A, T)
\end{aligned}$$

Let $H_{\rm o}$ be a *sample history* of *known size* $z_{\rm o} = |H_{\rm o}| > 0$ in the *known sample variables*, $V_{\rm o}$, which has a *known histogram* $A_{\rm o} = \mathrm{histogram}(H_{\rm o}) + V_{\rm o}^{\rm CZ} \in \mathcal{A}_{U,{\rm i},V_{\rm o},z_{\rm o}}$. In *classical specialising induction*, where the *history probability function* is the *specialising history probability function*, $P = P_{U,X,{\rm G},T_{\rm o},{\rm H}}$, given some *substrate transform* in the *sample variables* $T_{\rm o} \in \mathcal{T}_{U,V_{\rm o}}$, the *log likelihood* of the *specialising probability distribution* (a) varies with the *size* scaled *underlying entropy*,

$$\ln \hat{Q}_{{\rm G},T_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o}) \ \sim \ z_{\rm o} \times \mathrm{entropy}(A_{\rm o})$$

(b) varies against the *possible derived volume* where $w'_{\rm o} < z_{\rm o}$, otherwise against the *size* scaled *log possible derived volume*, $z_{\rm o} \ln w'_{\rm o}$

$$\ln \hat{Q}_{{\rm G},T_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o}) \ \sim \ -\big((w'_{\rm o} \ : \ w'_{\rm o} < z_{\rm o}) \ + \ (z_{\rm o} \ln w'_{\rm o} \ : \ w'_{\rm o} \geq z_{\rm o})\big)$$

(c) varies against the *size* scaled *derived entropy*

$$\ln \hat{Q}_{{\rm G},T_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o}) \ \sim \ -z_{\rm o} \times \mathrm{entropy}(A_{\rm o} * T_{\rm o})$$

(d) varies with the *size* scaled *component size cardinality cross entropy*

$$\ln \hat{Q}_{{\rm G},T_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o}) \ \sim \ z_{\rm o} \times \mathrm{entropyCross}(A_{\rm o} * T_{\rm o}, V_{\rm o}^{\rm C} * T_{\rm o})$$

and (e) varies with the *size* scaled *size expected component entropy*,

$$\ln \hat{Q}_{{\rm G},T_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o}) \ \sim \ z_{\rm o} \times \mathrm{entropyComponent}(A_{\rm o}, T_{\rm o})$$

So the *specialising log likelihood* is maximised when (a) the *possible derived volume* is minimised, (b) the *underlying entropy* is maximised, (c) the *derived entropy* is minimised, (d) high *size components* are low *cardinality components* and low *size components* are high *cardinality components*, and (e) the *expected component entropy* is maximised.

In the case where the *model* is *unary*, $T_{\mathrm{o}} = T_{\mathrm{u}}$, (i) the *derived entropy* is zero, $\mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{u}}) = 0$, (ii) the *cross entropy* is zero, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{u}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{u}}) = 0$, so (iii) the *relative entropy* is zero. The *size expected component entropy* equals the *underlying entropy*. The *log likelihood* of the *specialising probability distribution* only varies with the *size* scaled *underlying entropy*,

$$\ln \hat{Q}_{\mathrm{G},T_{\mathrm{u}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}})$$

In the case where the *model* is *self*, $T_{\mathrm{o}} = T_{\mathrm{s}}$, (i) the *derived entropy* equals the *underlying entropy*, $\mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{s}}) = \mathrm{entropy}(A_{\mathrm{o}})$, (ii) the *cross entropy* equals the *volume space*, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{s}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{s}}) = \ln v_{\mathrm{o}}$, so (iii) the *relative entropy* varies against the *underlying entropy*. The *size expected component entropy* is zero. The *log likelihood* of the *specialising probability distribution* only varies against the *underlying volume* where $v_{\mathrm{o}} < z_{\mathrm{o}}$, otherwise against the *size* scaled *log underlying volume*, $z_{\mathrm{o}} \ln v_{\mathrm{o}}$

$$\ln \hat{Q}_{\mathrm{G},T_{\mathrm{s}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -((v_{\mathrm{o}} \ : \ v_{\mathrm{o}} < z_{\mathrm{o}}) \ + \ (z_{\mathrm{o}} \ln v_{\mathrm{o}} \ : \ v_{\mathrm{o}} \geq z_{\mathrm{o}}))$$

Although the *specialising history probability function*, $P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, is not derived from a conditional *draw* from a *distribution history*, it does have a physical analogy in isolated thermodynamic systems.

Let $Y \subset \mathcal{X}$ be a set of unweighted micro-states. Consider a system of n distinguishable particles each in a micro-state. The states of the system is the set of micro-state functions of particle identifier, $\{1 \ldots n\} :\to Y$. Each state implies a distribution of particles over micro-states,

$$I = \{(R, \{(x, |C|) : (x, C) \in R^{-1}\}) : R \in \{1 \ldots n\} :\to Y\}$$

so the cardinality of states for each distribution is

$$W = \{(N, |D|) : (N, D) \in I^{-1}\} = \{(N, \frac{n!}{\prod_{(x,\cdot) \in N} N_x!}) : (N, \cdot) \in I^{-1}\}$$

Let E be an energy valued function of micro-state, $E \in Y :\to \mathbf{R}_{\geq 0}$. Consider the subset of the states that have total energy ϵ,

$$\{R : R \in \{1 \ldots n\} :\to Y, \ \sum_{(\cdot,x) \in R} E_x = \epsilon\}$$

If a state is chosen at random from this subset, the modal distribution $N_{n,E,\epsilon}$ has the greatest cardinality,

$$\begin{aligned} N_{n,E,\epsilon} \;\in\; \text{maxd}(\{(N, \frac{n!}{\prod_{(x,\cdot)\in N} N_x!}) : N \in Y \to \{1 \ldots n\}, \\ \sum_{(x,\cdot)\in N} N_x = n,\; \sum_{(x,\cdot)\in N} N_x E_x = \epsilon\}) \end{aligned}$$

The states of the modal distribution, $I^{-1}(N_{n,E,\epsilon}) \subseteq \{1 \ldots n\} :\to Y$, are said to be at thermodynamic equilibrium. The thermodynamic entropy of these states is $S_{n,E,\epsilon} = k \ln W(N_{n,E,\epsilon})$, where k is the Boltzmann constant. The logarithm of the multinomial coefficient approximates to the scaled entropy, so the modal distribution probability function $P_{n,E,\epsilon}$ is

$$\begin{aligned} P_{n,E,\epsilon} \;\in\; \text{maxd}(\{(P, \text{entropy}(P)) : P \in Y :\to \mathbf{R}_{[0,1]},\; \sum_{x\in Y} P_x = 1, \\ n \sum_{x\in Y} P_x E_x = \epsilon\}) \end{aligned}$$

The solution for this probability function is the Boltzmann distribution,

$$\begin{aligned} P_{n,E,\epsilon} &= \{(x, \frac{\exp(-E_x/k\tau_{n,E,\epsilon})}{\sum_{y\in Y} \exp(-E_y/k\tau_{n,E,\epsilon})}) : x \in Y\} \\ &= \{(x, \exp(-E_x/k\tau_{n,E,\epsilon})) : x \in Y\}^{\wedge} \end{aligned}$$

where $\tau_{n,E,\epsilon}$ is the temperature. The inverted temperature is the sensitivity of the equilibrium entropy to energy, $1/\tau_{n,E,\epsilon} = \partial S_{n,E,\epsilon}/\partial\epsilon$. The thermodynamic entropy at equilibrium varies with the probability distribution entropy, $S_{n,E,\epsilon} \sim nk \times \text{entropy}(P_{n,E,\epsilon})$. The Boltzmann distribution is the solution that maximises the entropy given the energy.

Conversely, let the distribution probability function $P_{n,s,E}$ be the probability function that minimises the energy given the entropy,

$$\begin{aligned} P_{n,s,E} \;\in\; \text{mind}(\{(P, \sum_{x\in Y} P_x E_x) : P \in Y :\to \mathbf{R}_{[0,1]},\; \sum_{x\in Y} P_x = 1, \\ nk \times \text{entropy}(P) = s\}) \end{aligned}$$

Now the temperature is the sensitivity of the equilibrium energy to entropy, $\tau_{n,s,E} = \partial\epsilon_{n,s,E}/\partial s$. In the case where the temperature is positive, so that the energy is monotonic with respect to entropy at equilibrium, these optimisations intersect, $P_{n,s,E} = P_{n,E,\epsilon}$, where $s = \text{entropy}(P_{n,E,\epsilon})$. That is, the

Boltzmann distribution is also the solution that minimises the energy given the entropy.

Together the Boltzmann distribution is the solution that minimises the ratio of (i) the energy to (ii) the temperature times the entropy at equilibrium,

$$\frac{\epsilon}{\tau_{n,E,\epsilon} \times S_{n,E,\epsilon}} = \frac{\epsilon_{n,s,E}}{\tau_{n,s,E} \times s}$$

Consider the probability of a thermodynamic particle being in micro-state x_2 given it is in either micro-state x_1 or x_2,

$$\frac{P_{n,E,\epsilon}(x_2)}{P_{n,E,\epsilon}(x_1) + P_{n,E,\epsilon}(x_2)} = \frac{1}{1 + \exp((E_{x_2} - E_{x_1})/k\tau_{n,E,\epsilon})}$$

As the temperature increases, the probability that the particle will be in the higher energy micro-state increases to a half. So both the entropy at equilibrium, $S_{n,E,\epsilon}$, and the energy at equilibrium, $\epsilon_{n,s,E}$, vary with temperature, $\tau_{n,E,\epsilon} = \tau_{n,s,E}$.

Consider treating the micro-state energy divided by the Boltzmann constant, E_x/k, as a continuous random variable. Let $P_{n,E,\epsilon,\lambda} = \{(u, \lambda \exp(-\lambda u)) : u \in \mathbf{R}_{\geq 0}\} \in \mathbf{R}_{\geq 0} :\to \mathbf{R}_{\geq 0}$ be the exponential distribution parameterised only by λ that is the nearest fit to the Boltzmann distribution, $\forall x \in Y\ (P_{n,E,\epsilon,\lambda}(E_x/k) \approx P_{n,E,\epsilon}(x))$. That is,

$$\forall x \in Y \left(\lambda \exp(-\lambda E_x/k) \approx \frac{\exp(-E_x/k\tau_{n,E,\epsilon})}{\sum_{y \in Y} \exp(-E_y/k\tau_{n,E,\epsilon})}\right)$$

For higher temperatures, $1/\lambda$ is between $\sum_{y \in Y} \exp(-E_y/k\tau_{n,E,\epsilon})$ and $\tau_{n,E,\epsilon}$. The variance of the exponential distribution, $P_{n,E,\epsilon,\lambda}$, is $1/\lambda^2$. Let the corresponding likelihood function be $L_{n,E,\epsilon}(u) = \{(\lambda, \lambda \exp(-\lambda u)) : \lambda \in \mathbf{R}_{\geq 0}\} \in \mathbf{R}_{\geq 0} :\to \mathbf{R}_{\geq 0}$. The exponential distribution is not centrally distributed, so the curvature of the likelihood function at the mode, $\partial^2(L_{n,E,\epsilon}(0))(\lambda)$, is not negative, and the Fisher information is not a measure of the sensitivity to parameter. Instead its sensitivity to parameter may be approximated by the negative gradient of the likelihood function at the mode, $-\partial(L_{n,E,\epsilon}(0))(\lambda) = \lambda^2 - 1$, which varies oppositely to the variance. That is, the sensitivity to parameter varies against up to the *temperature* squared,

$$\begin{aligned} -\partial(L_{n,E,\epsilon}(0))(\lambda) &= \lambda^2 - 1 \\ &\sim 1/\tau_{n,E,\epsilon}^2 \end{aligned}$$

Given a set of *variables* $V = \text{und}(T)$ and a *size* z, the *specialising history probability function* $P_{U,X,\text{G},T,\text{H},z}$ is defined

$$P_{U,X,\text{G},T,\text{H},z} \quad := \quad \{(H, \exp(-C_{\text{G},V,\text{T},\text{H}}(T)^{\text{s}}(H))) : H \in \mathcal{H}_{U,V,z,X}\}^{\wedge}$$

where $\mathcal{H}_{U,V,z,X} = \{H : H \in \mathcal{H}_{U,X},\ \text{vars}(H) = V,\ |H| = z\}$.

Mapping the *specialising history probability function*, $P_{U,X,\text{G},T,\text{H},z}$, to the Boltzmann distribution, $P_{n,E,\epsilon}$, implies that the energy of the micro-state, E_x, is proportional to the *specialising space* of the *history*, $C_{\text{G},V,\text{T},\text{H}}(T)^{\text{s}}(H)$. The *thermodynamic energy* $\epsilon_{n,U,X,T,z}$ is proportional to the *thermodynamic temperature* $\tau_{n,U,X,T,z}$ times the *expected specialising space*,

$$\epsilon_{n,U,X,T,z} \quad = \quad nk \times \tau_{n,U,X,T,z} \times \text{expected}(P_{U,X,\text{G},T,\text{H},z})(C_{\text{G},V,\text{T},\text{H}}(T)^{\text{s}})$$

The *thermodynamic entropy* at equilibrium $S_{n,U,X,T,z}$ is proportional to the entropy of the *specialising history probability function*,

$$S_{n,U,X,T,z} \quad \sim \quad nk \times \text{entropy}(P_{U,X,\text{G},T,\text{H},z})$$

The *specialising history probability function* is such that the *thermodynamic entropy*, $S_{n,U,X,T,z}$, is maximised for given *thermodynamic energy*, $\epsilon_{n,U,X,T,z}$,

$$\begin{aligned} &P_{U,X,\text{G},T,\text{H},z} \in \\ &\quad \text{maxd}(\{(P, nk \times \text{entropy}(P)) : P \in (\mathcal{H}_{U,V,z,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}, \\ &\qquad\qquad nk\tau_{n,U,X,T,z} \times \text{expected}(P)(C_{\text{G},V,\text{T},\text{H}}(T)^{\text{s}}) = \epsilon_{n,U,X,T,z}\}) \end{aligned}$$

and also such that the *thermodynamic energy*, $\epsilon_{n,U,X,T,z}$, is minimised for given *thermodynamic entropy*, $S_{n,U,X,T,z}$,

$$\begin{aligned} &P_{U,X,\text{G},T,\text{H},z} \in \\ &\quad \text{mind}(\{(P, nk\tau_{n,U,X,T,z} \times \text{expected}(P)(C_{\text{G},V,\text{T},\text{H}}(T)^{\text{s}})) : \\ &\qquad\qquad P \in (\mathcal{H}_{U,V,z,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P},\ nk \times \text{entropy}(P) = S_{n,U,X,T,z}\}) \end{aligned}$$

That is, *specialising induction* is balanced such that the total *specialising space* is minimised while the *specialising entropy* is maximised.

Physical thermodynamic systems usually define a topology or measure on the states of the system implied by the interactions between particles that conserve energy, so that dynamic processes starting in low entropy states generally evolve towards neighbouring high entropy states. *Specialising induction*, however, does not have parallels for these physical properties.

Note that, conversely, the mapping of the energy of the micro-state, E_x, to the *specialising space* of the *history*, $C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H)$, suggests that the micro-state energy, E_x, of any thermodynamical system can be viewed as the logarithm of the bounding integer of the integer encoding of the micro-state in some *coder* $C \in \mathrm{coders}(Y)$.

While *classical specialising induction*, $P = P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, is completely defined given a *system*, (U,X), and a *model*, T_{o}, the definition of *classical derived induction*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}$, also requires a *distribution history*, $H_{\mathrm{h}} \in \mathcal{H}_{U,X}$. In *derived induction*, the *sample history* is *drawn* from the *distribution history*,

$$H_{\mathrm{o}} \in \{H : H \in \mathrm{P}(H_{\mathrm{h}}\%V_{\mathrm{o}}),\ |H| = z_{\mathrm{o}},\ \hat{A}_H * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}\}$$

An analogy to the *distribution history*, H_{h}, is implied by *specialising induction*. Let $R_{\mathrm{G,H}}$ be a *thermodynamic state*, $R_{\mathrm{G,H}} \in \{1 \ldots n\} :\to (\mathcal{H}_{U,X} \setminus \{\emptyset\})$. The *thermodynamic state*, $R_{\mathrm{G,H}}$, is at *equilibrium* for arbitrary *draw*, $(V_{\mathrm{o}}, z_{\mathrm{o}})$, so $\{(H, |C|/n) : (H,C) \in R_{\mathrm{G,H}}^{-1}\} = P_{U,X,\mathrm{G},T_{\mathrm{o}}} \setminus \{(\emptyset, 0)\}$. Now the *sample history* is the *history* of a *particle* in the *thermodynamic state*,

$$H_{\mathrm{o}} \in \{H : (\cdot, H) \in R_{\mathrm{G,H}},\ V_H = V_{\mathrm{o}},\ |H| = z_{\mathrm{o}}\}$$

The ratio of the *thermodynamic energy* to the *thermodynamic temperature* times the *thermodynamic entropy* at equilibrium approximates to the ratio of the *expected specialising space* to the *specialising entropy*,

$$\frac{\epsilon_{n,U,X,T,z}}{\tau_{n,U,X,T,z} \times S_{n,U,X,T,z}} \sim \frac{\mathrm{expected}(P_{U,X,\mathrm{G},T,\mathrm{H},z})(C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}})}{\mathrm{entropy}(P_{U,X,\mathrm{G},T,\mathrm{H},z})}$$

This ratio is minimised at equilibrium. As the ratio tends to one, the *specialising coder* tends to an *entropy coder* and so the *degree of structure* tends to one. The *specialising structure* over all *variables* and *sizes* varies (i) against *thermodynamic energy*, (ii) with *thermodynamic temperature* and (iii) with *thermodynamic entropy*,

$$\begin{aligned}
&\mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}}, C_{\mathrm{G,T,H}}(T)) \\
&\quad := \frac{\mathrm{canonical}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}}) - \mathrm{expected}(P_{U,X,\mathrm{G},T,\mathrm{H}})(C_{\mathrm{G,T,H}}(T)^{\mathrm{s}})}{\mathrm{canonical}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}}) - \mathrm{entropy}(P_{U,X,\mathrm{G},T,\mathrm{H}})} \\
&\quad \sim \frac{\mathrm{canonical}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}}) - \epsilon_{n,U,X,T}/(nk \times \tau_{n,U,X,T})}{\mathrm{canonical}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}}) - S_{n,U,X,T}/(nk)}
\end{aligned}$$

In the case where the *model* is *unary*, $T_{\mathrm{o}} = T_{\mathrm{u}}$, the *expanded specialising derived substrate history space* equals the *index space*, $C_{\mathrm{G,T,H}}(T_{\mathrm{u}})^{\mathrm{s}}(H) =$

$C^{\mathrm{s}}_{\mathrm{H}}(H)$, so the *degree of structure* is zero or negative,

$$\mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T_{\mathrm{u}},\mathrm{H}}, C_{\mathrm{G,T,H}}(T_{\mathrm{u}})) \leq 0$$

In the case where the *model* is *self*, $T_{\mathrm{o}} = T_{\mathrm{s}}$, the *expanded specialising derived substrate history space* is at least the *index space* or the *classification space*, $C_{\mathrm{G,T,H}}(T_{\mathrm{s}})^{\mathrm{s}}(H) \geq \mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}(H), C^{\mathrm{s}}_{\mathrm{G}}(H))$, so the *degree of structure* is zero or negative,

$$\mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T_{\mathrm{s}},\mathrm{H}}, C_{\mathrm{G,T,H}}(T_{\mathrm{s}})) \leq 0$$

The *specialising history probability function*, $P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, is the *history probability function* P that maximises the *degree of structure* with respect to the *expanded specialising derived history coder*, $C_{\mathrm{G,T,H}}(T_{\mathrm{o}})$, given arbitrary *draw*, (V_H, z_H),

$$\begin{aligned}
&P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}} \in \\
&\quad \mathrm{maxd}(\{(P, \mathrm{structure}(U,X)(P, C_{\mathrm{G,T,H}}(T_{\mathrm{o}}))) : \\
&\qquad P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}, \\
&\qquad\quad |\{(\textstyle\sum P_H : H \in \mathcal{H}_{U,X},\ \mathrm{vars}(H) = V_H,\ |H| = z_H) : \\
&\qquad\qquad\qquad V_H \subseteq \mathrm{vars}(U),\ z_H \in \{1 \ldots |X|\}\}| = 1, \\
&\qquad\quad P_{\emptyset} = 0\})
\end{aligned}$$

The *model*, T, may be mapped to the continuous parameter of *temperature*, $\tau_{n,U,X,T}$, by choosing the micro-states such that the energy divided by the Boltzmann constant, E_x/k, approximates to the *canonical space*,

$$\forall H \in \mathcal{H}_{U,V,z,X}\ (C_{\mathrm{G},V,\mathrm{T,H}}(T)^{\mathrm{s}}(H) \approx \mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H},V}, C^{\mathrm{s}}_{\mathrm{G},V})(H)/\tau_{n,U,X,T,z})$$

Insofar as the approximation holds, the micro-states do not depend on the *model*, T, and so all of the dependency on *model* is encapsulated by the *temperature*, $\tau_{n,U,X,T}$. Under this mapping, the *temperature* is the ratio of the *canonical space* to the *expected specialising space*

$$\tau_{n,U,X,T} = \frac{\mathrm{canonical}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}})}{\mathrm{expected}(P_{U,X,\mathrm{G},T,\mathrm{H}})(C_{\mathrm{G,T,H}}(T)^{\mathrm{s}})}$$

and the *thermodynamic energy* $\epsilon_{n,U,X,T}$ is proportional to the *canonical space*,

$$\epsilon_{n,U,X,T} = nk \times \mathrm{canonical}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}})$$

Given a *modal history* $H \in \mathrm{maxd}(P_{U,X,\mathrm{G},T,\mathrm{H},z})$, the *temperature* varies with the *component size cardinality relative entropy*,

$$\tau_{n,U,X,T,z} \sim \mathrm{entropyRelative}(A_H * T, V^{\mathrm{C}} * T)$$

That is, at high *temperatures* the *modal history*, H, is such that high *size components* are low *cardinality components* and low *size components* are high *cardinality components*.

Formally, under the mapping such that the *thermodynamic energy* is proportional to the *canonical space*,

$$\epsilon_{n,U,X,T_\mathrm{o}} \quad = \quad nk \times \mathrm{canonical}(U,X)(P_{U,X,\mathrm{G},T_\mathrm{o},\mathrm{H}})$$

the *specialising structure* varies principally with *thermodynamic temperature*,

$$\mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T_\mathrm{o},\mathrm{H}}, C_{\mathrm{G,T,H}}(T_\mathrm{o})) \quad \sim \quad \tau_{n,U,X,T_\mathrm{o}}$$

The exponential distribution function, continuously parameterised only by λ_o, which is the best fit to the Boltzmann distribution is

$$P_{n,U,X,T_\mathrm{o},\lambda_\mathrm{o}} = \{(u, \lambda_\mathrm{o} \exp(-\lambda_\mathrm{o} u)) : u \in \mathbf{R}_{\geq 0}\} \in \mathbf{R}_{\geq 0} :\to \mathbf{R}_{\geq 0}$$

Its corresponding likelihood function at the mode is $L_{n,U,X,T_\mathrm{o}}(0) \in \mathbf{R}_{\geq 0} :\to \mathbf{R}_{\geq 0}$. The sensitivity to parameter is defined as the negative gradient of the likelihood function at the mode, $-\partial(L_{n,U,X,T_\mathrm{o}}(0))(\lambda_\mathrm{o}) = \lambda_\mathrm{o}^2 - 1$, which varies against up to the *temperature* squared, $-\partial(L_{n,U,X,T_\mathrm{o}}(0))(\lambda_\mathrm{o}) \sim 1/\tau_{n,U,X,T_\mathrm{o}}^2$. The *specialising structure* varies with *thermodynamic temperature*,

$$\mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T_\mathrm{o},\mathrm{H}}, C_{\mathrm{G,T,H}}(T_\mathrm{o})) \quad \sim \quad \tau_{n,U,X,T_\mathrm{o}}$$

so the sensitivity to parameter varies against the *specialising structure*,

$$-\partial(L_{n,U,X,T_\mathrm{o}}(0))(\lambda_\mathrm{o}) \quad \sim \quad -\ \mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T_\mathrm{o},\mathrm{H}}, C_{\mathrm{G,T,H}}(T_\mathrm{o}))$$

Although the *specialising entropy*, $\mathrm{entropy}(P_{U,X,\mathrm{G},T,\mathrm{H}})$, is maximised and the *expected specialising space*, $\mathrm{expected}(P_{U,X,\mathrm{G},T,\mathrm{H}})(C_{\mathrm{G,T,H}}(T)^\mathrm{s})$, is minimised but sometimes not minimal, the *modal specialising space*, $C_{\mathrm{G,T,H}}(T)^\mathrm{s}(H)$, of a *modal history* $H \in \mathrm{maxd}(P_{U,X,\mathrm{G},T,\mathrm{H}})$, is always minimal.

At high *thermodynamic temperatures* the sensitivity of *thermodynamic entropy* to *thermodynamic energy* is low, $1/\tau_{n,U,X,T} = \partial S_{n,U,X,T}/\partial \epsilon_{n,U,X,T} \approx 0$, and the *degree of structure* varies proportionately against the *expected specialising space*,

$$\mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}}, C_{\mathrm{G,T,H}}(T)) \quad \sim \quad -\ \mathrm{expected}(P_{U,X,\mathrm{G},T,\mathrm{H}})(C_{\mathrm{G,T,H}}(T)^\mathrm{s})$$

Let H be a *modal history*, $H \in \mathrm{maxd}(P_{U,X,\mathrm{G},T,\mathrm{H}})$. At high *thermodynamic temperatures*, as the *mean specialising space*, $\mathrm{expected}(P_{U,X,\mathrm{G},T,\mathrm{H}})(C_{\mathrm{G,T,H}}(T)^\mathrm{s})$, decreases to the *modal specialising space*, $C_{\mathrm{G,T,H}}(T)^{\mathrm{s}}(H)$, the *degree of structure* tends to maximal, $\mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}}, C_{\mathrm{G,T,H}}(T)) \approx 1$. Conjecture that as the *mean specialising space* decreases, for constant *specialising entropy*, the logarithm of the *modal probability*, or *log likelihood*, increases,

$$\ln P_{U,X,\mathrm{G},T,\mathrm{H}}(H) \quad \sim \quad -\ \mathrm{expected}(P_{U,X,\mathrm{G},T,\mathrm{H}})(C_{\mathrm{G,T,H}}(T)^{\mathrm{s}})$$

so

$$\ln P_{U,X,\mathrm{G},T,\mathrm{H}}(H) \quad \sim \quad \mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T,\mathrm{H}}, C_{\mathrm{G,T,H}}(T))$$

Conjecture further that this correlation is always positive regardless of *thermodynamic temperature*. Formally, in *classical specialising induction*, where (i) the *history probability function* is the *specialising history probability function*, $P = P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, and (ii) the *sample history* is *modal*, $H_{\mathrm{o}} \in \mathrm{maxd}(P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}})$, the *log likelihood* of the *specialising probability distribution* varies with the *degree of structure*,

$$\ln \hat{Q}_{\mathrm{G},T_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}, C_{\mathrm{G,T,H}}(T_{\mathrm{o}}))$$

Note that the *degree of structure* is a property only of the *system* and *model*, not of the *sample*. The *sample* is itself implied by the *system* and *model* because it is *modal*.

In the case where the *model*, T_{o}, is *known*, '*likelihood*' is an abuse of terminology because there is no *distribution histogram* nor other *unknown* parameterisation of the *probability function*, $\hat{Q}_{\mathrm{G},T_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}}) \in \mathcal{P}$.

In the case, however, where the *model*, T_{o}, is *unknown*, the *maximum likelihood estimate* $\tilde{T}_{\mathrm{o}}$ can be defined as an optimisation of the *specialising probability* given the *sample*,

$$\tilde{T}_{\mathrm{o}} \ \in \ \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

That is, the *probability function* is now parameterised by the *unknown model*, T_{o}, so a *model* is the argument to the corresponding *likelihood function* parameterised by the *sample histogram*, A_{o}.

In *specialising induction*, which is such that the *history probability function* is $P = P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, given some *unknown substrate transform* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, occurs at the minimisation of the *specialising space* of the *sample history*,

$$\tilde{T}_{\mathrm{o}} \ \in \ \mathrm{mind}(\{(T, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T)^{\mathrm{s}}(H_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

The *maximum likelihood estimate* for the *model*, $\tilde{T}_{\rm o}$, is non-trivial, $\tilde{T}_{\rm o} \notin \{T_{\rm s}, T_{\rm u}\}$, if there exists a *model* for which the *specialising derived substrate history space* is less than either the *index space* or the *classification space*, $C_{\rm G,V_o,T,H}(\tilde{T}_{\rm o})^{\rm s}(H_{\rm o}) < \text{minimum}(C^{\rm s}_{{\rm H},V_{\rm o}}(H_{\rm o}), C^{\rm s}_{{\rm G},V_{\rm o}}(H_{\rm o}))$. This the case if the *degree of structure* is greater than zero,

$$\text{structure}(U,X)(P_{U,X,{\rm G},T_{\rm o},{\rm H}}, C_{\rm G,T,H}(\tilde{T}_{\rm o})) \quad > \quad 0$$

and the *component size cardinality relative entropy* of the modal *sample* is greater than zero,

$$\text{entropyRelative}(A_{\rm o} * \tilde{T}_{\rm o}, V_{\rm o}^{\rm C} * \tilde{T}_{\rm o}) \quad > \quad 0$$

An example of a non-trivial *model* is where the *histogram* is *sparse*, $\text{trim}(A_{\rm o}) = A_{\rm o}^{\rm F}$ or $A_{\rm o} * A_{\rm o}^{\rm F} = A_{\rm o}^{\rm F}$. In this case, the *under-fitted effective binary transform* is the solution to the optimisation,

$$A_{\rm o} * A_{\rm o}^{\rm F} = A_{\rm o}^{\rm F} \quad \implies \quad \tilde{T}_{\rm o} = \{A_{\rm o}^{\rm FS},\ V_{\rm o}^{\rm CS} \setminus A_{\rm o}^{\rm FS}\}^{\rm T}$$

The *effective binary transform* has a *component* for *effective states*, $A_{\rm o}^{\rm FS}$, and a remainder *component* for the *ineffective states*, $V_{\rm o}^{\rm CS} \setminus A_{\rm o}^{\rm FS}$. The *derived volume* is 2.

The *sensitivity* to *model* is defined as the negative logarithm of the cardinality of the *maximum likelihood estimate models*,

$$-\ \ln |\text{max}(\{(T, \hat{Q}_{{\rm G},T,{\rm H},U}(z_{\rm o})(A_{\rm o})) : T \in \mathcal{T}_{U,{\rm V_o}}\})|$$

Conjecture that the cardinality of the modes varies against the negative gradient of the likelihood function of the singly parameterised exponential function fitted to the Boltzmann distribution, so the *sensitivity* to *model* varies with the negative gradient,

$$-\ \ln |\text{max}(\{(T, \hat{Q}_{{\rm G},T,{\rm H},U}(z_{\rm o})(A_{\rm o})) : T \in \mathcal{T}_{U,{\rm V_o}}\})| \quad \sim \quad -\partial(L_{n,U,X,T_{\rm o}}(0))(\lambda_{\rm o})$$

Hence, given the *canonical* mapping, the *sensitivity* to *model* varies against the *specialising structure*,

$$-\ \ln |\text{max}(\{(T, \hat{Q}_{{\rm G},T,{\rm H},U}(z_{\rm o})(A_{\rm o})) : T \in \mathcal{T}_{U,{\rm V_o}}\})| \ \sim \\ -\ \text{structure}(U,X)(P_{U,X,{\rm G},T_{\rm o},{\rm H}}, C_{\rm G,T,H}(T_{\rm o}))$$

As shown above, the *degree of structure* varies with the *log likelihood* of the *specialising probability distribution*,

$$\text{structure}(U,X)(P_{U,X,{\rm G},T_{\rm o},{\rm H}}, C_{\rm G,T,H}(T_{\rm o})) \quad \sim \quad \ln \hat{Q}_{{\rm G},T_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o})$$

and so the *sensitivity* to *model* varies against the *log likelihood*,

$$- \ \ln |\max(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})| \quad \sim \quad - \ \ln \hat{Q}_{\mathrm{G},T_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})$$

That is, maximisation of the *log likelihood* also tends to minimise the *sensitivity* to *model*.

In the cases where there exists a modal *transform* which is either *self* or *unary*, $\tilde{T}_{\mathrm{o}} \in \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$, the *degree of structure* is negative and the modal *specialising space* is high, $C_{\mathrm{G,V_o,T,H}}(\tilde{T}_{\mathrm{o}})^{\mathrm{s}}(H_{\mathrm{o}}) \geq \mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H},V_{\mathrm{o}}}(H_{\mathrm{o}}), C^{\mathrm{s}}_{\mathrm{G},V_{\mathrm{o}}}(H_{\mathrm{o}}))$, so the *log likelihood* is low and the *sensitivity* to *model* is high. All *transforms* must have at least *canonical space*, $\forall T \in \mathcal{T}_{U,V_{\mathrm{o}}}\ (C_{\mathrm{G,V_o,T,H}}(T)^{\mathrm{s}}(H_{\mathrm{o}}) \geq C_{\mathrm{G,V_o,T,H}}(\tilde{T}_{\mathrm{o}})^{\mathrm{s}}(H_{\mathrm{o}}))$.

The conclusion that the *sensitivity* to *model* varies against the *log likelihood* is rather counter-intuitive but has other evidence that is more direct than the analysis of parameterisation above. Consider a pair of modal *transforms* $T, T' \in \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z)(A)) : T \in \mathcal{T}_{U,V}\})$. The two *transforms* are equal except that a *sub-component* $C'_1 \subset C_1$ is transferred from *component* C_1 to C_2 such that (i) the *derived counts* are unchanged, $(A * T')_{R_1} = (A * T)_{R_2}$, and $(A * T')_{R_2} = (A * T)_{R_1}$, (ii) the *cartesian derived counts* are unchanged, $(V^{\mathrm{C}} * T')_{R_1} = (V^{\mathrm{C}} * T)_{R_2}$, and $(V^{\mathrm{C}} * T')_{R_2} = (V^{\mathrm{C}} * T)_{R_1}$, and (iii) the other *components* are unchanged, $\mathrm{ran}((T')^{-1}) \setminus \{C_1, C_2\} = \mathrm{ran}(T^{-1}) \setminus \{C_1, C_2\}$, where $(R_1, C_1), (R_2, C_2) \in T^{-1}$ and $(R_1, C_1 \setminus C'_1), (R_2, C_2 \cup C'_1) \in (T')^{-1}$. In this case, the *specialising spaces* are equal, $C_{\mathrm{G},V,\mathrm{T,H}}(T')^{\mathrm{s}}(H) = C_{\mathrm{G},V,\mathrm{T,H}}(T)^{\mathrm{s}}(H)$. The cardinality of these pairs with respect to *transform* T is the cardinality of

$$\begin{aligned}
\{(C_1, C_2, C'_1) : \ & (R_1, C_1), (R_2, C_2) \in T^{-1}, \\
& R_1 \neq R_2,\ C'_1 \subset C_1,\ 0 < |C'_1| < |C_1|, \\
& (A * T)_{R_1} = (A * T)_{R_2} + \mathrm{size}(A * C'_1), \\
& (V^{\mathrm{C}} * T)_{R_1} = (V^{\mathrm{C}} * T)_{R_2} + |C'_1|\}
\end{aligned}$$

which is the intersection,

$$\begin{aligned}
& \{(C_1, C_2, C'_1) : (R_1, C_1), (R_2, C_2) \in T^{-1}, \\
& \qquad R_1 \neq R_2,\ C'_1 \subset C_1,\ 0 < |C'_1| < |C_1|, \\
& \qquad (A * T)_{R_1} = (A * T)_{R_2} + \mathrm{size}(A * C'_1)\} \\
\cap\ & \{(C_1, C_2, C'_1) : (R_1, C_1), (R_2, C_2) \in T^{-1}, \\
& \qquad R_1 \neq R_2,\ C'_1 \subset C_1,\ 0 < |C'_1| < |C_1|, \\
& \qquad (V^{\mathrm{C}} * T)_{R_1} = (V^{\mathrm{C}} * T)_{R_2} + |C'_1|\}
\end{aligned}$$

The logarithm of the cardinality of the *derived* term of the intersection varies against the *derived entropy*,

$$\ln |\{(A * T)_{R_1} - (A * T)_{R_2} : (R_1, \cdot), (R_2, \cdot) \in T^{-1}\}| \;\sim\; -\,\mathrm{entropy}(A * T)$$

and the logarithm of the cardinality of the *cartesian derived* term of the intersection varies against the *cartesian derived entropy*,

$$\ln |\{V^{\mathrm{C}} * T)_{R_1} - (V^{\mathrm{C}} * T)_{R_2} : (R_1, \cdot), (R_2, \cdot) \in T^{-1}\}| \;\sim\; -\,\mathrm{entropy}(V^{\mathrm{C}} * T)$$

Given the *derived entropy* and *cartesian derived entropy*, the cardinality of the intersection decreases as the correlation between the *sub-component size*, $\mathrm{size}(A * C'_1)$, and the *sub-component cardinality*, $|C'_1|$, increases. That is, the intersection is more constrained when the *sub-component size* and *cardinality* are synchronised, so the logarithm of the cardinality of the intersection varies with the *cross entropy*, $\mathrm{entropy}(A * T + V^{\mathrm{C}} * T)$. Overall, the logarithm of the cardinality of these pairs varies with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}
&\ln |\{(C_1, C_2, C'_1) : (R_1, C_1), (R_2, C_2) \in T^{-1}, \\
&\qquad R_1 \neq R_2,\ C'_1 \subset C_1,\ 0 < |C'_1| < |C_1|, \\
&\qquad\qquad (A * T)_{R_1} = (A * T)_{R_2} + \mathrm{size}(A * C'_1), \\
&\qquad\qquad (V^{\mathrm{C}} * T)_{R_1} = (V^{\mathrm{C}} * T)_{R_2} + |C'_1|\}| \\
&\sim\; (z + v) \times \mathrm{entropy}(A * T + V^{\mathrm{C}} * T) \\
&\qquad\qquad -z \times \mathrm{entropy}(A * T)\ - v \times \mathrm{entropy}(V^{\mathrm{C}} * T)
\end{aligned}$$

Hence the *sensitivity* to *model* varies against the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}
&-\ \ln |\max(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})| \;\sim \\
&\qquad -\big((z_{\mathrm{o}} + v_{\mathrm{o}}) \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}} + V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) \\
&\qquad\qquad -z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}})\ - v_{\mathrm{o}} \times \mathrm{entropy}(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})\big)
\end{aligned}$$

The *size* scaled *component size cardinality relative entropy* approximates to the *size-volume* scaled *component size cardinality sum relative entropy*, especially where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$,

$$\begin{aligned}
&z_{\mathrm{o}} \times \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) \;\approx \\
&\qquad (z_{\mathrm{o}} + v_{\mathrm{o}}) \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}} + V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) \\
&\qquad\qquad -z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}})\ - v_{\mathrm{o}} \times \mathrm{entropy}(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})
\end{aligned}$$

but the *log-likelihood* varies with the *size* scaled *component size cardinality relative entropy*,

$$\ln \hat{Q}_{\mathrm{G},T_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad z_{\mathrm{o}} \times \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})$$

so, again, the *sensitivity* to *model* varies against the *log likelihood*,

$$-\ \ln |\max(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})| \quad \sim \quad -\ \ln \hat{Q}_{\mathrm{G},T_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})$$

It is shown above in *classical uniform possible modelled induction*, where the *history probability function* is *uniform possible iso-derived historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}$, that, in the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the *possible derived volume*, w'_{o}, where the *possible derived volume* is less than the *size*, $w'_{\mathrm{o}} < z_{\mathrm{o}}$, otherwise against the *size* scaled *log possible derived volume*, $z_{\mathrm{o}} \ln w'_{\mathrm{o}}$,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ ((w'_{\mathrm{o}}\ :\ w'_{\mathrm{o}} < z_{\mathrm{o}})\ +\ (z_{\mathrm{o}} \ln w'_{\mathrm{o}}\ :\ w' \geq z_{\mathrm{o}}))$$

(b) varies with the *size* scaled *component size cardinality relative entropy*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad z_{\mathrm{o}} \times \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})$$

so (c) varies against the *specialising derived substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

In *classical specialising induction*, where the *history probability function* is the *specialising history probability function*, $P = P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, the *specialising history probability function* is specifically defined such that the *log likelihood* of the *specialising probability distribution* is proportional to the *classification space* of the *underlying histogram* less the *specialising space* of the corresponding *history*,

$$\ln \hat{Q}_{\mathrm{G},T_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \propto \quad \mathrm{spaceClassification}(A_{\mathrm{o}})\ -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

All *iso-derived* subsets of the *distribution history* for a given set of *variables* and *size* are defined as equally *iso-derived conditional probable*, and, because the *specialising space* is the same for all members of an *iso-derived*, all *iso-derived* subsets of the *distribution history* for a given set of *variables* and *size* are also equally *specialising probable*,

$$\forall V \subseteq V_{\mathrm{h}}\ \forall H, G \subseteq H_{\mathrm{h}}\%V$$
$$(A_G * T_{\mathrm{o}} = A_H * T_{\mathrm{o}} \implies P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}(G) = P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}(H))$$

Therefore, in the *near-natural* case, insofar as the *uniform possible iso-derived history probability function* approximates to the *specialising history probability function*, $P_{U,X,H_\mathrm{h},\mathrm{d},\mathrm{p},T_\mathrm{o}} \approx P_{U,X,\mathrm{G},T_\mathrm{o},\mathrm{H}}$, conjecture that (a) the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* varies with the *log likelihood* of the *specialising probability distribution*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o}) \;\sim\; \ln \hat{Q}_{\mathrm{G},T_\mathrm{o},\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})$$

and (b) the *degree of structure* of the *uniform possible iso-derived history probability function* with respect to the *specialising coder* varies with *specialising degree of structure*,

$$\mathrm{structure}(U,X)(P_{U,X,H_\mathrm{h},\mathrm{d},\mathrm{p},T_\mathrm{o}}, C_{\mathrm{G},\mathrm{T},\mathrm{H}}(T_\mathrm{o})) \;\sim\; \mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T_\mathrm{o},\mathrm{H}}, C_{\mathrm{G},\mathrm{T},\mathrm{H}}(T_\mathrm{o}))$$

So conjecture that, in the case where the *sample history* is modal, $H_\mathrm{o} \in \mathrm{maxd}(P_{U,X,H_\mathrm{h},\mathrm{d},\mathrm{p},T_\mathrm{o}})$, the *log-likelihood* of the *iso-derived conditional stuffed historical probability distribution* also varies with its *degree of structure*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o}) \;\sim\; \mathrm{structure}(U,X)(P_{U,X,H_\mathrm{h},\mathrm{d},\mathrm{p},T_\mathrm{o}}, C_{\mathrm{G},\mathrm{T},\mathrm{H}}(T_\mathrm{o}))$$

In other words, when the *iso-derived log-likelihood* is high, the expected *space* of the *specialising coder* is low and so the *compression* of the *coder* with respect to the *iso-derived historically distributed history probability function* is high.

Further, conjecture that the *sensitivity* to *model* also varies against the *log likelihood*,

$$-\ \ln |\max(\{(T, \hat{Q}_{\mathrm{h},\mathrm{d},T,U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o})) : T \in \mathcal{T}_{U,V_\mathrm{o}},\ A_\mathrm{o} \approx A_\mathrm{o} * T * T^{\dagger}\})| \;\sim\; -\ \ln \hat{Q}_{\mathrm{h},\mathrm{d},T_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o})$$

In the case where the *relative entropy* is high, $\mathrm{entropyCross}(A_\mathrm{o} * T_\mathrm{o}, V_\mathrm{o}^\mathrm{C} * T_\mathrm{o}) > \ln |T_\mathrm{o}^{-1}|$, the *sum sensitivity* varies against the *log likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\mathrm{d},T_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o}))) \;\sim\; -\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o})$$

So in this case, in *classical uniform possible modelled induction*, both (a) the *sensitivity* to *distribution histogram*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\mathrm{d},T_\mathrm{o},U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})))$$

and (b) the *sensitivity* to *model*,

$$-\ \ln |\max(\{(T, \hat{Q}_{\mathrm{h},\mathrm{d},T,U}(A_{\mathrm{o},z_\mathrm{h}}, z_\mathrm{o})(A_\mathrm{o})) : T \in \mathcal{T}_{U,V_\mathrm{o}},\ A_\mathrm{o} \approx A_\mathrm{o} * T * T^{\dagger}\})|$$

vary against the *log-likelihood*. That is, in *classical modelled induction* in some circumstances, the optimisation of the *log-likelihood* tends to minimise the *sensitivity* to parameter.

5.5.9 Specialising functional definition set induction

Again, consider extending the *model* for *specialising induction* from *transforms* to *functional definition sets*.

Consider the *specialising functional definition set history probability function* $P_{U,X,\mathrm{G},F_\mathrm{o},\mathrm{H}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ which is defined such that the probability of a *history* is inversely proportional to the bounding integer, for which the *space* is the logarithm, of the integer encoding of the *history* in the *specialising fud coder*, given a non-empty *known substrate fud* in the *sample variables* $F_\mathrm{o} \in \mathcal{F}_{U,V_\mathrm{o}} \setminus \{\emptyset\}$,

$$\begin{aligned} &P_{U,X,\mathrm{G},F_\mathrm{o},\mathrm{H}} := \\ &\quad \Big(\bigcup \{\{(H, \exp(-C_{\mathrm{G,F,H}}(F_\mathrm{o})^\mathrm{s}(H))) : \\ &\qquad\qquad H \in \mathcal{H}_{U,X},\ \mathrm{vars}(H) = V_H,\ |H| = z_H\}^\wedge : \\ &\qquad\qquad\qquad V_H \subseteq \mathrm{vars}(U),\ z_H \in \{1 \ldots |X|\}\}\Big)^\wedge \cup \\ &\quad \{(\emptyset, 0)\} \end{aligned}$$

where the *specialising fud substrate history coder* is

$$C_{\mathrm{G},V,\mathrm{F,H}}(F) \;=\; \mathrm{coderHistorySubstrateFudSpecialising}(U, X, F, D_\mathrm{S}, D_\mathrm{X})$$

and the *expanded specialising fud history coder* $C_{\mathrm{G,F,H}}(F) \in \mathrm{coders}(\mathcal{H}_{U,X})$ is derived from the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F,H}}$,

$$\begin{aligned} C_{\mathrm{G,F,H}}(F)^\mathrm{s}(H) \;=\; & (C_{\mathrm{G},V_H,\mathrm{F,H}}(F^{V_H})^\mathrm{s}(H) + s_{|V_H|} \;:\; V_H \supseteq V) \;+ \\ & (C^\mathrm{s}_\mathrm{H}(H) \;:\; V_H \not\supseteq V) \end{aligned}$$

where F^V is the *expansion* that adds a *unary transform* in the remaining *underlying variables*, $F \cup \{\{(V \setminus \mathrm{und}(F))^\mathrm{CS}\}^\mathrm{T}\}$, and $s_n = \mathrm{spaceVariables}(U)(n)$.

All *non-empty histories* are *possible* in *specialising fud induction*, $\forall H \in \mathcal{H}_{U,X} \setminus \{\emptyset\}\ (P_{U,X,\mathrm{G},F_\mathrm{o},\mathrm{H}}(H) > 0)$.

All *histories* having the same *specialising space* for a given set of *variables* and *size* are defined as equally probable,

$$\begin{aligned} &\forall H, G \in \mathcal{H}_{U,V,z,X} \\ &\quad (C_{\mathrm{G,F,H}}(F_\mathrm{o})^\mathrm{s}(G) = C_{\mathrm{G,F,H}}(F_\mathrm{o})^\mathrm{s}(H) \implies P_{U,X,\mathrm{G},F_\mathrm{o},\mathrm{H}}(G) = P_{U,X,\mathrm{G},F_\mathrm{o},\mathrm{H}}(H)) \end{aligned}$$

In *specialising fud induction* there is no *distribution histogram*, so the *drawn history* is parameterised only by *substrate variables* and *size*, (V_H, z_H). Nor is any *sample* constrained to equal its *fud-independent*, $A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$. If a *history* is *possible* in *uniform possible fud induction*, then it is *possible* in *specialising fud induction*,

$$\forall H \in \mathcal{H}_{U,X}\ (P_{U,X,H_{\mathrm{h}},\mathrm{d,p},F_{\mathrm{o}}}(H) > 0 \implies P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}(H) > 0)$$

The *specialising space* is the same for all members of an *iso-fud*, $\forall B \in D_{U,\mathrm{i},F,z}^{-1}(D_{U,F,z}(A))\ (C_{\mathrm{G,F,H}}(F)^{\mathrm{s}}(H_B) = C_{\mathrm{G,F,H}}(F)^{\mathrm{s}}(H_A))$, so all *iso-fud* subsets of the *distribution history* for a given set of *variables* and *size* are not only equally *iso-fud probable*,

$$\begin{aligned}&\forall V \subseteq V_{\mathrm{h}}\ \forall H, G \subseteq H_{\mathrm{h}}\%V \\ &\quad (\forall T \in F_{\mathrm{o}}\ (A_G * T_{F_{\mathrm{o}}} = A_H * T_{F_{\mathrm{o}}}) \implies P_{U,X,H_{\mathrm{h}},\mathrm{d,p},F_{\mathrm{o}}}(G) = P_{U,X,H_{\mathrm{h}},\mathrm{d,p},F_{\mathrm{o}}}(H))\end{aligned}$$

but also equally *specialising probable*,

$$\begin{aligned}&\forall V \subseteq V_{\mathrm{h}}\ \forall H, G \subseteq H_{\mathrm{h}}\%V \\ &\quad (\forall T \in F_{\mathrm{o}}\ (A_G * T_{F_{\mathrm{o}}} = A_H * T_{F_{\mathrm{o}}}) \implies P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}(G) = P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}(H))\end{aligned}$$

where $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$.

Given a *history* $H \in \mathcal{H}_{U,X}$, such that $H \neq \emptyset$, the *specialised historical probability* of *histogram* $A_H = \mathrm{histogram}(H) + V_H^{\mathrm{CZ}} \in \mathcal{A}_{U,\mathrm{i},V_H,z_H}$ is

$$\hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_H)(A_H) \ \propto\ \sum(P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *specialising fud probability distribution* is defined

$$\begin{aligned}&\hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z) \ := \\ &\quad \{(A, \frac{z!}{\prod_{S \in A^{\mathrm{FS}}} A_S!} \times \exp(-C_{\mathrm{G,F,H}}(F)^{\mathrm{s}}(H_A))) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}^{\wedge}\end{aligned}$$

where $V = \mathrm{und}(F)$ and $H_A = \mathrm{history}(A)$.

The *log likelihood* is proportional to the *classification space* of the *underlying histogram* less the *specialising space* of the corresponding *history*,

$$\ln \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z)(A) \ \propto\ \mathrm{spaceClassification}(A) \ -\ \mathrm{space}(C_{\mathrm{G},V,\mathrm{F,H}}(F))(H_A)$$

In the *law-like* case where the *fud* has a *top transform*, $\exists T \in F\ (W_T = \mathrm{der}(F))$, the *space* of the *specialising coder* is

$$\begin{aligned}&\mathrm{space}(C_{\mathrm{G},V,\mathrm{F,H}}(F))(H) = \\ &\qquad \mathrm{spaceIds}(|X|, |H|) + \\ &\qquad \mathrm{spaceCountsDerived}(U)(A, F^{\mathrm{T}}) + \\ &\qquad \mathrm{spaceClassification}(A * F^{\mathrm{T}}) + \\ &\qquad \sum_{T \in F} \mathrm{spaceEventsPartition}(A * \mathrm{dep}(F, V_T)^{\mathrm{T}}, T)\end{aligned}$$

where $V_T = \text{und}(T)$, $W_T = \text{der}(T)$, and $\text{dep} = \text{depends}$.

The *space* of the *specialising fud substrate history coder*, $C_{\text{G},V,\text{F},\text{H}}(F)$, varies (i) with the *possible fud derived volume*, $w' = |(F^{\text{T}})^{-1}|$, or the *size* scaled *log possible fud derived volume*, $z \ln w'$, (ii) with the *size* scaled *fud transform derived entropy* and (iii) against the sum of the *size* scaled *component size cardinality cross entropies* of the *transforms* of the *fud*,

$$\begin{aligned} C_{\text{G},V,\text{F},\text{H}}(F)^{\text{s}}(H) \ \sim \\ &(w' \ : \ w' < z) \ + \ (z \ln w' \ : \ w' \geq z) \\ &+ z \times \text{entropy}(A * F^{\text{T}}) \\ &- z \times \sum_{T \in F} \text{entropyCross}(A * T_F, V_T^{\text{C}} * T) \end{aligned}$$

The *specialising-canonical space difference*, $2C_{\text{G},V,\text{F},\text{H}}(F)^{\text{s}}(H) - C^{\text{s}}_{\text{H},V}(H) - C^{\text{s}}_{\text{G},V}(H)$, varies (i) with twice the total *possible derived volume* or twice the total *size* scaled *log possible derived volume*, (ii) with the sum of the *size* scaled *derived entropies*, (iii) against twice the sum of the *size* scaled *component size cardinality cross entropies* and (iv) against the sum of the *size* scaled *size expected component entropies*,

$$\begin{aligned} 2C_{\text{G},V,\text{F},\text{H}}(F)^{\text{s}}(H) - C^{\text{s}}_{\text{H},V}(H) - C^{\text{s}}_{\text{G},V}(H) \ \sim \\ &\sum_{T \in F} 2\big((w'_T \ : \ w'_T < z) \ + \ (z \ln w'_T \ : \ w'_T \geq z)\big) \\ &+ \sum_{T \in F} z \times \text{entropy}(A * T_F) \\ &- \sum_{T \in F} 2z \times \text{entropyCross}(A * T_F, V_T^{\text{C}} * T) \\ &- \sum_{T \in F} z \times \text{entropyComponent}(A * \text{dep}(F, V_T)^{\text{T}}, T) \end{aligned}$$

where $w'_T = |T^{-1}|$, and $T_F = \text{depends}(F, W_T)^{\text{T}}$.

The *specialising log likelihood* varies (a) with the *size* scaled *underlying en-*

tropy and (b) against the *specialising-canonical space difference*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z)(A) \ \sim \\ &-\sum_{T \in F} 2\big((w'_T \ : \ w'_T < z) \ + \ (z \ln w'_T \ : \ w'_T \geq z)\big) \\ &+ z \times \mathrm{entropy}(A) \\ &-\sum_{T \in F} z \times \mathrm{entropy}(A * T_F) \\ &+\sum_{T \in F} 2z \times \mathrm{entropyCross}(A * T_F, V_T^{\mathrm{C}} * T) \\ &+\sum_{T \in F} z \times \mathrm{entropyComponent}(A * \mathrm{dep}(F, V_T)^{\mathrm{T}}, T)\end{aligned}$$

Let H_{o} be a *sample history* of *known size* $z_{\mathrm{o}} = |H_{\mathrm{o}}| > 0$ in the *known sample variables*, V_{o}, which has a *known histogram* $A_{\mathrm{o}} = \mathrm{histogram}(H_{\mathrm{o}}) + V_{\mathrm{o}}^{\mathrm{CZ}} \in \mathcal{A}_{U,\mathrm{i},V_{\mathrm{o}},z_{\mathrm{o}}}$. In *classical specialising fud induction*, where the *history probability function* is the *specialising fud history probability function*, $P = P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}$, given some *substrate fud* in the *sample variables* $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, the *log likelihood* of the *specialising fud probability distribution* (a) varies with the *size* scaled *underlying entropy*,

$$\ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}})$$

(b) varies against the total *possible derived volume* or *size* scaled *log possible derived volume*,

$$\ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ -\sum_{T \in F_{\mathrm{o}}} \big((w'_T \ : \ w'_T < z_{\mathrm{o}}) \ + \ (z_{\mathrm{o}} \ln w'_T \ : \ w'_T \geq z_{\mathrm{o}})\big)$$

(c) varies against the total *size* scaled *derived entropy*

$$\ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ -z_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropy}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}})$$

(d) varies with the total *size* scaled *component size cardinality cross entropy*

$$\ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ z_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropyCross}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_T^{\mathrm{C}} * T)$$

and (e) varies with the total *size* scaled *size expected component entropy*,

$$\ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ z_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropyComponent}(A_{\mathrm{o}} * \mathrm{dep}(F_{\mathrm{o}}, V_T)^{\mathrm{T}}, T)$$

So the *specialising log likelihood* is maximised when (a) the total *possible derived volume* is minimised, (b) the *underlying entropy* is maximised, (c) the total *derived entropy* is minimised, (d) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for each *transform*, and (e) the total *expected component entropy* is maximised. It is also conjectured that, (i) the *derived entropy* decreases up the *layers*, (ii) the *possible derived volume* decreases up the *layers*, (iii) the *expected component entropy* increases up the *layers*, and (iv) the *component size cardinality cross entropy* increases up the *layers*.

The ratio of the *expected specialising space* to the *specialising entropy*,

$$\frac{\text{expected}(P_{U,X,\text{G},F,\text{H},z})(C_{\text{G,F,H}}(F)^{\text{s}})}{\text{entropy}(P_{U,X,\text{G},F,\text{H},z})}$$

is minimised at equilibrium. As the ratio tends to one, the *specialising coder* tends to an *entropy coder* and so the *degree of structure* tends to one,

$$\begin{aligned}&\text{structure}(U,X)(P_{U,X,\text{G},F,\text{H}},C_{\text{G,F,H}}(F))\\ &:= \frac{\text{canonical}(U,X)(P_{U,X,\text{G},F,\text{H}}) - \text{expected}(P_{U,X,\text{G},F,\text{H}})(C_{\text{G,F,H}}(F)^{\text{s}})}{\text{canonical}(U,X)(P_{U,X,\text{G},F,\text{H}}) - \text{entropy}(P_{U,X,\text{G},F,\text{H}})}\end{aligned}$$

The *specialising fud history probability function*, $P_{U,X,\text{G},F_{\text{o}},\text{H}}$, is the *history probability function* P that maximises the *degree of structure* with respect to the *expanded specialising fud history coder*, $C_{\text{G,F,H}}(F_{\text{o}})$, given arbitrary *draw*, (V_H, z_H),

$$\begin{aligned}&P_{U,X,\text{G},F_{\text{o}},\text{H}} \in\\ &\quad \text{maxd}(\{(P,\text{structure}(U,X)(P,C_{\text{G,F,H}}(F_{\text{o}}))) :\\ &\qquad P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P},\\ &\qquad\quad |\{(\textstyle\sum P_H : H \in \mathcal{H}_{U,X},\ \text{vars}(H) = V_H,\ |H| = z_H) :\\ &\qquad\qquad\qquad V_H \subseteq \text{vars}(U),\ z_H \in \{1 \ldots |X|\}\}| = 1,\\ &\qquad\quad P_\emptyset = 0\})\end{aligned}$$

Although the *specialising entropy*, $\text{entropy}(P_{U,X,\text{G},F,\text{H}})$, is maximised and the *expected specialising space*, $\text{expected}(P_{U,X,\text{G},F,\text{H}})(C_{\text{G,F,H}}(F)^{\text{s}})$, is minimised but sometimes not minimal, the *modal specialising space*, $C_{\text{G,F,H}}(F)^{\text{s}}(H)$, of a *modal history* $H \in \text{maxd}(P_{U,X,\text{G},F,\text{H}})$, is always minimal. The *degree of structure* varies proportionately against the *expected specialising space*,

$$\text{structure}(U,X)(P_{U,X,\text{G},F,\text{H}},C_{\text{G,F,H}}(F)) \sim -\text{expected}(P_{U,X,\text{G},F,\text{H}})(C_{\text{G,F,H}}(F)^{\text{s}})$$

Conjecture that as the *mean specialising space* decreases, the *log likelihood* increases. So, in *classical specialising fud induction*, where (i) the *history probability function* is the *specialising fud history probability function*, $P = P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}$, given some *substrate fud* in the *sample variables* $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, and (ii) the *sample history* is *modal*, $H_{\mathrm{o}} \in \mathrm{maxd}(P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}})$, the *log likelihood* of the *specialising fud probability distribution* varies with the *degree of structure*,

$$\ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \mathrm{structure}(U,X)(P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}, C_{\mathrm{G,F,H}}(F_{\mathrm{o}}))$$

In the case where the *model*, F_{o}, is *known*, '*likelihood*' is an abuse of terminology because there is no *distribution histogram* nor other *unknown* parameterisation of the *probability function*, $\hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}}) \in \mathcal{P}$.

In the case, however, where the *model*, F_{o}, is *unknown*, the *maximum likelihood estimate* $\tilde{F}_{\mathrm{o}}$ can be defined as an optimisation of the *specialising probability* given the *sample*,

$$\tilde{F}_{\mathrm{o}} \;\in\; \mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})$$

That is, the *probability function* is now parameterised by the *unknown model*, F_{o}, so a *model* is the argument to the corresponding *likelihood function* parameterised by the *sample histogram*, A_{o}.

In *specialising fud induction*, which is such that the *history probability function* is $P = P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}$, given some *unknown substrate fud* $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, the *maximum likelihood estimate* of the *model*, $\tilde{F}_{\mathrm{o}}$, occurs at the minimisation of the *specialising space* of the *sample history*,

$$\tilde{F}_{\mathrm{o}} \;\in\; \mathrm{mind}(\{(F, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F,H}}(F^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})$$

The *maximum likelihood estimate* for the *model*, $\tilde{F}_{\mathrm{o}}$, is non-trivial, $\tilde{F}_{\mathrm{o}} \notin \{\{T_{\mathrm{s}}\},\{T_{\mathrm{u}}\}\}$, if there exists a *model* for which the *specialising fud substrate history space* is less than either the *index space* or the *classification space*, $C_{\mathrm{G,V_o,F,H}}(\tilde{F}_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}}) < \mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H},V_{\mathrm{o}}}(H_{\mathrm{o}}), C^{\mathrm{s}}_{\mathrm{G},V_{\mathrm{o}}}(H_{\mathrm{o}}))$. This the case if the *degree of structure* is greater than zero,

$$\mathrm{structure}(U,X)(P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}, C_{\mathrm{G,F,H}}(\tilde{F}_{\mathrm{o}})) \;>\; 0$$

In the case where the *histogram* is *sparse*, $\mathrm{trim}(A_{\mathrm{o}}) = A_{\mathrm{o}}^{\mathrm{F}}$ or $A_{\mathrm{o}} * A_{\mathrm{o}}^{\mathrm{F}} = A_{\mathrm{o}}^{\mathrm{F}}$, the *maximum likelihood estimate* for the *model* in *specialising transform induction* is the *under-fitted effective binary transform*,

$$A_{\mathrm{o}} * A_{\mathrm{o}}^{\mathrm{F}} = A_{\mathrm{o}}^{\mathrm{F}} \quad\implies\quad \tilde{T}_{\mathrm{o}} = \{A_{\mathrm{o}}^{\mathrm{FS}},\ V_{\mathrm{o}}^{\mathrm{CS}} \setminus A_{\mathrm{o}}^{\mathrm{FS}}\}^{\mathrm{T}}$$

In *specialising fud induction*, however, *under-fitted* or *over-fitted models* may sometimes be avoided if there exist *reductions* that are not *sparse*, $\exists K \subset V_{\mathrm{o}}$ $(\mathrm{trim}(A_{\mathrm{o}}\%K) \neq (A_{\mathrm{o}}\%K)^{\mathrm{F}})$. In these cases the *fud* contains a *transform* $T \in \tilde{F}_{\mathrm{o}}$ on a subset of the *substrate*, $\mathrm{und}(T) = K$, such that $T \neq \{(A_{\mathrm{o}}\%K)^{\mathrm{FS}},\ K^{\mathrm{CS}} \setminus (A_{\mathrm{o}}\%K)^{\mathrm{FS}}\}^{\mathrm{T}}$ and $T \neq ((A_{\mathrm{o}}\%K)^{\mathrm{FS}\{\}} \cup \{K^{\mathrm{CS}} \setminus (A_{\mathrm{o}}\%K)^{\mathrm{FS}}\})^{\mathrm{T}}$.

The *sensitivity* to *model* is defined as the negative logarithm of the cardinality of the *maximum likelihood estimate models*,

$$-\ \ln |\max(\{(F, \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})|$$

Conjecture that the cardinality of the modes varies against the negative gradient of the likelihood function of the singly parameterised exponential function fitted to the Boltzmann distribution, so the *sensitivity* to *model* varies with the negative gradient. Hence the *sensitivity* to *model* varies against the *specialising structure*,

$$\begin{aligned} -\ \ln |\max(\{(F, \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})|\ \sim& \\ -\ \mathrm{structure}(U,X)(P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}, C_{\mathrm{G,F,H}}(F_{\mathrm{o}}))& \end{aligned}$$

As shown above, the the *degree of structure* varies with the *log likelihood* of the *specialising fud probability distribution*,

$$\mathrm{structure}(U,X)(P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}, C_{\mathrm{G,F,H}}(F_{\mathrm{o}}))\ \sim\ \ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})$$

so the *sensitivity* to *model* varies against the *log likelihood*,

$$-\ \ln |\max(\{(F, \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})|\quad \sim\ -\ \ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})$$

That is, maximisation of the *log likelihood* also tends to minimise the *sensitivity* to *model*.

It is shown above in *classical uniform possible fud induction*, where the *history probability function* is *uniform possible iso-fud historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}$, that, in the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-independent*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})}$, the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the sum of the *possible derived volumes* or *size* scaled *log possible derived volumes*,

$$\begin{aligned} &\ln \hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})\ \sim \\ &\quad -\sum_{T \in F_{\mathrm{o}}} \big((w'_{T_{F_{\mathrm{o}}}}\ :\ w'_{T_{F_{\mathrm{o}}}} < z_{\mathrm{o}})\ +\ (z_{\mathrm{o}} \ln w'_{T_{F_{\mathrm{o}}}}\ :\ w'_{T_{F_{\mathrm{o}}}} \geq z_{\mathrm{o}})\big) \end{aligned}$$

(b) varies with the *size* scaled *component size cardinality relative entropies*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad z_{\mathrm{o}} \times \sum_{T \in F_{\mathrm{o}}} \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}})$$

so (c) varies against the *specialising fud substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F,H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})$$

and (d) varies against the *specialising-canonical space difference*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \\ -(2C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F,H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}}) - C_{\mathrm{H},V_{\mathrm{o}}}^{\mathrm{s}}(H_{\mathrm{o}}) - C_{\mathrm{G},V_{\mathrm{o}}}^{\mathrm{s}}(H_{\mathrm{o}}))$$

In *classical specialising fud induction*, where the *history probability function* is the *specialising fud history probability function*, $P = P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}$, the *specialising fud history probability function* is specifically defined such that the *log likelihood* of the *specialising fud probability distribution* is proportional to the *classification space* of the *underlying histogram* less the *specialising space* of the corresponding *history*,

$$\ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \propto \quad \mathrm{spaceClassification}(A_{\mathrm{o}}) \ - \ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F,H}}(F_{\mathrm{o}}))(H_{\mathrm{o}})$$

All *iso-fud* subsets of the *distribution history* for a given set of *variables* and *size* are defined as equally *iso-fud conditional probable*, and, because the *specialising space* is the same for all members of an *iso-fud*, all *iso-fud* subsets of the *distribution history* for a given set of *variables* and *size* are also equally *specialising probable*,

$$\forall V \subseteq V_{\mathrm{h}}\ \forall H, G \subseteq H_{\mathrm{h}} \% V \\ (\forall T \in F_{\mathrm{o}}\ (A_G * T_{F_{\mathrm{o}}} = A_H * T_{F_{\mathrm{o}}}) \implies P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}(G) = P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}(H))$$

Therefore, in the *near-fud-independent* case, insofar as the *uniform possible iso-fud history probability function* approximates to the *specialising fud history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d,p},F_{\mathrm{o}}} \approx P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}$, conjecture that (a) the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* varies with the *log likelihood* of the *specialising fud probability distribution*,

$$\ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})$$

and (b) the *degree of structure* of the *uniform possible iso-fud history probability function* with respect to the *specialising coder* varies with *specialising degree of structure*,

$$\mathrm{structure}(U, X)(P_{U,X,H_{\mathrm{h}},\mathrm{d,p},F_{\mathrm{o}}}, C_{\mathrm{G,F,H}}(F_{\mathrm{o}})) \ \sim \\ \mathrm{structure}(U, X)(P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}, C_{\mathrm{G,F,H}}(F_{\mathrm{o}}))$$

So conjecture that, in the case where the *sample history* is modal, $H_{\rm o} \in \mathrm{maxd}(P_{U,X,H_{\rm h},{\rm d},{\rm p},F_{\rm o}})$, the *log-likelihood* of the *iso-fud conditional stuffed historical probability distribution* also varies with its *degree of structure*,

$$\ln \hat{Q}_{{\rm h},{\rm d},F_{\rm o},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o}) \quad \sim \quad \mathrm{structure}(U, X)(P_{U,X,H_{\rm h},{\rm d},{\rm p},F_{\rm o}}, C_{\rm G,F,H}(F_{\rm o}))$$

In other words, when the *iso-fud log-likelihood* is high, the expected *space* of the *specialising coder* is low and so the *compression* of the *coder* with respect to the *iso-fud historically distributed history probability function* is high.

Further, conjecture that the *sensitivity* to *model* also varies against the *log likelihood*,

$$\begin{aligned} -\ \ln |\max(\{(F, \hat{Q}_{{\rm h},{\rm d},F,U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o})) :& \\ F \in \mathcal{F}_{U,V_{\rm o}},\ \exists T \in F\ (W_T = W_F),\ A_{\rm o} \approx A_{\rm o}^{{\rm E_F}(F)}\})|& \ \sim \\ -\ \ln \hat{Q}_{{\rm h},{\rm d},F_{\rm o},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o})& \end{aligned}$$

In the case where the *relative entropies* are high, $\forall T \in F_{\rm o}\ (\mathrm{entropyCross}(A_{\rm o} * T_{F_{\rm o}}, V_{\rm o}^{\rm C} * T_{F_{\rm o}}) > \ln |T_{F_{\rm o}}^{-1}|)$, the *sum sensitivity* varies against the *log likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm h},{\rm d},F_{\rm o},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o}))) \quad \sim \quad -\ln \hat{Q}_{{\rm h},{\rm d},F_{\rm o},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o})$$

So in this case, in *classical uniform possible modelled induction*, both (a) the *sensitivity* to *distribution histogram*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm h},{\rm d},F_{\rm o},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})))$$

and (b) the *sensitivity* to *model*,

$$\begin{aligned} -\ \ln |\max(\{(F, \hat{Q}_{{\rm h},{\rm d},F,U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o})) :& \\ F \in \mathcal{F}_{U,V_{\rm o}},\ \exists T \in F\ (W_T = W_F),\ A_{\rm o} \approx A_{\rm o}^{{\rm E_F}(F)}\})|& \end{aligned}$$

vary against the *log-likelihood*. That is, in *classical modelled induction* in some circumstances, the optimisation of the *log-likelihood* tends to minimise the *sensitivity* to parameter.

5.5.10 Specialising functional definition set decomposition induction

Again, consider extending the *model* for *specialising induction* from *functional definition sets* to *functional definition set decompositions*.

Consider the *specialising functional definition set decomposition history probability function* $P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ which is defined such that the probability of a *history* is inversely proportional to the bounding integer, for which the *space* is the logarithm, of the integer encoding of the *history* in the *specialising fud decomposition coder*, given a non-empty *known substrate fud decomposition* in the *sample variables* $D_{\mathrm{o}} \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\}$,

$$
\begin{aligned}
&P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}} := \\
&\quad \Big(\bigcup \{\{(H, \exp(-C_{\mathrm{G,D,F,H}}(D_{\mathrm{o}})^{\mathrm{s}}(H))) : \\
&\qquad\qquad H \in \mathcal{H}_{U,X},\ \mathrm{vars}(H) = V_H,\ |H| = z_H\}^{\wedge} : \\
&\qquad\qquad\qquad V_H \subseteq \mathrm{vars}(U),\ z_H \in \{1 \ldots |X|\}\}\Big)^{\wedge} \cup \\
&\quad \{(\emptyset, 0)\}
\end{aligned}
$$

where the *specialising fud decomposition substrate history coder* is

$$
\begin{aligned}
C_{\mathrm{G},V,\mathrm{D,F,H}}(D) &= \\
&\quad \mathrm{coderHistorySubstrateFudDecompSpecialising}(U, X, F, D_{\mathrm{S}}, D_{\mathrm{X}})
\end{aligned}
$$

and the *expanded specialising fud decomposition history coder* $C_{\mathrm{G,D,F,H}}(D) \in \mathrm{coders}(\mathcal{H}_{U,X})$ is derived from the *specialising fud decomposition substrate history coder*, $C_{\mathrm{G},V,\mathrm{D,F,H}}$,

$$
\begin{aligned}
C_{\mathrm{G,D,F,H}}(D)^{\mathrm{s}}(H) &= (C_{\mathrm{G},V_H,\mathrm{D,F,H}}(D^{V_H})^{\mathrm{s}}(H) + s_{|V_H|} \ :\ V_H \supseteq V) \ + \\
&\quad\ (C^{\mathrm{s}}_{\mathrm{H}}(H) \ :\ V_H \not\supseteq V)
\end{aligned}
$$

where D^V is the *expansion* that adds a *unary transform* in the remaining *underlying variables* to the leaf *fuds* in the *decomposition* tree such that the *fud* of each path of the *application* tree has complete coverage of the *substrate*,

$$
\forall L \in \mathrm{paths}(D^*)\ (\bigcup_{(\cdot,(F,\cdot)) \in L} V_F = V)
$$

where $V_F = \mathrm{und}(D)$, and $s_n = \mathrm{spaceVariables}(U)(n)$.

All *non-empty histories* are *possible* in *specialising fud decomposition induction*, $\forall H \in \mathcal{H}_{U,X} \setminus \{\emptyset\}\ (P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}}(H) > 0)$.

All *histories* having the same *specialising space* for a given set of *variables* and *size* are defined as equally probable,

$$
\begin{aligned}
&\forall H, G \in \mathcal{H}_{U,V,z,X} \\
&\quad (C_{\mathrm{G,D,F,H}}(D_{\mathrm{o}})^{\mathrm{s}}(G) = C_{\mathrm{G,D,F,H}}(D_{\mathrm{o}})^{\mathrm{s}}(H) \implies P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}}(G) = P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}}(H))
\end{aligned}
$$

Given a *history* $H \in \mathcal{H}_{U,X}$, such that $H \neq \emptyset$, the *specialised historical probability* of *histogram* $A_H = \text{histogram}(H) + V_H^{\text{CZ}} \in \mathcal{A}_{U,\text{i},V_H,z_H}$ is

$$\hat{Q}_{\text{G},D_{\text{o}},\text{H},U}(z_H)(A_H) \quad \propto \quad \sum(P_{U,X,\text{G},D_{\text{o}},\text{H}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *specialising fud decomposition probability distribution* is defined

$$\begin{aligned}\hat{Q}_{\text{G},D,\text{H},U}(z) \ &:= \\ &\{(A, \frac{z!}{\prod_{S \in A^{\text{FS}}} A_S!} \times \exp(-C_{\text{G,D,F,H}}(D)^{\text{s}}(H_A))) : A \in \mathcal{A}_{U,\text{i},V,z}\}^{\wedge}\end{aligned}$$

where $V = \text{und}(D)$ and $H_A = \text{history}(A)$.

The *log likelihood* is proportional to the *classification space* of the *underlying histogram* less the *specialising space* of the corresponding *history*,

$$\ln \hat{Q}_{\text{G},D,\text{H},U}(z)(A) \quad \propto \quad \text{spaceClassification}(A) \ - \ \text{space}(C_{\text{G},V,\text{D,F,H}}(D))(H_A)$$

Let H_{o} be a *sample history* of *known size* $z_{\text{o}} = |H_{\text{o}}| > 0$ in the *known sample variables*, V_{o}, which has a *known histogram* $A_{\text{o}} = \text{histogram}(H_{\text{o}}) + V_{\text{o}}^{\text{CZ}} \in \mathcal{A}_{U,\text{i},V_{\text{o}},z_{\text{o}}}$. In *classical specialising fud decomposition induction*, where the *history probability function* is the *specialising fud decomposition history probability function*, $P = P_{U,X,\text{G},D_{\text{o}},\text{H}}$, given some *substrate fud decomposition* in the *sample variables* $D_{\text{o}} \in \mathcal{D}_{\text{F},U,V_{\text{o}}}$, the *log likelihood* of the *specialising fud decomposition probability distribution* (a) varies with the *size* scaled *underlying entropy*,

$$\ln \hat{Q}_{\text{G},D_{\text{o}},\text{H},U}(z_{\text{o}})(A_{\text{o}}) \quad \sim \quad z_{\text{o}} \times \text{entropy}(A_{\text{o}})$$

(b) varies against the total *possible derived volume* or *size* scaled *log possible derived volume*

$$\begin{aligned}&\ln \hat{Q}_{\text{G},D_{\text{o}},\text{H},U}(z_{\text{o}})(A_{\text{o}}) \ \sim \\ &\quad - \sum_{(C,F) \in \text{cont}(D_{\text{o}})} \sum_{T \in F} \Big((|T_F^{-1}| \ : \ |T_F^{-1}| < z_{A_{\text{o}} * C}) \ + \\ &\qquad\qquad\qquad (z_{A_{\text{o}} * C} \ln |T_F^{-1}| \ : \ |T_F^{-1}| \geq z_{A_{\text{o}} * C}) \Big)\end{aligned}$$

(c) varies against the total *size* scaled *derived entropy*

$$\ln \hat{Q}_{\text{G},D_{\text{o}},\text{H},U}(z_{\text{o}})(A_{\text{o}}) \quad \sim \quad - \sum_{(C,F) \in \text{cont}(D_{\text{o}})} \Big(z_{A_{\text{o}} * C} \times \sum_{T \in F} \text{entropy}(A_{\text{o}} * C * T_F) \Big)$$

(d) varies with the total *size* scaled *component size cardinality cross entropy*

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{G},D_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \\ &\quad \sum_{(C,F) \in \mathrm{cont}(D_{\mathrm{o}})} \big(z_{A_{\mathrm{o}} * C} \times \sum_{T \in F} \mathrm{entropyCross}(A_{\mathrm{o}} * C * T_F, C * T)\big)\end{aligned}$$

and (e) varies with the total *size* scaled *size expected component entropy*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{G},D_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \\ &\quad \sum_{(C,F) \in \mathrm{cont}(D_{\mathrm{o}})} \big(z_{A_{\mathrm{o}} * C} \times \sum_{T \in F} \mathrm{entropyComponent}(A_{\mathrm{o}} * C * \mathrm{dep}(F, V_T)^{\mathrm{T}}, T)\big)\end{aligned}$$

where $\mathrm{cont}(D) := \mathrm{elements}(\mathrm{contingents}(D))$ and $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$.

So the *specialising log likelihood* is maximised when (a) the total *possible derived volume* is minimised, (b) the *underlying entropy* is maximised, (c) the total *derived entropy* is minimised, (d) high *size components* are low *cardinality components* and low *size components* are high *cardinality components* for each *transform* for all *slices*, and (e) the total *expected component entropy* is maximised. It is also conjectured that, for all *fuds*, (i) the *derived entropy* decreases up the *layers*, (ii) the *possible derived volume* decreases up the *layers*, (iii) the *expected component entropy* increases up the *layers*, and (iv) the *component size cardinality cross entropy* increases up the *layers*.

The *specialising fud decomposition history probability function*, $P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}}$, is the *history probability function* P that maximises the *degree of structure* with respect to the *expanded specialising fud decomposition history coder*, $C_{\mathrm{G,D,F,H}}(D_{\mathrm{o}})$, given arbitrary *draw*, (V_H, z_H),

$$\begin{aligned}&P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}} \in \\ &\quad \mathrm{maxd}(\{(P, \mathrm{structure}(U,X)(P, C_{\mathrm{G,D,F,H}}(D_{\mathrm{o}}))) : \\ &\qquad P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}, \\ &\qquad\quad |\{(\sum P_H : H \in \mathcal{H}_{U,X},\ \mathrm{vars}(H) = V_H,\ |H| = z_H) : \\ &\qquad\qquad V_H \subseteq \mathrm{vars}(U),\ z_H \in \{1 \ldots |X|\}\}| = 1, \\ &\qquad\quad P_{\emptyset} = 0\})\end{aligned}$$

In *classical specialising fud decomposition induction*, where (i) the *history probability function* is the *specialising fud decomposition history probability function*, $P = P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}}$, given some *substrate fud decomposition* in the *sample variables* $D_{\mathrm{o}} \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}}$, and (ii) the *sample history* is *modal*, $H_{\mathrm{o}} \in$

$\mathrm{maxd}(P_{U,X,\mathrm{G},D_\mathrm{o},\mathrm{H}})$, the *log likelihood* of the *specialising fud decomposition probability distribution* varies with the *degree of structure*,

$$\ln \hat{Q}_{\mathrm{G},D_\mathrm{o},\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o}) \sim \mathrm{structure}(U,X)(P_{U,X,\mathrm{G},D_\mathrm{o},\mathrm{H}}, C_{\mathrm{G,D,F,H}}(D_\mathrm{o}))$$

In the case where the *model*, D_o, is *unknown*, the *maximum likelihood estimate* $\tilde{D}_\mathrm{o}$ can be defined as an optimisation of the *specialising probability* given the *sample*,

$$\tilde{D}_\mathrm{o} \in \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{G},D,\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})) : D \in \mathcal{D}_{\mathrm{F},U,V_\mathrm{o}}\})$$

In *specialising fud decomposition induction*, which is such that the *history probability function* is $P = P_{U,X,\mathrm{G},D_\mathrm{o},\mathrm{H}}$, given some *unknown substrate fud decomposition* $D_\mathrm{o} \in \mathcal{D}_{\mathrm{F},U,V_\mathrm{o}}$, the *maximum likelihood estimate* of the *model*, $\tilde{D}_\mathrm{o}$, occurs at the minimisation of the *specialising space* of the *sample history*,

$$\tilde{D}_\mathrm{o} \in \mathrm{mind}(\{(D, C_{\mathrm{G},\mathrm{V}_\mathrm{o},\mathrm{D,F,H}}(D^{V_\mathrm{o}})^\mathrm{s}(H_\mathrm{o})) : D \in \mathcal{D}_{\mathrm{F},U,V_\mathrm{o}}\})$$

The *maximum likelihood estimate* for the *model*, $\tilde{D}_\mathrm{o}$, is non-trivial, $\tilde{D}_\mathrm{o} \notin \{\{((\emptyset,\{T_\mathrm{s}\}),\emptyset)\},\{((\emptyset,\{T_\mathrm{u}\}),\emptyset)\}\}$, if there exists a *model* for which the *specialising fud decomposition substrate history space* is less than either the *index space* or the *classification space*,

$$C_{\mathrm{G},\mathrm{V}_\mathrm{o},\mathrm{D,F,H}}(\tilde{D}_\mathrm{o}^{V_\mathrm{o}})^\mathrm{s}(H_\mathrm{o}) < \mathrm{minimum}(C^\mathrm{s}_{\mathrm{H},V_\mathrm{o}}(H_\mathrm{o}), C^\mathrm{s}_{\mathrm{G},V_\mathrm{o}}(H_\mathrm{o}))$$

Conjecture that the *sensitivity* to *model* varies against the *specialising structure*,

$$\begin{aligned} -\ \ln |\max(\{(D, \hat{Q}_{\mathrm{G},D,\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})) : D \in \mathcal{D}_{\mathrm{F},U,V_\mathrm{o}}\})| &\sim \\ &-\ \mathrm{structure}(U,X)(P_{U,X,\mathrm{G},D_\mathrm{o},\mathrm{H}}, C_{\mathrm{G,D,F,H}}(D_\mathrm{o})) \end{aligned}$$

so the *sensitivity* to *model* varies against the *log likelihood*,

$$\begin{aligned} -\ \ln |\max(\{(D, \hat{Q}_{\mathrm{G},D,\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})) : D \in \mathcal{D}_{\mathrm{F},U,V_\mathrm{o}}\})| &\sim \\ &-\ \ln \hat{Q}_{\mathrm{G},D_\mathrm{o},\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o}) \end{aligned}$$

That is, maximisation of the *log likelihood* also tends to minimise the *sensitivity* to *model*.

It is shown above in *classical uniform possible fud decomposition induction*, where the *history probability function* is *uniform possible iso-fud-decomposition historically distributed*, $P = P_{U,X,H_\mathrm{h},\mathrm{d},\mathrm{p},D_\mathrm{o}}$, that, in the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-decomposition-independent*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E}_{\mathrm{D,F}}(D_{\mathrm{o}})}$, the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* (a) varies against the sum of the *possible derived volumes* or *size* scaled *log possible derived volumes* of the *slices*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\; \\ -\sum_{(C,F) \in \mathrm{cont}(D_{\mathrm{o}})} \sum_{T \in F} \big((|T_F^{-1}| \;:\; |T_F^{-1}| < z_{A_{\mathrm{o}}*C}) \;+ \\ (z_{A_{\mathrm{o}}*C} \ln |T_F^{-1}| \;:\; |T_F^{-1}| \geq z_{A_{\mathrm{o}}*C})\big)\end{aligned}$$

(b) varies with the *size* scaled *component size cardinality relative entropies* of all *transforms* for all *slices*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\; \\ \sum_{(C,F) \in \mathrm{cont}(D_{\mathrm{o}})} \big(z_{A_{\mathrm{o}}*C} \times \sum_{T \in F} \mathrm{entropyRelative}(A_{\mathrm{o}} * C * T_F, C * T_F)\big)\end{aligned}$$

so (c) varies against the *specialising fud decomposition substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\; -\, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{D,F,H}}(D_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})$$

and (d) varies against the *specialising-canonical space difference*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\; \\ -(2C_{\mathrm{G},V_{\mathrm{o}},\mathrm{D,F,H}}(D_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{H},V_{\mathrm{o}}}(H_{\mathrm{o}}) - C^{\mathrm{s}}_{\mathrm{G},V_{\mathrm{o}}}(H_{\mathrm{o}}))\end{aligned}$$

In the *near-fud-decomposition-independent* case, insofar as the *uniform possible iso-fud-decomposition history probability function* approximates to the *specialising fud decomposition history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d,p},D_{\mathrm{o}}} \approx P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}}$, conjecture that (a) the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* varies with the *log likelihood* of the *specialising fud decomposition probability distribution*,

$$\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\; \ln \hat{Q}_{\mathrm{G},D_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})$$

and (b) the *degree of structure* of the *uniform possible iso-fud-decomposition history probability function* with respect to the *specialising coder* varies with *specialising degree of structure*,

$$\begin{aligned}\mathrm{structure}(U, X)(P_{U,X,H_{\mathrm{h}},\mathrm{d,p},D_{\mathrm{o}}}, C_{\mathrm{G,D,F,H}}(D_{\mathrm{o}})) \;\sim\; \\ \mathrm{structure}(U, X)(P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}}, C_{\mathrm{G,D,F,H}}(D_{\mathrm{o}}))\end{aligned}$$

So conjecture that, in the case where the *sample history* is modal, $H_{\rm o} \in \mathrm{maxd}(P_{U,X,H_{\rm h},{\rm d},{\rm p},D_{\rm o}})$, the *log-likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* also varies with its *degree of structure*,

$$\ln \hat{Q}_{{\rm h},{\rm d},D_{\rm o},U}(A_{{\rm o},z_{\rm h}},z_{\rm o})(A_{\rm o}) \quad \sim \quad \mathrm{structure}(U,X)(P_{U,X,H_{\rm h},{\rm d},{\rm p},D_{\rm o}}, C_{\rm G,D,F,H}(D_{\rm o}))$$

In other words, when the *iso-fud-decomposition log-likelihood* is high, the expected *space* of the *specialising coder* is low and so the *compression* of the *coder* with respect to the *iso-fud-decomposition historically distributed history probability function* is high.

Further, conjecture that the *sensitivity* to *model* also varies against the *log likelihood*,

$$\begin{aligned}
- \ \ln |\mathrm{max}(\{(D, \hat{Q}_{{\rm h},{\rm d},D,U}(A_{{\rm o},z_{\rm h}},z_{\rm o})(A_{\rm o})) :& \\
D \in \mathcal{D}_{{\rm F},U,V_{\rm o}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F),& \\
A_{\rm o} \approx A_{\rm o}^{{\rm E}_{\rm D,F}(D)}\})| \ \sim& \\
- \ \ln \hat{Q}_{{\rm h},{\rm d},D_{\rm o},U}(A_{{\rm o},z_{\rm h}},z_{\rm o})(A_{\rm o})&
\end{aligned}$$

In the case where the *relative entropies* are high, $\forall (C,F) \in \mathrm{cont}(D_{\rm o})\ \forall T \in F\ (\mathrm{entropyCross}(A_{\rm o} * C * T_F, C * T_F) > \ln |T_F^{-1}|)$, the *sum sensitivity* varies against the *log likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm h},{\rm d},D_{\rm o},U}(A_{{\rm o},z_{\rm h}},z_{\rm o}))) \quad \sim \quad -\ln \hat{Q}_{{\rm h},{\rm d},D_{\rm o},U}(A_{{\rm o},z_{\rm h}},z_{\rm o})(A_{\rm o})$$

So in this case, in *classical uniform possible modelled induction*, both (a) the *sensitivity* to *distribution histogram*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm h},{\rm d},D_{\rm o},U}(A_{{\rm o},z_{\rm h}},z_{\rm o})))$$

and (b) the *sensitivity* to *model*,

$$\begin{aligned}
- \ \ln |\mathrm{max}(\{(D, \hat{Q}_{{\rm h},{\rm d},D,U}(A_{{\rm o},z_{\rm h}},z_{\rm o})(A_{\rm o})) :& \\
D \in \mathcal{D}_{{\rm F},U,V_{\rm o}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F),& \\
A_{\rm o} \approx A_{\rm o}^{{\rm E}_{\rm D,F}(D)}\})|&
\end{aligned}$$

vary against the *log-likelihood*. That is, in *classical modelled induction* in some circumstances, the optimisation of the *log-likelihood* tends to minimise the *sensitivity* to parameter.

5.5.11 Tractable transform induction

It was noted at the beginning of section 'Specialising induction' that, although the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, is computable for *uniform possible derived induction*,

$$\tilde{T}_{\mathrm{o}} \;\in\; \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger}\})$$

the computation is not tractable. Insofar as the *uniform possible iso-derived history probability function* approximates to the *specialising history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d,p},T_{\mathrm{o}}} \approx P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, consider instead computing the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, for *specialising induction*,

$$\tilde{T}_{\mathrm{o}} \;\in\; \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

or

$$\tilde{T}_{\mathrm{o}} \;\in\; \mathrm{mind}(\{(T, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T)^{\mathrm{s}}(H_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

This computation is more tractable, because there is no need to compute the *derived* function, $D_{U,\mathrm{i},T,z_{\mathrm{o}}} \in \mathcal{A}_{U,\mathrm{i},V_{\mathrm{o}},z_{\mathrm{o}}} :\to \mathcal{A}_{U,\mathrm{i},W,z_{\mathrm{o}}}$. However, it is still necessary to compute the set of *substrate transforms*, $\mathcal{T}_{U,V_{\mathrm{o}}}$, and so the computation of the minimum *coder space* is still intractable.

It is conjectured in section 'Inducers and Compression', above, that, although the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T,H}}$, is defined completely separately of the notions of *alignment* and *independence*, the properties of the minimum *coder space* are similar in many ways to the properties of the maximum *summed alignment valency-density* of the *tractable limited-models summed alignment valency-density substrate aligned non overlapping infinite-layer fud decomposition inducer*,

$$I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}} \in \mathrm{inducers}(z)$$

Given *non-independent substrate histogram* $A \in \mathcal{A}_z \setminus \{A^{\mathrm{X}}\}$, the *midising, idealising fud decomposition inducer* is defined,

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A) = \\
&\quad \{(D, I^{*}_{\approx\mathbf{R}}(\mathrm{algnValDensSum}(U_A)(A, D^{\mathrm{D}}))) : \\
&\qquad\qquad D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})), \\
&\qquad\qquad\qquad \forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A * C * F^{\mathrm{T}}) > 0)\}
\end{aligned}$$

where (i) the *limited-models fuds*, $\mathcal{F}_{\mathrm{q}}$ is the intersection of *limited-breadth*, *limited-layer*, *limited-underlying* and *limited-derived fuds*, $\mathcal{F}_{\mathrm{q}} = \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}} \cap$

$\mathcal{F}_{\mathrm{h}} \cap \mathcal{F}_{\mathrm{b}}$, (ii) $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$, (iii) $()^{\mathrm{D}} \in \mathcal{D}_{\mathrm{F}} \to \mathcal{D}$, and (iv) the *summed derived alignment valency density* $\mathrm{algnValDensSum}(U) \in \mathcal{A} \times \mathcal{D} \to \mathbf{R}$ is defined as

$$\begin{aligned}&\mathrm{algnValDensSum}(U)(A,D) := \\ &\quad \sum_{(C,T) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * T)/\mathrm{capacityValency}(U)((A * C * T)^{\mathrm{FS}})\end{aligned}$$

The *fud decomposition minimum space specialising derived search function* for *history* $H \in \mathcal{H}_{U,X}$ is defined in terms of the *expanded specialising derived history coder*, $C_{\mathrm{G,T,H}}(T) \in \mathrm{coders}(\mathcal{H}_{U,X})$, as

$$Z_{\mathrm{D,F,P,m,G,T,H}}(H) \quad = \quad \{(D, -C_{\mathrm{G,T,H}}(D^{\mathrm{T}})^{\mathrm{s}}(H)) : D \in \mathcal{D}_{\mathrm{F},U,\mathrm{P}}\}$$

It is maximised by finding the *fud decomposition* $D \in \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$ which minimises the *specialising derived substrate history coder space*, $C_{\mathrm{G},V,\mathrm{T,H}}(D^{\mathrm{PVT}})^{\mathrm{s}}(H)$ where $V = \mathrm{vars}(H)$.

The *summed alignment valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, application also defines a *fud decomposition search function*, but restricted to the *limited-models non-overlapping fud decompositions*, $\mathcal{D}_{\mathrm{F},U,\mathrm{P}} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})) \subseteq \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$. Define the *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*

$$\begin{aligned}&Z_{\mathrm{D,F,P,n,q,Sd}}(H) = \\ &\quad \{(D, I^{*}_{\approx\mathbf{R}}(\textstyle\sum \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F} : (C,F) \in \mathrm{cont}(D))) : \\ &\qquad\qquad D \in \mathcal{D}_{\mathrm{F},U,\mathrm{P}} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})),\ \mathrm{und}(D) \subseteq V, \\ &\qquad\qquad\qquad \forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A * C * F^{\mathrm{T}}) > 0)\} \ \cup \\ &\quad \{(D_{\mathrm{u}}, 0)\}\end{aligned}$$

where $W_F = \mathrm{der}(F)$, $w_F = |W_F^{\mathrm{C}}|$, $m_F = |W_F|$, $V = \mathrm{vars}(H)$, $A = \mathrm{histogram}(H)$, the *unary fud decomposition* $D_{\mathrm{u}} = \{((\emptyset, \{T_{\mathrm{u}}\}), \emptyset)\}$, and the *unary transform* $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$.

The *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}(H)$, is maximised by searching for the *fud decomposition* $D \in \mathrm{maxd}(Z_{\mathrm{D,F,P,n,q,Sd}}(H)) \subset \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$, which maximises *summed alignment valency-density*,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \quad = \quad \sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F}$$

In section 'Inducers and Compression', it is conjectured that for all finite *systems* and finite *event identifier sets* there exists a class of *limited-models fuds* such that the *search functions* are positively correlated for uniform *history probability function*,

$$\begin{aligned}&\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ (|\mathcal{H}_{U,X}| < \infty \implies\\&\quad \exists \mathcal{F}_{\mathrm{q}} \subset \mathcal{F}\ (\mathrm{covariance}(\mathcal{H}_{U,X} \times \{1/|\mathcal{H}_{U,X}|\})\\&\qquad\qquad (\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,T,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,n,q,Sd}}) \geq 0))\end{aligned}$$

The discussion considers the relations between the *summed alignment valency-density* and the *specialising space*. In particular, it is shown that the *summed alignment valency-density* (a) varies against the *derived entropy* of the *nullable transform*,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \quad \sim \quad -\ \mathrm{entropy}(A * D^{\mathrm{T}})$$

(b) varies against the *possible derived volume* $w' = |(D^{\mathrm{T}})^{-1}|$,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \quad \sim \quad 1/w'$$

(c) varies with the *expected component entropy*,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \quad \sim \quad \mathrm{entropyComponent}(A, D^{\mathrm{T}})$$

and (d) varies with the *component size cardinality relative entropy*,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \quad \sim \quad \mathrm{entropyRelative}(A * D^{\mathrm{T}}, V^{\mathrm{C}} * D^{\mathrm{T}})$$

With regard to this last relation, note that although the maximisation of the *midisation alignment* tends to minimise the *mid component size cardinality relative entropy*, $\mathrm{entropyRelative}(A * C * F^{\mathrm{T}}, C * F^{\mathrm{T}}) \approx 0$ where $(C, F) \in \mathrm{cont}(D)$, the subsequent maximisation of the *idealisation alignment* in the *super-decomposition* tends to increase the overall *relative entropy*.

Given this evidence for the correlation between the *fud decomposition minimum space specialising derived search function*, $Z_{\mathrm{D,F,P,m,G,T,H}}$, and the *tractable limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, conjecture that, in the case where the *model*, T_{o}, is *unknown*, the *maximum likelihood estimate* for the *model* for *specialising induction*,

$$\tilde{T}_{\mathrm{o}} \ \in \ \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

or

$$\tilde{T}_{\mathrm{o}} \ \in \ \mathrm{mind}(\{(T, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T)^{\mathrm{s}}(H_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

can be tractably approximated by the maximisation of the *tractable limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer*, $I_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}^{'}$,

$$\tilde{T}_{\mathrm{o}} \ \approx \ D_{\mathrm{o,Sd}}^{\mathrm{T}}$$

where

$$D_{\mathrm{o,Sd}} \ \in \ \mathrm{maxd}(I_{z_{\mathrm{o}},\mathrm{Sd,D,F},\infty,\mathrm{n,q}}^{'*}(A_{\mathrm{o}}))$$

and $A_{\mathrm{o}} \neq A_{\mathrm{o}}^{\mathrm{X}}$. The *tractable model*, $D_{\mathrm{o,Sd}}$, is defined explicitly,

$$\begin{aligned}
D_{\mathrm{o,Sd}} \ \in \ \mathrm{maxd}(\{(D, I_{\approx\mathbf{R}}^{*}(\mathrm{algnValDensSum}(U)(A_{\mathrm{o}}, D^{\mathrm{D}}))) :& \\
D \in \mathcal{D}_{\mathrm{F},\infty,U,V_{\mathrm{o}}} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})),& \\
\forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A_{\mathrm{o}} * C * F^{\mathrm{T}}) > 0)\})&
\end{aligned}$$

The approximation, $D_{\mathrm{o,Sd}}^{\mathrm{T}}$, can be compared to the *maximum likelihood estimate*, $\tilde{T}_{\mathrm{o}}$, by computing the relative entropy between *derived*,

$$\mathrm{entropyRelative}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}}, A_{\mathrm{o}} * D_{\mathrm{o,Sd}}^{\mathrm{T}})$$

The approximation improves as the relative entropy decreases.

The *accuracy* of the approximation can be defined as the ratio of the *tractable model specialising likelihood* to the *maximum model specialising likelihood*,

$$0 \ < \ \frac{\hat{Q}_{\mathrm{G},D_{\mathrm{o,Sd}}^{\mathrm{T}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \ \leq \ 1$$

The *accuracy* is computable, though not tractable and so not necessarily practicable. The definition of *accuracy* is consistent with the gradient of the likelihood function at the mode,

$$\frac{\hat{Q}_{\mathrm{G},D_{\mathrm{o,Sd}}^{\mathrm{T}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \ \sim \ \partial(L_{n,U,X,T_{\mathrm{o}}}(0))(\lambda_{\mathrm{o}})$$

So the *accuracy* varies against the *sensitivity* to *model*,

$$\begin{aligned}
\frac{\hat{Q}_{\mathrm{G},D_{\mathrm{o,Sd}}^{\mathrm{T}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \ &\sim \ -(-\partial(L_{n,U,X,T_{\mathrm{o}}}(0))(\lambda_{\mathrm{o}})) \\
&\sim \ \mathrm{structure}(U,X)(P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}, C_{\mathrm{G,T,H}}(T_{\mathrm{o}})) \\
&\sim \ -(-\ln|\mathrm{max}(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})|)
\end{aligned}$$

It was noted above that, in *specialising induction*, where $P = P_{U,X,\mathrm{G},T_\mathrm{o},\mathrm{H}}$, the maximisation of the *log likelihood* also tends to minimise the *sensitivity* to *model*,

$$-\ \ln |\max(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})) : T \in \mathcal{T}_{U,V_\mathrm{o}}\})| \quad \sim \quad -\ \ln \hat{Q}_{\mathrm{G},T_\mathrm{o},\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})$$

So, although the *maximum model specialising likelihood*, $\hat{Q}_{\mathrm{G},\tilde{T}_\mathrm{o},\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})$, appears in the denominator of the *accuracy*, the *accuracy* of the *tractable model* in fact varies with the *log-likelihood*,

$$\frac{\hat{Q}_{\mathrm{G},D^\mathrm{T}_{\mathrm{o},\mathrm{Sd}},\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})}{\hat{Q}_{\mathrm{G},\tilde{T}_\mathrm{o},\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})} \ \sim\ \ln \hat{Q}_{\mathrm{G},T_\mathrm{o},\mathrm{H},U}(z_\mathrm{o})(A_\mathrm{o})$$

That is, although the *model* obtained from the *tractable summed alignment valency-density inducer* is merely an approximation, in the cases where the *log-likelihood* is high, and so the *sensitivity* to *model* is low, the approximation may be reasonably close nonetheless.

Consider the *tractable model* obtained by maximisation of the *derived alignment valency-density* of the *tractable limited-models derived alignment valency-density substrate non-overlapping infinite-layer fud inducer*,

$$I'_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q}} \in \mathrm{inducers}(z)$$

Given *non-independent substrate histogram* $A \in \mathcal{A}_z \setminus \{A^\mathrm{X}\}$, the *midising fud inducer* is defined,

$$I'^{*}_{z,\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q}}(A) \quad = \quad \{(F, I^{*}_{\approx\mathbf{R}}(\mathrm{algn}(A * F^\mathrm{T})/w^{1/m})) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_\mathrm{n} \cap \mathcal{F}_\mathrm{q}\}$$

Let the *tractable derived alignment valency-density fud* be

$$F_{\mathrm{o},\mathrm{ad}} \in \mathrm{maxd}(I'^{*}_{z_\mathrm{o},\mathrm{ad},\mathrm{F},\infty,\mathrm{n},\mathrm{q}}(A_\mathrm{o}))$$

In order for the *inducer* to be *alignment-bounded* while tractably *lifting* to *derived transform*, it is necessary to maximise *formal-abstract equality* by the maximisation of *midisation alignment*, which is approximated by the maximisation of *derived alignment valency-density.* The maximisation of the *midisation alignment*, however, tends to minimise the *mid component size cardinality relative entropy*,

$$\mathrm{entropyRelative}(A * F^\mathrm{T}_{\mathrm{o},\mathrm{ad}}, V^\mathrm{C}_\mathrm{o} * F^\mathrm{T}_{\mathrm{o},\mathrm{ad}}) \ \approx\ 0$$

whereas the *fud decomposition*, $D_{\mathrm{o,Sd}}$, which has the *fud*, $F_{\mathrm{o,ad}}$, in its root, $\{((\emptyset, F_{\mathrm{o,ad}}), \cdot)\} = D_{\mathrm{o,Sd}}$, restores the *relative entropy* by maximisation of the *idealisation alignment* during *decomposition*,

$$\mathrm{entropyRelative}(A * D_{\mathrm{o,Sd}}^{\mathrm{T}}, V_{\mathrm{o}}^{\mathrm{C}} * D_{\mathrm{o,Sd}}^{\mathrm{T}}) \quad > \quad 0$$

so the *tractable derived alignment valency-density fud accuracy* is less than *tractable summed alignment valency-density fud decomposition accuracy*,

$$\frac{\hat{Q}_{\mathrm{G},F_{\mathrm{o,ad}}^{\mathrm{T}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \quad < \quad \frac{\hat{Q}_{\mathrm{G},D_{\mathrm{o,Sd}}^{\mathrm{T}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}$$

Consider the *tractable model* obtained by maximisation of the *derived alignment* of the *tractable limited-models derived alignment substrate non-overlapping infinite-layer fud inducer*,

$$I_{z,\mathrm{a,F},\infty,\mathrm{n,q}}^{'} \in \mathrm{inducers}(z)$$

Given *non-independent substrate histogram* $A \in \mathcal{A}_z \setminus \{A^{\mathrm{X}}\}$, the *fud inducer* is defined,

$$I_{z,\mathrm{a,F},\infty,\mathrm{n,q}}^{'*}(A) \quad = \quad \{(F, I_{\approx\mathbf{R}}^{*}(\mathrm{algn}(A * F^{\mathrm{T}}))) : F \in \mathcal{F}_{\infty,U_A,V_A} \cap \mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}}\}$$

Let the *tractable derived alignment fud* be

$$F_{\mathrm{o,a}} \in \mathrm{maxd}(I_{z_{\mathrm{o}},\mathrm{a,F},\infty,\mathrm{n,q}}^{'*}(A_{\mathrm{o}}))$$

The *component size cardinality relative entropy* of the *derived alignment fud* is sometimes higher than that of the *derived alignment valency-density fud*,

$$\mathrm{entropyRelative}(A * F_{\mathrm{o,ad}}^{\mathrm{T}}, V_{\mathrm{o}}^{\mathrm{C}} * F_{\mathrm{o,ad}}^{\mathrm{T}}) \quad < \quad \mathrm{entropyRelative}(A * F_{\mathrm{o,a}}^{\mathrm{T}}, V_{\mathrm{o}}^{\mathrm{C}} * F_{\mathrm{o,a}}^{\mathrm{T}})$$

although the *derived entropy* is sometimes higher,

$$\mathrm{entropy}(A * F_{\mathrm{o,ad}}^{\mathrm{T}}) \quad < \quad \mathrm{entropy}(A * F_{\mathrm{o,a}}^{\mathrm{T}})$$

so conjecture that the *accuracy* is sometimes also greater,

$$\frac{\hat{Q}_{\mathrm{G},F_{\mathrm{o,ad}}^{\mathrm{T}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \quad < \quad \frac{\hat{Q}_{\mathrm{G},F_{\mathrm{o,a}}^{\mathrm{T}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}$$

but the *derived alignment fud inducer*, $I_{z,\mathrm{a,F},\infty,\mathrm{n,q}}^{'}$, has limited *derived volume* with respect to the *summed alignment valency-density fud decomposition inducer*, $I_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}^{'}$, so conjecture that the *accuracy* is still less than that of the *fud decomposition inducer*,

$$\frac{\hat{Q}_{\mathrm{G},F_{\mathrm{o,a}}^{\mathrm{T}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \quad < \quad \frac{\hat{Q}_{\mathrm{G},D_{\mathrm{o,Sd}}^{\mathrm{T}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}$$

Consider the *practicable model* obtained by maximisation of the *summed shuffle content alignment valency-density* of the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*,

$$I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}} \in \mathrm{inducers}(z)$$

Given *substrate histogram* $A \in \mathcal{A}_z$, the *practicable fud decomposition inducer* is defined in section 'Optimisation', above, as

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_{\emptyset}, 0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D,F,d}})),\ \{D\} = Q
\end{aligned}$$

Let the *practicable fud decomposition* be

$$D_{\mathrm{o,Scsd},P} \in \mathrm{maxd}(I'^{*}_{z_{\mathrm{o}},\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}(A_{\mathrm{o}}))$$

The *practicable fud decomposition inducer* imposes a sequence on the search and other constraints that do not apply to the *tractable summed alignment valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, corresponding to the *limited-models summed alignment valency-density aligned non-overlapping fud decomposition search function*, $Z_{\mathrm{D,F,P,n,q,Sd}}$, so conjecture that the *accuracy* is less than that of the *tractable fud decomposition inducer*,

$$\frac{\hat{Q}_{\mathrm{G},D^{\mathrm{T}}_{\mathrm{o,Scsd},P},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \quad < \quad \frac{\hat{Q}_{\mathrm{G},D^{\mathrm{T}}_{\mathrm{o,Sd}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}$$

It is shown above in *classical uniform possible modelled induction*, where the *history probability function* is *uniform possible iso-derived historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d,p},T_{\mathrm{o}}}$, that, in the case where (i) the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, and (ii) the *maximum likelihood estimate relative entropy* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, (a) the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising derived substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

(b) the *sensitivity* to *distribution* varies against the *log likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad -\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

and (c) the *sensitivity* to *model* varies against the *log likelihood*,

$$\begin{aligned} -\ \ln|\max(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})| \ \sim \\ -\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \end{aligned}$$

Insofar as the *uniform possible iso-derived history probability function* approximates to the *specialising history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d,p},T_{\mathrm{o}}} \approx P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, conjecture that the *model*, $D_{\mathrm{o,Sd}}^{\mathrm{T}}$, obtained by the maximisation of the *tractable summed alignment valency-density inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is also a tractable approximation to the *maximum likelihood estimate* for the *model* for *uniform possible iso-derived induction*,

$$\tilde{T}_{\mathrm{o}} \ \in \ \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})$$

That is, in the *near-natural*, high *relative entropy* case, a tractable *maximum likelihood estimate* for the *model* may be obtained for *classical modelled induction* by optimisation of the *summed alignment valency-density inducer*,

$$\tilde{T}_{\mathrm{o}} \ \approx \ D_{\mathrm{o,Sd}}^{\mathrm{T}}$$

The *accuracy* of the approximation can be defined as the ratio of the *tractable model uniform possible iso-derived likelihood* to the *maximum model uniform possible iso-derived likelihood*,

$$0 \ < \ \frac{\hat{Q}_{\mathrm{h,d},D_{\mathrm{o,Sd}}^{\mathrm{T}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \ \leq \ 1$$

Just as the *tractable model specialising accuracy* varies with the *log-likelihood*, so too does the *tractable model uniform possible iso-derived accuracy*,

$$\frac{\hat{Q}_{\mathrm{h,d},D_{\mathrm{o,Sd}}^{\mathrm{T}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \ \sim \ \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

Again, in the cases where the *log-likelihood* is high, and so the *sensitivity* to *model* is low, the tractable approximation in the *near-natural*, high *relative entropy* case may be reasonably close.

5.5.12 Tractable functional definition set induction

In section 'Uniform possible derived functional definition set induction' it is shown that the *maximum likelihood estimate* for the *model*, $\tilde{F}_{\mathrm{o}}$, is computable, if an approximation is used for the *fud-independent*, $A_{\mathrm{o}}^{\mathrm{E_F}(F)}$,

$$\begin{aligned}\tilde{F}_{\mathrm{o}} \in \ & \mathrm{maxd}(\{(F, \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},F,z_{\mathrm{o}}})|} \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in D^{-1}_{U,\mathrm{i},F,z_{\mathrm{o}}}(D_{U,F,z_{\mathrm{o}}}(A_{\mathrm{o}}))} Q_{\mathrm{m},U}(A_{\mathrm{o}},z_{\mathrm{o}})(B)}) : \\ & F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F),\ A_{\mathrm{o}} = Z_{1/|F|} * \sum_{T \in F} A_{\mathrm{o}} * T_F * T_F^{\dagger}\})\end{aligned}$$

However, the computation is not tractable. Insofar as the *uniform possible iso-fud history probability function* approximates to the *specialising fud history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}} \approx P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}$, consider instead computing the *maximum likelihood estimate* for the *model*, $\tilde{F}_{\mathrm{o}}$, for *specialising fud induction*,

$$\tilde{F}_{\mathrm{o}} \in \mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})$$

or

$$\tilde{F}_{\mathrm{o}} \in \mathrm{mind}(\{(F, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F},\mathrm{H}}(F^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})$$

This computation is more tractable, because there is no need to compute the *derived set* function, $D_{U,\mathrm{i},F,z_{\mathrm{o}}}$. However, it is still necessary to compute the set of *substrate fuds*, $\mathcal{F}_{U,V_{\mathrm{o}}}$, and so the computation of the minimum *coder space* is still intractable.

It is conjectured in section 'Artificial neural networks and Compression', above, that the properties of the minimum *coder space* of the *specialising fud substrate history coder*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}$, are similar, in some supervised cases of search parameters $P \in \mathcal{L}(\mathcal{X})$ and *histogram* $A \in \mathcal{A}_{U,\mathrm{i},V,z}$, to the properties of the minimum least squares loss of the *least squares gradient descent substrate net tree searcher*, $Z_{P,A,\mathrm{gr},\mathrm{lsq}}$.

Let the set of artificial neural networks be defined

$$\mathrm{nets} := \{G : G \in \mathcal{L}(\mathrm{P}(\mathcal{V}) \times \mathcal{V} \times \mathcal{L}(\mathbf{R})),\ \forall (\cdot,(V,\cdot,Q)) \in G\ (|Q| = |V|+1)\}$$

Define the graph, $\mathrm{graph} \in \mathrm{nets} \to \mathcal{L}(\mathrm{P}(\mathcal{V}) \times \mathcal{V})$ as

$$\mathrm{graph}(G) := \{(i,(V,w)) : (i,(V,w,\cdot)) \in G\}$$

Define the real weights, $\mathrm{weights} \in \mathrm{nets} \to \mathcal{L}(\mathbf{R})$ as

$$\mathrm{weights}(G) := \mathrm{concat}(\{(i,Q) : (i,(\cdot,\cdot,Q)) \in G\})$$

Define the set of *transforms*, $\mathrm{fud}(\sigma) \in \mathrm{nets} \to \mathrm{P}(\mathcal{T}_{\mathrm{f}})$ as

$$\begin{aligned}
&\mathrm{fud}(\sigma)(G) \ := \\
&\quad \{(\{S^V \cup \{(w, \sigma(\sum_{i \in \{1 \ldots n\}} Q_i S_i \ + \ Q_{n+1}))\} : S \in \mathbf{R}^n\} \times \{1\}, \{w\}) : \\
&\qquad\qquad\qquad\qquad\qquad\qquad (\cdot, (V, w, Q)) \in G, \ n = |V|\}
\end{aligned}$$

where the activation function is $\sigma \in \mathbf{R} :\to \mathbf{R}$.

The *least squares gradient descent substrate net tree searcher* is defined

$$Z_{P,A,\mathrm{gr},\mathrm{lsq}} = \mathrm{searchTreer}(\mathrm{nets}(U,K,\sigma), P_{P,A,\mathrm{gr},\mathrm{lsq}}, \{G_R\})$$

where (i) the *substrate net set* is $\mathrm{nets}(U,V,\sigma) = \{G : G \in \mathrm{nets},\ \mathrm{fud}(\sigma)(G) \in \mathcal{F}_{\infty,U,V}\}$, (ii) the initial *substrate net* is $G_R \in \mathrm{nets}(U,K,\sigma)$, (iii) the neighbourhood function is

$$\begin{aligned}
P_{P,A,\mathrm{gr},\mathrm{lsq}}(G) = \{&G' : \mathrm{lsq}(\sigma)(A,G,K) > t, \\
&G' \in \mathrm{nets}(U,K,\sigma),\ \mathrm{graph}(G') = \mathrm{graph}(G), \\
&Q = \mathrm{weights}(G),\ Q' = \mathrm{weights}(G'), \\
&Q' = \{(i,\ Q_i - r \times \mathrm{dlsq}(\sigma)(i)(A,G,K)(Q)) : i \in \{1 \ldots |Q|\}\}\}
\end{aligned}$$

(iv) the loss threshold is $t \in \mathrm{set}(P)$, (v) the rate of descent is $r \in \mathrm{set}(P)$, (vi) the activation function is $\sigma \in \mathrm{set}(P)$, (vii) the query *variables* are $K \in \mathrm{set}(P)$, (viii) the least squares loss function of the *fud* $\mathrm{lsq} \in \mathcal{A} \times \mathcal{F} \times \mathrm{P}(\mathcal{V}) \to \mathbf{R}$ is

$$\mathrm{lsq}(A,F,K) \ := \ \sum_{(S,c) \in A * X_F} \big(c \times \sum_{i \in \{1 \ldots m\}} ((S\%W_F)^{[]}_i - (S\%(V \setminus K))^{[]}_i)^2\big)$$

(ix) the *derived dimension* is $m = |W_F| = |(V \setminus K)|$, (x) the least squares loss function of the net $\mathrm{lsq}(\sigma) \in \mathcal{A} \times \mathrm{nets} \times \mathrm{P}(\mathcal{V}) \to \mathbf{R}$ is

$$\mathrm{lsq}(\sigma)(A,G,K) \ := \ \mathrm{lsq}(A, \mathrm{fud}(\sigma)(G), K)$$

and (xi) the derivative with respect to the i-th weight is $\mathrm{dlsq}(\sigma)(i) \in \mathcal{A} \times \mathrm{nets} \times \mathrm{P}(\mathcal{V}) \to (\mathcal{L}(\mathbf{R}) \to \mathbf{R})$.

The *fud minimum space specialising fud search function* for *history* $H \in \mathcal{H}_{U,X}$ is defined in terms of the *expanded specialising fud history coder*, $C_{\mathrm{G,F,H}}(F) \in \mathrm{coders}(\mathcal{H}_{U,X})$, as

$$Z_{\mathrm{F,P,m,G,F,H}}(H) \ = \ \{(F, -C_{\mathrm{G,F,H}}(F)^{\mathrm{s}}(H)) : F \in \mathcal{F}_{U,\mathrm{P}}\}$$

It is maximised by finding the *fud* $F \in \mathcal{F}_{U,\mathrm{P}}$ which minimises the *specialising fud substrate history coder space*, $C_{\mathrm{G},V,\mathrm{F},\mathrm{H}}(F^V)^{\mathrm{s}}(H)$ where $V = \mathrm{vars}(H)$.

The *least squares gradient descent substrate net tree searcher*, $Z_{P,A,\mathrm{gr},\mathrm{lsq}}$, also defines a *fud search function*, but restricted to the *neural net substrate fud set*, $\mathcal{F}_{\infty,U,V,\sigma} = \mathcal{F}_{\infty,U,V} \cap (\mathrm{fud}(\sigma) \circ \mathrm{nets})$. Let *history* $H \in \mathcal{H}_{U,X}$ be such that its *histogram* $A = \mathrm{histogram}(H)$ satisfies the supervised constraints, of (i) real *valued variables*, (ii) *causal histogram*, and (iii) a *literal frame*, imposed by the search parameters P of the *least squares gradient descent substrate net tree searcher*, $Z_{P,A,\mathrm{gr},\mathrm{lsq}}$. Define the *least squares gradient descent fud search function* as

$$\begin{aligned}
&Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}(H) = \\
&\quad \{(\mathrm{fud}(\sigma)(G), -\mathrm{lsq}(\sigma)(A,G,K)) : Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{gr},\mathrm{lsq}})),\ \{G\} = Q\}
\end{aligned}$$

In section 'Artificial neural networks and Compression' it is conjectured that, given search parameters P, there sometimes exists a subset of *histories* $\mathcal{H}_{U,X,P} \subset \mathcal{H}_{U,X}$ satisfying the constraints of (i) real *valued variables*, (ii) *causal histogram*, (iii) a *literal frame*, and (iv) clustered *histogram* such that there is a positive correlation between the *least squares gradient descent fud search function*, $Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}$, and the *fud minimum space specialising fud search function*, $Z_{\mathrm{F},\mathrm{P},\mathrm{m},\mathrm{G},\mathrm{F},\mathrm{H}}$,

$$\mathrm{covariance}(P_{U,X,P})(\mathrm{maxr} \circ Z_{\mathrm{F},\mathrm{P},\mathrm{m},\mathrm{G},\mathrm{F},\mathrm{H}}, \mathrm{maxr} \circ Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}) \quad \geq \quad 0$$

where $P_{U,X,P} = \mathcal{H}_{U,X,P} \times \{1/|\mathcal{H}_{U,X,P}|\}$. The generalisation of a correlation to all cases of finite *systems* and finite *event identifier sets* cannot be made because the *history*, H, is not independent of the search parameters, P. Least squares gradient descent supervised neural net optimisation requires specific configuration for each *history*.

The discussion considers the relations between the negative least squares loss and the *specialising space*. In the computations of *alignment* and *entropy* that follow, the *derived variables* are *discretised* to the *values* of the label *variables*, $D = \cup\{U_v : v \in (V \setminus K)\}$. It is shown that the negative least squares loss (a) varies against the *derived entropy* of the *fud transform*,

$$-\ \mathrm{lsq}(A, F_D, K) \quad \sim \quad -\ \mathrm{entropy}(A * F_D^{\mathrm{T}})$$

(b) varies against the *effective derived volume*

$$-\ \mathrm{lsq}(A, F_D, K) \quad \sim \quad -\ |(A * F_D^{\mathrm{T}})^{\mathrm{F}}|$$

(c) varies with the *expected component entropy*,

$$-\ \mathrm{lsq}(A, F_D, K) \ \sim\ \mathrm{entropyComponent}(A, F_D^{\mathrm{T}})$$

and (d) varies with the *component size cardinality relative entropy*,

$$-\ \mathrm{lsq}(A, F_D, K) \ \sim\ \mathrm{entropyRelative}(A * F_D^{\mathrm{T}}, V^{\mathrm{C}} * F_D^{\mathrm{T}})$$

This last property only holds where the *histogram* is clustered by the label *variables*, which requires *alignment* within the query *variables*, $\mathrm{algn}(A\%K) > 0$.

The discussion goes on to consider the relations between the negative least squares loss and the *specialising fud space* with regard to the *entropy* properties by *layer*. That is, the *least squares gradient descent fud search function* is also such that (a) the *derived entropy* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (\mathrm{entropy}(A * F_{\{1\ldots i\},D}^{\mathrm{T}}) < \mathrm{entropy}(A * F_{\{1\ldots i-1\},D}^{\mathrm{T}}))$$

(b) the *effective derived volume* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (|(A * F_{\{1\ldots i\},D}^{\mathrm{T}})^{\mathrm{F}}| < |(A * F_{\{1\ldots i-1\},D}^{\mathrm{T}})^{\mathrm{F}}|)$$

(c) the *expected component entropy* increases up the *layers*,

$$\begin{aligned}&\forall i \in \{2 \ldots l\}\\ &\quad (\mathrm{entropyComponent}(A, F_{\{1\ldots i\},D}^{\mathrm{T}})\ >\ \mathrm{entropyComponent}(A, F_{\{1\ldots i-1\},D}^{\mathrm{T}}))\end{aligned}$$

and (d) the *component size cardinality relative entropy* increases up the *layers*,

$$\begin{aligned}&\forall i \in \{2 \ldots l\}\\ &\quad (\mathrm{entropyRelative}(A * F_{\{1\ldots i\},D}^{\mathrm{T}}, V_D^{\mathrm{C}} * F_{\{1\ldots i\},D}^{\mathrm{T}})\ >\\ &\qquad\qquad \mathrm{entropyRelative}(A * F_{\{1\ldots i-1\},D}^{\mathrm{T}}, V_D^{\mathrm{C}} * F_{\{1\ldots i-1\},D}^{\mathrm{T}}))\end{aligned}$$

Given this evidence for the correlation in some cases between the *fud minimum space specialising fud search function*, $Z_{\mathrm{F,P,m,G,F,H}}$, and the *least squares gradient descent fud search function*, $Z_{\mathrm{F,P},P,\mathrm{gr,lsq}}$, conjecture that, in the case where the *model*, F_{o}, is *unknown*, the *maximum likelihood estimate* for the *model* for *specialising induction*,

$$\tilde{F}_{\mathrm{o}}\ \in\ \mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})$$

or

$$\tilde{F}_{\mathrm{o}} \in \mathrm{mind}(\{(F, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F},\mathrm{H}}(F^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})$$

can be tractably approximated by the maximisation of the *least squares gradient descent fud search function*,

$$\tilde{F}_{\mathrm{o}} \approx F_{\mathrm{o,gr,lsq}}$$

where

$$F_{\mathrm{o,gr,lsq}} \in \mathrm{maxd}(Z_{\mathrm{F},\mathrm{P},P_{\mathrm{o}},\mathrm{gr,lsq}}(H_{\mathrm{o}}))$$

and the search parameters, P_{o}, are configured for the given *sample history*, H_{o}.

The *tractable model* is defined explicitly, $F_{\mathrm{o,gr,lsq}} = \mathrm{fud}(\sigma)(G)$ where $\{G\} = \mathrm{leaves}(\mathrm{tree}(Z_{P_{\mathrm{o}},A_{\mathrm{o}},\mathrm{gr,lsq}}))$.

The *accuracy* of the approximation can be defined as the ratio of the *tractable model specialising likelihood* to the *maximum model specialising likelihood*,

$$0 < \frac{\hat{Q}_{\mathrm{G},F_{\mathrm{o,gr,lsq}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{F}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \leq 1$$

The *accuracy* varies against the *sensitivity* to *model*,

$$\frac{\hat{Q}_{\mathrm{G},F_{\mathrm{o,gr,lsq}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{F}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \sim -(-\ln|\max(\{(F, \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})|)$$

and varies with the *log-likelihood*,

$$\frac{\hat{Q}_{\mathrm{G},F_{\mathrm{o,gr,lsq}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{F}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \sim \ln \hat{Q}_{\mathrm{G},F_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})$$

That is, although the *model* obtained from the *least squares gradient descent fud search function* is merely an approximation, in the cases where the *log-likelihood* is high, and so the *sensitivity* to *model* is low, the approximation may be reasonably close nonetheless.

It is shown above in *classical uniform possible fud induction*, where the *history probability function* is *uniform possible iso-fud historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}}$, that, in the case where (i) the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-independent*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E}_{\mathrm{F}}(F_{\mathrm{o}})}$, and (ii) the *maximum likelihood estimate relative entropies* are high, $\forall T \in F_{\mathrm{o}}\ (\mathrm{entropyCross}(A_{\mathrm{o}} * T_{F_{\mathrm{o}}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{F_{\mathrm{o}}}) > \ln |T_{F_{\mathrm{o}}}^{-1}|)$, (a) the *log likelihood* of the *iso-fud conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising fud substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim\ -\ C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F},\mathrm{H}}(F_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})$$

(b) the *sensitivity* to *distribution* varies against the *log likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \ \sim\ -\ln \hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

and (c) the *sensitivity* to *model* varies against the *log likelihood*,

$$\begin{aligned} -\ \ln |\max(\{(F, \hat{Q}_{\mathrm{h},\mathrm{d},F,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) :\ & \\ F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F),\ A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E}_{\mathrm{F}}(F)}\})| \ \sim & \\ -\ \ln \hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) & \end{aligned}$$

Insofar as the *uniform possible iso-fud history probability function* approximates to the *specialising history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},F_{\mathrm{o}}} \approx P_{U,X,\mathrm{G},F_{\mathrm{o}},\mathrm{H}}$, conjecture that the *model*, $F_{\mathrm{o},\mathrm{gr},\mathrm{lsq}}$, obtained by the maximisation of the *least squares gradient descent fud search function*, $Z_{\mathrm{F},\mathrm{P},P,\mathrm{gr},\mathrm{lsq}}$, is also a tractable approximation to the *maximum likelihood estimate* for the *model* for *uniform possible iso-fud induction*,

$$\begin{aligned} \tilde{F}_{\mathrm{o}} \ \in\ \mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{h},\mathrm{d},F,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) :\ & \\ F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F),\ A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E}_{\mathrm{F}}(F)}\}) & \end{aligned}$$

That is, in the *near-natural*, high *relative entropy* case, a tractable *maximum likelihood estimate* for the *model* may be obtained for *classical modelled induction* by optimisation of the *least squares gradient descent fud search*,

$$\tilde{F}_{\mathrm{o}} \ \approx\ F_{\mathrm{o},\mathrm{gr},\mathrm{lsq}}$$

The *accuracy* of the approximation can be defined as the ratio of the *tractable model uniform possible iso-fud likelihood* to the *maximum model uniform possible iso-fud likelihood*,

$$0 \ <\ \frac{\hat{Q}_{\mathrm{h},\mathrm{d},F_{\mathrm{o},\mathrm{gr},\mathrm{lsq}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h},\mathrm{d},\tilde{F}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \ \leq\ 1$$

Just as the *tractable model specialising accuracy* varies with the *log-likelihood*, so too does the *tractable model uniform possible iso-fud accuracy*,

$$\frac{\hat{Q}_{\mathrm{h,d},F_{\mathrm{o,gr,lsq}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{F}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \sim \ln \hat{Q}_{\mathrm{h,d},F_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

Again, in the cases where the *log-likelihood* is high, and so the *sensitivity* to *model* is low, the tractable approximation in the *near-natural*, high *relative entropy* case may be reasonably close.

5.5.13 Tractable functional definition set decomposition induction

In section 'Uniform possible derived functional definition set decomposition induction' it is shown that the *maximum likelihood estimate* for the *model*, $\tilde{D}_{\mathrm{o}}$, is computable, if an approximation is used for the *fud-decomposition-independent*, $A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)}$,

$$\begin{aligned}
\tilde{D}_{\mathrm{o}} \in\ & \\
& \mathrm{maxd}(\{(D, \frac{1}{|\mathrm{ran}(D_{U,\mathrm{i},D,\mathrm{F},z_{\mathrm{o}}})|} \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in D_{U,\mathrm{i},D,\mathrm{F},z_{\mathrm{o}}}^{-1}(D_{U,D,\mathrm{F},z_{\mathrm{o}}}(A_{\mathrm{o}}))} Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(B)}) : \\
& \quad D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F), \\
& \quad A_{\mathrm{o}} = Z_{z_{\mathrm{o}}} * \left(\sum_{(C,F) \in \mathrm{cont}(D)} \left(Z_{1/|F|} * \sum_{T \in F} A_{\mathrm{o}} * C * T_F * T_F^{\dagger} \right) \right)^{\wedge} \})
\end{aligned}$$

However, the computation is not tractable. Insofar as the *uniform possible iso-fud-decomposition history probability function* approximates to the *specialising fud decomposition history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d,p},D_{\mathrm{o}}} \approx P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}}$, consider instead computing the *maximum likelihood estimate* for the *model*, $\tilde{D}_{\mathrm{o}}$, for *specialising fud decomposition induction*,

$$\tilde{D}_{\mathrm{o}} \in \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{G},D,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}}\})$$

or

$$\tilde{D}_{\mathrm{o}} \in \mathrm{mind}(\{(D, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{D,F,H}}(D^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})) : D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}}\})$$

This computation is more tractable, because there is no need to compute the *component derived set* function, $D_{U,\mathrm{i},D,\mathrm{F},z_{\mathrm{o}}}$. However, it is still necessary to compute the set of *substrate fud decompositions*, $\mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}}$, and so the computation of the minimum *coder space* is still intractable.

Section ‘Tractable transform induction’, above, also considers tractable approximations to the *model* for *uniform possible derived induction*, where the *model* is a *transform* instead of a *fud decomposition*. There it is shown that there are *tractable* and *practicable inducers* that have *entropy* properties similar to the *entropy* properties of the *specialising coder*. The *tractable models* then approximate to the *maximum likelihood estimate model* in *specialising induction* and thence *uniform possible derived induction*.

Section ‘Tractable functional definition set induction’, above, extended the *model* from *transforms* to *fuds*. Rather than approximating to *tractable models* derived from *inducers*, it is shown that, in some cases, artificial neural networks can provide approximations to the *model* in *specialising fud induction* and thence *uniform possible derived fud induction*.

Now the *model* is extended to *fud decompositions*. *Tractable functional definition set decomposition induction* is more closely related to *tractable transform induction* than *tractable fud induction* because again *tractable* and *practicable inducers* are shown to provide approximations to the *model* in *specialising fud decomposition induction* and thence *uniform possible derived fud decomposition induction*.

It is conjectured in section ‘Inducers and Compression’, above, that the properties of the minimum *coder space* of the *specialising fud decomposition substrate history coder*, $C_{\mathrm{G},V,\mathrm{D},\mathrm{F},\mathrm{H}}$, are similar in many ways to the properties of the maximum *summed shuffle content alignment valency-density* of the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*,

$$I'_{z,\mathrm{Scsd},\mathrm{D},\mathrm{F},\infty,\mathrm{q},P,\mathrm{d}} \in \mathrm{inducers}(z)$$

Given parameter tuple $P \in \mathcal{L}(\mathcal{X})$ and *substrate histogram* $A \in \mathcal{A}_z$, the *practicable fud decomposition inducer* is defined in section ‘Optimisation’, above, as

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd},\mathrm{D},\mathrm{F},\infty,\mathrm{q},P,\mathrm{d}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_{\emptyset}, 0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D},\mathrm{F},\mathrm{d}})),\ \{D\} = Q
\end{aligned}$$

where the *summed shuffle content alignment valency-density computer* $I_{\mathrm{Scsd}} \in$ computers is defined as

$$\begin{aligned}
&I^{*}_{\mathrm{Scsd}}((A,D)) = \\
&\quad \sum (I^{*}_{\mathrm{a}}(A * C * F^{\mathrm{T}}) - I^{*}_{\mathrm{a}}((A * C)_{R(A*C)} * F^{\mathrm{T}}))/I^{*}_{\mathrm{cvl}}(F) : (C,F) \in \mathrm{cont}(D)
\end{aligned}$$

The *fud decomposition minimum space specialising fud decomposition search function* is defined in terms of the *expanded specialising fud decomposition history coder* $C_{\mathrm{G,D,F,H}}(D) \in \mathrm{coders}(\mathcal{H}_{U,X})$,

$$Z_{\mathrm{D,F,P,m,G,D,F,H}}(H) \;=\; \{(D, -C_{\mathrm{G,D,F,H}}(D)^{\mathrm{s}}(H)) : D \in \mathcal{D}_{\mathrm{F},U,\mathrm{P}}\}$$

The *search function* is maximised by finding the *fud decomposition* $D \in \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$ which minimises the *specialising fud decomposition substrate history coder space*, $C_{\mathrm{G},V,\mathrm{D,F,H}}(D^V)^{\mathrm{s}}(H)$ where $V = \mathrm{vars}(H)$.

The *highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}$, also defines a *fud decomposition search function*. Define the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*

$$\begin{aligned}
&Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}(H) = \\
&\quad \{(D, I^*_{\mathrm{Scsd}}((A_H, D))) : Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A_H,\mathrm{D,F,d}})),\ Q \neq \emptyset,\ \{D\} = Q\}\ \cup \\
&\quad \{(D_{\mathrm{u}}, 0)\}
\end{aligned}$$

where *unary fud decomposition* $D_{\mathrm{u}} = \{((\emptyset, \{T_{\mathrm{u}}\}), \emptyset)\}$.

In section 'Inducers and Compression', it is conjectured that for all finite *systems* and finite *event identifier sets* there exists a tuple of parameters such that the *search functions* are positively correlated for uniform *history probability function*,

$$\begin{aligned}
&\forall U \in \mathcal{U}\ \forall X \subset \mathcal{X}\ (|\mathcal{H}_{U,X}| < \infty \implies \\
&\quad \exists P \in \mathcal{L}(\mathcal{X})\ (\mathrm{covariance}(P_{U,X}) \\
&\qquad\qquad (\mathrm{maxr} \circ Z_{\mathrm{D,F,P,m,G,D,F,H}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}) \geq 0))
\end{aligned}$$

The discussion considers the relations between the *summed shuffle content alignment valency-density* and the *specialising space*.

Depending on the parameters, P, which imply a set of *limited-models*, $\mathcal{F}_{\mathrm{q}} \subset \mathcal{F}$, there is a high correlation between the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer* and the *tractable summed alignment valency-density decomposition inducer*,

$$\mathrm{covariance}(P_{U,X})(\mathrm{maxr} \circ Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,n,q,Sd}})$$

so, given the relations between the *summed alignment valency-density* and the *specialising space*, it is conjectured that the *summed shuffle content alignment valency-density* (a) varies against the *derived entropy* of the *nullable* *transform*,

$$I^{*}_{\mathrm{Scsd}}((A,D)) \sim -\,\mathrm{entropy}(A * D^{\mathrm{T}})$$

(b) varies against the *possible derived volume* $w' = |(D^{\mathrm{T}})^{-1}|$,

$$I^{*}_{\mathrm{Scsd}}((A,D)) \sim 1/w'$$

(c) varies with the *expected component entropy*,

$$I^{*}_{\mathrm{Scsd}}((A,D)) \sim \mathrm{entropyComponent}(A, D^{\mathrm{T}})$$

and (d) varies with the *component size cardinality relative entropy*,

$$I^{*}_{\mathrm{Scsd}}((A,D)) \sim \mathrm{entropyRelative}(A * D^{\mathrm{T}}, V^{\mathrm{C}} * D^{\mathrm{T}})$$

The discussion goes on to consider the relations between the *summed shuffle content alignment valency-density* and the *specialising fud decomposition space* with regard to the *entropy* properties by *layer* and *slice*. That is, the *summed shuffle content alignment valency-density* is such that within each *slice*, $(C,F) \in \mathrm{cont}(D)$, (a) the *derived entropy* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (\mathrm{entropy}(A * C * F^{\mathrm{T}}_{\{1 \ldots i\}}) < \mathrm{entropy}(A * C * F^{\mathrm{T}}_{\{1 \ldots i-1\}}))$$

(b) the *derived volume* decreases up the *layers*,

$$\forall i \in \{2 \ldots l\}\ (|W^{\mathrm{C}}_{F,i}| < |W^{\mathrm{C}}_{F,i-1}|)$$

(c) the *expected component entropy* increases up the *layers*,

$$\begin{aligned}&\forall i \in \{2 \ldots l\}\\ &\quad (\mathrm{entropyComponent}(A * C, F^{\mathrm{T}}_{\{1 \ldots i\}}) > \mathrm{entropyComponent}(A * C, F^{\mathrm{T}}_{\{1 \ldots i-1\}}))\end{aligned}$$

and (d) the *component size cardinality relative entropy* increases up the *layers*,

$$\begin{aligned}&\forall i \in \{2 \ldots l\}\\ &\quad (\mathrm{entropyRelative}(A * C * F^{\mathrm{T}}_{\{1 \ldots i\}}, C * F^{\mathrm{T}}_{\{1 \ldots i\}}) >\\ &\qquad\qquad \mathrm{entropyRelative}(A * C * F^{\mathrm{T}}_{\{1 \ldots i-1\}}, C * F^{\mathrm{T}}_{\{1 \ldots i-1\}}))\end{aligned}$$

Given this evidence for the correlation between the *fud decomposition minimum space specialising fud decomposition search function*, $Z_{\mathrm{D,F,P,m,G,D,F,H}}$, and the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, conjecture that, in the case where the *model*, $D_{\rm o}$, is *unknown*, the *maximum likelihood estimate* for the *model* for *specialising fud decomposition induction*,

$$\tilde{D}_{\rm o} \ \in \ \mathrm{maxd}(\{(D, \hat{Q}_{\rm G,\mathit{D},H,\mathit{U}}(z_{\rm o})(A_{\rm o})) : D \in \mathcal{D}_{{\rm F},U,V_{\rm o}}\})$$

or

$$\tilde{D}_{\rm o} \ \in \ \mathrm{mind}(\{(D, C_{{\rm G},V_{\rm o},{\rm D,F,H}}(D^{V_{\rm o}})^{\rm s}(H_{\rm o})) : D \in \mathcal{D}_{{\rm F},U,V_{\rm o}}\})$$

can be tractably approximated by the maximisation of the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,{\rm Scsd,D,F},\infty,{\rm q},P,{\rm d}}$,

$$\tilde{D}_{\rm o} \ \approx \ D_{{\rm o,Scsd},P}$$

where

$$D_{{\rm o,Scsd},P} \ \in \ \mathrm{maxd}(I'^{*}_{z_{\rm o},{\rm Scsd,D,F},\infty,{\rm q},P,{\rm d}}(A_{\rm o}))$$

The *tractable model* is defined explicitly, $\{D_{{\rm o,Scsd},P}\} = \mathrm{leaves}(\mathrm{tree}(Z_{P,A_{\rm o},{\rm D,F,d}}))$.

The *accuracy* of the approximation can be defined as the ratio of the *tractable model specialising likelihood* to the *maximum model specialising likelihood*,

$$0 \ < \ \frac{\hat{Q}_{{\rm G},D_{{\rm o,Scsd},P},{\rm H},U}(z_{\rm o})(A_{\rm o})}{\hat{Q}_{{\rm G},\tilde{D}_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o})} \ \leq \ 1$$

The *accuracy* varies against the *sensitivity* to *model*,

$$\frac{\hat{Q}_{{\rm G},D_{{\rm o,Scsd},P},{\rm H},U}(z_{\rm o})(A_{\rm o})}{\hat{Q}_{{\rm G},\tilde{D}_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o})} \ \sim \ -(-\ln|\max(\{(D, \hat{Q}_{{\rm G},D,{\rm H},U}(z_{\rm o})(A_{\rm o})) : D \in \mathcal{D}_{{\rm F},U,V_{\rm o}}\})|)$$

and varies with the *log-likelihood*,

$$\frac{\hat{Q}_{{\rm G},D_{{\rm o,Scsd},P},{\rm H},U}(z_{\rm o})(A_{\rm o})}{\hat{Q}_{{\rm G},\tilde{D}_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o})} \ \sim \ \ln \hat{Q}_{{\rm G},D_{\rm o},{\rm H},U}(z_{\rm o})(A_{\rm o})$$

That is, although the *model* obtained from the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer* is merely an approximation, in the cases where the *log-likelihood* is high, and so the *sensitivity* to *model* is low, the approximation may be reasonably close nonetheless.

Given that, depending on the parameters, P, there is a high correlation between the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer* and the *tractable summed alignment valency-density decomposition inducer*,

$$\mathrm{covariance}(P_{U,X})(\mathrm{maxr} \circ Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}, \mathrm{maxr} \circ Z_{\mathrm{D,F,P,n,q,Sd}})$$

consider the *tractable model* obtained by maximisation of the *summed alignment valency-density* of the *tractable summed alignment valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$,

$$D_{\mathrm{o,Sd}} \in \mathrm{maxd}(I'^{*}_{z_{\mathrm{o}},\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A_{\mathrm{o}}))$$

The *tractable model*, $D_{\mathrm{o,Sd}}$, is defined explicitly,

$$\begin{aligned} D_{\mathrm{o,Sd}} \in \mathrm{maxd}(\{(D, I^{*}_{\approx\mathbf{R}}(\mathrm{algnValDensSum}(U)(A_{\mathrm{o}}, D^{\mathrm{D}}))) :\\ D \in \mathcal{D}_{\mathrm{F},\infty,U,V_{\mathrm{o}}} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})),\\ \forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A_{\mathrm{o}} * C * F^{\mathrm{T}}) > 0)\}) \end{aligned}$$

The *entropy* properties of the *tractable fud decomposition inducer* do not depend directly on the *transforms* of the *fuds* of the *decomposition*, only on the *transform* of the *decomposition*, $D_{\mathrm{o,Sd}}^{\mathrm{T}}$. There is no constraint that the *derived entropy* and the *possible derived volume* decreases up the *layers*, nor any constraint that the *expected component entropy* increases up the *layers*. There is no sense that the *fuds* are built *layer* by *layer* in sequence. Hence conjecture that the *accuracy* is less than that of the *practable fud decomposition inducer*,

$$\frac{\hat{Q}_{\mathrm{G},D_{\mathrm{o,Sd}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{D}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} < \frac{\hat{Q}_{\mathrm{G},D_{\mathrm{o,Scsd},P},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{D}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}$$

It is shown above in *classical uniform possible fud decomposition induction*, where the *history probability function* is *uniform possible iso-fud-decomposition historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d,p},D_{\mathrm{o}}}$, that, in the case where (i) the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *fud-decomposition-independent*, $A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D_{\mathrm{o}})}$, and (ii) the *maximum likelihood estimate relative entropies* are high, $\forall (C,F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (\mathrm{entropyCross}(A_{\mathrm{o}} * C * T_F, C * T_F) > \ln |T_F^{-1}|)$, (a) the *log likelihood* of the *iso-fud-decomposition conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising fud substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim -\ C_{\mathrm{G},V_{\mathrm{o}},\mathrm{D,F,H}}(D_{\mathrm{o}}^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})$$

(b) the *sensitivity* to *distribution* varies against the *log likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim -\ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

and (c) the *sensitivity* to *model* varies against the *log likelihood*,

$$\begin{aligned}
-\ \ln |\max(\{(D, \hat{Q}_{\mathrm{h,d},D,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) :\\
D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F),\\
A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)}\})|\ \sim\\
-\ \ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})
\end{aligned}$$

Insofar as the *uniform possible iso-fud-decomposition history probability function* approximates to the *specialising history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d,p},D_{\mathrm{o}}} \approx P_{U,X,\mathrm{G},D_{\mathrm{o}},\mathrm{H}}$, conjecture that the *model*, $D_{\mathrm{o,Scsd},P}$, obtained by the maximisation of the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}$, is also a tractable approximation to the *maximum likelihood estimate* for the *model* for *uniform possible iso-fud-decomposition induction*,

$$\begin{aligned}
\tilde{D}_{\mathrm{o}} \in \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{h,d},D,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) :\\
D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F),\\
A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)}\})
\end{aligned}$$

That is, in the *near-natural*, high *relative entropy* case, a tractable *maximum likelihood estimate* for the *model* may be obtained for *classical modelled induction* by optimisation of the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*,

$$\tilde{D}_{\mathrm{o}} \approx D_{\mathrm{o,Scsd},P}$$

The *accuracy* of the approximation can be defined as the ratio of the *tractable model uniform possible iso-fud-decomposition likelihood* to the *maximum model uniform possible iso-fud-decomposition likelihood*,

$$0 < \frac{\hat{Q}_{\mathrm{h,d},D_{\mathrm{o,Scsd},P},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{D}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \leq 1$$

Just as the *tractable model specialising accuracy* varies with the *log-likelihood*, so too does the *tractable model uniform possible iso-fud-decomposition accuracy*,

$$\frac{\hat{Q}_{\mathrm{h,d},D_{\mathrm{o,Scsd},P},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{D}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \sim \ln \hat{Q}_{\mathrm{h,d},D_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

Again, in the cases where the *log-likelihood* is high, and so the *sensitivity* to *model* is low, the tractable approximation in the *near-natural*, high *relative entropy* case may be reasonably close.

5.6 Aligned induction

Having considered the case of *classical non-modelled induction*, where the *history probability function* is *historically distributed*, $P = P_{U,X,H_{\mathrm{h}}}$, now consider the special case of *aligned non-modelled induction*.

In *aligned induction*, while the *distribution probability histogram*, $\hat{E}_{\mathrm{h}}$, remains *unknown*, the *independent distribution probability histogram*, $\hat{E}_{\mathrm{h}}^{\mathrm{X}}$, is *necessary*. Now the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have an *independent probability histogram* equal to the *reduced independent distribution probability histogram*, $\hat{A}_H^{\mathrm{X}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}}\%V_H$. Define the *iso-independent historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{y}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$,

$$
\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\mathrm{y}} := \\
&\quad \Big(\bigcup \{\{(H,1) : H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H,\ \hat{A}_H^{\mathrm{X}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}}\%V_H\}^{\wedge} : \\
&\qquad\qquad\qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ \hat{A}_H^{\mathrm{X}} \neq \hat{E}_{\mathrm{h}}^{\mathrm{X}}\%V_H\} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \nsubseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset,0)\}
\end{aligned}
$$

That is, *drawn histories necessarily* have normalised *independent histogram* equal to that of the *distribution histogram*, $\forall H \in \mathcal{H}_{U,X}\ (P_{U,X,H_{\mathrm{h}},\mathrm{y}}(H) > 0 \implies \hat{A}_H^{\mathrm{X}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}}\%V_H)$.

In *aligned induction* the *history probability function* is *iso-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y}}$.

The *independent distribution probability histogram reduced* to *observation variables*, $\hat{E}_{\mathrm{o}}^{\mathrm{X}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}}\%V_{\mathrm{o}}$, is *known*, $\hat{E}_{\mathrm{o}}^{\mathrm{X}} = \hat{A}_{\mathrm{o}}^{\mathrm{X}}$.

Given a *drawn history* $H \in \mathcal{H}_{U,X}$, where $P_{U,X,H_{\mathrm{h}},\mathrm{y}}(H) > 0$, the *iso-independent historical probability* of *histogram* $A_H = \mathrm{histogram}(H) + V_H^{\mathrm{CZ}} \in \mathcal{A}_{U,\mathrm{i},V_H,z_H}$ is now conditional,

$$
\frac{Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H)}{\sum_{B \in Y_{U,\mathrm{i},V_{\mathrm{o}},z_H}^{-1}(A_H^{\mathrm{X}})} Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(B)} = \frac{\sum P_{U,X,H_{\mathrm{h}},\mathrm{y}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\mathrm{y}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}
$$

The *iso-derived historical probability* may be expressed in terms of a *histogram distribution* which is not explicitly conditional on the *necessary independent*,

$\hat{E}_{\mathrm{o}}^{\mathrm{X}}$,

$$\hat{Q}_{\mathrm{h,y},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) \quad \propto \quad \sum(P_{U,X,H_{\mathrm{h}},\mathrm{y}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *iso-independent conditional stuffed historical probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\mathrm{h,y},U}(E,z) \\
&:= \quad \{(A, \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\
&\qquad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \not\leq E\}
\end{aligned}$$

which is defined if $z \leq \mathrm{size}(E)$. The *independent histogram* valued *integral histogram* function $Y_{U,\mathrm{i},V,z}$ is defined

$$Y_{U,\mathrm{i},V,z} = \{(A, A^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\} \subset \mathrm{independent}$$

The finite set of *iso-independents* of *independent histogram* A^{X} is

$$Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} = A^{\mathrm{X}}\}$$

In the case where all the *independent* are *possible*,

$$\forall A' \in \mathrm{ran}(Y_{U,\mathrm{i},V,z})\ \exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ ((A^{\mathrm{X}} = A') \ \wedge\ (A \leq E))$$

the normalisation of the *iso-derived conditional stuffed historical probability distribution* is a fraction $1/|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|$,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h,y},U}(E,z) \\
&= \quad \{(A, \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|} \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}
\end{aligned}$$

The corresponding *iso-independent conditional generalised multinomial probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\mathrm{m,y},U}(E,z) \\
&:= \quad \{(A, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \leq E^{\mathrm{F}}\}^{\wedge} \cup \\
&\qquad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \not\leq E^{\mathrm{F}}\}
\end{aligned}$$

which is defined if $\mathrm{size}(E) > 0$.

The case where all the *independent* are *possible* is weaker than for *historical*,

$$\forall A' \in \mathrm{ran}(Y_{U,\mathrm{i},V,z})\ \exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ ((A^{\mathrm{X}} = A')\ \wedge\ (A^{\mathrm{F}} \leq E^{\mathrm{F}}))$$

In this case the *iso-independent conditional generalised multinomial probability distribution* is

$$\begin{aligned}&\hat{Q}_{\mathrm{m,y},U}(E,z)\\ &\quad = \ \{(A, \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|}\frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}\end{aligned}$$

Assume that the *distribution history size*, z_{h}, is large with respect to the *sample size* $z_{\mathrm{o}} = \mathrm{size}(A_{\mathrm{o}})$, so that, in the limit, the *iso-independent conditional stuffed historical probability*, $\hat{Q}_{\mathrm{h,y},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$, approximates to the *iso-independent conditional multinomial probability*, $\hat{Q}_{\mathrm{m,y},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$. That is, if $z_{\mathrm{o}} \ll z_{\mathrm{h}}$ then

$$\hat{Q}_{\mathrm{h,y},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \approx \hat{Q}_{\mathrm{m,y},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

The *iso-independent multinomial parameterised probability density function*, $\mathrm{myppdf}(z) \in \mathrm{ppdfs}(v,v)$, and *iso-independent multinomial likelihood function*, $\mathrm{mylf}(z) \in \mathrm{lfs}(v,v)$, corresponding to the *iso-independent multinomial probability distribution*, $\hat{Q}_{\mathrm{m,y},U}$, are not given explicitly here, but are such that

$$\mathrm{myppdf}(z)(\hat{E}^{\square})(A^{\square}) = \mathrm{mylf}(z)(A^{\square})(\hat{E}^{\square}) = \hat{Q}_{\mathrm{m,y},U}(E,z)(A)$$

Now in the case of *aligned induction* the *real maximum likelihood estimate* $\tilde{E}'_{\mathrm{o}} \in \mathbf{R}^{v_{\mathrm{o}}}_{(0,1)}$ for the parameter of the *iso-independent multinomial parameterised probability density function* is

$$\{\tilde{E}'_{\mathrm{o}}\} = \mathrm{maxd}(\mathrm{mylf}(z_{\mathrm{o}})(A^{\square}_{\mathrm{o}}))$$

which is such that $\forall i \in \{1 \ldots v_{\mathrm{o}}\}\ (\partial_i(\mathrm{mylf}(z_{\mathrm{o}})(A^{\square}_{\mathrm{o}}))(\tilde{E}'_{\mathrm{o}}) = 0)$. The *maximum likelihood estimate* $\tilde{E}'_{\mathrm{o}}$ is only defined in the case where the *sample histogram* is *completely effective*, $A^{\mathrm{F}}_{\mathrm{o}} = V^{\mathrm{C}}_{\mathrm{o}} \implies \hat{A}^{\square}_{\mathrm{o}} \in \mathbf{R}^{v_{\mathrm{o}}}_{(0,1)}$, because the *binomial likelihood function* is only defined for the open set. That is, $\mathrm{d}(\mathrm{blf}(z_{\mathrm{o}})(0))$ is undefined and so the derivative of the *iso-independent multinomial parameterised probability density function* is undefined where there are *ineffective states*.

In the case of *completely effective sample histogram*, $A_{\rm o}^{\rm F} = V_{\rm o}^{\rm C}$, the maximisation of the *iso-independent conditional generalised multinomial probability* parameterised by the *complete congruent histograms* of unit *size* is a singleton of the *rational maximum likelihood estimate*

$$\{\tilde{E}_{\rm o}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\rm m,y,\mathit{U}}(E, z_{\rm o})(A_{\rm o})) : E \in \mathcal{A}_{U,V_{\rm o},1}\})$$

The *real maximum likelihood estimate*, $\tilde{E}'_{\rm o}$, is not necessarily a rational coordinate, $\mathbf{R}_{(0,1)}^{v_{\rm o}} \supset \mathbf{Q}_{(0,1)}^{v_{\rm o}}$, and so the *rational maximum likelihood estimate* is not necessarily equal to the *real maximum likelihood estimate*. However, it is conjectured that the maximisation of the *distribution* approximates to the maximisation of the *likelihood function*,

$$\tilde{E}_{\rm o}^{\square} \approx \tilde{E}'_{\rm o}$$

In the case where the *sample histogram* is not *completely effective*, $A_{\rm o}^{\rm F} < V_{\rm o}^{\rm C}$, the maximisation of the *iso-independent conditional generalised multinomial probability distribution* is well defined, unlike the *parameterised probability density function*, but is not necessarily a singleton

$$|\max(\{(E, \hat{Q}_{\rm m,y,\mathit{U}}(E, z_{\rm o})(A_{\rm o})) : E \in \mathcal{A}_{U,V_{\rm o},1}\})| \geq 1$$

In the case where the maximisation of the *iso-independent conditional generalised multinomial probability distribution* is a singleton, it is equal to the *normalised dependent*, $\tilde{E}_{\rm o} = \hat{A}_{\rm o}^{\rm Y}$, where the *dependent* $A^{\rm Y} \in \mathcal{A}_{U,V,z}$ is defined in 'Likely histograms', above, as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-independent*,

$$\{A^{\rm Y}\} = \mathrm{maxd}(\{(D, \frac{Q_{\rm m,\mathit{U}}(D, z)(A)}{\sum Q_{\rm m,\mathit{U}}(D, z)(B) : B \in Y_{U,{\rm i},V,z}^{-1}(A^{\rm X})}) : D \in \mathcal{A}_{U,V,z}\})$$

The *dependent*, $A^{\rm Y}$, is sometimes not computable. The finite approximation to the *dependent* is

$$\{A_k^{\rm Y}\} = \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\rm m,\mathit{U}}(D, z)(A)}{\sum Q_{\rm m,\mathit{U}}(D, z)(B) : B \in Y_{U,{\rm i},V,z}^{-1}(A^{\rm X})}) : D \in \mathcal{A}_{U,{\rm i},V,kz}\})$$

The approximation, $A_k^{\rm Y} \approx A^{\rm Y}$, improves as the scaling factor, k, increases.

Unlike in *classical non-modelled induction* where the *maximum likelihood* *estimate*, $\tilde{E}_{\mathrm{o}}$, is equal to the *sample probability histogram*, $\hat{A}_{\mathrm{o}}$, in *aligned non-modelled induction* the *maximum likelihood estimate* is not necessarily equal to the *sample probability histogram*. In the case where the *sample histogram* is *independent* the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{X}} \implies A_{\mathrm{o}}^{\mathrm{Y}} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

In general, the overall *maximum likelihood estimate*, which is the *dependent*, is near the *histogram*, $\tilde{E}_{\mathrm{o}} \sim \hat{A}_{\mathrm{o}}$, only in as much as it is far from the *independent*, $\tilde{E}_{\mathrm{o}} \nsim \hat{A}_{\mathrm{o}}^{\mathrm{X}}$.

In section 'Iso-sets', above, the degree to which an *integral iso-set* $I \subseteq \mathcal{A}_{U,\mathrm{i},V,z}$, where $A \in I$, is said to be *aligned-like*, or the *iso-independence*, is defined as

$$\frac{|I\ \cap\ Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})|}{|I\ \cup\ Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})|}$$

In the case of *aligned non-modelled induction* the *integral iso-set* is the *integral iso-independents*, $I = Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})$, and so *aligned non-modelled induction* is maximally *aligned-like*.

The requirement that the *distribution history* itself be *drawable*, $P_{U,X,H_{\mathrm{h}},\mathrm{y}}(H_{\mathrm{h}}) > 0$, has been ignored so far. This requirement modifies the maximisation to add the constraint that the *maximum likelihood estimate* be *iso-independent*, $\tilde{E}_{\mathrm{o}}^{\mathrm{X}} = \hat{A}_{\mathrm{o}}^{\mathrm{X}}$,

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m,y},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ E^{\mathrm{X}} = \hat{A}_{\mathrm{o}}^{\mathrm{X}}\})$$

So, strictly speaking, the *maximum likelihood estimate* is only approximately equal to the *normalised dependent*, $\tilde{E}_{\mathrm{o}} \approx \hat{A}_{\mathrm{o}}^{\mathrm{Y}}$, if the *dependent* is not *iso-independent*, $A_{\mathrm{o}}^{\mathrm{YX}} \neq A_{\mathrm{o}}^{\mathrm{X}}$.

Consider the *maximum likelihood estimate* of the *iso-independent conditional generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m,y},U}$. In section 'Likely histograms', above, the logarithm of the *maximum conditional probability* with respect to the *dependent-analogue* is conjectured to vary with the *relative space* with respect to the *independent-analogue*, which, in the case of *iso-independent conditional*, is the *alignment*,

$$\begin{aligned} \ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}}, z)(B) : B \in Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})} &\sim \mathrm{spaceRelative}(A^{\mathrm{X}})(A) \\ &= \mathrm{algn}(A) \end{aligned}$$

In *aligned induction*, where (i) the *history probability function* is *iso-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y}}$, (ii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iii) the *scaled estimate distribution histogram* is *integral*, $\tilde{E}_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, the *log likelihood* of the *iso-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *alignment*

$$\ln \hat{Q}_{\mathrm{h,y},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \mathrm{algn}(A_{\mathrm{o}})$$

It is conjectured that if the *independent* is *integral*, the *relative space* of the *histogram* is positive and less than or equal to the *relative space* of the *dependent*,

$$A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies 0 \leq \mathrm{spaceRelative}(A^{\mathrm{X}})(A) \leq \mathrm{spaceRelative}(A^{\mathrm{X}})(A^{\mathrm{Y}})$$

In 'Dependent alignment', above, it is conjectured that in the case where the *independent* of the *dependent* equals the *independent*, $A^{\mathrm{YX}} = A^{\mathrm{X}}$, then the inequality is

$$A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}} \implies 0 \leq \mathrm{algn}(A) \leq \mathrm{algn}(A^{\mathrm{Y}})$$

and that if the *histogram* is at *maximum alignment* the *dependent* equals the *histogram*,

$$\mathrm{algn}(A) = \mathrm{algnMax}(U)(V, z) \implies A^{\mathrm{Y}} = A$$

where algnMax = alignmentMaximum.

So conjecture that the *scaled maximum likelihood estimate*, $Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}}$, is at least as *aligned* as the *sample histogram*, A_{o},

$$\mathrm{algn}(Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}}) \geq \mathrm{algn}(A_{\mathrm{o}})$$

where $Z_{\mathrm{o}} = \mathrm{scalar}(z_{\mathrm{o}})$.

This may be compared to *classical induction* in which the *alignments* are equal,

$$\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}} \implies \mathrm{algn}(Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}}) = \mathrm{algn}(A_{\mathrm{o}})$$

It is conjectured that it is also in the case where the *sample alignment* is maximised that the *maximum likelihood estimate* equals the *sample probability histogram*,

$$\mathrm{algn}(A_{\mathrm{o}}) = \mathrm{algnMax}(U)(V_{\mathrm{o}}, z_{\mathrm{o}}) \implies A_{\mathrm{o}}^{\mathrm{Y}} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

That is, in *aligned non-modelled induction* there are two cases where the *maximum likelihood estimate* equals the *sample probability histogram*, $\tilde{E}_{\rm o} = \hat{A}_{\rm o}$, which are (i) *minimum alignment*, $\mathrm{algn}(A_{\rm o}) = 0$, and (ii) *maximum alignment*, $\mathrm{algn}(A_{\rm o}) = \mathrm{algnMax}(U)(V_{\rm o}, z_{\rm o})$.

In *aligned induction*, where (i) the *history probability function* is *iso-independent historically distributed*, $P = P_{U,X,H_{\rm h},{\rm y}}$, (ii) the *volume* is non-singleton, $v_{\rm o} > 1$, (iii) the *sample histogram* is *completely effective*, $A_{\rm o}^{\rm F} = V_{\rm o}^{\rm C}$, if (iv) the *sample alignment* is minimised, $\mathrm{algn}(A_{\rm o}) = 0$, or maximised, $\mathrm{algn}(A_{\rm o}) = \mathrm{algnMax}(U)(V_{\rm o}, z_{\rm o})$, then the *maximum likelihood estimate*, $\tilde{E}_{\rm o}$, of the *unknown distribution probability histogram*, $\hat{E}_{\rm o}$, in the *iso-independent conditional stuffed historical probability distribution*, $\hat{Q}_{{\rm h},{\rm y},U}(E_{\rm o}, z_{\rm o})$, is

$$\tilde{E}_{\rm o} = \hat{A}_{\rm o}$$

Now consider the *iso-independent conditional generalised multinomial distribution sum sensitivity* at the *maximum likelihood estimate*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm m},{\rm y},U}(\tilde{E}_{\rm o}, z_{\rm o})))$$

In section 'Iso-sets', above, it is conjectured that the cardinality of the *iso-independents* corresponding to $A^{\rm X}$ varies with the *entropy* of the *independent*, $A^{\rm X}$

$$\ln |Y_{U,{\rm i},V,z}^{-1}(A^{\rm X})| \quad \sim \quad z \times \mathrm{entropy}(A^{\rm X})$$

so, as the *independent entropy*, $\mathrm{entropy}(A^{\rm X})$, increases, the set of *iso independents*, $Y_{U,{\rm i},V,z}^{-1}(A^{\rm X})$, tends to the set of *substrate histograms*, $\mathcal{A}_{U,{\rm i},V,z}$, and the *sum sensitivity* of the denominator decreases, increasing the overall *sum sensitivity*. Also, as shown above for *classical induction*, the *sum sensitivity* of the numerator varies against the *scaled entropy*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm m},U}(E, z))) \quad \sim \quad -z \times \mathrm{entropy}(E)$$

The *alignment* approximates to the difference in *entropies*,

$$\mathrm{algn}(A) \approx z \times \mathrm{entropy}(A^{\rm X}) - z \times \mathrm{entropy}(A)$$

So at low *alignments* where the *maximum likelihood estimate* approximates to the *histogram*, $\mathrm{algn}(A) \approx 0 \implies \tilde{E} \approx \hat{A}$, the *sum sensitivity* varies with the *sample alignment*,

$$\begin{aligned}&\mathrm{algn}(A) \approx 0 \implies \\ &\quad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm m},{\rm y},U}(\tilde{E}, z))) \sim \mathrm{algn}(A)\end{aligned}$$

However, as the *alignment* of the *scaled maximum likelihood estimate*, $\text{algn}(Z * \tilde{E})$, increases, the *probability* of the *independent* term, $Q_{\text{m},U}(\tilde{E},z)(A^{\text{X}})$, decreases and the *sensitivity* of the denominator tends to be correlated with the numerator, lowering the overall *sensitivity*. Therefore conjecture that at high *alignments* where the *maximum likelihood estimate* approximates to the *histogram*, $\text{algn}(A) \approx \text{algnMax}(U)(V,z) \implies \tilde{E} \approx \hat{A}$, the *sum sensitivity* varies against the *sample alignment*,

$$\begin{aligned}&\text{algn}(A) \approx \text{algnMax}(U)(V,z) \implies \\ &\quad \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m,y},U}(\tilde{E},z))) \sim -\ \text{algn}(A)\end{aligned}$$

This implies that there is some intermediate *alignment* where the *sum sensitivity* is constant,

$$\begin{aligned}&0 \ll \text{algn}(A) \ll \text{algnMax}(U)(V,z) \implies \\ &\quad \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m,y},U}(\tilde{E},z))) = c\end{aligned}$$

where c is a constant.

In *aligned induction*, where (i) the *history probability function* is *iso-independent historically distributed*, $P = P_{U,X,H_{\text{h}},\text{y}}$, (ii) the *volume* is non-singleton, $v_{\text{o}} > 1$, (iii) the *sample histogram* is *completely effective*, $A^{\text{F}}_{\text{o}} = V^{\text{C}}_{\text{o}}$, if (iv) the *distribution history size* is large with respect to the *sample size*, $z_{\text{h}} \gg z_{\text{o}}$, and such that (v) the *scaled probability sample histogram* is *integral*, $A_{\text{o},z_{\text{h}}} \in \mathcal{A}_{\text{i}}$ where $A_{\text{o},z_{\text{h}}} = \text{scalar}(z_{\text{h}}) * \hat{A}_{\text{o}}$, then (a) the *iso-independent conditional stuffed historical distribution sum sensitivity* at the *maximum likelihood estimate* varies with the *sample alignment* in the case where the *sample alignment* is small,

$$\begin{aligned}&\text{algn}(A_{\text{o}}) \approx 0 \implies \\ &\quad \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,y},U}(A_{\text{o},z_{\text{h}}},z_{\text{o}}))) \sim \text{algn}(A_{\text{o}})\end{aligned}$$

and (b) the *iso-independent conditional stuffed historical distribution sum sensitivity* at the *maximum likelihood estimate* varies against the *sample alignment* in the case where the *sample alignment* is large,

$$\begin{aligned}&\text{algn}(A_{\text{o}}) \approx \text{algnMax}(U)(V_{\text{o}},z_{\text{o}}) \implies \\ &\quad \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{h,y},U}(A_{\text{o},z_{\text{h}}},z_{\text{o}}))) \sim -\ \text{algn}(A_{\text{o}})\end{aligned}$$

5.7 Idealisation induction

Having considered (i) *classical modelled induction*, which requires *necessary derived*, and (ii) *aligned non-modelled induction*, which requires *necessary independent*, now consider (iii) *idealisation induction*, which is a stricter intersection between the two, requiring *necessary idealisation*.

Given some *known substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, the *idealisation histogram* of the *distribution probability histogram* is $\hat{E}_{\mathrm{h}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger E_{\mathrm{h}}}$. In *idealisation induction*, while the *distribution probability histogram*, $\hat{E}_{\mathrm{h}}$, remains *unknown*, the *idealisation distribution probability histogram*, $\hat{E}_{\mathrm{h}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger E_{\mathrm{h}}}$, is *known* and *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have a *idealisation probability histogram* equal to the *known idealisation distribution probability histogram*, $\hat{A}_H * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_H} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger E_{\mathrm{h}}} \% V_H$. Define the *iso-idealisation historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$,

$$\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}} := \\
&\quad \Big(\bigcup \{\{(H,1) : H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H, \\
&\qquad\qquad \hat{A}_H * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_H} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger E_{\mathrm{h}}}\%V_H\}^{\wedge} : \\
&\qquad\qquad\qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \dots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ \hat{A}_H * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_H} \neq \hat{E}_{\mathrm{h}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger E_{\mathrm{h}}}\%V_H\} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \nsubseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset,0)\}
\end{aligned}$$

For *drawn histories* the *idealisation probability histogram* is *necessary*, $\forall H \in \mathcal{H}_{U,X}\ (P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H) > 0 \implies \hat{A}_H * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_H} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger E_{\mathrm{h}}}\%V_H)$. Not all *sizes* and sets of *variables* are necessarily *drawable*. That is, in some cases, $\exists z \in \{1 \dots z_{\mathrm{h}}\}\ \exists V \subseteq V_{\mathrm{h}}\ \forall H \in \mathcal{H}_{U,X}\ ((z_H = z) \wedge (V_H = V) \implies P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H) = 0)$. The *distribution history* can always be *drawn*, so the *probability function* is not a *weak probability function*, $\sum_{H \in \mathcal{H}_{U,X}} P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H) = 1$.

All *iso-idealisation* subsets of the *distribution history* for a given set of *variables* and *size* are defined as equally probable,

$$\begin{aligned}
&\forall V \subseteq V_{\mathrm{h}}\ \forall H, G \subseteq H_{\mathrm{h}}\%V \\
&\quad (A_G * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_G} = A_H * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_H} \implies P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(G) = P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H))
\end{aligned}$$

In *idealisation induction* the *history probability function* is *iso-idealisation historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}$.

Given a *drawn history* $H \in \mathcal{H}_{U,X}$, where $P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H) > 0$, the *iso-idealisation historical probability* of *histogram* $A_H = \mathrm{histogram}(H) + V_H^{\mathrm{CZ}} \in$

$\mathcal{A}_{U,\mathrm{i},V_H,z_H}$ is now conditional,

$$\frac{Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H)}{\sum_{B \in \mathrm{isoi}(U)(T_{\mathrm{o}},A_{\mathrm{o}})} Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(B)} = \frac{\sum P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}$$

where $\mathrm{isoi}(U)(T,A) := Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A*T*T^{\dagger A})$ and the *independent components* valued *histogram* function $Y_{U,\mathrm{i},T,\dagger,z}$ is defined

$$Y_{U,\mathrm{i},T,\dagger,z} = \{(A, A*T*T^{\dagger A}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The finite set of *iso-idealisations* of *independent components* $\{(A*C^{\mathrm{U}})^{\mathrm{X}} : C \in T^{\mathrm{P}}\}$ is

$$Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A*T*T^{\dagger A}) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B*T*T^{\dagger B} = A*T*T^{\dagger A}\}$$

The *iso-idealisation historical probability* may be expressed in terms of a *histogram distribution* which is not explicitly conditional on the *necessary idealisation*, $\hat{E}_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger E_{\mathrm{o}}}$,

$$\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) \ \ \propto\ \ \sum(P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *iso-idealisation conditional stuffed historical probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\dagger,T,U}(E,z) \\
&\quad := \ \{(A, \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in \mathrm{isoi}(U)(T,A)} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\
&\qquad\ \ \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \not\leq E\}
\end{aligned}$$

which is defined if $z \leq \mathrm{size}(E)$.

In the case where all the *idealisations* are *possible*,

$$\forall A' \in \mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})\ \exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ ((A*T*T^{\dagger A} = A') \ \wedge\ (A \leq E))$$

the normalisation of the *iso-idealisation conditional stuffed historical probability distribution* is a fraction $1/|\mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})|$,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\dagger,T,U}(E,z) \\
&\quad = \ \{(A, \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})|}\frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in \mathrm{isoi}(U)(T,A)} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}
\end{aligned}$$

In the case of a *full functional transform*, $T_{\mathrm{s}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}}$, the *iso-idealisation* is a singleton of the *sample histogram*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{s}},\dagger,z}(A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger A}) = \{A * T_{\mathrm{s}} * T_{\mathrm{s}}^{\dagger A}\} = \{A\}$, and so the denominator equals the numerator, $\sum(Q_{\mathrm{h},U}(E,z)(B) : B \in \{A\}) = Q_{\mathrm{h},U}(E,z)(A)$. Thus the *iso-idealisation historically distributed history probability* is a constant, $\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{s}},U}(E,z)(Z * \hat{E}) = 1/|\mathcal{A}_{U,\mathrm{i},V,z}|$, where $Z = \mathrm{scalar}(z)$. In this case, the *distribution probability histogram*, $\hat{E}$, is *known*, because $\hat{E}*T_{\mathrm{s}}*T_{\mathrm{s}}^{\dagger E}$ is *known*, and so everything is *known*.

At the other extreme of a *unary transform*, $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, the set of *iso-idealisations* equals the *iso-independents*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{u}},\dagger,z}(A*T_{\mathrm{u}}*T_{\mathrm{u}}^{\dagger A}) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$. Thus the *iso-idealisation conditional stuffed historical probability distribution* equals the *iso-independent conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{u}},U}(E,z) = \hat{Q}_{\mathrm{h,y},U}(E,z)$. In this case *idealisation induction* reduces to *aligned non-modelled induction*.

The *iso-idealisation conditional generalised multinomial probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\mathrm{m},\dagger,T,U}(E,z) \\
&\quad := \{(A, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathrm{isoi}(U)(T,A)} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \leq E^{\mathrm{F}}\}^{\wedge} \cup \\
&\qquad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \nleq E^{\mathrm{F}}\}
\end{aligned}$$

which is defined if $\mathrm{size}(E) > 0$.

The case where all the *idealisations* are *possible* is weaker than for *historical*,

$$\forall A' \in \mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})\ \exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ ((A * T * T^{\dagger A} = A')\ \wedge\ (A^{\mathrm{F}} \leq E^{\mathrm{F}}))$$

In this case the *iso-idealisation conditional generalised multinomial probability distribution* is

$$\begin{aligned}
&\hat{Q}_{\mathrm{m},\dagger,T,U}(E,z) \\
&\quad = \{(A, \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})|} \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathrm{isoi}(U)(T,A)} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}
\end{aligned}$$

It is assumed that the *distribution history size*, z_{h}, is large with respect to the *sample size* $z_{\mathrm{o}} = \mathrm{size}(A_{\mathrm{o}})$, so that, in the limit, the *iso-idealisation historical probability*, $\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$, approximates to the *iso-idealisation multinomial probability*, $\hat{Q}_{\mathrm{m},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$. That is, if $z_{\mathrm{o}} \ll z_{\mathrm{h}}$ then

$$\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}}) \approx \hat{Q}_{\mathrm{m},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$$

In the case of *completely effective sample histogram*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the maximisation for *known transform*, T_{o}, of the *iso-idealisation conditional generalised multinomial probability* parameterised by the *complete congruent histograms* of unit *size* is a singleton of the *rational maximum likelihood estimate*

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m},\dagger,T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

In the case where the *sample histogram* is not *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} < V_{\mathrm{o}}^{\mathrm{C}}$, the maximisation of the *iso-idealisation conditional generalised multinomial probability distribution* for *known transform* is not necessarily a singleton

$$|\max(\{(E, \hat{Q}_{\mathrm{m},\dagger,T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})| \geq 1$$

In the case where the maximisation of the *iso-idealisation conditional generalised multinomial probability distribution* is a singleton, it is equal to the *normalised idealisation-dependent*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\dagger(T_{\mathrm{o}})}$, where the *idealisation-dependent* $A^{\dagger(T)} \in \mathcal{A}_{U,V,z}$ is defined in 'Likely histograms', above, as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-idealisation*,

$$\{A^{\dagger(T)}\} = \\ \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D, z)(A)}{\sum Q_{\mathrm{m},U}(D, z)(B) : B \in \mathrm{isoi}(U)(T, A)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *idealisation-dependent*, $A^{\dagger(T)}$, is sometimes not computable. The finite approximation to the *idealisation-dependent* is

$$\{A_k^{\dagger(T)}\} = \\ \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D, z)(A)}{\sum Q_{\mathrm{m},U}(D, z)(B) : B \in \mathrm{isoi}(U)(T, A)}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})$$

The approximation, $A_k^{\dagger(T)} \approx A^{\dagger(T)}$, improves as the scaling factor, k, increases.

Unlike in *classical non-modelled induction* where the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, is equal to the *sample probability histogram*, $\hat{A}_{\mathrm{o}}$, in *idealisation induction* the *maximum likelihood estimate* is not necessarily equal to the *sample probability histogram*. It is only in the case where the *sample histogram* is *ideal* that the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}} \implies A_{\mathrm{o}}^{\dagger(T_{\mathrm{o}})} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

Otherwise, the overall *maximum likelihood estimate*, which is the *idealisation-dependent*, is near the *histogram*, $\tilde{E}_{\rm o} \sim \hat{A}_{\rm o}$, only in as much as it is far from the *idealisation*, $\tilde{E}_{\rm o} \nsim \hat{A}_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger A_{\rm o}}$.

The requirement that the *distribution history* itself be *drawable*, $P_{U,X,H_{\rm h},\dagger,T_{\rm o}}(H_{\rm h}) > 0$, has been ignored so far. This requirement modifies the maximisation to add the constraint that the *maximum likelihood estimate* be an *iso-idealisation*, $\tilde{E}_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger \tilde{E}_{\rm o}} = \hat{A}_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger \hat{A}_{\rm o}}$,

$$\begin{aligned}\{\tilde{E}_{\rm o}\} = \mathrm{maxd}(\{(E, \hat{Q}_{{\rm m},\dagger,T_{\rm o},U}(E,z_{\rm o})(A_{\rm o})) : E \in \mathcal{A}_{U,V_{\rm o},1},\\ E * T_{\rm o} * T_{\rm o}^{\dagger E} = \hat{A}_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger \hat{A}_{\rm o}}\})\end{aligned}$$

So, strictly speaking, the *maximum likelihood estimate* is only approximately equal to the *normalised idealisation-dependent*, $\tilde{E}_{\rm o} \approx \hat{A}_{\rm o}^{\dagger(T_{\rm o})}$, if the *idealisation-dependent* is not an *iso-idealisation*, $A_{\rm o}^{\dagger(T_{\rm o})} * T_{\rm o} * T_{\rm o}^{\dagger A_{\rm o}^{\dagger(T_{\rm o})}} \neq \hat{A}_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger \hat{A}_{\rm o}}$. In the special case, however, where the *sample histogram* is *ideal*, the *maximum likelihood estimate* is exactly equal to the *normalised idealisation-dependent*, $A_{\rm o} = A_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger A_{\rm o}} \implies \tilde{E}_{\rm o} = \hat{A}_{\rm o}^{\dagger(T_{\rm o})} = \hat{A}_{\rm o}$.

In *idealisation induction*, where (i) the *history probability function* is *iso-idealisation historically distributed*, $P = P_{U,X,H_{\rm h},\dagger,T_{\rm o}}$, given some *substrate transform* in the *sample variables* $T_{\rm o} \in \mathcal{T}_{U,V_{\rm o}}$, if it is the case that (ii) the *sample histogram* is *ideal*, $A_{\rm o} = A_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger A_{\rm o}}$, then the *maximum likelihood estimate*, $\tilde{E}_{\rm o}$, of the *unknown distribution probability histogram*, $\hat{E}_{\rm o}$, in the *iso-idealisation conditional stuffed historical probability distribution*, $\hat{Q}_{{\rm h},\dagger,T_{\rm o},U}(E_{\rm o}, z_{\rm o})$, is

$$\tilde{E}_{\rm o} = \hat{A}_{\rm o}$$

The set of *iso-idealisations* is a subset of the *iso-deriveds*, so it is a *law-like iso-set* of the *histogram*, A,

$$Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \ \subseteq\ D^{-1}_{U,T,z}(A * T)$$

The *iso-derivedence* or degree of *law-likeness* is

$$\frac{|Y^{-1}_{U,{\rm i},T,\dagger,z}(A * T * T^{\dagger A})|}{|D^{-1}_{U,{\rm i},T,z}(A * T)|} \ \leq\ 1$$

So *idealisation induction* is not maximally *law-like* if the *iso-deriveds* is a proper superset of the *iso-idealisations*, $D^{-1}_{U,T,z}(A * T) \supset Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A})$.

The set of *iso-idealisations* is a subset of the *iso-abstracts*, so it is an *entity-like iso-set* of the *histogram*, A,

$$Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \;\subseteq\; Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

The *iso-abstractence* or degree of *entity-likeness* is less than or equal to the *iso-derivedence*

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \;\leq\; \frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|}$$

so *idealisation induction* is less *entity-like* and more *law-like*.

The set of *iso-idealisations* is a subset of the *iso-independents*, $Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \subseteq Y^{-1}_{U,V,z}(A^{\mathrm{X}})$, so the degree to which the *iso-idealisations* is *aligned-like*, or the *iso-independence*, is $|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|/|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|$.

In some cases the *iso-independence* of the *iso-idealisations* is greater than or equal to the *iso-independence* of the *iso-derived*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} \;\geq\; \frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T) \;\cap\; Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|D^{-1}_{U,\mathrm{i},T,z}(A * T) \;\cup\; Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

and so *idealisation induction* may be said to be more *aligned-like* than *classical modelled induction*.

As the *iso-independence* increases, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, which equals the *idealisation-dependent*, $\hat{A}_{\mathrm{o}}^{\dagger(T_{\mathrm{o}})}$, tends to the *dependent*, $\hat{A}_{\mathrm{o}}^{\mathrm{Y}}$, which is independent of the *model*, T_{o}, because the *independent analogue*, $A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, tends to the *independent*, $A_{\mathrm{o}}^{\mathrm{X}}$, which is also independent of the *model*, as the *transform* becomes more *unary*.

Given the *known substrate transform*, T_{o}, consider the *maximum likelihood estimate* of the *iso-idealisation conditional generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m},\dagger,T_{\mathrm{o}},U}$. In section 'Likely histograms', above, the logarithm of the *maximum conditional probability* with respect to the *dependent-analogue* is conjectured to vary with the *relative space* with respect to the *independent-analogue*. In the case of *iso-idealisation conditional*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\dagger(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\dagger(T)}, z)(B) : B \in \mathrm{isoi}(U)(T,A)} \;\sim\; \mathrm{spaceRelative}(A * T * T^{\dagger A})(A)$$

where the *distribution-relative multinomial space* is defined, in section 'Likely histograms', above, as

$$\mathrm{spaceRelative}(E)(A) \ := \ -\ln \frac{\mathrm{mpdf}(U)(E,z)(A)}{\mathrm{mpdf}(U)(E,z)(E)}$$

In section 'Transform alignment', above, because the set of *iso-idealisations* is *law-like*, it is shown that, in the case where the *dependent analogue* is in the *iso-set*, the difference in *relative space* between the *histogram* and the *dependent* must be in the *relative spaces* of the *components*,

$$\begin{aligned} A^{\dagger(T)} \in D_{U,T,z}^{-1}(A * T) \implies \\ \sum_{(\cdot,C) \in T^{-1}} \mathrm{spaceRelative}(A * T * T^{\dagger A} * C)(A * C) \\ \leq \sum_{(\cdot,C) \in T^{-1}} \mathrm{spaceRelative}(A * T * T^{\dagger A} * C)(A^{\dagger(T)} * C) \end{aligned}$$

So, in the case of the *idealisation-dependent*, the *component alignments* must be greater than or equal to the *component alignments* of the *histogram*,

$$\begin{aligned} A^{\dagger(T)} \in D_{U,T,z}^{-1}(A * T) \implies \\ \sum_{(\cdot,C) \in T^{-1}} \mathrm{algn}(A * C) \ \leq \ \sum_{(\cdot,C) \in T^{-1}} \mathrm{algn}(A^{\dagger(T)} * C) \end{aligned}$$

The *idealisation-dependent* varies with the *histogram*, $\tilde{E}_{\mathrm{o}} \sim \hat{A}_{\mathrm{o}}$, so conjecture that in the case where the *sample* is not *ideal*, $A \neq A * T * T^{\dagger A} \implies \mathrm{spaceRelative}(A * T * T^{\dagger A})(A) > 0$, the *log-likelihood* varies with the sum of the *component alignments*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\dagger(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\dagger(T)},z)(B) : B \in \mathrm{isoi}(U)(T,A)} \ \sim \ \sum_{(\cdot,C) \in T^{-1}} \mathrm{algn}(A * C)$$

In *idealisation induction*, where (i) the *history probability function* is *iso-idealisation historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample histogram* is not *ideal*, $A_{\mathrm{o}} \neq A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iv) the *scaled estimate distribution histogram* is *integral*, $\tilde{E}_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-idealisation conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *relative space* of the *sample* with respect to the *idealisation*,

$$\ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ \mathrm{spaceRelative}(A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}})(A_{\mathrm{o}})$$

and varies with the sum of the *component alignments* of the *sample components*,

$$\ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \sum_{(\cdot,C) \in T_{\mathrm{o}}^{-1}} \mathrm{algn}(A_{\mathrm{o}} * C)$$

The set of *iso-idealisations* is a subset of the intersection of the *iso-independents* and *iso-deriveds* which is a subset of the *iso-liftisations* which is a subset of the *iso-transform-independents*,

$$\begin{aligned}
&Y_{U,T,\dagger,z}^{-1}(A * T * T^{\dagger A}) \\
&\qquad \subseteq Y_{U,V,z}^{-1}(A^{\mathrm{X}}) \cap D_{U,T,z}^{-1}(A * T) \\
&\qquad \subseteq Y_{U,T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \cap D_{U,T,z}^{-1}(A * T) \\
&\qquad \subseteq Y_{U,T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \cap Y_{U,T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})
\end{aligned}$$

Since the *idealisation entropy* is conjectured (i) to be less than or equal to the *independent entropy*, $\mathrm{entropy}(A * T * T^{\dagger A}) \leq \mathrm{entropy}(A^{\mathrm{X}})$, and (ii) to be less than or equal to the *naturalisation entropy*, $\mathrm{entropy}(A * T * T^{\dagger A}) \leq \mathrm{entropy}(A * T * T^{\dagger})$, the *idealisation relative space* is conjectured (a) to be less than or equal to the *independent relative space*, which is the *alignment*,

$$\mathrm{spaceRelative}(A * T * T^{\dagger A})(A) \quad \leq \quad \mathrm{spaceRelative}(A^{\mathrm{X}})(A) = \mathrm{algn}(A)$$

and (b) to be less than or equal to the *naturalisation relative space*,

$$\mathrm{spaceRelative}(A * T * T^{\dagger A})(A) \quad \leq \quad \mathrm{spaceRelative}(A * T * T^{\dagger})(A)$$

Given the *known substrate transform*, T_{o}, consider the *log likelihood* of the *iso-idealisation conditional generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m},\dagger,T_{\mathrm{o}},U}$, at the *maximum likelihood estimate*, in the special case where the *histogram* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\dagger(T_{\mathrm{o}})} = \hat{A}_{\mathrm{o}}$.

The set of *iso-idealisations* is a subset of the intersection of the *iso-independents* and *iso-deriveds*,

$$Y_{U,T,\dagger,z}^{-1}(A * T * T^{\dagger A}) \quad \subseteq \quad Y_{U,V,z}^{-1}(A^{\mathrm{X}}) \cap D_{U,T,z}^{-1}(A * T)$$

In section 'Iso-sets', above, it is conjectured that the cardinality of the *integral iso-independents* varies with the *scaled entropy* of the *independent*, A^{X},

$$\ln |Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})| \quad \sim \quad z \times \mathrm{entropy}(A^{\mathrm{X}})$$

Alignment is approximately the difference in the *scaled entropies* of the *independent* and the *histogram*,

$$\text{algn}(A) \approx z \times \text{entropy}(A^{\text{X}}) - z \times \text{entropy}(A)$$

so the cardinality of the *iso-independents* varies with the *alignment*,

$$\ln |Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})| \sim \text{algn}(A)$$

At high *alignments*, the *integral iso-independents*, $Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})$, tends to the *integral substrate histograms*, $\mathcal{A}_{U,\text{i},V,z}$, so the *iso-independence* of the set of *iso-idealisations*, $|Y^{-1}_{U,\text{i},T,\dagger,z}(A * T * T^{\dagger A})|/|Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})|$, decreases, and the *iso-derivedence*, $|Y^{-1}_{U,\text{i},T,\dagger,z}(A * T * T^{\dagger A})|/|D^{-1}_{U,\text{i},T,z}(A * T)|$, increases. The *sample* is the *independent-analogue*, so it equals the *dependent-analogue* and the *maximum likelihood estimate* is just the *sample probability histogram*, $\tilde{E} = \hat{A}$. So the numerator of the *iso-idealisation probability* for *ideal sample* equals the numerator of the *iso-derived probability* for *natural sample*, $Q_{\text{m},U}(A, z)(A)$. At high *alignments*, the set of *iso-idealisations* of the denominator approximates to the *iso-deriveds*, $Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \approx D^{-1}_{U,T,z}(A * T)$, so conjecture that the *iso-idealisation conditional generalised multinomial probability* varies with the *iso-derived conditional generalised multinomial probability*,

$$\ln \frac{Q_{\text{m},U}(A, z)(A)}{\sum Q_{\text{m},U}(A, z)(B) : B \in \text{isoi}(U)(T, A)} \sim \ln \frac{Q_{\text{m},U}(A, z)(A)}{\sum Q_{\text{m},U}(A, z)(B) : B \in D^{-1}_{U,\text{i},T,z}(A * T)}$$

and that the *iso-idealisation log likelihood* varies with the *iso-derived log likelihood*,

$$\ln \hat{Q}_{\text{m},\dagger,T,U}(A, z)(A) \sim \ln \hat{Q}_{\text{m},\text{d},T,U}(A, z)(A)$$

So the *log likelihood* varies with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}\ln \hat{Q}_{\text{m},\dagger,T,U}(A, z)(A) \sim &\\ &(z + v) \times \text{entropy}(A * T + V^{\text{C}} * T)\\ &\quad -z \times \text{entropy}(A * T) - v \times \text{entropy}(V^{\text{C}} * T)\end{aligned}$$

The set of *iso-idealisations* is a subset of the intersection of the *iso-independents* and *iso-deriveds*, $Y^{-1}_{U,T,\dagger,z}(A*T*T^{\dagger A}) \subseteq Y^{-1}_{U,V,z}(A^{\text{X}}) \cap D^{-1}_{U,T,z}(A*T)$, so conjecture that the *iso-idealisation conditional multinomial distribution sum sensitivity* varies with the *iso-independent sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},\dagger,T,U}(A,z))) \sim \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,y},U}(A,z)))$$

and the *iso-derived sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},\dagger,T,U}(A,z))) \sim \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,d},T,U}(A,z)))$$

In section 'Aligned induction', above, it is conjectured that at low *alignments*, $\mathrm{algn}(A) \approx 0$, the *sum sensitivity* varies with the *sample alignment*,

$$\begin{aligned}&\mathrm{algn}(A) \approx 0 \implies \\ &\qquad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,y},U}(A,z))) \sim \mathrm{algn}(A)\end{aligned}$$

but at high *alignments*, $\mathrm{algn}(A) \approx \mathrm{algnMax}(U)(V,z)$, the *sum sensitivity* varies against the *sample alignment*,

$$\begin{aligned}&\mathrm{algn}(A) \approx \mathrm{algnMax}(U)(V,z) \implies \\ &\qquad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,y},U}(A,z))) \sim -\ \mathrm{algn}(A)\end{aligned}$$

and so there is some intermediate *alignment* where the *sum sensitivity* is independent of the *alignment*,

$$\begin{aligned}&0 \ll \mathrm{algn}(A) \ll \mathrm{algnMax}(U)(V,z) \implies \\ &\qquad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,y},U}(A,z))) = c\end{aligned}$$

where c is a constant. So in the case of intermediate *alignment* the *iso-independent sum sensitivity* is constant and the *iso-idealisation sum sensitivity* varies only with the *iso-derived sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},\dagger,T,U}(A,z))) \sim \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,d},T,U}(A,z)))$$

In section 'Classical modelled induction', above, it is shown that, given *necessary derived*, in the special case where the *histogram* is *natural*, $A = A * T * T^{\dagger} \implies \tilde{E} = \hat{A}^{\mathrm{D}(T)} = \hat{A}$, and the *component size cardinality cross entropy* is greater than the logarithm of the *possible derived volume*, $\mathrm{entropyCross}(A * T, V^{\mathrm{C}} * T) > \ln w'$, so the *relative entropy* is high, the *iso-derived conditional multinomial probability* at the *maximum likelihood estimate* varies with the *underlying-derived relative multinomial probability*,

$$\frac{\hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)} \hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B)} \sim \frac{\hat{Q}_{\mathrm{m},U}(A * T * T^{\dagger}, z)(A * T * T^{\dagger})}{\hat{Q}_{\mathrm{m},U}(A * T, z)(A * T)}$$

Similarly, given *necessary idealisation*, if (i) the *sample* is *ideal*, $A = A * T * T^{\dagger A} \implies E = A^{\dagger(T)} = A$, (ii) the *relative entropy* is high, $\text{entropyCross}(A * T, V^{\text{C}} * T) > \ln w'$, and (iii) the *alignment* is intermediate, $\text{algn}(A) \approx \text{algnMax}(U)(V,z)/2$, then the *iso-idealisation conditional multinomial probability* at the *maximum likelihood estimate* varies with the *underlying-derived relative multinomial probability*,

$$\frac{\hat{Q}_{\text{m},U}(A * T * T^{\dagger A}, z)(A * T * T^{\dagger A})}{\sum_{B \in \text{isoi}(U)(T,A)} \hat{Q}_{\text{m},U}(A * T * T^{\dagger A}, z)(B)} \sim \frac{\hat{Q}_{\text{m},U}(A * T * T^{\dagger A}, z)(A * T * T^{\dagger A})}{\hat{Q}_{\text{m},U}(A * T, z)(A * T)}$$

Thus, at intermediate *alignments* where the *sample* is *ideal*, the *sum sensitivity* of the *iso-idealisation conditional multinomial probability distribution*, $\hat{Q}_{\text{m},\dagger,T,U}$, is conjectured to vary with the *unknown-known multinomial probability distribution sum sensitivity difference*,

$$\begin{aligned}&\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},\dagger,T,U}(A,z))) \sim \\ &\quad \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(A,z))) - \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(A * T,z)))\end{aligned}$$

and so the *sum sensitivity* of the *iso-idealisation conditional multinomial probability distribution* is sometimes less than or equal to the *sum sensitivity* of the *multinomial probability distribution*,

$$\text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},\dagger,T,U}(A,z))) \leq \text{sum}(\text{sensitivity}(U)(\hat{Q}_{\text{m},U}(A,z)))$$

Note that in *classical modelled induction* the denominator,

$$\sum_{B \in D^{-1}_{U,\text{i},T,z}(A * T)} \hat{Q}_{\text{m},U}(A * T * T^{\dagger}, z)(B)$$

is *lifted* to the *derived*, $\hat{Q}_{\text{m},U}(A * T, z)(A * T)$, while in *idealisation induction* a rather different denominator,

$$\sum_{B \in \text{isoi}(U)(T,A)} \hat{Q}_{\text{m},U}(A * T * T^{\dagger A}, z)(B)$$

is *lifted* to the same *derived*. It is assumed that, although the set of *iso-idealisations* is only a subset of the *iso-deriveds* because of the intersections of the *iso-independents*, these intersections are arbitrary with respect to the *iso-derived* at intermediate *alignments*. That is, the set of *iso-idealisations*, $\text{isoi}(U)(T,A)$, is assumed to be a representative subset of the set of *iso-derived*, $D^{-1}_{U,\text{i},T,z}(A * T)$, and so can be *lifted*.

The *iso-idealisation sum sensitivity* varies with the *unknown-known sum sensitivity difference* similarly to the *iso-derived sum sensitivity*, and so has similar characteristics. However, note that the *expected component entropy* of the *idealisation components* is less than or equal to that of the *naturalisation*,

$$\begin{aligned}
&\mathrm{entropyComponent}(A * T * T^{\dagger A}, T) \\
= \;& \mathrm{expected}(\hat{A} * T)(\{(R, \mathrm{entropy}((A * C)^{\mathrm{X}})) : (R,C) \in T^{-1}\}) \\
\leq \;& \mathrm{expected}(\hat{A} * T)(\{(R, \mathrm{entropy}(C)) : (R,C) \in T^{-1}\}) \\
= \;& \mathrm{entropyComponent}(A * T * T^{\dagger}, T)
\end{aligned}$$

though, because the *components* are *independent*, the *expected component entropy* of the *idealisation* is greater than or equal to the *expected component entropy* where the *sample* is not *ideal*,

$$\begin{aligned}
&\mathrm{entropyComponent}(A, T) \\
= \;& \mathrm{expected}(\hat{A} * T)(\{(R, \mathrm{entropy}(A * C)) : (R,C) \in T^{-1}\}) \\
\leq \;& \mathrm{expected}(\hat{A} * T)(\{(R, \mathrm{entropy}((A * C)^{\mathrm{X}})) : (R,C) \in T^{-1}\}) \\
= \;& \mathrm{entropyComponent}(A * T * T^{\dagger A}, T)
\end{aligned}$$

Overall, in the case where the *histogram* is *ideal* and the *alignment* is intermediate or higher, the properties of *necessary idealisation induction* are expected to be similar to those of *necessary derived induction*.

In *idealisation induction*, where (i) the *history probability function* is *iso-idealisation historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample histogram* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-idealisation conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iii) the *alignment* is intermediate or high, $\mathrm{algn}(A_{\mathrm{o}}) \geq \mathrm{algnMax}(U)(V_{\mathrm{o}}, z_{\mathrm{o}})/2$, (iv) the *relative entropy* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, (v) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (vi) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *iso-idealisation conditional stuffed historical probability distribution* at the *maximum likelihood estimate* is such that (a) the *log likelihood* varies with the *iso-derived log likelihood*,

$$\ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\; \ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

(b) the *log likelihood* varies with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ &\sim \\ &(z_{\mathrm{o}} + v_{\mathrm{o}}) \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}} + V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) \\ &\quad -z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}}) \ - v_{\mathrm{o}} \times \mathrm{entropy}(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})\end{aligned}$$

and (c) the *sum sensitivity* is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \\ &\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\end{aligned}$$

If, in addition, (vii) the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *idealisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, then (d) the *log likelihood* varies with the *scaled component size cardinality relative entropy*,

$$\ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ z_{\mathrm{o}} \times \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})$$

and (e) the *log likelihood* varies against the *specialising derived substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ - \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

where

$$C_{\mathrm{G},V,\mathrm{T},\mathrm{H}}(T) = \mathrm{coderHistorySubstrateDerivedSpecialising}(U, X, T, D_{\mathrm{S}}, D_{\mathrm{X}})$$

So (f) the *sum sensitivity* varies against the *log-likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \ \sim \ - \ \ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

and (g) the *sensitivity* to *model* also varies against the *log likelihood*,

$$\begin{aligned}- \ \ln |\max(\{(T, \hat{Q}_{\mathrm{h},\dagger,T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\})| \ &\sim \\ - \ \ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})&\end{aligned}$$

Note that the anti-correlation between the *log-likelihood* and *specialising space*,

$$\ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ - \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

is conjectured to be weaker than that of *classical modelled induction*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ - \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

because the *expected component entropy* of the *idealisation* is less than or equal to that of the *naturalisation*.

In section 'Classical modelled induction', above, it is shown that the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$, can be related to queries on the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, in the special case where the *sample histogram* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$. The given *substrate transform* must be such that its *contraction* has *underlying variables* that are a subset of the query *variables*, $\mathrm{und}(T_{\mathrm{o}}^{\%}) \subseteq K$. In the case where the query *histogram* consists of one *effective state*, $Q = \{(S_Q, 1)\}$, the application of the query in terms of a modified *sample histogram* is

$$\begin{aligned}
&(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K) = \\
&\quad \{(N, (\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}))^{1/z_{\mathrm{o}}}) : N \in (V_{\mathrm{o}} \setminus K)^{\mathrm{CS}}, \\
&\qquad A_{Q,N} = A_{\mathrm{o}} - (A_{\mathrm{o}} * C_Q) + ((A_{\mathrm{o}} * C_Q) \ \% \ K * \{N\}^{\mathrm{U}})\}^{\wedge}
\end{aligned}$$

where $\{R_Q\} = (Q * T_{\mathrm{o}}^{\%})^{\mathrm{FS}}$, $C_Q = T_{\mathrm{o}}^{-1}(R_Q)$ and $\mathrm{his} = \mathrm{histogram} \in \mathcal{T} \to \mathcal{A}$. If the *sample histogram* is *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the modified *sample histogram*, $A_{Q,N}$, can be *drawn* from the *distribution*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}) > 0$, because its *derived* is equal to the *known derived*, $A_{Q,N} * T_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}}$. That is, the modified *sample histogram* is in the *iso-deriveds*, $A_{Q,N} \in D^{-1}_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}}(A_{\mathrm{o}} * T_{\mathrm{o}})$.

However, in the case of *idealisation induction*, where the *idealisation* is *necessary*, the modified *sample histogram* is not in the *iso-idealisations*, $A_{Q,N} \notin \mathrm{isoi}(U)(T_{\mathrm{o}}, A_{\mathrm{o}})$, if the *volume* of the label *variables* is non-singular, $|(V_{\mathrm{o}} \setminus K)^{\mathrm{C}}| > 1$. That is, strictly speaking, even if the *transform* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, the application of the query via the *model*, $Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}}$, cannot be expressed in terms of the *iso-idealisation conditional stuffed historical probability distribution* at the *maximum likelihood estimate*, $\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$.

However, it is conjectured that, especially in the case of small label *volume*, $|(V_{\mathrm{o}} \setminus K)^{\mathrm{C}}| \approx 2$, the *query sensitivity* to the *distribution histogram* varies as the *iso-idealisation sum sensitivity* divided by the *size*

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}$$

Although the query application via the *model* is sometimes not equal to the *estimated transformed conditional product*, the *query sensitivity* to the *distribution histogram* is sometimes lower,

$$\begin{aligned}
&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}} \\
&\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}
\end{aligned}$$

Similarly, the *model likelihood* of the *distribution histogram* is sometimes higher,

$$\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{Q,N}) \;\;\geq\;\; \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{Q,N})$$

In other words, querying via the *known derived* of the *model* sometimes reduces the *sensitivity* to the *unknown* and increases the *likelihood* at the cost of modifying the query. Note that the degree to which this is case is lower in *idealisation induction* than it is in *classical modelled induction*.

In the discussion above, the *model*, $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, is *known* and the *idealisation*, $\hat{E}_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger \hat{E}_{\mathrm{o}}}$, is both *necessary* and *known*. Optimisation can be done to find the *maximum likelihood estimate* of the *distribution histogram* for *known model*,

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m},\dagger,T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

Just as in the discussion above of *classical modelled induction*, consider the case where the *idealisation*, $\hat{E}_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger \hat{E}_{\mathrm{o}}}$, is still *necessary* but the *model*, T_{o}, is *unknown* and so the *idealisation* is *unknown*. Again, the *maximum likelihood estimate* for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$ can be defined as an optimisation of the *multinomial probability* conditional on the *iso-idealisations* where both the *distribution histogram* and *transform* are treated as arguments to a likelihood function,

$$\begin{aligned} &\{(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})\} \\ &\quad = \;\; \mathrm{maxd}(\{((E,T), \frac{Q_{\mathrm{m},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in \mathrm{isoi}(U)(T,A_{\mathrm{o}})} Q_{\mathrm{m},U}(E, z_{\mathrm{o}})(B)}) : \\ &\qquad\qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\; T \in \mathcal{T}_{U,V_{\mathrm{o}}}\}) \end{aligned}$$

There is, however, no singular solution to this optimisation,

$$\begin{aligned} &\mathrm{maxd}(\{((E,T), \frac{Q_{\mathrm{m},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in \mathrm{isoi}(U)(T,A_{\mathrm{o}})} Q_{\mathrm{m},U}(E, z_{\mathrm{o}})(B)}) : \\ &\qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\; T \in \mathcal{T}_{U,V_{\mathrm{o}}}\}) \supseteq \mathcal{A}_{U,V_{\mathrm{o}},1} \times \{T_{\mathrm{s}}\} \end{aligned}$$

where T_{s} is a *self transform*. That is, the maximisation does not yield a single solution for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$. Similarly to *classical modelled induction* in the case where the *derived* is *necessary* but *unknown*, in the case where the *idealisation* is *necessary* but *unknown*, the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, is just the *self transform*, $\tilde{T}_{\mathrm{o}} = T_{\mathrm{s}}$, which is the trivial case where everything is *known*.

Again, this singular solution for *unknown transform* can be addressed by relaxing the constraint that the *sample* be *necessarily drawn* from the *iso-idealisations* of the *distribution* to the constraint that the *sample* be *possibly drawn* from the *iso-idealisations* of the *distribution*. This is equivalent to assuming that the *sample* is *drawn* from the *uniform possible iso-idealisation historically distributed history probability function*, $P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}$, which is defined as the solution to

$$
\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}} := \\
&\quad \big(\bigcup \{\{(H, 1/\sum(P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}(G) : \\
&\qquad\qquad G \subseteq H_{\mathrm{h}}\%V_H,\ |G| = z_H,\ A_G * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_G} = A_H * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_H})) : \\
&\qquad\qquad\qquad H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H\}^{\wedge} : \\
&\qquad\qquad\qquad\qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\big)^{\wedge} \cup \\
&\quad \{(H, 0) : H \in \mathcal{H}_{U,X},\ H \not\subseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset, 0)\}
\end{aligned}
$$

The *uniform possible iso-idealisation historically distributed history probability function* is such that given a *drawn history* $H \in \mathcal{H}_{U,X}$

$$
\begin{aligned}
&\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) = \\
&\qquad \frac{\sum P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}
\end{aligned}
$$

The *possible history probability function*, $P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}$, is related to the *iso-idealisation conditional historical probability distribution*, $\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)$, in the same way as for the *necessary* case, $P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}$, except that the normalising fraction is restored. In the case where all *idealisations* are *possible* the normalising fraction is $1/|\mathrm{ran}(Y_{U,\mathrm{i},T_{\mathrm{o}},\dagger,z_H})|$. Any *historically drawn history* is *possible*,

$$\forall H \subseteq H_{\mathrm{h}}\%V_H\ (P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}(H) > 0)$$

but sometimes the *probability* is lower than in the *necessary* case,

$$\forall H \subseteq H_{\mathrm{h}}\%V_H\ (P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H) > 0 \iff P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}(H) \leq P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H))$$

The *uniform possible log likelihood* has similar properties to the *necessary log likelihood* but restores the normalising fraction,

$$\ln \hat{Q}_{\mathrm{m},\dagger,T,U}(E,z)(A) \;=\; \ln \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathrm{isoi}(U)(T,A)} Q_{\mathrm{m},U}(E,z)(B)} - \ln |\mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})|$$

The cardinality of the *idealisations*, $|\mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})|$, varies with the cardinality of the *derived*, $|\mathrm{ran}(D_{U,\mathrm{i},T,z})|$, which is equal to the cardinality of the *possible* *derived substrate histograms*,

$$\begin{aligned} |\mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})| &\sim |\mathrm{ran}(D_{U,\mathrm{i},T,z})| \\ &= \frac{(z+w'-1)!}{z!\ (w'-1)!} \end{aligned}$$

where $w' = |T^{-1}|$. So the additional term in the *uniform possible log likelihood*, $-\ln|\mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})|$, varies against the *possible derived volume*, w', where the *possible derived volume* is less than the *size*, $w' < z$, otherwise against the *size* scaled *log possible derived volume*, $z \ln w'$,

$$\begin{aligned} -\ln|\mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z})| &\sim -\ \mathrm{spaceCountsDerived}(U)(A,T) \\ &\sim -\ ((w'\ :\ w' < z)\ +\ (z \ln w'\ :\ w' \geq z)) \end{aligned}$$

In *uniform possible idealisation induction*, where (i) the *history probability function* is *uniform possible iso-idealisation historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample histogram* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-idealisation conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, so, if it is also the case that (iii) the *alignment* is intermediate or high, $\mathrm{algn}(A_{\mathrm{o}}) \geq \mathrm{algnMax}(U)(V_{\mathrm{o}}, z_{\mathrm{o}})/2$, (iv) the *relative entropy* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln|T_{\mathrm{o}}^{-1}|$, (v) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (vi) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *iso-idealisation conditional stuffed historical probability distribution* at the *maximum likelihood estimate* is such that in addition to the properties for *necessary idealisation induction*, formally stated above, the *log likelihood* varies against the *possible derived volume*, w'_{o}, where the *possible derived volume* is less than the *size*, $w'_{\mathrm{o}} < z_{\mathrm{o}}$, otherwise against the *size* scaled *log possible derived volume*, $z_{\mathrm{o}} \ln w'_{\mathrm{o}}$,

$$\ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim\ -\ ((w'_{\mathrm{o}}\ :\ w'_{\mathrm{o}} < z_{\mathrm{o}})\ +\ (z_{\mathrm{o}} \ln w'_{\mathrm{o}}\ :\ w'_{\mathrm{o}} \geq z_{\mathrm{o}}))$$

If, in addition, (vii) the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *idealisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, then conjecture that, in the case where the *sample history* is modal, $H_{\mathrm{o}} \in \mathrm{maxd}(P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}})$, the *log-likelihood* of the *iso-idealisation conditional stuffed historical probability distribution* varies with its *degree of structure* with respect to the *expanded specialising derived history coder*, $C_{\mathrm{G,T,H}}$,

$$\ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim\ \mathrm{structure}(U,X)(P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}, C_{\mathrm{G,T,H}}(T_{\mathrm{o}}))$$

Note that this correlation is conjectured to be weaker than that of *classical modelled induction*,

$$\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \mathrm{structure}(U,X)(P_{U,X,H_{\mathrm{h}},\mathrm{d,p},T_{\mathrm{o}}}, C_{\mathrm{G,T,H}}(T_{\mathrm{o}}))$$

because the correlation between the *log-likelihood* and the *specialising coder space* is weaker.

The *maximum likelihood estimate* for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$ in the *possible* case is

$$\{(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})\} \quad = \quad \mathrm{maxd}(\{((E,T), \hat{Q}_{\mathrm{m},\dagger,T,U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

if there is a unique maximum. This can be rewritten in terms of the *idealisation-dependent*,

$$\{\tilde{T}_{\mathrm{o}}\} \quad = \quad \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{m},\dagger,T,U}(A_{\mathrm{o}}^{\dagger(T)}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

Strictly speaking, this is only the case for the subset of *substrate transforms*, $\mathcal{T}_{U,V_{\mathrm{o}}}$, for which the *idealisation-dependent histogram*, $A_{\mathrm{o}}^{\dagger(T)}$, is defined.

If the optimisation is restricted to *ideal transforms*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}} \implies A_{\mathrm{o}}^{\dagger(T)} = A_{\mathrm{o}}$, then the optimisation is

$$\{\tilde{T}_{\mathrm{o}}\} \quad = \quad \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{m},\dagger,T,U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\})$$

In this case, all the *idealisations* are *possible* because the *distribution* equals the *sample*, so the optimisation is

$$\begin{aligned} &\{\tilde{T}_{\mathrm{o}}\} \\ &= \mathrm{maxd}(\{(T, \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},T,\dagger,z_{\mathrm{o}}})|} \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in \mathrm{isoi}(U)(T,A_{\mathrm{o}})} Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(B)}) : \\ &\qquad T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\}) \end{aligned}$$

Now, if the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, is unique, it is computable.

In this case, the numerator, $Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$, is constant.

The *maximum likelihood estimate* for the *model* is not *self*, $\tilde{T}_{\mathrm{o}} \neq T_{\mathrm{s}}$, if

$$\frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},\tilde{T}_{\mathrm{o}},\dagger,z_{\mathrm{o}}})|} \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in \mathrm{isoi}(U)(\tilde{T}_{\mathrm{o}},A_{\mathrm{o}})} Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(B)} \quad > \quad \frac{1}{|\mathrm{ran}(\mathcal{A}_{U,\mathrm{i},V_{\mathrm{o}},z_{\mathrm{o}}})|}$$

which is the case if the *iso-idealisation conditional multinomial probability* is greater than the inverted average cardinality of the *iso-idealisations*,

$$\frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in \mathrm{isoi}(U)(\tilde{T}_{\mathrm{o}},A_{\mathrm{o}})} Q_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}})(B)} \quad > \quad \frac{|\mathrm{ran}(Y_{U,\mathrm{i},\tilde{T}_{\mathrm{o}},\dagger,z_{\mathrm{o}}})|}{|\mathrm{dom}(Y_{U,\mathrm{i},\tilde{T}_{\mathrm{o}},\dagger,z_{\mathrm{o}}})|}$$

The *sample* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, so the *maximum likelihood estimate* for the *model* is not *unary*, $\tilde{T}_{\mathrm{o}} \neq T_{\mathrm{u}}$, if the *sample* is not *cartesian*, $\hat{A}_{\mathrm{o}} \neq \hat{V}_{\mathrm{o}}^{\mathrm{C}}$.

In some cases the *maximum likelihood estimate* for the *model* is neither *self* nor *unary*, $\tilde{T}_{\mathrm{o}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$.

In *idealisation induction*, where (i) the *history probability function* is *uniform possible iso-idealisation historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}$, given some *unknown substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample histogram* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, then the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-idealisation conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, and, if it is also the case that (iii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iv) the *scaled probability sample histogram* is *integral*, $A_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, in the *iso-idealisation conditional stuffed historical probability distribution* at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is

$$\tilde{T}_{\mathrm{o}} \quad \in \quad \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h},\dagger,T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\})$$

and in some cases the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, is non-trivial,

$$\tilde{T}_{\mathrm{o}} \quad \notin \quad \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$$

5.8 Abstract induction

In *classical modelled induction*, above, it was shown that if the *model*, $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, were *unknown* then for *necessary derived* there would be no unique solution for the *maximum likelihood estimate* for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$,

$$\mathrm{maxd}(\{((E,T), \frac{Q_{\mathrm{m},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in D_{U,\mathrm{i},T,z_{\mathrm{o}}}^{-1}(A_{\mathrm{o}} * T)} Q_{\mathrm{m},U}(E, z_{\mathrm{o}})(B)}) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ T \in \mathcal{T}_{U,V_{\mathrm{o}}}\}) \neq \{(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})\}$$

If the *iso-derived* condition is weakened from *necessary* to *possible*,

$$\mathrm{maxd}(\{((E,T),\hat{Q}_{\mathrm{h,d},T,U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

then in some cases the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, is non-trivial, $\tilde{T}_{\mathrm{o}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$, but at the cost of lower *likelihood*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H) > 0 \implies P_{U,X,H_{\mathrm{h}},\mathrm{d,p},T_{\mathrm{o}}}(H) \leq P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H)$. Similarly in *idealisation induction*, above, where instead of *necessary derived* the stricter *necessary idealisation* is required, the *iso-idealisation* condition must also be weakened from *necessary* to *possible* at the cost of lower *likelihood*, $P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H) > 0 \implies P_{U,X,H_{\mathrm{h}},\dagger,\mathrm{p},T_{\mathrm{o}}}(H) \leq P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H)$. Now, instead of weakening the condition from *necessary derived* to *possible derived*, consider weakening the condition to *necessary abstract*. Now there is a unique solution for the *maximum likelihood estimate* for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$.

First, however, consider the case where the given *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, is *known*.

Given some *known substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, the *abstract histogram* of the *distribution probability histogram* is $(\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$. In *abstract induction*, while the *distribution probability histogram*, $\hat{E}_{\mathrm{h}}$, remains *unknown*, the *abstract distribution probability histogram*, $(\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$, is *known* and *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have a *abstract probability histogram* equal to the *known abstract distribution probability histogram*, $(\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$. Define the *iso-abstract historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\rightarrow \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$,

$$\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}} := \\
&\quad \Big(\bigcup\{\{(H,1) : H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H, \\
&\qquad\qquad (\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}\}^{\wedge} : \\
&\qquad\qquad\qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ (\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} \neq (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}\} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \nsubseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset,0)\}
\end{aligned}$$

For *drawn histories* the *abstract probability histogram* is *necessary*, $\forall H \in \mathcal{H}_{U,X}\ (P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(H) > 0 \implies (\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}})$. Not all *sizes* and sets of *variables* are necessarily *drawable*. That is, in some cases, $\exists z \in \{1 \ldots z_{\mathrm{h}}\}\ \exists V \subseteq V_{\mathrm{h}}\ \forall H \in \mathcal{H}_{U,X}\ ((z_H = z) \wedge (V_H = V) \implies P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(H) = 0)$. The *distribution history* can always be *drawn*, so the *probability function* is not a *weak probability function*, $\sum_{H \in \mathcal{H}_{U,X}} P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(H) = 1$.

All *iso-abstract* subsets of the *distribution history* for a given set of *variables* and *size* are defined as equally probable,

$$\begin{aligned}&\forall V \subseteq V_{\mathrm{h}}\ \forall H,G \subseteq H_{\mathrm{h}}\%V \\ &\qquad ((A_G * T_{\mathrm{o}})^{\mathrm{X}} = (A_H * T_{\mathrm{o}})^{\mathrm{X}} \implies P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(G) = P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(H))\end{aligned}$$

In *abstract induction* the *history probability function* is *iso-abstract historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}$.

Given a *drawn history* $H \in \mathcal{H}_{U,X}$, where $P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(H) > 0$, the *iso-abstract historical probability* of *histogram* $A_H = \mathrm{histogram}(H) + V_H^{\mathrm{CZ}} \in \mathcal{A}_{U,\mathrm{i},V_H,z_H}$ is now conditional,

$$\begin{aligned}&\frac{Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},T_{\mathrm{o}},\mathrm{W},z_H}((A_{\mathrm{o}}*T_{\mathrm{o}})^{\mathrm{X}})} Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(B)} = \\ &\qquad\qquad \frac{\sum P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}\end{aligned}$$

where the *abstract* valued *histogram* function $Y_{U,\mathrm{i},T,\mathrm{W},z}$ is defined

$$Y_{U,\mathrm{i},T,\mathrm{W},z} = \{(A, (A * T)^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}$$

The finite set of *integral iso-abstracts* of *abstract* $(A * T)^{\mathrm{X}}$ is

$$Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}$$

The *iso-abstract historical probability* may be expressed in terms of a *histogram distribution* which is not explicitly conditional on the *necessary abstract*, $(\hat{E}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$,

$$\hat{Q}_{\mathrm{h},\mathrm{w},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) \ \ \propto \ \ \sum(P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *iso-abstract conditional stuffed historical probability distribution* is defined

$$\begin{aligned}&\hat{Q}_{\mathrm{h},\mathrm{w},T,U}(E, z) \\ &\quad := \ \ \{(A, \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\ &\qquad\quad \{(A, 0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \not\leq E\}\end{aligned}$$

which is defined if $z \leq \text{size}(E)$.

In the case where all the *abstracts* are *possible*,

$$\forall A' \in \text{ran}(Y_{U,\text{i},T,\text{W},z})\ \exists A \in \mathcal{A}_{U,\text{i},V,z}\ (((A * T)^{\text{X}} = A')\ \wedge\ (A \leq E))$$

the normalisation of the *iso-abstract conditional stuffed historical probability distribution* is a fraction $1/|\text{ran}(Y_{U,\text{i},T,\text{W},z})|$,

$$\begin{aligned}
&\hat{Q}_{\text{h,w},T,U}(E,z) \\
&\quad = \ \{(A, \frac{1}{|\text{ran}(Y_{U,\text{i},T,\text{W},z})|} \frac{Q_{\text{h},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\text{i},T,\text{W},z}((A*T)^{\text{X}})} Q_{\text{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\text{i},V,z}\}
\end{aligned}$$

In the case of a *full functional transform*, $T_{\text{f}} = \{\{w\}^{\text{CS}\{\}V\text{T}} : w \in V\}^{\text{T}}$, the *iso-abstracts* equals the *iso-independents*, $Y^{-1}_{U,\text{i},T_{\text{f}},\text{W},z}((A * T_{\text{f}})^{\text{X}}) = Y^{-1}_{U,\text{i},V,z}(A^{\text{X}})$. Thus the *iso-abstract conditional stuffed historical probability distribution* equals the *iso-independent conditional stuffed historical probability distribution*, $\hat{Q}_{\text{h,w},T_{\text{f}},U}(E,z) = \hat{Q}_{\text{h,y},U}(E,z)$. In this case *abstract induction* reduces to *aligned non-modelled induction*.

At the other extreme of a *unary transform*, $T_{\text{u}} = \{V^{\text{CS}}\}^{\text{T}}$, the set of *iso-abstracts* equals the *substrate histograms*, $Y^{-1}_{U,\text{i},T_{\text{u}},\text{W},z}((A * T_{\text{u}})^{\text{X}}) = \mathcal{A}_{U,\text{i},V,z}$. Thus the *iso-abstract conditional stuffed historical probability distribution* equals the *stuffed historical probability distribution*, $\hat{Q}_{\text{h,w},T_{\text{f}},U}(E,z) = \hat{Q}_{\text{h},U}(E,z)$. In this case *abstract induction* reduces to *classical non-modelled induction*.

The *iso-abstract conditional generalised multinomial probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\text{m,w},T,U}(E,z) \\
&\quad := \ \{(A, \frac{Q_{\text{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\text{i},T,\text{W},z}((A*T)^{\text{X}})} Q_{\text{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\text{i},V,z},\ A^{\text{F}} \leq E^{\text{F}}\}^{\wedge} \cup \\
&\qquad\quad \{(A, 0) : A \in \mathcal{A}_{U,\text{i},V,z},\ A^{\text{F}} \nleq E^{\text{F}}\}
\end{aligned}$$

which is defined if $\text{size}(E) > 0$.

The case where all the *abstracts* are *possible* is weaker than for *historical*,

$$\forall A' \in \text{ran}(Y_{U,\text{i},T,\text{W},z})\ \exists A \in \mathcal{A}_{U,\text{i},V,z}\ (((A * T)^{\text{X}} = A')\ \wedge\ (A^{\text{F}} \leq E^{\text{F}}))$$

In this case the *iso-abstract conditional generalised multinomial probability distribution* is

$$\begin{aligned}
&\hat{Q}_{\mathrm{m,w},T,U}(E,z) \\
&= \{(A, \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},T,\mathrm{W},z})|} \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}
\end{aligned}$$

It is assumed that the *distribution history size*, z_{h}, is large with respect to the *sample size* $z_{\mathrm{o}} = \mathrm{size}(A_{\mathrm{o}})$, so that, in the limit, the *iso-abstract historical probability*, $\hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$, approximates to the *iso-abstract multinomial probability*, $\hat{Q}_{\mathrm{m,w},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$. That is, if $z_{\mathrm{o}} \ll z_{\mathrm{h}}$ then

$$\hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(E_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}}) \approx \hat{Q}_{\mathrm{m,w},T_{\mathrm{o}},U}(E_{\mathrm{o}},z_{\mathrm{o}})(A_{\mathrm{o}})$$

In the case of *completely effective sample histogram*, $A^{\mathrm{F}}_{\mathrm{o}} = V^{\mathrm{C}}_{\mathrm{o}}$, the maximisation for *known transform*, T_{o}, of the *iso-abstract conditional generalised multinomial probability* parameterised by the *complete congruent histograms* of unit *size* is a singleton of the *rational maximum likelihood estimate*

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m,w},T_{\mathrm{o}},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

In the case where the *sample histogram* is not *completely effective*, $A^{\mathrm{F}}_{\mathrm{o}} < V^{\mathrm{C}}_{\mathrm{o}}$, the maximisation of the *iso-abstract conditional generalised multinomial probability distribution* for *known transform* is not necessarily a singleton

$$|\mathrm{max}(\{(E, \hat{Q}_{\mathrm{m,w},T_{\mathrm{o}},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})| \geq 1$$

In the case where the maximisation of the *iso-abstract conditional generalised multinomial probability distribution* is a singleton, it is equal to the *normalised abstract-dependent*, $\tilde{E}_{\mathrm{o}} = \hat{A}^{\mathrm{W}(T_{\mathrm{o}})}_{\mathrm{o}}$, where the *abstract-dependent* $A^{\mathrm{W}(T)} \in \mathcal{A}_{U,V,z}$ is defined in 'Likely histograms', above, as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-abstract*,

$$\begin{aligned}
&\{A^{\mathrm{W}(T)}\} = \\
&\quad \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})}) : D \in \mathcal{A}_{U,V,z}\})
\end{aligned}$$

The *abstract-dependent*, $A^{\mathrm{W}(T)}$, is sometimes not computable. The finite approximation to the *abstract-dependent* is

$$\begin{aligned}
&\{A^{\mathrm{W}(T)}_k\} = \\
&\quad \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})
\end{aligned}$$

The approximation, $A_k^{\mathrm{W}(T)} \approx A^{\mathrm{W}(T)}$, improves as the scaling factor, k, increases.

Unlike in *classical non-modelled induction* where the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, is equal to the *sample probability histogram*, $\hat{A}_{\mathrm{o}}$, in *abstract induction* the *maximum likelihood estimate* is not necessarily equal to the *sample probability histogram*. It is only in the case where the *sample histogram* is *naturalised abstract* that the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger} \implies A_{\mathrm{o}}^{\mathrm{W}(T_{\mathrm{o}})} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

Otherwise, the overall *maximum likelihood estimate*, which is the *abstract-dependent*, is near the *histogram*, $\tilde{E}_{\mathrm{o}} \sim \hat{A}_{\mathrm{o}}$, only in as much as it is far from the *naturalised abstract*, $\tilde{E}_{\mathrm{o}} \nsim (\hat{A}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger}$.

The requirement that the *distribution history* itself be *drawable*, $P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(H_{\mathrm{h}}) > 0$, has been ignored so far. This requirement modifies the maximisation to add the constraint that the *maximum likelihood estimate* be an *iso-abstract*, $(\tilde{E}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{A}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$,

$$\begin{aligned}\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m,w},T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\\ (E * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{A}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}\})\end{aligned}$$

So, strictly speaking, the *maximum likelihood estimate* is only approximately equal to the *normalised abstract-dependent*, $\tilde{E}_{\mathrm{o}} \approx \hat{A}_{\mathrm{o}}^{\mathrm{W}(T_{\mathrm{o}})}$, if the *abstract-dependent* is not an *iso-abstract*, $(A_{\mathrm{o}}^{\mathrm{W}(T_{\mathrm{o}})} * T_{\mathrm{o}})^{\mathrm{X}} \neq (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$. In the special case, however, where the *sample histogram* is *naturalised abstract*, the *maximum likelihood estimate* is exactly equal to the *normalised abstract-dependent*, $A_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{W}(T_{\mathrm{o}})} = \hat{A}_{\mathrm{o}}$.

In *abstract induction*, where (i) the *history probability function* is *iso-abstract historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample histogram* is *naturalised abstract*, $A_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-abstract conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is

$$\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

Classical modelled induction is termed *law-like* because the set of *iso-deriveds*, $D_{U,T,z}^{-1}(A * T)$, where the *derived*, $A * T$, is *necessary*, is defined as *law-like*. All *iso-sets* that are subsets of the *iso-deriveds* are also *law-like* because the *derived* is still *necessary*. So *idealisation induction* is also termed *law-like*, because the set of *iso-idealisations* is a subset of the set of *iso-deriveds*, $Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \subseteq D^{-1}_{U,T,z}(A * T)$.

The *iso-derivedence*, or degree of *law-likeness*, of the *iso-abstracts* equals the *iso-abstractence* of the *iso-deriveds*,

$$\frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \leq 1$$

So *abstract induction* is not maximally *law-like* if the *iso-deriveds* is a proper subset of the *iso-abstracts*, $D^{-1}_{U,T,z}(A * T) \subset Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$.

Abstract induction is termed *entity-like* because the set of *iso-abstracts*, $Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$, where the *abstract*, $(A * T)^{\mathrm{X}}$, is *necessary*, is defined as *entity-like*. That is, the *iso-abstractence* or degree of *entity-likeness* of the *iso-abstracts* is one, and so *abstract induction* is maximally *entity-like*. *Law-like iso-sets* are subsets of the set of *iso-abstracts*, $D^{-1}_{U,T,z}(A * T) \subseteq Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$, and so are also *entity-like*. So *abstract induction* may be said to be more *entity-like* and less *law-like* than either *classical modelled induction* or *idealisation induction*.

As conditions are added to *abstract induction* that increase the *law-likeness*, or the degree to which the *derived* is *necessary*, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, tends from the *abstract-dependent*, $\hat{A}^{\mathrm{W}(T_{\mathrm{o}})}_{\mathrm{o}}$, to the *derived-dependent*, $\hat{A}^{\mathrm{D}(T_{\mathrm{o}})}_{\mathrm{o}}$. That is, given *iso-set* $I \subset \mathcal{A}_{U,\mathrm{i},V,z}$, which is such that $D^{-1}_{U,\mathrm{i},T,z}(A*T) \subseteq I \subseteq Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})$, as the *iso-derivedence* increases and the *iso-abstractence* decreases, the type of *induction* moves from *entity-like* to *law-like*, implying a more *classical dependent analogue*, $\hat{A}^{\mathrm{I}(T_{\mathrm{o}})}_{\mathrm{o}} \approx \hat{A}^{\mathrm{D}(T_{\mathrm{o}})}_{\mathrm{o}}$, and so a more *classical maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}} \approx \hat{A}^{\mathrm{D}(T_{\mathrm{o}})}_{\mathrm{o}}$.

Also, even if there are no additional conditions and the *iso-set* remains equal to the *iso-abstracts*, $I = Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$, constraints on the *sample* can make the denominator, $\sum Q_{\mathrm{m},U}(A^{\mathrm{W}(T)}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})$, more approximate to the *iso-derived* denominator, $\sum Q_{\mathrm{m},U}(A^{\mathrm{D}(T)}, z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)$. In this way also *abstract induction* can sometimes be more like *classical modelled induction* even if the *iso-derivedence* remains unchanged.

The *iso-independence* of the *iso-abstracts* is

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \ \cup \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

In some cases the *iso-independence* of the *iso-idealisations* is greater than or equal to the *iso-independence* of the *iso-abstracts*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} \ \geq \ \frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \ \cup \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

and so *abstract induction* may be said to be less *aligned-like* than *idealisation induction*. However, the *derived iso-independence* of the *integral lifted iso-abstracts* is necessarily greater than or equal to the *derived iso-independence* of any *law-like iso-set*,

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|} \ \geq \ \frac{1}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

and so *abstract induction* may be said to be more *derived aligned-like* than either *classical modelled induction* or *idealisation induction*.

As the *iso-independence* increases, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, which equals the *abstract-dependent*, $\hat{A}_{\mathrm{o}}^{\mathrm{W}(T_{\mathrm{o}})}$, tends to the *dependent*, $\hat{A}_{\mathrm{o}}^{\mathrm{Y}}$, which is independent of the *model*, T_{o}, because the *independent analogue*, $(A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger}$, tends to the *independent*, $A_{\mathrm{o}}^{\mathrm{X}}$, which is also independent of the *model*, as the *transform* tends to *full functional*. As the *derived iso-independence* increases, however, the *lifted independent analogue*, $A_{\mathrm{o}}^{\mathrm{U}(T_{\mathrm{o}})'}$, tends to the *abstract*, $(A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, which is not independent of the *model*, T_{o}.

Given the *known substrate transform*, T_{o}, consider the *maximum likelihood estimate* of the *iso-abstract conditional generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m,w},T_{\mathrm{o}},U}$.

The *independent-analogue* or *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, is the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of membership of the *iso-abstracts*,

$$\begin{aligned}&\{(A * T)^{\mathrm{X}} * T^{\dagger}\} = \\ &\quad \mathrm{maxd}(\{(D, \textstyle\sum(Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}))) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The corresponding *dependent-analogue* or *abstract-dependent*, $A^{\mathrm{W}(T)}$, is the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-abstract*,

$$\begin{aligned}\{(A^{\mathrm{W}(T)}, \frac{Q_{\mathrm{m},U}(A^{\mathrm{W}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{W}(T)},z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})})\} = \\ \max(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})}) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

In section 'Likely histograms', above, the logarithm of the *maximum conditional probability* with respect to the *dependent-analogue* is conjectured to vary with the *relative space* with respect to the *independent-analogue*. In the case of *iso-abstract conditional*,

$$\begin{aligned}\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{W}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{W}(T)},z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})} \sim \\ \mathrm{spaceRelative}((A*T)^{\mathrm{X}} * T^{\dagger})(A)\end{aligned}$$

where the *distribution-relative multinomial space* is defined, in section 'Likely histograms', above, as

$$\mathrm{spaceRelative}(E)(A) \quad := \quad -\ln \frac{\mathrm{mpdf}(U)(E,z)(A)}{\mathrm{mpdf}(U)(E,z)(E)}$$

The set of *iso-abstracts* is *entity-like* so the *derived*, $A*T$, and the *dependent derived*, $A^{\mathrm{W}(T)} * T$, are not necessarily equal to each other and nor are they necessarily equal to the *abstract*, $(A*T)^{\mathrm{X}}$. In section 'Transform alignment', above, it is conjectured that the relation between the *relative spaces*,

$$\begin{aligned} 0 &= \mathrm{spaceRelative}((A*T)^{\mathrm{X}} * T^{\dagger})((A*T)^{\mathrm{X}} * T^{\dagger}) \\ &\leq \mathrm{spaceRelative}((A*T)^{\mathrm{X}} * T^{\dagger})(A) \\ &\leq \mathrm{spaceRelative}((A*T)^{\mathrm{X}} * T^{\dagger})(A^{\mathrm{W}(T)})\end{aligned}$$

can be *lifted* and so the *dependent analogue derived alignment* is conjectured to be greater than or equal to the *derived alignment* which in turn is greater than or equal to the *independent analogue derived alignment*,

$$0 = \mathrm{algn}((A*T)^{\mathrm{X}}) \leq \mathrm{algn}(A*T) \leq \mathrm{algn}(A^{\mathrm{W}(T)} * T)$$

The *abstract-dependent* varies with the *histogram*, $A^{\mathrm{W}(T)} \sim A$, so conjecture that the *log-likelihood* varies with the *derived alignment*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{W}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{W}(T)},z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A*T)^{\mathrm{X}})} \quad \sim \quad \mathrm{algn}(A*T)$$

The derivation of this correlation can be seen more clearly in terms of a decomposition into three separate correlations. First, conjecture that the logarithm of the *iso-abstract conditional multinomial probability* of the *histogram*, A, with respect to the *dependent analogue* or *abstract-dependent*, $A^{\mathrm{W}(T)}$, varies against the logarithm of the *iso-abstract conditional multinomial probability* with respect to the *independent analogue* or *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{W}(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{W}(T)}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})} \sim$$
$$- \ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(A)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})}$$

This relation is called the *dependent-independent anti-correlation.* As shown in 'Likely histograms', above, the strength of the *dependent-independent anti-correlation* depends on the *relative space* of the *histogram* with respect to the *independent analogue*, $\mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A)$.

Second, conjecture that the negative logarithm of the *iso-abstract conditional multinomial probability* of the *histogram*, A, with respect to the *independent analogue* or *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, varies with the negative logarithm of the *lifted iso-abstract conditional multinomial probability* of the *derived*, $A * T$, with respect to the *lifted independent analogue* or *abstract*, $(A * T)^{\mathrm{X}}$,

$$- \ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(A)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})} \sim$$
$$- \ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(B') : B' \in \mathrm{isowl}(U)(T, A)}$$

where the *lifted integral iso-abstracts* is abbreviated

$$\mathrm{isowl}(U)(T, A) := \{B * T : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\}$$

This correlation is called the *underlying-lifted correlation. Lifting* the *iso-abstracts* is functional,

$$\{(A * T, (A * T)^{\mathrm{X}}) : A \in \mathcal{A}_{U,V,z}\} \;\; \in \;\; \mathcal{A}_{U,W,z} \to \mathcal{A}_{U,W,z}$$

and

$$\{(A * T, Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,V,z}\} \;\; \in \;\; \mathcal{A}_{U,W,z} \to \mathrm{P}(\mathcal{A}_{U,V,z})$$

so the *underlying-lifted correlation* is expected to be positive.

Third, conjecture that, in the case where the *abstract* is *integral*, $(A * T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, the denominator of the *lifted iso-abstract conditional multinomial probability* is dominated by the *abstract* term, $Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)((A * T)^{\mathrm{X}})$, and similar terms, and so the negative logarithm of the *lifted iso-abstract conditional multinomial probability* with respect to the *lifted independent analogue* or *abstract*, $(A*T)^{\mathrm{X}}$, varies with the negative logarithm of the *relative multinomial probability* with respect to the *abstract*, $(A*T)^{\mathrm{X}}$, which is the *relative space* with respect to the *abstract*, which is the *derived alignment*,

$$\begin{aligned}
&-\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(B') : B' \in \mathrm{isowl}(U)(T, A)} \\
&\qquad\sim\ -\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)((A * T)^{\mathrm{X}})} \\
&\qquad=\ \mathrm{spaceRelative}((A * T)^{\mathrm{X}})(A * T) \\
&\qquad=\ \mathrm{algn}(A * T)
\end{aligned}$$

This correlation is called the *conditional-relative correlation*. The strength of the *conditional-relative correlation* increases with the *derived iso-independence* of the *integral lifted iso-abstracts*,

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

As the *derived iso-independence* increases, the *lifted abstract-independent*, $A^{\mathrm{U}(T)'}$, tends to the *abstract*, $(A * T)^{\mathrm{X}}$, and the *lifted abstract-independent* term, $Q_{\mathrm{m},U}((A*T)^{\mathrm{X}}, z)(A^{\mathrm{U}(T)'})$, tends to equal the *abstract* term, $Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)((A*T)^{\mathrm{X}})$, in the case where both are *integral*, $A^{\mathrm{U}(T)'}, (A*T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$. In the limit, the *lifted iso-abstract conditional multinomial probability*, with respect to the *independent-analogue*, $(A * T)^{\mathrm{X}}$, equals the *iso-independent conditional multinomial probability*, with respect to the *independent*, $(A*T)^{\mathrm{X}}$, of the *layer* above, which is where the *substrate histogram* is the *derived histogram*, $A * T$,

$$\begin{aligned}
&-\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(B') : B' \in \mathrm{isowl}(U)(T, A)} \\
&\qquad\sim\ -\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})}
\end{aligned}$$

The corresponding *iso-independent conditional multinomial probability*, with respect to the *dependent*, $(A * T)^{\mathrm{Y}}$, of the *derived layer* is shown in section

'Aligned induction', above, to vary with the *alignment* of the *derived layer's substrate histogram*, $A * T$,

$$\begin{aligned}
-\ln \frac{Q_{\mathrm{m},U}((A*T)^{\mathrm{X}},z)(A*T)}{\sum Q_{\mathrm{m},U}((A*T)^{\mathrm{X}},z)(B) : B \in Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})} \\
\sim \ \ln \frac{Q_{\mathrm{m},U}((A*T)^{\mathrm{Y}},z)(A*T)}{\sum Q_{\mathrm{m},U}((A*T)^{\mathrm{Y}},z)(B) : B \in Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})} \\
\sim \ \mathrm{algn}(A*T)
\end{aligned}$$

That is, in the case where the *derived iso-independence* is high, *abstract induction* may be viewed as *aligned non-modelled induction* of the *derived.*

In *abstract induction*, where (i) the *history probability function* is *iso-abstract historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iii) the *scaled estimate distribution histogram* is *integral*, $\tilde{E}_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-abstract conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *relative space* of the *sample* with respect to the *naturalised abstract*,

$$\ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ \mathrm{spaceRelative}((A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger})(A_{\mathrm{o}})$$

and varies with the *derived alignment*,

$$\ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

The *derived alignment* of the *maximum likelihood estimate* is greater than or equal to that of the *sample*,

$$\mathrm{algn}(Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}} * T) \ \geq \ \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

In section 'Classical modelled induction', above, it is shown that the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})$, can be related to queries on the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, in the special case where the *sample histogram* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$. The given *substrate transform* must be such that its *contraction* has *underlying variables* that are a subset of the query *variables*, $\mathrm{und}(T_{\mathrm{o}}^{\%}) \subseteq K$. In the case where the query *histogram* consists of one *effective state*, $Q = \{(S_Q, 1)\}$, the application of the query in terms of a modified *sample histogram* is

$$\begin{aligned}&(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K) = \\ &\quad \{(N, (\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}))^{1/z_{\mathrm{o}}}) : N \in (V_{\mathrm{o}} \setminus K)^{\mathrm{CS}}, \\ &\qquad A_{Q,N} = A_{\mathrm{o}} - (A_{\mathrm{o}} * C_Q) + ((A_{\mathrm{o}} * C_Q)\ \%\ K * \{N\}^{\mathrm{U}})\}^{\wedge}\end{aligned}$$

where $\{R_Q\} = (Q * T_{\mathrm{o}}^{\%})^{\mathrm{FS}}$, $C_Q = T_{\mathrm{o}}^{-1}(R_Q)$ and $\mathrm{his} = \mathrm{histogram} \in \mathcal{T} \to \mathcal{A}$. If the *sample histogram* is *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the modified *sample histogram*, $A_{Q,N}$, can be *drawn* from the *distribution*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}) > 0$, because its *derived* is equal to the *known derived*, $A_{Q,N} * T_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}}$. That is, the modified *sample histogram* is in the *iso-deriveds*, $A_{Q,N} \in D_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}}^{-1}(A_{\mathrm{o}} * T_{\mathrm{o}})$.

However, in the case of *abstract induction*, where the *abstract* is *necessary*, although the modified *sample histogram* is in the *iso-abstracts*, $A_{Q,N} \in Y_{U,\mathrm{i},T_{\mathrm{o}},\mathrm{W},z}^{-1}((A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}})$, the modified *derived*, $\hat{A}_{Q,N} * T_{\mathrm{o}}$, is not *necessarily* equal to that of the *distribution*, $\hat{E}_{\mathrm{h}} * T_{\mathrm{o}}$. That is, in some cases $\hat{A}_{Q,N} * T_{\mathrm{o}} \neq \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}$. Only the modified *abstract* is *necessary*, $(\hat{A}_{Q,N} * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$. Furthermore, even if the *sample* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{W}(T)}$, is not necessarily equal to the *sample*, $\tilde{E}_{\mathrm{o}} \neq \hat{A}_{\mathrm{o}}$. So it cannot be assumed that application of the query via the *model* of the *sample* is equal to the query via the *model* of the *distribution*, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K) \neq (Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * E_{\mathrm{h}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$. Nor can the query via the *model* of the *sample*, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$, be expressed in terms of the *iso-abstract conditional stuffed historical probability distribution* at the *scaled naturalised sample*, $\hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$.

Consider the constraints that may be added to *abstract induction* to increase the resemblance to *classical modelled induction*, so that queries via the *model* of the *sample* are more approximate to queries via the *model* of the *distribution*, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K) \approx (Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * E_{\mathrm{h}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$.

The set of *law-like iso-deriveds* are a subset of the set of *entity-like iso-abstracts*, $D_{U,T,z}^{-1}(A * T) \subseteq Y_{U,T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})$, so conjecture that the logarithm of the fraction of the sum of the *iso-abstract multinomial probabilities*, with respect to the *naturalisation*, $A * T * T^{\dagger}$, that are *iso-derived* varies as the *relative space* of the *naturalisation* with respect to the *naturalised abstract*,

$$\begin{aligned}
\ln \frac{\sum Q_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)}{\sum Q_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})} \\
\sim\ -\ \mathrm{spaceRelative}(A * T * T^{\dagger})((A * T)^{\mathrm{X}} * T^{\dagger}) \\
\sim\ \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A * T * T^{\dagger})
\end{aligned}$$

If the *relative space* is high, the elements of the *iso-abstracts* which are not *iso-deriveds* and so do not have the same *derived* as the *naturalisation*, $A * T * T^{\dagger} * T = A * T$, have low *multinomial probability* with respect to the *naturalisation*,

$$\sum (Q_{\mathrm{m},U}(A * T * T^{\dagger}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \setminus D^{-1}_{U,\mathrm{i},T,z}(A * T)) \ \approx\ 0$$

If the *sample* is *known* to be *naturalised*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, then as the *relative space* of the *sample* with respect to the *naturalised sample abstract*, $\mathrm{spaceRelative}((A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger})(A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger})$, increases, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, which is the *abstract-dependent*, $\hat{A}_{\mathrm{o}}^{\mathrm{W}(T_{\mathrm{o}})}$, tends to the *derived-dependent* which equals the *naturalisation*, $\hat{A}_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, and away from the *naturalised abstract*, $(A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger}$. Thus, the *maximum likelihood estimate* is more *classical* if the *sample* is *known* to be *naturalised* and the *relative space* is high.

The *relative space* of the *histogram* with respect to the *naturalised abstract* varies with the *lifted relative space*, which equals the *derived alignment*,

$$\begin{aligned}
&\mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A) \\
&\quad \sim\ \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger} * T)(A * T) \\
&\quad =\ \mathrm{spaceRelative}((A * T)^{\mathrm{X}})(A * T) \\
&\quad =\ \mathrm{algn}(A * T)
\end{aligned}$$

depending on the *underlying-lifted correlation* and the *conditional-relative correlation*.

The *conditional-relative correlation* improves as the *derived iso-independence* increases and the *lifted abstract-independent*, $A^{\mathrm{U}(T)'}$, tends to the *abstract*, $(A*T)^{\mathrm{X}}$. In the case where the *formal* is *independent*, $A^{\mathrm{X}}*T = (A^{\mathrm{X}}*T)^{\mathrm{X}}$, the *possible derived volume* equals the *derived volume*, $w' = w$ where $w' = |T^{-1}|$ and $w = |W^{\mathrm{C}}|$. As shown in 'Deltas and Perturbations', above, any subset of the *integral congruent deltas* $Q_A \subset \mathcal{A}_{\mathrm{i}} \times \mathcal{A}_{\mathrm{i}}$ which conserves *iso-independence*,

$\forall (D,I) \in Q_A\ (A - D + I \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}))$, is a linear *sum* of *circuit deltas*, where the *circuit deltas* are defined as the subset of *iso-independent deltas* having *size* less than or equal to two, $C_A = \{(D,I) : (D,I) \in Q_A,\ \mathrm{size}(I) \leq 2\}$. The set of *lifted iso-independent deltas*, $\forall (D',I') \in Q_{A'}\ (A * T - D' + I' \in \{B * T : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\})$, must be smaller than the set of *derived iso-independent deltas*, $\forall (D',I') \in Q_{A*T}\ (A*T - D' + I' \in Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}}))$, if the *possible derived volume* is less than the *derived volume*, $w' < w \implies |Q_{A'}| < |Q_{A*T}|$, because *circuit deltas* cannot be constructed on *impossible states*. Therefore in the case of *independent formal*, where the *possible derived volume* equals the *derived volume*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}} \implies w' = w$, the *derived iso-independence* is greater than would otherwise be the case, and the *lifted abstract-independent* approximates to the *abstract*, $A^{\mathrm{U}(T)'} \approx (A * T)^{\mathrm{X}}$.

If the *sample* is *known* to have *independent formal*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$, the correlation between the *relative space* of the *histogram* with respect to the *naturalised abstract* and the *derived alignment*,

$$\mathrm{spaceRelative}((A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger})(A_{\mathrm{o}}) \quad \sim \quad \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

is higher than would be the case if there was *formal alignment*, $\mathrm{algn}(A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) > 0$.

So the *maximum likelihood estimate* is more *classical* and less *formal* if (i) the *sample* is *naturalised*, (ii) the *sample* has *independent formal* and (iii) the *derived alignment* is high.

In *abstract induction*, where (i) the *history probability function* is *iso-abstract historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, (iii) the *sample formal* is *independent*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$, (iv) the *derived alignment* is high, $\mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}}) \gg 0$, (v) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (vi) the *scaled estimate distribution histogram* is *integral*, $\tilde{E}_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-abstract conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *naturalisation*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

the *formal alignment* of the *maximum likelihood estimate* is small,

$$\mathrm{algn}(Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) \quad \approx \quad 0$$

and the *derived* of the *maximum likelihood estimate* approximates to the *normalised sample derived*,

$$\tilde{E}_{\mathrm{o}} * T_{\mathrm{o}} \;\;\approx\;\; \hat{A}_{\mathrm{o}} * T_{\mathrm{o}}$$

In this case queries via the *model* of the *sample* approximate to queries via the *model* of the *distribution*,

$$(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge} \;\%\; (V_{\mathrm{o}} \setminus K) \;\;\approx\;\; (Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * E_{\mathrm{h}})^{\wedge} \;\%\; (V_{\mathrm{o}} \setminus K)$$

That is, at high *derived alignments* where the *sample* is *known* to be *natural* and the *sample formal* is *known* to be *independent*, *abstract induction* has similar properties to *non-formal classical modelled induction*.

If the *relative entropy* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *naturalisation* varies with the *derived entropy*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \;\;\sim\;\; z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

Note, however, that because the *abstract induction* is more *derived aligned-like* than *classical modelled induction*,

$$\frac{|\{B * T : B \in Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})\}|}{|Y_{U,\mathrm{i},W,z}^{-1}((A * T)^{\mathrm{X}})|} \;\;\geq\;\; \frac{1}{|Y_{U,\mathrm{i},W,z}^{-1}((A * T)^{\mathrm{X}})|}$$

the *sum sensitivity* of the *iso-abstract conditional stuffed historical probability distribution* at the *maximum likelihood estimate* may be expected rather to vary against the *derived alignment*,

$$\begin{aligned}
&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \\
&\qquad \sim z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}}) - z_{\mathrm{o}} \times \mathrm{entropy}((A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}) \\
&\qquad \approx -\ \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})
\end{aligned}$$

So the *log likelihood* of the *iso-abstract conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *sum sensitivity* of the *iso-abstract conditional stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\;&\sim\;\; \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}}) \\
&\sim\;\; -\ \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))
\end{aligned}$$

In the case of high *relative entropy*, the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* is conjectured to vary with the *unknown-known multinomial probability distribution sum sensitivity difference*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim \\ &\quad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}}))) - \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}} * T_{\mathrm{o}}, z_{\mathrm{o}})))\end{aligned}$$

so the *sum sensitivity* of the *iso-abstract conditional stuffed historical probability distribution* is also conjectured to vary with the *unknown-known multinomial probability distribution sum sensitivity difference*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \sim \\ &\quad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}}, z_{\mathrm{o}}))) - \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}} * T_{\mathrm{o}}, z_{\mathrm{o}})))\end{aligned}$$

the *sum sensitivity* of the *iso-abstract conditional stuffed historical probability distribution* at the *maximum likelihood estimate* is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))\end{aligned}$$

and the *log likelihood* of the *iso-abstract conditional stuffed historical probability distribution* at the *maximum likelihood estimate* is greater than or equal to the *log likelihood* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{\mathrm{o}}) \quad \geq \quad \ln \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{\mathrm{o}})$$

In the case where (i) the *sample* is *natural*, (ii) the *sample formal* is *independent*, and (iii) the *relative entropy* is high, as the *derived alignment* increases (a) the *non-formal classical log-likelihood* increases and (b) the *underlying-derived sum sensitivity difference* decreases.

If, in addition, the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, then the *log likelihood* varies against the *specialising derived substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

where

$$C_{\mathrm{G},V,\mathrm{T,H}}(T) = \mathrm{coderHistorySubstrateDerivedSpecialising}(U, X, T, D_{\mathrm{S}}, D_{\mathrm{X}})$$

So the *sum sensitivity* varies against the *log-likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}}))) \sim -\ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}})$$

the *sensitivity* to *model* also varies against the *log likelihood*,

$$\begin{aligned}- \ln|\max(\{(T,\hat{Q}_{\mathrm{h,w},T,U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}}^{\mathrm{X}} * T)^{\mathrm{X}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})| \sim \\ - \ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}})\end{aligned}$$

and the *log-likelihood* varies with its *degree of structure* with respect to the *expanded specialising derived history coder*, $C_{\mathrm{G,T,H}}$,

$$\ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \mathrm{structure}(U,X)(P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}},C_{\mathrm{G,T,H}}(T_{\mathrm{o}}))$$

Note that, although the added constraint of *known natural sample*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, can increase the resemblance to *classical induction*, the *induction* remains *abstract induction* because the condition of *necessary abstract* has not changed and so neither the *iso-set*, $Y^{-1}_{U,\mathrm{i},T_{\mathrm{o}},\mathrm{W},z_{\mathrm{o}}}((A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}})$, nor the *iso-derivedence* have changed. That is, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, does not move away from the *abstract-dependent*, $\hat{A}_{\mathrm{o}}^{\mathrm{W}(T_{\mathrm{o}})}$, to the *derived-dependent*, $\hat{A}_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})}$, but rather both the *maximum likelihood estimate* and the *abstract-dependent* move together towards the *derived-dependent*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{W}(T_{\mathrm{o}})} \approx \hat{A}_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})}$.

Note also that the assumption of high *derived alignment*, $\mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}}) \gg 0$, is not well defined, although there is an upper bound, $\mathrm{algnMax}(U)(W_{\mathrm{o}},z_{\mathrm{o}})$. A more formal method of expression would be to say that the correlation between the *iso-abstract conditional stuffed historical probability distribution* and the *iso-derived conditional stuffed historical probability distribution* is itself correlated to the *derived alignment*,

$$[\ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}})] \sim \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

More formal still would be to define this relation in terms of the correlations of functions of the *sized cardinal substrate histograms*, $\mathcal{A}_z$, given the *renormalised geometry-weighted probability function*, $\mathrm{corr}(z) \in (\mathcal{A}_z \to \mathbf{R}) \times (\mathcal{A}_z \to \mathbf{R}) \to \mathbf{R}$, as in section 'Substrate structures alignment', above.

In the discussion above, the *model*, $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, is *known*, and the *abstract*, $(\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$, is both *necessary* and *known*. Optimisation can be done to find the *maximum likelihood estimate* of the *distribution histogram* for *known model*,

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E,\hat{Q}_{\mathrm{m,w},T_{\mathrm{o}},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

Just as in the discussion above of *classical modelled induction*, consider the case where the *abstract*, $(\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$, is still *necessary* but the *model*, T_{o}, is *unknown* and so the *abstract* is *unknown*. Again, the *maximum likelihood estimate* for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$ can be defined as an optimisation of the *multinomial probability* conditional on the *iso-abstracts* where both the *distribution histogram* and *transform* are treated as arguments to a likelihood function,

$$\begin{aligned}
&\{(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})\} \\
&\quad= \mathrm{maxd}(\{((E,T), \frac{Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z_{\mathrm{o}}}((A_{\mathrm{o}}*T)^{\mathrm{X}})} Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(B)}) : \\
&\qquad\qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})
\end{aligned}$$

It is conjectured that in *abstract induction* there are some cases in which there is a unique solution for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$. This is because in *entity-like induction*, but not *law-like induction*, the denominator does not necessarily reduce to equal the numerator, so avoiding degeneracy. In the case where there is a unique solution then the maximisation can be rewritten in terms of the *abstract-dependent*,

$$\{\tilde{T}_{\mathrm{o}}\} = \mathrm{maxd}(\{(T, \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{W}(T)},z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z_{\mathrm{o}}}((A_{\mathrm{o}}*T)^{\mathrm{X}})} Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{W}(T)},z_{\mathrm{o}})(B)}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

The *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, is sometimes not computable because the *abstract-dependent*, $A_{\mathrm{o}}^{\mathrm{W}(\tilde{T}_{\mathrm{o}})}$, is sometimes not computable. A finite approximation to arbitrary accuracy for the *abstract-dependent*, $A_k^{\mathrm{W}(T)} \approx A^{\mathrm{W}(T)}$, is computable. However, even an approximation is not *tractable*. The *abstract* function, $Y_{U,\mathrm{i},T,\mathrm{W},z} \in \mathcal{A}_{U,\mathrm{i},V,z} :\to \mathcal{A}_{U,W,z}$, is *intractable* because its computation requires the *intractable* computation of its domain of the *substrate histograms*, $\mathcal{A}_{U,\mathrm{i},V,z}$.

In *abstract induction*, where the *history probability function* is *iso-abstract historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}$, given some *unknown substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, in some cases the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is non-trivial,

$$\tilde{T}_{\mathrm{o}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$$

Consider how an approximation to the optimisation may be made more *tractable*. It is conjectured in section 'Likely histograms', above, that the *log-likelihood* with respect to the *dependent-analogue* varies with the *relative space* with respect to the *independent-analogue*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{W}(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{W}(T)}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})} \sim \mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A)$$

and conjectured further in section 'Transform alignment', above, that the *relative space* with respect to the *naturalised abstract* varies with the *derived alignment*,

$$\mathrm{spaceRelative}((A * T)^{\mathrm{X}} * T^{\dagger})(A) \quad \sim \quad \mathrm{algn}(A * T)$$

This correlation was decomposed in the discussion above into three separate correlations, (i) the *dependent-independent anti-correlation*, (ii) the *underlying-lifted correlation* and (iii) the *conditional-relative correlation*. Now consider how the optimisation of the terms of these relations may form the definition of *induction* assumptions.

The *maximum likelihood estimate* for the *unknown model*, $\tilde{T}_{\mathrm{o}}$, with respect to the *dependent-analogue* is

$$\{\tilde{T}_{\mathrm{o}}\} = \mathrm{maxd}(\{(T, \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{W}(T)}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{W}(T)}, z_{\mathrm{o}})(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z_{\mathrm{o}}}((A_{\mathrm{o}} * T)^{\mathrm{X}})}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

First, given the *dependent-independent anti-correlation*, assume that the *maximum likelihood estimate* of the *iso-abstract conditional multinomial probability* with respect to the *dependent-analogue* or *abstract-dependent*, $A_{\mathrm{o}}^{\mathrm{W}(T)}$, is also the *minimum likelihood estimate* of the *iso-abstract conditional multinomial probability* with respect to the *independent-analogue* or *naturalised abstract*, $(A_{\mathrm{o}} * T)^{\mathrm{X}} * T^{\dagger}$,

$$\{\tilde{T}_{\mathrm{o}}\} = \mathrm{mind}(\{(T, \frac{Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}} * T^{\dagger}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}} * T^{\dagger}, z_{\mathrm{o}})(B) : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z_{\mathrm{o}}}((A_{\mathrm{o}} * T)^{\mathrm{X}})}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

This assumption is the *iso-abstract dependent-independent anti-optimisation assumption*. It relies on the monotonicity of the *dependent-independent anti-correlation*.

Second, given the *underlying-lifted correlation*, assume that the *minimum likelihood estimate* of the *iso-abstract conditional multinomial probability* with respect to the *independent-analogue* or *naturalised abstract*, $(A_o * T)^X * T^\dagger$, is also the *minimum likelihood estimate* of the *lifted iso-abstract conditional multinomial probability* with respect to the *lifted independent-analogue* or *abstract*, $(A_o * T)^X$,

$$\{\tilde{T}_o\} = \\ \text{mind}(\{(T, \frac{Q_{m,U}((A_o * T)^X, z_o)(A_o * T)}{\sum Q_{m,U}((A_o * T)^X, z_o)(B') : B' \in \text{isowl}(U)(T, A_o)}) : T \in \mathcal{T}_{U,V_o}\})$$

where the *lifted integral iso-abstracts* is abbreviated

$$\text{isowl}(U)(T, A) := \{B * T : B \in Y^{-1}_{U,i,T,W,z}((A * T)^X)\}$$

This assumption is the *iso-abstract underlying-lifted optimisation assumption.* It relies on the monotonicity of the *underlying-lifted correlation.*

Third, given the *conditional-relative correlation*, assume that the *minimum likelihood estimate* of the *lifted iso-abstract conditional multinomial probability* with respect to the *lifted independent-analogue* or *abstract*, $(A_o * T)^X$, is also the *minimum likelihood estimate* of the *relative multinomial probability* with respect to the *lifted independent-analogue* or *abstract*, $(A_o * T)^X$,

$$\{\tilde{T}_o\} = \text{mind}(\{(T, \frac{Q_{m,U}((A_o * T)^X, z_o)(A_o * T)}{Q_{m,U}((A_o * T)^X, z_o)((A_o * T)^X)}) : T \in \mathcal{T}_{U,V_o}\})$$

The negative logarithm of the *relative multinomial probability* is the *relative space* of the *derived* with respect to the *abstract*, which is the *derived alignment*,

$$\begin{aligned} -\ln \frac{Q_{m,U}((A * T)^X, z)(A * T)}{Q_{m,U}((A * T)^X, z)((A * T)^X)} &= \text{spaceRelative}((A * T)^X)(A * T) \\ &= \text{algn}(A * T) \end{aligned}$$

So the third assumption is that the *minimum likelihood estimate* of the *lifted iso-abstract conditional multinomial probability* with respect to the *abstract*, $(A_o * T)^X$, is also the *maximum likelihood estimate* with respect to the *derived alignment*,

$$\{\tilde{T}_o\} = \text{maxd}(\{(T, \text{algn}(A_o * T)) : T \in \mathcal{T}_{U,V_o}\})$$

This assumption is the *iso-abstract conditional-relative optimisation assumption.* It relies on the monotonicity of the *conditional-relative correlation.*

A finite approximation to arbitrary accuracy of the *derived alignment*, $\mathrm{algn}(A_{\mathrm{o}} * T)$, is computable by means of an approximation to the gamma function. The computation of the *derived alignment* is *tractable* given limits on the *derived volume*, $|T^{-1}|$. So the optimisation of *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, at least for a limited subset of the *substrate transforms*, is *tractable.*

In *abstract induction*, where (i) the *history probability function* is *iso-abstract historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}$, given some *unknown substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *iso-abstract dependent-independent anti-optimisation assumption* is true, (iii) the *iso-abstract underlying-lifted optimisation assumption* is true, and (iv) the *iso-abstract conditional-relative optimisation assumption* is true, then the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is

$$\{\tilde{T}_{\mathrm{o}}\} \quad = \quad \mathrm{maxd}(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

It is shown in the *known transform* case above that the *maximum likelihood estimate* is more *classical* and less *formal* if (i) the *sample* is *naturalised*, (ii) the *sample* has *independent formal* and (iii) the *derived alignment* is high. This is the case for *unknown transform* too. In fact, if the three *iso-abstract optimisation assumptions* are true, then the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, occurs at the maximisation of the *derived alignment*, implying that the *derived alignment* is as high as possible, $\forall T \in \mathcal{T}_{U,V_{\mathrm{o}}}\ ((A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger}) \wedge (A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}}^{\mathrm{X}} * T)^{\mathrm{X}}) \implies \mathrm{algn}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}}) \geq \mathrm{algn}(A_{\mathrm{o}} * T))$.

In *abstract induction*, where (i) the *history probability function* is *iso-abstract historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}$, given some *unknown substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *iso-abstract dependent-independent anti-optimisation assumption* is true, (iii) the *iso-abstract underlying-lifted optimisation assumption* is true, (iv) the *iso-abstract conditional-relative optimisation assumption* is true, (v) the *sample* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, (vi) the *sample formal* is *independent*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$, then (a) the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is

$$\begin{aligned}\{\tilde{T}_{\mathrm{o}}\} \quad = \quad & \mathrm{maxd}(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \\ & \qquad A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger},\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}}^{\mathrm{X}} * T)^{\mathrm{X}}\})\end{aligned}$$

(b) the *log likelihood* of the *iso-abstract conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *derived alignment*,

$$\ln \hat{Q}_{\mathrm{h,w},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

(c) the *log likelihood* of the *iso-abstract conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *naturalisation*,

$$\ln \hat{Q}_{\mathrm{h,w},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}})$$

(d) the *formal alignment* of the *maximum likelihood estimate* is small,

$$\mathrm{algn}(Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}}^{\mathrm{X}} * \tilde{T}_{\mathrm{o}}) \approx 0$$

and (e) the *derived* of the *maximum likelihood estimate* approximates to the *normalised sample derived*,

$$\tilde{E}_{\mathrm{o}} * \tilde{T}_{\mathrm{o}} \approx \hat{A}_{\mathrm{o}} * T_{\mathrm{o}}$$

In this case queries via the *model* of the *sample* approximate to queries via the *model* of the *distribution*,

$$(Q * \tilde{T}_{\mathrm{o}}^{\%} * \mathrm{his}(\tilde{T}_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K) \approx (Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * E_{\mathrm{h}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K)$$

If, in addition, (vii) the *component size cardinality relative entropy* of the *maximum likelihood estimate* for the *model* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, then the *sum sensitivity* varies against the *log-likelihood*,

$$\begin{aligned} \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,w},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}}))) &\sim -\ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \\ &\sim -\mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}}) \end{aligned}$$

If, further, (viii) the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, then the *sensitivity* to *model* also varies against the *log likelihood*,

$$\begin{aligned} -\ \ln |\max(\{(T, \hat{Q}_{\mathrm{h,w},T,U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}}^{\mathrm{X}} * T)^{\mathrm{X}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})| \sim \\ -\ \ln \hat{Q}_{\mathrm{h,w},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \end{aligned}$$

or

$$- \ln |\max(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},$$
$$A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}}^{\mathrm{X}} * T)^{\mathrm{X}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})| \sim$$
$$- \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

So (a) by weakening the *induction* condition from *law-like necessary derived* to *entity-like necessary abstract* and (b) by strengthening the constraints on the *sample* to be *natural* and have *independent formal*, it is found that in some cases the *abstract induction maximum likelihood estimate* of the *model* is non-trivial, $\tilde{T}_{\mathrm{o}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$, but retains properties of *classical induction* such as allowing query via the *model*, minimising *sensitivity* to the *unknown underlying* and minimising *sensitivity* to the *model*. Furthermore, the optimisation is *tractable* depending on the limits on the searched subset of the *substrate transforms*.

5.9 Aligned modelled induction

The case of *classical modelled induction*, where the *derived* is *necessary*, may be termed *law-like* because the set of *iso-deriveds* is *law-like*. All *drawn histories* $H \in \mathcal{H}_{U,X}$, where $P_{U,X,H_{\mathrm{h}},\mathrm{d},T_{\mathrm{o}}}(H) > 0$, are such that their *normalised derived histograms* are fixed, $\hat{A}_H * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}$. That is, in *law-like induction* the relationship between the *derived variables* is unchanging,

$$\forall R_1, R_2 \in (A_H * T_{\mathrm{o}})^{\mathrm{FS}}\ \left(\frac{(A_H * T_{\mathrm{o}})_{R_2}}{(A_H * T_{\mathrm{o}})_{R_1}} = \frac{(E_{\mathrm{h}} * T_{\mathrm{o}})_{R_2}}{(E_{\mathrm{h}} * T_{\mathrm{o}})_{R_1}}\right)$$

Idealisation induction is also *law-like* because the *derived* is still *necessary*, $\hat{A}_H * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}$ where $P_{U,X,H_{\mathrm{h}},\dagger,T_{\mathrm{o}}}(H) > 0$. In fact, *idealisation induction* is stricter because it also imposes the constraint that the *independent components* be *necessary*, $\forall C \in T_{\mathrm{o}}^{\mathrm{P}}\ ((A_H * C^{\mathrm{U}})^{\mathrm{X}\wedge} = (E_{\mathrm{o}} * C^{\mathrm{U}})^{\mathrm{X}\wedge})$.

However, as is shown above, in the case where the *model*, T_{o}, is *unknown*, neither of the *law-like induction* types, *necessary derived* and *necessary idealisation*, have a singular solution for the *maximum likelihood estimate* of the *distribution-model* pair, $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$. It is necessary to relax the condition to *possible derived* and *possible idealisation* to obtain a singular solution.

Also discussed above is the case of *abstract induction*. This case, where the *abstract* is *necessary*, may be termed *entity-like* because the set of *iso-abstracts* is *entity-like*. All *drawn histories* $H \in \mathcal{H}_{U,X}$, where $P_{U,X,H_{\mathrm{h}},\mathrm{w},T_{\mathrm{o}}}(H) >$

0, are such that their *normalised abstract histograms* are fixed, $(\hat{A}_H * T_\mathrm{o})^\mathrm{X} = (\hat{E}_\mathrm{h} * T_\mathrm{o})^\mathrm{X}$. If the *model* is a *substrate transform*, $T_\mathrm{o} \in \mathcal{T}_{U,V_\mathrm{o}}$, then *necessary abstract* is equivalent to *necessary derived variables*, $(\hat{A}_H * T_\mathrm{o})^\mathrm{X} = (\hat{E}_\mathrm{h} * T_\mathrm{o})^\mathrm{X} \iff \forall P \in W_\mathrm{o}\ (\hat{A}_H * P^\mathrm{T} = \hat{E}_\mathrm{h} * P^\mathrm{T})$. In *entity-like induction* the relations between the *derived variables* are not *necessary*, but only relations between the *values* within each *derived variable* separately are *necessary*,

$$\begin{aligned}&\forall P \in W_\mathrm{o}\ \forall u_1, u_2 \in P\\ &\quad\left(\frac{(A_H * T_\mathrm{o})\%\{P\}(\{(P,u_2)\})}{(A_H * T_\mathrm{o})\%\{P\}(\{(P,u_1)\})} = \frac{(E_\mathrm{h} * T_\mathrm{o})\%\{P\}(\{(P,u_2)\})}{(E_\mathrm{h} * T_\mathrm{o})\%\{P\}(\{(P,u_1)\})}\right)\end{aligned}$$

That is, the *derived variables* are separately *necessary*. In *entity-like induction* it is sometimes the case that the *sample derived* is not equal to the *distribution derived*, $(\hat{A}_H * T_\mathrm{o})^\mathrm{X} = (\hat{E}_\mathrm{h} * T_\mathrm{o})^\mathrm{X} \impliedby \hat{A}_H * T_\mathrm{o} = \hat{E}_\mathrm{h} * T_\mathrm{o}$.

In the case of *unknown model*, T_o, *abstract induction* has a singular solution for the *maximum likelihood estimate* of the *distribution-model* pair, $(\tilde{E}_\mathrm{o}, \tilde{T}_\mathrm{o})$. *Abstract induction* is not *law-like* but only *entity-like*, so it is sometimes the case that the *estimated derived* does not equal the *estimated distribution derived*, $\hat{A}_H * \tilde{T}_\mathrm{o} \neq \hat{E}_\mathrm{h} * \tilde{T}_\mathrm{o}$.

However, in the case where (i) the *sample* is *natural*, $A_\mathrm{o} = A_\mathrm{o} * T_\mathrm{o} * T_\mathrm{o}^\dagger$, and (ii) the *sample formal* is *independent*, $A_\mathrm{o}^\mathrm{X} * T_\mathrm{o} = (A_\mathrm{o}^\mathrm{X} * T_\mathrm{o})^\mathrm{X}$, then the *estimated derived alignment* is maximised, $\forall T \in \mathcal{T}_{U,V_\mathrm{o}}\ ((A_\mathrm{o} = A_\mathrm{o} * T * T^\dagger) \wedge (A_\mathrm{o}^\mathrm{X} * T = (A_\mathrm{o}^\mathrm{X} * T)^\mathrm{X}) \implies \mathrm{algn}(A_\mathrm{o} * \tilde{T}_\mathrm{o}) \geq \mathrm{algn}(A_\mathrm{o} * T))$, and, as *derived alignment*, $\mathrm{algn}(A_\mathrm{o} * \tilde{T}_\mathrm{o})$, increases, *abstract induction* tends to *classical induction*, where the *normalised sample derived* approximates to the *derived* of the *maximum likelihood estimate*, $\hat{A}_\mathrm{o} * \tilde{T}_\mathrm{o} \approx \tilde{E}_\mathrm{o} * \tilde{T}_\mathrm{o}$. Moreover, the computation of the *maximum likelihood estimate* of the *model*, $\tilde{T}_\mathrm{o}$, may be made *tractable* if limits are imposed on the optimisation.

Although *abstract induction* can provide a non-trivial solution for the *maximum likelihood estimate* for the *model*, $\tilde{T}_\mathrm{o}$, and constrained *abstract induction* can do so such that the *maximum likelihood estimate* for the *distribution histogram*, $\tilde{E}_\mathrm{o}$, is approximately *classical*, $\tilde{E}_\mathrm{o} \approx \hat{A}_\mathrm{o}^{\mathrm{D}(\tilde{T}_\mathrm{o})}$, *abstract induction* is neutral with respect to *formal alignment*. That is, the set of *iso-abstracts* is conditional on neither the *formal*, $A^\mathrm{X} * T$, nor the *formal independent*, $(A^\mathrm{X} * T)^\mathrm{X}$, so the *abstract dependent*, $A^{\mathrm{W}(T)}$, is neutral with respect to the *formal* and the *formal independent*, and nothing can be said of its *formal alignment*, $\mathrm{algn}(A^{\mathrm{W}(T)\mathrm{X}} * T)$. Indeed in some cases the *abstract dependent* may be purely *formal*, $A^{\mathrm{W}(T)} * T = A^{\mathrm{W}(T)\mathrm{X}} * T \implies \mathrm{algn}(A^{\mathrm{W}(T)} * T) = \mathrm{algn}(A^{\mathrm{W}(T)\mathrm{X}} * T)$.

So, in some cases, the *model estimate*, $\tilde{T}_{\rm o}$, can be *tautological* or otherwise *overlapping*, $\mathrm{overlap}(\tilde{T}_{\rm o}) \implies A_{\rm o}^{\rm X} * \tilde{T}_{\rm o} \neq (A_{\rm o}^{\rm X} * \tilde{T}_{\rm o})^{\rm X} \implies \mathrm{algn}(A_{\rm o}^{\rm X} * \tilde{T}_{\rm o}) > 0$. Even in the case of constrained *abstract induction* where the *sample formal* is *known* to be *independent*, $A_{\rm o}^{\rm X} * T_{\rm o} = (A_{\rm o}^{\rm X} * T_{\rm o})^{\rm X}$, the other members of the *iso-abstracts* may have *formal alignment*, $\forall B \in Y^{-1}_{U,T_{\rm o},{\rm W},z_{\rm o}}((A_{\rm o} * T_{\rm o})^{\rm X})\ (\mathrm{algn}(B^{\rm X} * T_{\rm o}) \geq 0)$, and the *abstract-dependent* may have *formal alignment*, $\mathrm{algn}(A_{\rm o}^{{\rm W}(T_{\rm o}){\rm X}} * T_{\rm o}) \geq 0$. To address this, consider strengthening *abstract induction* to *partition induction* where the condition of *necessary formal independent*, $(A_{\rm o}^{\rm X} * T_{\rm o})^{\rm X}$, is added to the condition of *necessary abstract*, $(A_{\rm o} * T_{\rm o})^{\rm X}$.

The *partition-independent*, $A^{{\rm P}(T)} \in \mathcal{A}_{U,V,z}$, is defined in section 'Likely histograms', above, as

$$\{A^{{\rm P}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{{\rm m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A))) : D \in \mathcal{A}_{U,V,z}\})$$

where the *integral iso-partition-independents* is abbreviated

$$\mathrm{isop}(U)(T,A) \ := \ Y^{-1}_{U,{\rm i},T,{\rm V},{\rm x},z}((A^{\rm X} * T)^{\rm X}) \ \cap \ Y^{-1}_{U,{\rm i},T,{\rm W},z}((A * T)^{\rm X})$$

and the *iso-partition-independents* is such that

$$\begin{aligned}&Y^{-1}_{U,T,{\rm V},{\rm x},z}((A^{\rm X} * T)^{\rm X}) \ \cap \ Y^{-1}_{U,T,{\rm W},z}((A * T)^{\rm X}) \\ &\quad = \ \{B : B \in \mathcal{A}_{U,V,z},\ (B^{\rm X} * T)^{\rm X} = (A^{\rm X} * T)^{\rm X},\ (B * T)^{\rm X} = (A * T)^{\rm X}\}\end{aligned}$$

The corresponding *dependent analogue* is the *partition-dependent*, $A^{{\rm R}(T)} \in \mathcal{A}_{U,V,z}$, defined

$$\{A^{{\rm R}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{{\rm m},U}(D,z)(A)}{\sum Q_{{\rm m},U}(D,z)(B) : B \in \mathrm{isop}(U)(T,A)}) : D \in \mathcal{A}_{U,V,z}\})$$

In *partition induction* the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories*, $P(H) > 0$, have (i) an *formal independent probability histogram* equal to the *formal independent distribution probability histogram*, $(\hat{A}_H^{\rm X} * T_{\rm o})^{\rm X} = (\hat{E}_{\rm h}^{\rm X} * T_{\rm o})^{\rm X}$, and (ii) an *abstract probability histogram* equal to the *abstract distribution probability histogram*, $(\hat{A}_H * T_{\rm o})^{\rm X} = (\hat{E}_{\rm h} * T_{\rm o})^{\rm X}$. In the case of *known transform*, $T_{\rm o}$, the *maximum likelihood estimate* for the *distribution histogram* is the *partition-dependent*, $\tilde{E}_{\rm o} = \hat{A}_{\rm o}^{{\rm R}(T_{\rm o})}$. *Partition induction* is *entity-like* because the *iso-partition-independents* is a subset of the *iso-abstracts*, so the *derived variables* are separately *necessary*,

$$\begin{aligned}&\forall P \in W_{\rm o}\ \forall u_1, u_2 \in P \\ &\quad \left(\frac{(A_H * T_{\rm o})\%\{P\}(\{(P,u_2)\})}{(A_H * T_{\rm o})\%\{P\}(\{(P,u_1)\})} = \frac{(E_{\rm h} * T_{\rm o})\%\{P\}(\{(P,u_2)\})}{(E_{\rm h} * T_{\rm o})\%\{P\}(\{(P,u_1)\})}\right)\end{aligned}$$

although the *iso-abstractence* is lower,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \ \leq \ 1$$

and so *partition induction* may be less *entity-like* than *abstract induction*.

Now, in addition, *formal variables* are separately *necessary*,

$$\forall P \in W_{\mathrm{o}} \ \forall u_1, u_2 \in P \\ \left(\frac{(A_H^{\mathrm{X}} * T_{\mathrm{o}})\%\{P\}(\{(P,u_2)\})}{(A_H^{\mathrm{X}} * T_{\mathrm{o}})\%\{P\}(\{(P,u_1)\})} = \frac{(E_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}})\%\{P\}(\{(P,u_2)\})}{(E_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}})\%\{P\}(\{(P,u_1)\})}\right)$$

That is, the stricter condition of *partition induction* requires that the *partition derived variables* are *necessary* with respect to both the *histogram* and its *independent*. If the *model* is a *substrate transform*, $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, then (i) *necessary abstract* is equivalent to *necessary derived variables*, $(\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}} \iff \forall P \in W_{\mathrm{o}} \ (\hat{A}_H * P^{\mathrm{T}} = \hat{E}_{\mathrm{h}} * P^{\mathrm{T}})$, and (ii) *necessary formal independent* is equivalent to *necessary formal variables*, $(\hat{A}_H^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}} \iff \forall P \in W_{\mathrm{o}} \ (\hat{A}_H^{\mathrm{X}} * P^{\mathrm{T}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}} * P^{\mathrm{T}})$.

Also, *partition induction* may be more *law-like* than *abstract induction* if the *iso-derivedence* is greater,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cap \ D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|(Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \ \cup \ D^{-1}_{U,\mathrm{i},T,z}(A * T)) \ \cap \ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} \ > \ \frac{|D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}$$

depending on the relative intersection cardinalities.

It is conjectured in 'Transform alignment', above, that the *formal alignment* of the *independent analogue* of the *iso-partition-independents*, $\mathrm{algn}(A^{\mathrm{P}(T)\mathrm{X}} * T)$, is less than or equal to the *formal alignment*, $\mathrm{algn}(A^{\mathrm{X}} * T)$, which in turn is less than or equal to the *dependent analogue formal alignment*, $\mathrm{algn}(A^{\mathrm{R}(T)\mathrm{X}} * T)$,

$$\mathrm{algn}(A^{\mathrm{P}(T)\mathrm{X}} * T) \leq \mathrm{algn}(A^{\mathrm{X}} * T) \leq \mathrm{algn}(A^{\mathrm{R}(T)\mathrm{X}} * T)$$

That is, the *formal alignment* of the *maximum likelihood estimate* of *partition induction*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{R}(T_{\mathrm{o}})}$, is greater than or equal to the *sample formal* *alignment*, $\mathrm{algn}(Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) \geq \mathrm{algn}(A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})$. So in some cases even if the *sample* is not purely *formal*, $A_{\mathrm{o}} * T_{\mathrm{o}} \neq A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}$, the *distribution histogram estimate* may be purely *formal*, $A_{\mathrm{o}}^{\mathrm{R}(T_{\mathrm{o}})} * T_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{R}(T_{\mathrm{o}})\mathrm{X}} * T_{\mathrm{o}}$.

If there is no *knowledge* regarding *formal alignment*, *partition induction* may be preferable to *abstract induction* because of its possibly higher degree of *law-likeness* or *iso-derivedence*. In this case the *partition induction maximum likelihood estimate* approximates more closely to *classical*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{R}(T_{\mathrm{o}})} \approx \hat{A}_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})}$, than *abstract induction*.

If, however, the *distribution histogram* is *known* to have small *formal alignment*, $\mathrm{algn}(Z_{\mathrm{o}} * \hat{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) \approx 0$, then, even if the *sample formal* is *independent*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$, the *distribution formal estimate* is unconstrained,

$$\mathrm{algn}(Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) \geq \mathrm{algn}(A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) = 0$$

Of course, this is also the case if the *distribution formal* is *known* to be *independent*, $E_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (E_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}} \implies \mathrm{algn}(Z_{\mathrm{o}} * \hat{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) = 0$.

Clearly, in the case of small or zero *distribution formal alignment*, the *partition induction* condition is insufficient. Now consider further strengthening the condition of *necessary formal independent*, $(\hat{A}_{H}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$, to *necessary formal*, where the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories*, $P(H) > 0$, have a *formal probability histogram* equal to the *formal distribution probability histogram*, $\hat{A}_{H}^{\mathrm{X}} * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}$.

In *aligned modelled induction*, also called *transform induction*, the condition is *necessary transform independent*, $\hat{A}_{H}^{\mathrm{X}(T_{\mathrm{o}})} = \hat{E}_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})}$, or *necessary formal* and *necessary abstract*, $(\hat{A}_{H}^{\mathrm{X}} * T_{\mathrm{o}}, (\hat{A}_{H} * T_{\mathrm{o}})^{\mathrm{X}}) = (\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}, (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}})$. The corresponding *iso-set* is the *iso-transform-independents*, which is the intersection of the *iso-abstracts* and the *iso-formals*, which in turn is a subset of the *iso-partition-independents* which in turn is a subset of the *iso-abstracts*,

$$\begin{aligned} Y_{U,T,z}^{-1}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) &= Y_{U,T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}}) \cap Y_{U,T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \\ &\subseteq Y_{U,T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}}) \cap Y_{U,T,\mathrm{V},\mathrm{x},z}^{-1}((A^{\mathrm{X}} * T)^{\mathrm{X}}) \\ &\subseteq Y_{U,T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}}) \end{aligned}$$

The *transform-independent*, $A^{\mathrm{X}(T)} \in \mathcal{A}_{U,V,z}$, is defined in section 'Likely histograms', above, as

$$\{A^{\mathrm{X}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A))) : D \in \mathcal{A}_{U,V,z}\})$$

where the *integral iso-transform-independents* is abbreviated

$$\begin{aligned}\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) &= Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \\ &= \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}\end{aligned}$$

The corresponding *dependent analogue* is the *transform-dependent*, $A^{\mathrm{Y}(T)} \in \mathcal{A}_{U,V,z}$, defined

$$\{A^{\mathrm{Y}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}) : D \in \mathcal{A}_{U,V,z}\})$$

In 'Transform alignment', above, it is conjectured that the *partition-dependent formal alignment* is greater than or equal to the *formal alignment*, which in turn is greater than or equal to the *transform-dependent formal alignment*

$$\mathrm{algn}(A^{\mathrm{R}(T)\mathrm{X}} * T) \ \geq\ \mathrm{algn}(A^{\mathrm{X}} * T) \ \geq\ \mathrm{algn}(A^{\mathrm{Y}(T)\mathrm{X}} * T)$$

So if the *sample formal alignment* is small, the *transform-dependent formal alignment* is also small,

$$\mathrm{algn}(A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) \ \approx\ 0 \implies \mathrm{algn}(A_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})\mathrm{X}} * T_{\mathrm{o}}) \ \approx\ 0$$

whereas the *partition-dependent formal alignment* remains unconstrained,

$$\mathrm{algn}(A_{\mathrm{o}}^{\mathrm{R}(T_{\mathrm{o}})\mathrm{X}} * T_{\mathrm{o}}) \ \geq\ \mathrm{algn}(A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) \ \approx\ 0$$

Therefore if it is *known* that the *distribution histogram formal alignment* is small, $\mathrm{algn}(Z_{\mathrm{o}} * \hat{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}) \approx 0$, then, although *partition induction* is stricter than *abstract induction*, *partition induction* is still insufficiently constrained compared to *transform induction*. Furthermore, if the *formal* is *necessarily independent*, $\forall H \in \mathcal{H}_{U,X}\ (P(H) > 0 \implies A_H^{\mathrm{X}} * T_{\mathrm{o}} = (A_H^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}})$, then this additional condition may be obtained in *transform induction* by constraining the *sample* to have *independent formal*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$, because the *iso-transform-independents necessarily* have the same *formal*,

$$\begin{aligned}&A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}} \implies \\ &\quad \forall B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \\ &\qquad\qquad (B^{\mathrm{X}} * T = A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}} = (B^{\mathrm{X}} * T)^{\mathrm{X}})\end{aligned}$$

Just as for *partition induction*, in *transform induction* the *derived variables* and *formal variables* are separately *necessary*. If the *model* is a *substrate transform*, $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, then

$$\forall P \in W_{\mathrm{o}}\ ((\hat{A}_H * P^{\mathrm{T}} = \hat{E}_{\mathrm{h}} * P^{\mathrm{T}}) \ \wedge\ (\hat{A}_H^{\mathrm{X}} * P^{\mathrm{T}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}} * P^{\mathrm{T}}))$$

If the *iso-derivedence* of the *integral iso-transform-independents* is greater than the *iso-derivedence* of the *integral iso-partition-independents*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)\ \cap\ D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|(Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)\ \cup\ D^{-1}_{U,\mathrm{i},T,z}(A * T))\ \cap\ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|} >$$
$$\frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cap\ D^{-1}_{U,\mathrm{i},T,z}(A * T)|}{|(Y^{-1}_{U,\mathrm{i},T,\mathrm{V},\mathrm{x},z}((A^{\mathrm{X}} * T)^{\mathrm{X}})\ \cup\ D^{-1}_{U,\mathrm{i},T,z}(A * T))\ \cap\ Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})|}$$

then *transform induction* is more *law-like* than *partition induction*, and so is more *classical*.

Like *abstract induction* and *partition induction*, in the case of *unknown model*, T_{o}, *transform induction* has a unique solution for the *maximum likelihood estimate* for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$.

If, in addition, the *sample* is considered to be special by assuming it is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, the *dependent analogue*, $A_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})}$, is closer to the *derived dependent*, $A_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})}$, and therefore more *law-like*. This is analogous to the special case in *abstract induction* where the *sample* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$.

Furthermore, if the *sample* is assumed to be such that the *formal* equals the *abstract*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, then the *model estimate* optimisation can be *lifted* into the *derived variables*, and so made *tractable* and thence *practicable*. This is analogous to the special case in *abstract induction* where the *sample* has *independent formal*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$.

First, however, consider the case where the given *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, is *known*.

In *aligned modelled induction*, while the *distribution probability histogram*, $\hat{E}_{\mathrm{h}}$, remains *unknown*, the *formal distribution probability histogram*, $\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}$, and the *abstract distribution probability histogram*, $(\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$, are *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have (i) a *formal probability histogram* equal to the *formal distribution probability histogram*, $\hat{A}_H^{\mathrm{X}} * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}$, and (ii) an *abstract probability histogram* equal to the *abstract distribution probability histogram*, $(\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$. Define the *iso-transform-independent historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}} \in$

$(\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$,

$$\begin{aligned}
&P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}} := \\
&\quad \Big(\bigcup \{\{(H,1) : H \subseteq H_{\mathrm{h}}\%V_H,\ |H| = z_H, \\
&\qquad\qquad \hat{A}_H^{\mathrm{X}} * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}},\ (\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}\}^{\wedge} : \\
&\qquad\qquad\qquad V_H \subseteq V_{\mathrm{h}},\ z_H \in \{1 \ldots z_{\mathrm{h}}\}\}\Big)^{\wedge} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ \hat{A}_H^{\mathrm{X}} * T_{\mathrm{o}} \neq \hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}\ \vee\ (\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} \neq (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}\} \cup \\
&\quad \{(H,0) : H \in \mathcal{H}_{U,X},\ H \nsubseteq H_{\mathrm{h}}\%V_H\} \cup \{(\emptyset,0)\}
\end{aligned}$$

For *drawn histories* the *formal probability histogram* and *abstract probability histogram* are *necessary*, $\forall H \in \mathcal{H}_{U,X}\ (P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}(H) > 0 \implies \hat{A}_H^{\mathrm{X}} * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}\ \wedge\ (\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}})$. Not all *sizes* and sets of *variables* are necessarily *drawable*. That is, in some cases, $\exists z \in \{1 \ldots z_{\mathrm{h}}\}\ \exists V \subseteq V_{\mathrm{h}}\ \forall H \in \mathcal{H}_{U,X}\ ((z_H = z) \wedge (V_H = V) \implies P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}(H) = 0)$. The *distribution history* can always be *drawn*, so the *probability function* is not a *weak probability function*, $\sum_{H \in \mathcal{H}_{U,X}} P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}(H) = 1$.

In *aligned modelled induction* the *history probability function* is *iso-transform-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}$.

Given a *drawn history* $H \in \mathcal{H}_{U,X}$, where $P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}(H) > 0$, the *iso-transform-independent probability* of *histogram* $A_H = \mathrm{histogram}(H) + V_H^{\mathrm{CZ}} \in \mathcal{A}_{U,\mathrm{i},V_H,z_H}$ is now conditional,

$$\frac{Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{o}},z_H}(A)} Q_{\mathrm{h},U}(E_{\mathrm{h}}\%V_H, z_H)(B)} = \frac{\sum P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H}{\sum P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ V_G = V_H,\ |G| = z_H}$$

where the *integral iso-transform-independents* is abbreviated

$$\mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) \quad = \quad Y_{U,\mathrm{i},T,z}^{-1}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$$

and the set of *integral iso-transform-independents* is the intersection of the *iso-formals* and *iso-abstracts*

$$\begin{aligned}
&Y_{U,\mathrm{i},T,z}^{-1}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \\
&\quad = \ Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T)\ \cap\ Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}}) \\
&\quad = \ \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T,\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}
\end{aligned}$$

The *iso-transform-independent historical probability* may be expressed in terms of a *histogram distribution* which is not explicitly conditional on the *necessary formal* and *necessary abstract*, $(\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}, (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}})$,

$$\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_H, z_H)(A_H) \quad \propto \quad \sum(P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

where the *iso-derived conditional stuffed historical probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\mathrm{h,y},T,U}(E, z) \\
&\quad := \quad \{(A, \frac{Q_{\mathrm{h},U}(E, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A)} Q_{\mathrm{h},U}(E, z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\
&\qquad\quad \{(A, 0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \not\leq E\}
\end{aligned}$$

which is defined if $z \leq \mathrm{size}(E)$. In the case where all the *formal-abstract* pairs are *possible*,

$$\forall A' \in \mathrm{ran}(Y_{U,\mathrm{i},T,z})\ \exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ (((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}) = A')\ \wedge\ (A \leq E))$$

the normalisation of the *iso-transform-independent conditional stuffed historical probability distribution* is a fraction $1/|\mathrm{ran}(Y_{U,\mathrm{i},T,z})|$,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h,y},T,U}(E, z) \\
&\quad = \quad \{(A, \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},T,z})|} \frac{Q_{\mathrm{h},U}(E, z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A)} Q_{\mathrm{h},U}(E, z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}
\end{aligned}$$

In the case of a *full functional transform*, $T_{\mathrm{f}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}}$, the *iso-transform-independents* equals the *iso-independents*, $\mathcal{A}_{U,\mathrm{i,y},T_{\mathrm{f}},z}(A) = Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})$, and so the case is the same as for *aligned non-modelled induction*, (i) the *maximum likelihood estimate* varies with the *sample probability histogram*, $\tilde{E}_{\mathrm{o}} \sim \hat{A}_{\mathrm{o}}$, and against the *independent sample probability histogram*, $\tilde{E}_{\mathrm{o}} \nsim \hat{A}_{\mathrm{o}}^{\mathrm{X}}$, and (ii) the *sum sensitivity* varies with the *sample alignment*, $\mathrm{algn}(A_{\mathrm{o}})$, at low *alignments* and against the *sample alignment*, $-\ \mathrm{algn}(A_{\mathrm{o}})$, at high *alignments*.

At the other extreme of a *unary transform*, $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, the *iso-transform-independents* equals the *substrate histograms*, $\mathcal{A}_{U,\mathrm{i,y},T_{\mathrm{u}},z}(A) = \mathcal{A}_{U,\mathrm{i},V,z}$, and so the case is the same as for *classical non-modelled induction*, (i) the *maximum likelihood estimate* equals the *sample probability histogram*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, and (ii) the *sum sensitivity* varies with the negative *scaled sample entropy*, $-z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}})$.

The *iso-transform-independent conditional generalised multinomial probability distribution* is defined

$$\begin{aligned}
&\hat{Q}_{\mathrm{m,y},T,U}(E,z) \\
&:= \{(A, \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A)} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \leq E^{\mathrm{F}}\}^{\wedge} \cup \\
&\quad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A^{\mathrm{F}} \not\leq E^{\mathrm{F}}\}
\end{aligned}$$

which is defined if $\mathrm{size}(E) > 0$.

The case where all the *formal-abstract* pairs are *possible* is weaker than for *historical*,

$$\forall A' \in \mathrm{ran}(Y_{U,\mathrm{i},T,z})\ \exists A \in \mathcal{A}_{U,\mathrm{i},V,z}\ (((A^{\mathrm{X}} * T, (A * T)^{\mathrm{X}}) = A')\ \wedge\ (A^{\mathrm{F}} \leq E^{\mathrm{F}}))$$

In this case the *iso-transform-independent conditional generalised multinomial probability distribution* is

$$\begin{aligned}
&\hat{Q}_{\mathrm{m,y},T,U}(E,z) \\
&= \{(A, \frac{1}{|\mathrm{ran}(Y_{U,\mathrm{i},T,z})|} \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A)} Q_{\mathrm{m},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\}
\end{aligned}$$

Assume that the *distribution history size*, z_{h}, is large with respect to the *sample size* $z_{\mathrm{o}} = \mathrm{size}(A_{\mathrm{o}})$, so that, in the limit, the *iso-transform-independent conditional stuffed historical probability*, $\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$, approximates to the *iso-transform-independent conditional multinomial probability*, $\hat{Q}_{\mathrm{m,y},T_{\mathrm{o}},U}(E_{\mathrm{h}}\%V_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$. That is, if $z_{\mathrm{o}} \ll z_{\mathrm{h}}$ then

$$\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \approx \hat{Q}_{\mathrm{m,y},T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

The *iso-transform-independent multinomial parameterised probability density function*, $\mathrm{mytppdf}(T,z) \in \mathrm{ppdfs}(v,v)$, and *iso-transform-independent multinomial likelihood function*, $\mathrm{mytlf}(T,z) \in \mathrm{lfs}(v,v)$, corresponding to the *iso-transform-independent multinomial probability distribution*, $\hat{Q}_{\mathrm{m,y},T,U}$, are not given explicitly here, but are such that

$$\mathrm{mytppdf}(T,z)(\hat{E}^{\square})(A^{\square}) = \mathrm{mytlf}(T,z)(A^{\square})(\hat{E}^{\square}) = \hat{Q}_{\mathrm{m,y},T,U}(E,z)(A)$$

Now in the case of *aligned modelled induction* where the *transform*, T_{o}, is *known*, the *real maximum likelihood estimate* $\tilde{E}'_{\mathrm{o}} \in \mathbf{R}^{v_{\mathrm{o}}}_{(0,1)}$ for the parameter of the *iso-transform-independent multinomial parameterised probability density function* is

$$\{\tilde{E}'_{\mathrm{o}}\} = \mathrm{maxd}(\mathrm{mytlf}(T_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}}^{\square}))$$

which is such that $\forall i \in \{1 \ldots v_{\mathrm{o}}\}\ (\partial_i(\mathrm{mytlf}(T_{\mathrm{o}}, z_{\mathrm{o}})(A_{\mathrm{o}}^{\square}))(\tilde{E}'_{\mathrm{o}}) = 0)$. The *maximum likelihood estimate* $\tilde{E}'_{\mathrm{o}}$ is only defined in the case where the *sample histogram* is *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}} \implies \hat{A}_{\mathrm{o}}^{\square} \in \mathbf{R}_{(0,1)}^{v_{\mathrm{o}}}$, because the *binomial likelihood function* is only defined for the open set. That is, $\mathrm{d}(\mathrm{blf}(z_{\mathrm{o}})(0))$ is undefined and so the derivative of the *iso-transform-independent multinomial parameterised probability density function* is undefined where there are *ineffective states*.

In the case of *completely effective sample histogram*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the maximisation for *known transform*, T_{o}, of the *iso-transform-independent conditional generalised multinomial probability* parameterised by the *complete congruent histograms* of unit *size* is a singleton of the *rational maximum likelihood estimate*

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m,y},T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

The *real maximum likelihood estimate*, $\tilde{E}'_{\mathrm{o}}$, is not necessarily a rational coordinate, $\mathbf{R}_{(0,1)}^{v_{\mathrm{o}}} \supset \mathbf{Q}_{(0,1)}^{v_{\mathrm{o}}}$, and so the *rational maximum likelihood estimate* is not necessarily equal to the *real maximum likelihood estimate*. However, it is conjectured that the maximisation of the *distribution* approximates to the maximisation of the *likelihood function*,

$$\tilde{E}_{\mathrm{o}}^{\square} \approx \tilde{E}'_{\mathrm{o}}$$

In the case where the *sample histogram* is not *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} < V_{\mathrm{o}}^{\mathrm{C}}$, the maximisation of the *iso-transform-independent conditional generalised multinomial probability distribution* for *known transform* is well defined, unlike the *parameterised probability density function*, but is not necessarily a singleton

$$|\mathrm{max}(\{(E, \hat{Q}_{\mathrm{m,y},T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})| \geq 1$$

In the case where the maximisation of the *iso-transform-independent conditional generalised multinomial probability distribution* is a singleton, it is equal to the *normalised transform-dependent*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})}$, where the *transform-dependent* $A^{\mathrm{Y}(T)} \in \mathcal{A}_{U,V,z}$ is defined in 'Likely histograms', above, as the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of the *histogram*, A, conditional that it is an *iso-transform-independent*,

$$\{A^{\mathrm{Y}(T)}\} = \mathrm{maxd}(\{(D, \frac{Q_{\mathrm{m},U}(D, z)(A)}{\sum Q_{\mathrm{m},U}(D, z)(B) : B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A)}) : D \in \mathcal{A}_{U,V,z}\})$$

The *transform-dependent*, $A^{\mathrm{Y}(T)}$, is sometimes not computable. The finite approximation to the *transform-dependent* is

$$\{A_k^{\mathrm{Y}(T)}\} = \\ \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}) : D \in \mathcal{A}_{U,\mathrm{i},V,kz}\})$$

The approximation, $A_k^{\mathrm{Y}(T)} \approx A^{\mathrm{Y}(T)}$, improves as the scaling factor, k, increases.

Unlike in *classical non-modelled induction*, where the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, is equal to the *sample probability histogram*, $\hat{A}_{\mathrm{o}}$, in *aligned modelled induction* the *maximum likelihood estimate* is not necessarily equal to the *sample probability histogram*. It is only in the case where the *sample histogram* equals the *transform-independent* that the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})} \implies A_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

where the *transform-independent* $A^{\mathrm{X}(T)} \in \mathcal{A}_{U,V,z}$ is defined in 'Likely histograms', above, as the *maximum likelihood estimate* for the *distribution histogram* of the sum of the *generalised multinomial probabilities* of the *integral iso-transform-independents* of the *histogram*, A,

$$\{A^{\mathrm{X}(T)}\} = \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A))) : D \in \mathcal{A}_{U,V,z}\})$$

Otherwise, the overall *maximum likelihood estimate*, which is the *transform-dependent*, is near the *histogram*, $\tilde{E}_{\mathrm{o}} \sim \hat{A}_{\mathrm{o}}$, only in as much as it is far from the *transform-independent*, $\tilde{E}_{\mathrm{o}} \nsim \hat{A}_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})}$.

The requirement that the *distribution history* itself be *drawable*, $P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}(H_{\mathrm{h}}) > 0$, has been ignored so far. This requirement modifies the maximisation to add the constraint that the *maximum likelihood estimate* be an *iso-transform-independent*, $((\tilde{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}), (\tilde{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}) = ((\hat{A}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}), (\hat{A}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}})$,

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m},\mathrm{y},T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}, \\ E^{\mathrm{X}} * T_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}},\ (E * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{A}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}\})$$

So, strictly speaking, the *maximum likelihood estimate* is only approximately equal to the *normalised transform-dependent*, $\tilde{E}_{\mathrm{o}} \approx \hat{A}_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})}$, if the *transform-dependent* is not an *iso-transform-independent*, $A_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})} \notin \mathcal{A}_{U,\mathrm{y},T_{\mathrm{o}},z_{\mathrm{o}}}(A_{\mathrm{o}})$. In the special case, however, where the *sample histogram* equals the *transform-independent*, the *maximum likelihood estimate* is exactly equal to the *normalised transform-dependent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})} = \hat{A}_{\mathrm{o}}$.

In *aligned modelled induction*, also known as *transform induction*, where (i) the *history probability function* is *iso-transform-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample histogram* equals the *transform-independent*, $A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})}$, then the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, of the *unknown distribution probability histogram*, $\hat{E}_{\mathrm{o}}$, in the *iso-transform-independent conditional stuffed historical probability distribution*, $\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(E_{\mathrm{o}}, z_{\mathrm{o}})$, is

$$\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

The set of *iso-transform-independents* is a subset of the *iso-abstracts*, $Y_{U,T,z}^{-1}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \subseteq Y_{U,T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})$, so *transform induction* is *entity-like*. The *iso-abstractence* or degree of *entity-likeness* is

$$\frac{|Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \ \cap \ Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})|}{|Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})|} \ \leq \ 1$$

If the set of *iso-transform-independents* is a proper subset of the *iso-abstracts*, then *transform induction* is less *entity-like* than *abstract induction*.

The set of *iso-transform-independents* is not a subset of the *iso-deriveds*, $Y_{U,T,z}^{-1}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \not\subseteq D_{U,T,z}^{-1}(A * T)$, so *transform induction* is not *law-like*, unlike *classical modelled induction* or *idealisation induction*. However, *transform induction* may be more *law-like* than *abstract induction* if the *iso-derivedence* or degree of *law-likeness* is greater,

$$\frac{|Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \ \cap \ D_{U,\mathrm{i},T,z}^{-1}(A * T)|}{|(Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) \ \cup \ D_{U,\mathrm{i},T,z}^{-1}(A * T)) \ \cap \ Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})|} \ > \ \frac{|D_{U,\mathrm{i},T,z}^{-1}(A * T)|}{|Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}})|}$$

which depends on the relative intersection cardinalities. That is, *transform induction* is sometimes less *entity-like* and more *law-like* than *abstract induction*.

Constraints on the *sample* can make the denominator, $\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(B) :$

$B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$, more approximate to the *iso-derived* denominator, $\sum Q_{\mathrm{m},U}(A^{\mathrm{D}(T)}, z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T)$. In this way *transform induction* can sometimes approximate to *classical modelled induction*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})} \approx \hat{A}_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})}$.

The degree to which the *iso-transform-independents* is said to be *aligned-like*, or the *iso-independence*, is

$$\frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} =$$

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|(Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}}) \ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})) \ \cap\ Y^{-1}_{U,\mathrm{i},T,\mathrm{V},z}(A^{\mathrm{X}} * T)|}$$

In some cases the *iso-independence* of the *iso-idealisations* is greater than or equal to the *iso-independence* of the *iso-transform-independents*,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|}{|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} \ \geq\ \frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \ \cap\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) \ \cup\ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}$$

and so *transform induction* is sometimes less *aligned-like* than *idealisation induction*. However, the *derived iso-independence* of the *integral lifted iso-transform-independents* is necessarily greater than or equal to the *derived iso-independence* of any *law-like iso-set*,

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|} \ \geq\ \frac{1}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

and so *transform induction* may be said to be more *derived aligned-like* than either *classical modelled induction* or *idealisation induction*. However, the *derived iso-independence* of the *integral lifted iso-transform-independents* is less than or equal to the *derived iso-independence* of the *integral lifted iso-abstracts*,

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|} \ \leq$$

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,\mathrm{W},z}((A * T)^{\mathrm{X}})\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

So *transform induction* is less *derived aligned-like* than *abstract induction*.

As the *iso-independence* increases, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, which equals the *transform-dependent*, $\hat{A}_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})}$, tends to the *dependent*, $\hat{A}_{\mathrm{o}}^{\mathrm{Y}}$, which is independent of the *model*, T_{o}, because the *independent analogue*, $\hat{A}_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})}$, tends to the *independent*, $A_{\mathrm{o}}^{\mathrm{X}}$, which is also independent of the *model*, as the *transform* tends to *full functional*. As the *derived iso-independence* increases, however, the *lifted independent analogue*, $A_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})'}$, tends to the *abstract*, $(A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, which is not independent of the *model*, T_{o}.

The finite set of *integral iso-formals* of $A^{\mathrm{X}} * T$ is

$$Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B^{\mathrm{X}} * T = A^{\mathrm{X}} * T\}$$

The *iso-independents* are a subset of the *iso-formals*,

$$Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}}) \subseteq Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T)$$

so the *iso-independent multinomial probability* is at least the *iso-formal multinomial probability*,

$$\frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})} Q_{\mathrm{m},U}(E,z)(B)} \geq \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y_{U,\mathrm{i},T,\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T)} Q_{\mathrm{m},U}(E,z)(B)}$$

That is, *necessary formal*, $\hat{A}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = \hat{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}$, is a weaker condition on the *drawable histories* than *necessary independent*, $\hat{A}_{\mathrm{o}}^{\mathrm{X}} = \hat{E}_{\mathrm{o}}^{\mathrm{X}}$.

The finite set of *integral iso-abstracts* of $(A * T)^{\mathrm{X}}$ is

$$Y_{U,\mathrm{i},T,\mathrm{W},z}^{-1}((A * T)^{\mathrm{X}}) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ (B * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}\}$$

In general, the *iso-independents* are neither a subset nor a superset of the *iso-abstracts*.

In the case of a *full functional transform*, $T_{\mathrm{f}} = \{\{w\}^{\mathrm{CS}\{\}V\mathrm{T}} : w \in V\}^{\mathrm{T}}$, both the *iso-formals* and the *iso-abstracts* equal the *iso-independents*, $Y_{U,\mathrm{i},T_{\mathrm{f}},\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T_{\mathrm{f}}) = Y_{U,\mathrm{i},T_{\mathrm{f}},\mathrm{W},z}^{-1}((A * T_{\mathrm{f}})^{\mathrm{X}}) = Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})$, and so the case is the same as for *aligned non-modelled induction*. At the other extreme of a *unary transform*, $T_{\mathrm{u}} = \{V^{\mathrm{CS}}\}^{\mathrm{T}}$, both the *iso-formals* and the *iso-abstracts* equal the *substrate histograms*, $Y_{U,\mathrm{i},T_{\mathrm{u}},\mathrm{V},z}^{-1}(A^{\mathrm{X}} * T_{\mathrm{u}}) = Y_{U,\mathrm{i},T_{\mathrm{u}},\mathrm{W},z}^{-1}((A * T_{\mathrm{u}})^{\mathrm{X}}) = \mathcal{A}_{U,\mathrm{i},V,z}$, and so the case is the same as for *classical non-modelled induction*.

This suggests that, in the case of *necessary formal*, $\hat{A}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = \hat{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}$, and *necessary abstract*, $(\hat{A}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, as the *transform*, T_{o}, ranges from *full functional*, T_{f}, to *unary*, T_{u}, the condition weakens from *necessary independent aligned induction* to unconditional *classical induction*, and so (i) the *maximum likelihood estimate* tends to the *sample probability histogram*, $\tilde{E}_{\mathrm{o}} \to \hat{A}_{\mathrm{o}}$, (ii) the *maximum likelihood alignment* tends to that of the *sample histogram*, $\mathrm{algn}(Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}}) \to \mathrm{algn}(A_{\mathrm{o}})$, and (iii) at low *alignments* the *sum sensitivity* tends to vary less with the *sample alignment* and more with the negative *scaled sample entropy*, $\mathrm{algn}(A_{\mathrm{o}}) \to -z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}})$.

In general, the *iso-independents* are neither a subset nor a superset of the *iso-transform-independents*, so the *iso-independent multinomial probability*,

$$\frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} Q_{\mathrm{m},U}(E,z)(B)}$$

is not bounded by the *iso-transform-independent multinomial probability*,

$$\frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} Q_{\mathrm{m},U}(E,z)(B)}$$

That is, *necessary formal* and *necessary abstract*, $\hat{A}^{\mathrm{X}}_{\mathrm{o}} * T_{\mathrm{o}} = \hat{E}^{\mathrm{X}}_{\mathrm{o}} * T_{\mathrm{o}} \ \wedge \ (\hat{A}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, is not necessarily a weaker condition on the *drawable histories* than *necessary independent*, $\hat{A}^{\mathrm{X}}_{\mathrm{o}} = \hat{E}^{\mathrm{X}}_{\mathrm{o}}$.

In the case of a *full functional transform*, T_{f}, the *iso-transform-independents* equals the *iso-independents*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{f}},z}(A) = Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})$, and so the case is the same as for *aligned non-modelled induction*. In the case of *unary transform*, T_{u}, the *iso-transform-independents* equals the *substrate histograms*, $\mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{u}},z}(A) = \mathcal{A}_{U,\mathrm{i},V,z}$, and so the case is the same as for *classical non-modelled induction*.

The *sample histogram* is in the intersection of the *iso-independents* and the *iso-transform-independents*,

$$A \ \in \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \ \cap \ Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}}))$$

so the *iso-independence* is non-zero,

$$\frac{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \ \cap \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|}{|Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T),(A * T)^{\mathrm{X}})) \ \cup \ Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|} \ > \ 0$$

and the *iso-independent multinomial probability* is correlated with the *iso-transform-independent multinomial probability*

$$\frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})} Q_{\mathrm{m},U}(E,z)(B)} \ \sim \ \frac{Q_{\mathrm{m},U}(E,z)(A)}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} Q_{\mathrm{m},U}(E,z)(B)}$$

In the case still of *necessary formal* and *necessary abstract*, but where the *transform* is constrained to be such that the *formal independent* equals the *abstract*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$, then the *independent* is an *iso-abstract*,

$$(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}} \in Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

If, in addition, the *independent* is *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, then both the *sample histogram* and the *independent sample histogram* are in the intersection of the *iso-independents* and the *iso-transform-independents*,

$$\{A, A^{\mathrm{X}}\} \subseteq Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}}) \cap Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$$

so the *iso-independence* may be expected to be higher, and the correlation between the *iso-independent multinomial probability* and the *iso-transform-independent multinomial probability* may be expected to be stronger.

This is also the case where the constraints on the *transform* are stricter. For example, if the *formal* equals the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies (A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$.

If the *formal histogram* equals the *abstract histogram* then the *lifted iso-transform-independents* contains the *abstract histogram*

$$(A * T)^{\mathrm{X}} = A^{\mathrm{X}} * T \in \{B * T : B \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\}$$

In this case, if the *abstract* is also *integral*, $(A * T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, the *derived iso-independence* of the *iso-transform-independents*,

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

is greater than would otherwise be the case because the *abstract* is in the intersection, $(A*T)^{\mathrm{X}} \in \{B*T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T), (A*T)^{\mathrm{X}}))\} \cap Y^{-1}_{U,\mathrm{i},W,z}((A*T)^{\mathrm{X}})$.

Note that it is only in the subset where the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}}*T = (A*T)^{\mathrm{X}}$, that the *lifted iso-transform-independent* relation is functional

$$\begin{aligned} &\{(A * T, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\} \\ \in\ & \mathcal{A}_{U,W,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z}) \end{aligned}$$

and

$$\begin{aligned}&\{(A * T, Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))) : A \in \mathcal{A}_{U,V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\} \\ \in\ &\mathcal{A}_{U,W,z} \to \mathrm{P}(\mathcal{A}_{U,\mathrm{i},V,z})\end{aligned}$$

Similarly it is only in the subset where the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, that the *formal* domained relation of the *iso-transform-independents* is functional

$$\begin{aligned}&\{(A^{\mathrm{X}} * T, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\} \\ \in\ &\mathcal{A}_{U,W,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})\end{aligned}$$

and

$$\begin{aligned}&\{(A^{\mathrm{X}} * T, Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))) : A \in \mathcal{A}_{U,V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\} \\ \in\ &\mathcal{A}_{U,W,z} \to \mathrm{P}(\mathcal{A}_{U,\mathrm{i},V,z})\end{aligned}$$

Given the *known substrate transform*, T_{o}, consider the *maximum likelihood estimate* of the *iso-transform-independent conditional generalised multinomial probability distribution*, $\hat{Q}_{\mathrm{m,y},T_{\mathrm{o}},U}$.

The *independent-analogue* or *transform-independent*, $A^{\mathrm{X}(T)}$, is the *maximum likelihood estimate* of the *distribution histogram* of the *multinomial probability* of membership of the *iso-transform-independents*,

$$\begin{aligned}&\{A^{\mathrm{X}(T)}\} = \\ &\qquad \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D, z)(B) : B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A))) : D \in \mathcal{A}_{U,V,z}\})\end{aligned}$$

The *lifted independent-analogue* or the *lifted transform-independent*, $A^{\mathrm{X}(T)'}$, is defined

$$\begin{aligned}&\{A^{\mathrm{X}(T)'}\} = \\ &\qquad \mathrm{maxd}(\{(D, \sum(Q_{\mathrm{m},U}(D, z)(B') : B \in \mathcal{A}'_{U,\mathrm{i,y},T,z}(A))) : D \in \mathcal{A}_{U,W,z}\})\end{aligned}$$

where the *lifted integral iso-transform-independents* is abbreviated

$$\mathcal{A}'_{U,\mathrm{i,y},T,z}(A) \;=\; \{B * T : B \in \mathcal{A}_{U,\mathrm{i,y},T,z}(A)\}$$

The corresponding *dependent-analogue* or *transform-dependent*, $A^{\mathrm{Y}(T)}$, is the *maximum likelihood estimate* of the *distribution histogram* of the *multino-*

mial probability of the *histogram*, A, conditional that it is an *iso-transform-independent*,

$$\{(A^{\mathrm{Y}(T)}, \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)})\} = \\ \max(\{(D, \frac{Q_{\mathrm{m},U}(D,z)(A)}{\sum Q_{\mathrm{m},U}(D,z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}) : D \in \mathcal{A}_{U,V,z}\})$$

In section 'Likely histograms', above, the logarithm of the *maximum conditional probability* with respect to the *dependent-analogue* is conjectured to vary with the *relative space* with respect to the *independent-analogue*. In the case of *iso-transform-independent conditional*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \sim \\ \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A)$$

where the *distribution-relative multinomial space* is defined, in section 'Likely histograms', above, as

$$\mathrm{spaceRelative}(E)(A) \quad := \quad -\ln \frac{\mathrm{mpdf}(U)(E,z)(A)}{\mathrm{mpdf}(U)(E,z)(E)}$$

The set of *iso-transform-independents* is *entity-like*, not *law-like*, so the *derived*, $A * T$, and the *transform-dependent derived*, $A^{\mathrm{Y}(T)} * T$, are not necessarily equal to each other and nor are they necessarily equal to the *transform-independent derived*, $A^{\mathrm{X}(T)} * T$. In section 'Transform alignment', above, it is conjectured that the relation between the *relative spaces*,

$$\begin{aligned} 0 &= \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{X}(T)}) \\ &\leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) \\ &\leq \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A^{\mathrm{Y}(T)}) \end{aligned}$$

can be *lifted* and so the *dependent analogue derived alignment* is conjectured to be greater than or equal to the *derived alignment* which in turn is greater than or equal to the *independent analogue derived alignment*,

$$\mathrm{algn}(A^{\mathrm{X}(T)} * T) \leq \mathrm{algn}(A * T) \leq \mathrm{algn}(A^{\mathrm{Y}(T)} * T)$$

The *transform-dependent* varies with the *histogram*, $A^{\mathrm{Y}(T)} \sim A$, so conjecture that the *log-likelihood* varies with the *derived alignment*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \quad \sim \quad \mathrm{algn}(A * T)$$

The derivation of this correlation can be seen more clearly in terms of a decomposition into three separate correlations. First, conjecture that the logarithm of the *iso-transform-independent conditional multinomial probability* of the *histogram*, A, with respect to the *dependent analogue* or *transform-dependent*, $A^{\mathrm{Y}(T)}$, varies against the logarithm of the *iso-transform-independent conditional multinomial probability* with respect to the *independent analogue* or *transform-independent*, $A^{\mathrm{X}(T)}$,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \sim \\ -\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}$$

This relation is called the *dependent-independent anti-correlation.* As shown in ‘Likely histograms’, above, the strength of the *dependent-independent anti-correlation* depends on the *relative space* of the *histogram* with respect to the *independent analogue*, $\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A)$.

Second, conjecture that the negative logarithm of the *iso-transform-independent conditional multinomial probability* of the *histogram*, A, with respect to the *independent analogue* or *transform-independent*, $A^{\mathrm{X}(T)}$, varies with the negative logarithm of the *lifted iso-transform-independent conditional multinomial probability* of the *derived*, $A * T$, with respect to the *lifted independent analogue* or *transform-independent derived*, $A^{\mathrm{X}(T)} * T$,

$$-\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{X}(T)},z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \sim \\ -\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)} * T,z)(A * T)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{X}(T)} * T,z)(B') : B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)}$$

This correlation is called the *underlying-lifted correlation.* As mentioned above in this section, *lifting* the *iso-transform-independents* is not functional,

$$\{(A * T, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,V,z}\} \quad \notin \quad \mathcal{A}_{U,W,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})$$

unless the *formal histogram* equals the *abstract histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$. The *underlying-lifted correlation* is expected to be weaker if the *lift* is not functional.

Third, conjecture that, in the case where the *lifted transform-independent* is *integral*, $A^{\mathrm{X}(T)'} \in \mathcal{A}_{\mathrm{i}}$, the denominator of the *lifted iso-transform-independent* *conditional multinomial probability* is dominated by the *lifted transform-independent* term, $Q_{\mathrm{m},U}(A^{\mathrm{X}(T)} * T, z)(A^{\mathrm{X}(T)'})$, and similar terms, and so the negative logarithm of the *lifted iso-transform-independent conditional multinomial probability* with respect to the *lifted independent analogue* or *transform-independent derived*, $A^{\mathrm{X}(T)} * T$, varies with the negative logarithm of the ratio of (i) the *multinomial probability* of the *derived*, $A * T$, with respect to the *transform-independent derived*, $A^{\mathrm{X}(T)} * T$, and (ii) the *multinomial probability* of the *lifted transform-independent*, $A^{\mathrm{X}(T)'}$, with respect to the *transform-independent derived*, $A^{\mathrm{X}(T)} * T$, which approximates to the *relative space* with respect to the *abstract*, $(A * T)^{\mathrm{X}}$, which is the *derived alignment*,

$$\begin{aligned}
-\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)} * T, z)(A * T)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{X}(T)} * T, z)(B') : B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} & \\
&\sim\ -\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)} * T, z)(A * T)}{Q_{\mathrm{m},U}(A^{\mathrm{X}(T)} * T, z)(A^{\mathrm{X}(T)'})} \\
&\approx\ -\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)((A * T)^{\mathrm{X}})} \\
&=\ \mathrm{spaceRelative}((A * T)^{\mathrm{X}})(A * T) \\
&=\ \mathrm{algn}(A * T)
\end{aligned}$$

This correlation is called the *conditional-relative correlation.* The strength of the *conditional-relative correlation* increases with the *derived iso-independence* of the *integral lifted iso-transform-independents*,

$$\frac{|\{B * T : B \in Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\}|}{|Y^{-1}_{U,\mathrm{i},W,z}((A * T)^{\mathrm{X}})|}$$

Insofar as the *transform-independent derived* approximates to the *abstract*, $A^{\mathrm{X}(T)} * T \approx (A * T)^{\mathrm{X}}$, as the *derived iso-independence* increases, the *lifted transform-independent*, $A^{\mathrm{X}(T)'}$, tends to the *abstract*, $(A * T)^{\mathrm{X}}$, and the *lifted transform-independent* term, $Q_{\mathrm{m},U}(A^{\mathrm{X}(T)} * T, z)(A^{\mathrm{X}(T)'})$, tends to the *abstract* term, $Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)((A * T)^{\mathrm{X}})$, in the case where both are *integral*, $A^{\mathrm{X}(T)'}, (A * T)^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$.

In *transform induction*, where (i) the *history probability function* is *iso-transform-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (iii) the *scaled estimate distribution histogram* is *integral*, $\tilde{E}_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then the *log likelihood* of the *iso-transform-independent* *conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *relative space* of the *sample* with respect to the *transform-independent*,

$$\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \mathrm{spaceRelative}(A_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})})(A_{\mathrm{o}})$$

and varies with the *derived alignment*,

$$\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

The *derived alignment* of the *maximum likelihood estimate* is greater than or equal to that of the *sample*,

$$\mathrm{algn}(Z_{\mathrm{o}} * \tilde{E}_{\mathrm{o}} * T) \quad \geq \quad \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

In section 'Classical modelled induction', above, it is shown that the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$, can be related to queries on the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$, in the special case where the *sample histogram* is *natural*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$. The given *substrate transform* must be such that its *contraction* has *underlying variables* that are a subset of the query *variables*, $\mathrm{und}(T_{\mathrm{o}}^{\%}) \subseteq K$. In the case where the query *histogram* consists of one *effective state*, $Q = \{(S_Q, 1)\}$, the application of the query in terms of a modified *sample histogram* is

$$\begin{aligned}
&(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K) = \\
&\quad \{(N, (\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}))^{1/z_{\mathrm{o}}}) : N \in (V_{\mathrm{o}} \setminus K)^{\mathrm{CS}}, \\
&\qquad\qquad A_{Q,N} = A_{\mathrm{o}} - (A_{\mathrm{o}} * C_Q) + ((A_{\mathrm{o}} * C_Q) \ \% \ K * \{N\}^{\mathrm{U}})\}^{\wedge}
\end{aligned}$$

where $\{R_Q\} = (Q * T_{\mathrm{o}}^{\%})^{\mathrm{FS}}$, $C_Q = T_{\mathrm{o}}^{-1}(R_Q)$ and $\mathrm{his} = \mathrm{histogram} \in \mathcal{T} \to \mathcal{A}$. If the *sample histogram* is *completely effective*, $A_{\mathrm{o}}^{\mathrm{F}} = V_{\mathrm{o}}^{\mathrm{C}}$, the modified *sample histogram*, $A_{Q,N}$, can be *drawn* from the *distribution*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{Q,N}) > 0$, because its *derived* is equal to the *known derived*, $A_{Q,N} * T_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}}$. That is, the modified *sample histogram* is in the *iso-deriveds*, $A_{Q,N} \in D_{U,\mathrm{i},T_{\mathrm{o}},z_{\mathrm{o}}}^{-1}(A_{\mathrm{o}} * T_{\mathrm{o}})$.

However, in the case of *transform induction*, where the *transform-independent* is *necessary*, the modified *sample histogram* is not necessarily in the *iso-transform-independents*, $A_{Q,N} \notin Y_{U,\mathrm{i},T_{\mathrm{o}},z}^{-1}((A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}}, (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}))$. Although the modified *sample histogram* is necessarily an *iso-abstract*, $(A_{Q,N} * T_{\mathrm{o}})^{\mathrm{X}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, in some cases the modified *sample histogram* is not an *iso-formal*,

$A^{\mathrm{X}}_{Q,N} * T_{\mathrm{o}} \neq A^{\mathrm{X}}_{\mathrm{o}} * T_{\mathrm{o}}$. Even if the modified *derived* is an *iso-transform-independent*, the modified *derived*, $\hat{A}_{Q,N} * T_{\mathrm{o}}$, is not *necessarily* equal to that of the *distribution*, $\hat{E}_{\mathrm{h}} * T_{\mathrm{o}}$. That is, in some cases $\hat{A}_{Q,N} * T_{\mathrm{o}} \neq \hat{E}_{\mathrm{h}} * T_{\mathrm{o}}$. So it cannot be assumed that application of the query via the *model* of the *sample* is equal to the query via the *model* of the *distribution*, $(Q*T_{\mathrm{o}}^{\%}*\mathrm{his}(T_{\mathrm{o}}^{\%})*A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K) \neq (Q*T_{\mathrm{o}}^{\%}*\mathrm{his}(T_{\mathrm{o}}^{\%})*E_{\mathrm{h}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$. Nor can the query via the *model* of the *sample*, $(Q*T_{\mathrm{o}}^{\%}*\mathrm{his}(T_{\mathrm{o}}^{\%})*A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$, be expressed in terms of the *iso-transform-independent conditional stuffed historical probability distribution* at the *scaled sample*, $\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$.

Consider the constraints that may be added to *transform induction* to increase the resemblance to *classical modelled induction*, so that queries via the *model* of the *sample* are more approximate to queries via the *model* of the *distribution*, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K) \approx (Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * E_{\mathrm{h}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$.

In *abstract induction*, above, it is conjectured that, because the *law-like iso-deriveds* are a subset of the set of *entity-like iso-abstracts*, $D^{-1}_{U,T,z}(A * T) \subseteq Y^{-1}_{U,T,\mathrm{W},z}((A*T)^{\mathrm{X}})$, the *maximum likelihood estimate* is more *classical* if (i) the *sample* is *known* to be equal to the *independent analogue*, or *naturalisation*, $A_{\mathrm{o}} = A_{\mathrm{o}}*T_{\mathrm{o}}*T_{\mathrm{o}}^{\dagger}$, and (ii) the *relative space* of the *sample* with respect to the *naturalised sample abstract*, $\mathrm{spaceRelative}((A_{\mathrm{o}}*T_{\mathrm{o}})^{\mathrm{X}}*T_{\mathrm{o}}^{\dagger})(A_{\mathrm{o}}*T_{\mathrm{o}}*T_{\mathrm{o}}^{\dagger})$, is high.

For *transform induction*, however, the set of *iso-deriveds* is not necessarily a subset of the *entity-like iso-transform-independents*, $|D^{-1}_{U,\mathrm{i},T,z}(A * T) \setminus Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}}*T),(A*T)^{\mathrm{X}})| \geq 0$. The set of *iso-liftisations* is a *law-like* subset of the *iso-transform-independents*,

$$Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T)\ \cap\ D^{-1}_{U,T,z}(A * T)\ \ \subseteq\ \ Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T)\ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})$$

but the *independent analogue*, which is the *liftisation*, $A^{\mathrm{K}(T)}$, is sometimes not computable. So consider instead the set of *iso-idealisations* which is a further *law-like* subset of the *iso-transform-independents*,

$$\begin{aligned}
Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) &= C^{-1}_{U,T,\mathrm{x},z}(\{(A * C^{\mathrm{U}})^{\mathrm{X}\wedge} : C \in T^{\mathrm{P}}\})\ \cap\ D^{-1}_{U,T,z}(A * T) \\
&\subseteq Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T)\ \cap\ D^{-1}_{U,T,z}(A * T) \\
&\subseteq Y^{-1}_{U,T,\mathrm{V},z}(A^{\mathrm{X}} * T)\ \cap\ Y^{-1}_{U,T,\mathrm{W},z}((A * T)^{\mathrm{X}})
\end{aligned}$$

Conjecture that the logarithm of the fraction of the sum of the *iso-transform-independent multinomial probabilities*, with respect to the *idealisation*, $A*T*$

$T^{\dagger A}$, that are *iso-idealisations* varies as the *relative space* of the *idealisation* with respect to the *transform-independent*,

$$\ln \frac{\sum Q_{\mathrm{m},U}(A * T * T^{\dagger A}, z)(B) : B \in Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})}{\sum Q_{\mathrm{m},U}(A * T * T^{\dagger A}, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}$$
$$\sim\ -\ \mathrm{spaceRelative}(A * T * T^{\dagger A})(A^{\mathrm{X}(T)})$$
$$\sim\ \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A * T * T^{\dagger A})$$

If the *relative space* is high, the elements of the *iso-transform-independents* which are not *iso-idealisations* and which sometimes do not have the same *derived* as the *idealisation*, $A * T * T^{\dagger A} * T = A * T$, have low *multinomial probability* with respect to the *idealisation*,

$$\sum (Q_{\mathrm{m},U}(A * T * T^{\dagger A}, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A) \setminus Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A}))\ \approx\ 0$$

If the *sample* is *known* to be *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, then with the increase of the *relative space* of the *sample* with respect to the *transform-independent*, $\mathrm{spaceRelative}(A_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})})(A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}})$, the *maximum likelihood estimate*, $\tilde{E}_{\mathrm{o}}$, which is the *transform-dependent*, $\hat{A}_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})}$, tends to the *idealisation-dependent* which equals the *idealisation*, $\hat{A}_{\mathrm{o}}^{\dagger(T_{\mathrm{o}})} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, and away from the *transform-independent*, $\hat{A}_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})}$. Consequently, *ideal sample* increases the correlation between the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* and the *log likelihood* of the *iso-idealisation conditional stuffed historical probability distribution* at the *idealisation*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})\ \sim\ \ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

Thus, the *maximum likelihood estimate* is more *classical* if the *sample* is *known* to be *ideal* and the *relative space* is high. That is, *transform induction* is analogous to *abstract induction* in this respect except that the stricter *ideal sample*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, is required instead of *natural sample*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$.

Even if the *sample* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, the *maximum likelihood estimate* of the *distribution histogram*, $\tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}^{\mathrm{Y}(T)}$, is not necessarily equal to the *sample*, $\tilde{E}_{\mathrm{o}} \neq \hat{A}_{\mathrm{o}}$. So it still cannot be assumed that application of the query via the *model* of the *sample* is equal to the query via the *model* of the *distribution*, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K) \neq (Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * E_{\mathrm{h}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$. Nor can the query via the *model* of the *sample*, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$, be expressed in terms of the *iso-transform-independent conditional stuffed historical probability distribution* at the *scaled idealised sample*, $\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$.

The *relative space* of the *histogram* with respect to the *independent analogue*, or *transform-independent*, $A^{\mathrm{X}(T)}$, varies with the *lifted relative space*, which varies with the *derived alignment*,

$$\begin{aligned}\mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A) &\sim \mathrm{spaceRelative}(A^{\mathrm{X}(T)} * T)(A * T) \\ &\approx \mathrm{spaceRelative}((A * T)^{\mathrm{X}})(A * T) \\ &= \mathrm{algn}(A * T)\end{aligned}$$

depending on the *underlying-lifted correlation* and the *conditional-relative correlation*.

The *conditional-relative correlation* improves as the *derived iso-independence* increases and the *lifted transform-independent*, $A^{\mathrm{X}(T)'}$, tends to the *abstract*, $(A * T)^{\mathrm{X}}$. As shown in 'Deltas and Perturbations' and 'Abstract induction', above, in the case where the *formal* is *independent*, $A^{\mathrm{X}} * T = (A^{\mathrm{X}} * T)^{\mathrm{X}}$, the *possible derived volume* equals the *derived volume*, $w' = w$ where $w' = |T^{-1}|$ and $w = |W^{\mathrm{C}}|$. In this case the *derived iso-independence* is greater than it would be otherwise, improving the approximation of the *lifted transform-independent* to the *abstract*, $A^{\mathrm{X}(T)'} \approx (A * T)^{\mathrm{X}}$.

In the stricter case where the *formal* equals the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, the *lifted iso-transform-independents* contains the *abstract histogram*

$$(A * T)^{\mathrm{X}} \in Y^{-1}_{U,W,z}((A * T)^{\mathrm{X}}) \cap \{B * T : B \in Y^{-1}_{U,T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))\}$$

and again the *derived iso-independence* is greater than it would be otherwise, strengthening the *conditional-relative correlation*.

Incidentally, the case where the *formal* equals the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, implies that the *formal independent* equals the *abstract*, $(A^{\mathrm{X}} * T)^{\mathrm{X}} = (A * T)^{\mathrm{X}}$, which implies that the *independent* is an *iso-transform-independent*,

$$A^{\mathrm{X}} \in Y^{-1}_{U,V,z}(A^{\mathrm{X}}) \cap Y^{-1}_{U,\mathrm{i},T,z}(((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}}))$$

and so the *underlying iso-independence* is also greater than it would be otherwise.

Furthermore, in the case where the *formal histogram* equals the *abstract* *histogram*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$, the *lifted iso-transform-independent* relation is functional

$$\begin{aligned}&\{(A * T, ((A^{\mathrm{X}} * T), (A * T)^{\mathrm{X}})) : A \in \mathcal{A}_{U,V,z},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\}\\ \in\ &\mathcal{A}_{U,W,z} \to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})\end{aligned}$$

and so the *underlying-lifted correlation* is also strengthened.

In the case where the *formal histogram* equals the *abstract histogram*, the *independent analogue*, or *transform-independent*, equals the computable *naturalised abstract*,

$$A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \implies A^{\mathrm{X}(T)} = (A * T)^{\mathrm{X}} * T^{\dagger}$$

So the three *correlations* now simplify. First, the *dependent-independent anti-correlation* between the logarithm of the *iso-transform-independent conditional multinomial probability* of the *histogram*, A, with respect to the *dependent analogue* or *transform-dependent*, $A^{\mathrm{Y}(T)}$, and the logarithm of the *iso-transform-independent conditional multinomial probability* with respect to the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, is,

$$\begin{aligned}\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \sim&\\ - \ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(A)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)}&\end{aligned}$$

Second, the *underlying-lifted correlation* between the negative logarithm of the *iso-transform-independent conditional multinomial probability* of the *histogram*, A, with respect to the *naturalised abstract*, $(A * T)^{\mathrm{X}} * T^{\dagger}$, and the negative logarithm of the *lifted iso-transform-independent conditional multinomial probability* of the *derived*, $A*T$, with respect to the *abstract*, $(A*T)^{\mathrm{X}}$, is,

$$\begin{aligned}- \ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(A)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \sim&\\ - \ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(B') : B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)}&\end{aligned}$$

Third, the *conditional-relative correlation* between the negative logarithm of the *lifted iso-transform-independent conditional multinomial probability* with respect to the *abstract*, $(A * T)^{\mathrm{X}}$, and the negative logarithm of the *relative* *multinomial probability* with respect to the *abstract*, $(A * T)^{\mathrm{X}}$, which is the *derived alignment*, is,

$$\begin{aligned}
&- \ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{\sum Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(B') : B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z}(A)} \\
&\qquad \sim \ - \ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A^{\mathrm{X}(T)'})} \\
&\qquad \approx \ - \ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)((A * T)^{\mathrm{X}})} \\
&\qquad = \ \mathrm{spaceRelative}((A * T)^{\mathrm{X}})(A * T) \\
&\qquad = \ \mathrm{algn}(A * T)
\end{aligned}$$

That is, if the *sample* is *known* to have *formal-abstract equivalence*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, the correlation between the *relative space* of the *histogram* with respect to the *naturalised abstract* and the *derived alignment*,

$$\mathrm{spaceRelative}((A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger})(A_{\mathrm{o}}) \ \sim \ \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

is higher than would otherwise be the case. So the correlation between the logarithm of the *iso-transform-independent conditional multinomial probability* with respect to the *transform-dependent*, and the *derived alignment*,

$$\ln \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T_{\mathrm{o}})}, z_{\mathrm{o}})(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{o}},z_{\mathrm{o}}}(A_{\mathrm{o}})} \ \sim \ \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

is also higher. Consequently, *formal-abstract equivalence* increases the correlation between the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* and the *derived alignment*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

Sample formal-abstract equivalence, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, implies stricter conditions for *transform induction*, (i) *necessary formal*, $\hat{A}_H^{\mathrm{X}} * T_{\mathrm{o}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}$, and *necessary independence of formal*, $\hat{A}_H^{\mathrm{X}} * T_{\mathrm{o}} = (\hat{A}_H^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$, and (ii) *necessary abstract*, $(\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$, and *necessary abstraction of independent*, $(\hat{A}_H^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{A}_H * T_{\mathrm{o}})^{\mathrm{X}}$. *Sample formal-abstract equivalence* also implies (i) *necessary independence of distribution formal*, $\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}} = (\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$, and (ii) *necessary abstraction of distribution independent*, $(\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$.

If the *sample* is *known* to be both (i) *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, and (ii) *formal-abstract equivalent*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, then the logarithm of the fraction of the sum of the *iso-transform-independent multinomial probabilities*, with respect to the *idealisation*, that are *iso-idealisations* also varies with the *derived alignment*,

$$\begin{aligned}\ln \frac{\sum Q_{\mathrm{m},U}(A_{\mathrm{o}},z_{\mathrm{o}})(B) : B \in Y^{-1}_{U,\mathrm{i},T_{\mathrm{o}},\dagger,z_{\mathrm{o}}}(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o}},z_{\mathrm{o}})(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T_{\mathrm{o}},z_{\mathrm{o}}}(A_{\mathrm{o}})} \\ &\sim \mathrm{spaceRelative}((A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}} * T_{\mathrm{o}}^{\dagger})(A_{\mathrm{o}}) \\ &\sim \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})\end{aligned}$$

So the *maximum likelihood estimate* becomes more *classical* as the *derived alignment* increases. As the *derived alignment* increases the more the query via the *model* of the *sample* approximates to the query via the *model* of the *distribution*, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K) \approx (Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * E_{\mathrm{h}})^{\wedge}\ \%\ (V_{\mathrm{o}} \setminus K)$.

Now *transform induction* is analogous to *abstract induction* except that (i) the stricter *ideal sample*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, is required instead of *natural sample*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, and (ii) the stricter *sample formal-abstract equality*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, is required instead of *independent sample formal*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$.

The set of *iso-idealisations* is a subset of the intersection of the *iso-independents* and *iso-deriveds*, $Y^{-1}_{U,T,\dagger,z}(A * T * T^{\dagger A}) \subseteq Y^{-1}_{U,V,z}(A^{\mathrm{X}}) \cap D^{-1}_{U,T,z}(A * T)$. In section 'Idealisation induction', above, it is shown that the cardinality of the *iso-independents* varies with the *alignment*,

$$\ln |Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})| \ \sim\ \mathrm{algn}(A)$$

At high *alignments*, the *iso-independence* of the *iso-idealisations*, $|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|/|Y^{-1}_{U,\mathrm{i},V,z}(A^{\mathrm{X}})|$, decreases, and the *iso-derivedence*, $|Y^{-1}_{U,\mathrm{i},T,\dagger,z}(A * T * T^{\dagger A})|/|D^{-1}_{U,\mathrm{i},T,z}(A * T)|$, increases. So the *iso-idealisation log likelihood* varies with the *iso-derived log likelihood*,

$$\ln \hat{Q}_{\mathrm{m},\dagger,T,U}(A,z)(A) \ \sim\ \ln \hat{Q}_{\mathrm{m},\mathrm{d},T,U}(A,z)(A)$$

The *iso-idealisation conditional multinomial distribution sum sensitivity* varies with the *iso-independent sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},\dagger,T,U}(A,z))) \ \sim\ \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},\mathrm{y},U}(A,z)))$$

and the *iso-derived sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},\dagger,T,U}(A,z))) \ \sim\ \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},\mathrm{d},T,U}(A,z)))$$

It is conjectured above that at intermediate *alignments*, $0 \ll \mathrm{algn}(A) \ll \mathrm{algnMax}(U)(V,z)$, the *iso-independent sum sensitivity* is constant and the *iso-idealisation sum sensitivity* varies only with the *iso-derived sum sensitivity*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},\dagger,T,U}(A,z))) \quad \sim \quad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m,d},T,U}(A,z)))$$

In *transform induction*, where (i) the *history probability function* is *iso-transform-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}$, given some *substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *sample* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, (iii) the *sample formal* equals the *sample abstract*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, (iv) the *alignment* is at least intermediate, $\mathrm{algn}(A_{\mathrm{o}}) > \mathrm{algnMax}(U)(V_{\mathrm{o}},z_{\mathrm{o}})/2$, (v) the *derived alignment* is high, $\mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}}) \gg 0$, (vi) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (vii) the *scaled estimate distribution histogram* is *integral*, $\tilde{E}_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then (a) the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *log likelihood* of the *iso-idealisation conditional stuffed historical probability distribution* at the *idealisation* which, in turn, varies with the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *sample*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \quad &\sim \quad \ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \\ &\sim \quad \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}})\end{aligned}$$

(b) so the *log likelihood* varies with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \\ &\quad (z_{\mathrm{o}}+v_{\mathrm{o}}) \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}} + V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) \\ &\qquad\qquad -z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}}) \ - v_{\mathrm{o}} \times \mathrm{entropy}(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})\end{aligned}$$

(c) the *formal alignment* of the *maximum likelihood estimate* is zero,

$$\begin{aligned}\tilde{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} \quad &= \quad \hat{A}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} \\ &= \quad (\hat{A}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}} \\ &= \quad (\tilde{E}_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}\end{aligned}$$

and (d) the *maximum likelihood estimate derived* approximates to the *normalised sample derived*,

$$\tilde{E}_{\mathrm{o}} * T_{\mathrm{o}} \quad \approx \quad \hat{A}_{\mathrm{o}} * T_{\mathrm{o}}$$

In this case queries via the *model* of the *sample* approximate to queries via the *model* of the *distribution*,

$$(Q * T_{\rm o}^{\%} * \mathrm{his}(T_{\rm o}^{\%}) * A_{\rm o})^{\wedge}\ \%\ (V_{\rm o} \setminus K) \quad \approx \quad (Q * T_{\rm o}^{\%} * \mathrm{his}(T_{\rm o}^{\%}) * E_{\rm h})^{\wedge}\ \%\ (V_{\rm o} \setminus K)$$

That is, at high *derived alignments*, and intermediate or high *underlying alignment*, where the *sample* is *known* to be *ideal* and the *sample formal* is *known* to be equal to the *sample abstract*, *aligned modelled induction* has similar properties to *classical modelled induction*.

If the *relative entropy* is high, $\mathrm{entropyCross}(A_{\rm o} * T_{\rm o}, V_{\rm o}^{\rm C} * T_{\rm o}) > \ln |T_{\rm o}^{-1}|$, the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* at the *naturalisation* varies with the *derived entropy*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\rm h,d,\mathit{T}_o,\mathit{U}}(A_{{\rm o},z_{\rm h}}, z_{\rm o}))) \quad \sim \quad z_{\rm o} \times \mathrm{entropy}(A_{\rm o} * T_{\rm o})$$

Note, however, that because the *transform induction* is more *derived aligned-like* than *classical modelled induction*,

$$\frac{|\{B * T : B \in \mathcal{A}_{U,\rm i,y,\mathit{T},\mathit{z}}(A)\}|}{|Y_{U,\rm i,\mathit{W},\mathit{z}}^{-1}((A * T)^{\rm X})|} \quad \geq \quad \frac{1}{|Y_{U,\rm i,\mathit{W},\mathit{z}}^{-1}((A * T)^{\rm X})|}$$

the *sum sensitivity* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* may be expected rather to vary against the *derived alignment*,

$$\begin{aligned} &\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\rm h,y,\mathit{T}_o,\mathit{U}}(\tilde{E}_{{\rm o},z_{\rm h}}, z_{\rm o}))) \\ &\qquad \sim z_{\rm o} \times \mathrm{entropy}(A_{\rm o} * T_{\rm o}) - z_{\rm o} \times \mathrm{entropy}((A_{\rm o} * T_{\rm o})^{\rm X}) \\ &\qquad \approx -\ \mathrm{algn}(A_{\rm o} * T_{\rm o}) \end{aligned}$$

So the *sum sensitivity* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned} \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\rm h,y,\mathit{T}_o,\mathit{U}}(\tilde{E}_{{\rm o},z_{\rm h}}, z_{\rm o}))) \quad &\sim \quad -\ \mathrm{algn}(A_{\rm o} * T_{\rm o}) \\ &\sim \quad -\ \ln \hat{Q}_{\rm h,y,\mathit{T}_o,\mathit{U}}(\tilde{E}_{{\rm o},z_{\rm h}}, z_{\rm o})(A_{\rm o}) \end{aligned}$$

In the case of high *relative entropy* the *sum sensitivity* of the *iso-derived conditional stuffed historical probability distribution* is conjectured to vary with the *unknown-known multinomial probability distribution sum sensitivity difference*,

$$\begin{aligned} &\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\rm h,d,\mathit{T}_o,\mathit{U}}(A_{{\rm o},z_{\rm h}}, z_{\rm o}))) \sim \\ &\quad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm m},U}(A_{\rm o}, z_{\rm o}))) - \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm m},U}(A_{\rm o} * T_{\rm o}, z_{\rm o}))) \end{aligned}$$

so the *sum sensitivity* of the *iso-transform-independent conditional stuffed historical probability distribution* is also conjectured to vary with the *unknown-known multinomial probability distribution sum sensitivity difference*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}}))) \sim \\ &\quad \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}},z_{\mathrm{o}}))) - \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{m},U}(A_{\mathrm{o}} * T_{\mathrm{o}},z_{\mathrm{o}})))\end{aligned}$$

the *sum sensitivity* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* is less than or equal to the *sum sensitivity* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}}))) \\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})))\end{aligned}$$

and the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* is greater than or equal to the *log likelihood* of the *stuffed historical probability distribution* at the *maximum likelihood estimate*,

$$\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z)(A_{\mathrm{o}}) \quad \geq \quad \ln \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}},z)(A_{\mathrm{o}})$$

That is, in the case where (i) the *sample* is *ideal*, (ii) the *sample formal* equals the *sample abstract*, and (iii) the *relative entropy* is high, as the *derived alignment* increases, (a) the *log-likelihood* increases and (b) the *underlying-derived sum sensitivity difference* decreases.

If, in addition, the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, then the *log likelihood* varies with the *scaled component size cardinality relative entropy*,

$$\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad z_{\mathrm{o}} \times \mathrm{entropyRelative}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})$$

and varies against the *specialising derived substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

where

$$C_{\mathrm{G},V,\mathrm{T,H}}(T) = \mathrm{coderHistorySubstrateDerivedSpecialising}(U,X,T,D_{\mathrm{S}},D_{\mathrm{X}})$$

Note that this correlation is conjectured to be weaker than that of *classical modelled induction*,

$$\ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

because the *expected component entropy* of the *idealisation* is less than or equal to that of the *naturalisation*.

Also note that, in the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the *integral idealisation* can be exactly equal to the *sample*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, but the *integral naturalisation* cannot, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$.

The *iso-transform-independent conditional stuffed historical probability distribution log-likelihood* is maximised and the *specialising derived substrate history coder space* is minimised by varying the *transform* such that (i) the *derived entropy* is low, (ii) the *possible derived volume* is small, (iii) the *underlying components* have high *entropy* and (iv) high *counts* are in low cardinality *components* and high cardinality *components* have low *counts*.

In the high *relative entropy* case, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, the *sum sensitivity* varies against the *log-likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad - \ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

In the case where the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, the *sensitivity* to *model* also varies against the *log likelihood*,

$$\begin{aligned} - \ \ln |\max(\{(T, \hat{Q}_{\mathrm{h,y},T,U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\})| \ \sim \\ - \ \ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \end{aligned}$$

and the *log-likelihood* varies with its *degree of structure* with respect to the *expanded specialising derived history coder*, $C_{\mathrm{G,T,H}}$,

$$\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \mathrm{structure}(U,X)(P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}, C_{\mathrm{G,T,H}}(T_{\mathrm{o}}))$$

Although (i) it cannot be assumed that the application of the query via the *model* of the *sample* is equal to the query via the *model* of the *distribution*, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K) \neq (Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * E_{\mathrm{h}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K)$, and (ii) the query via the *model* of the *sample*, $(Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K)$, cannot be expressed in terms of the *iso-transform-independent conditional stuffed historical probability distribution* at the *scaled ideal sample*, $\hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$, in the case where the *sample* is *ideal* and the *sample formal* equals the *sample abstract* the *maximum likelihood estimate* approximates to the *normalised sample derived*,

$$\tilde{E}_{\mathrm{o}} * T_{\mathrm{o}} \quad \approx \quad \hat{A}_{\mathrm{o}} * T_{\mathrm{o}}$$

and so queries via the *model* of the *sample* approximate to queries via the *model* of the *distribution*,

$$(Q * T_{\rm o}^{\%} * \mathrm{his}(T_{\rm o}^{\%}) * A_{\rm o})^{\wedge}\ \%\ (V_{\rm o} \setminus K) \quad \approx \quad (Q * T_{\rm o}^{\%} * \mathrm{his}(T_{\rm o}^{\%}) * E_{\rm h})^{\wedge}\ \%\ (V_{\rm o} \setminus K)$$

So it may be conjectured that (a) the *query sensitivity* to the *distribution histogram* varies as the *iso-transform-independent sum sensitivity* divided by the *size*

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\rm h,y,T_o,U}(\tilde{E}_{{\rm o},z_{\rm h}}, z_{\rm o})))/z_{\rm o}$$

(b) although the query application via the *model* is sometimes not equal to the *estimated transformed conditional product*, the *query sensitivity* to the *distribution histogram* is sometimes lower,

$$\begin{aligned}\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\rm h,y,T_o,U}(\tilde{E}_{{\rm o},z_{\rm h}}, z_{\rm o})))/z_{\rm o}&\\ \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{{\rm h},U}(A_{{\rm o},z_{\rm h}}, z_{\rm o})))/z_{\rm o}&\end{aligned}$$

and (c) the *model likelihood* of the *distribution histogram* is sometimes higher,

$$\hat{Q}_{\rm h,y,T_o,U}(\tilde{E}_{{\rm o},z_{\rm h}}, z)(A_{Q,N}) \quad \geq \quad \hat{Q}_{{\rm h},U}(A_{{\rm o},z_{\rm h}}, z)(A_{Q,N})$$

In other words, querying via the *known derived* of the *model* sometimes reduces the *sensitivity* to the *unknown* and increases the *likelihood* at the cost of modifying the query. Note that the degree to which this is case is lower in *aligned modelled induction* than it is in *classical modelled induction*.

Note that, although the added constraint of *known ideal sample*, $A_{\rm o} = A_{\rm o} * T_{\rm o} * T_{\rm o}^{\dagger A_{\rm o}}$, can increase the resemblance to *classical induction*, the *induction* remains *transform induction* because the condition of *necessary formal* and *necessary abstract* has not changed and so neither the *iso-set*, $Y^{-1}_{U,{\rm i},T_{\rm o},z_{\rm o}}(((A_{\rm o}^{\rm X} * T_{\rm o}), (A_{\rm o} * T_{\rm o})^{\rm X}))$, nor the *iso-derivedence* have changed. That is, the *maximum likelihood estimate*, $\tilde{E}_{\rm o}$, does not move away from the *transform-dependent*, $\hat{A}_{\rm o}^{{\rm Y}(T_{\rm o})}$, to the *idealisation-dependent*, $\hat{A}_{\rm o}^{\dagger(T_{\rm o})}$, but rather both the *maximum likelihood estimate* and the *transform-dependent* move together towards the *idealisation-dependent*, $\tilde{E}_{\rm o} = \hat{A}_{\rm o}^{{\rm Y}(T_{\rm o})} \approx \hat{A}_{\rm o}^{\dagger(T_{\rm o})}$, and so both the *maximum likelihood estimate* the *transform-dependent* move together towards the *derived-dependent*, $\tilde{E}_{\rm o} = \hat{A}_{\rm o}^{{\rm Y}(T_{\rm o})} \approx \hat{A}_{\rm o}^{{\rm D}(T_{\rm o})}$.

Note also that the assumption of high *derived alignment*, $\mathrm{algn}(A_{\rm o} * T_{\rm o}) \gg 0$, is not well defined, although there is an upper bound, $\mathrm{algnMax}(U)(W_{\rm o}, z_{\rm o})$. A more formal method of expression would be to say that the correlation between the *iso-transform-independent conditional stuffed historical probability* *distribution* and the *iso-derived conditional stuffed historical probability distribution* is itself correlated to the *derived alignment*,

$$[\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})] \quad \sim \quad \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

More formal still would be to define this relation in terms of the correlations of functions of the *sized cardinal substrate histograms*, $\mathcal{A}_z$, given the *renormalised geometry-weighted probability function*, $\mathrm{corr}(z) \in (\mathcal{A}_z \to \mathbf{R}) \times (\mathcal{A}_z \to \mathbf{R}) \to \mathbf{R}$, as in section 'Substrate structures alignment', above.

Similarly, the assumption of intermediate *underlying alignment*, $0 \ll \mathrm{algn}(A_{\mathrm{o}}) \ll \mathrm{algnMax}(U)(V_{\mathrm{o}}, z_{\mathrm{o}})$, is also ill-defined. Note, however, that the two assumptions are linked. Intermediate *underlying alignment* tends to imply at least intermediate *derived alignment*, $0 \ll \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}}) < \mathrm{algnMax}(U)(W_{\mathrm{o}}, z_{\mathrm{o}})$, because the *formal alignment* is zero, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}})^{\mathrm{X}}$. The relation is stronger as the *transform* tends to *full functional*.

In the discussion above, the *model*, $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, is *known*, and both the *formal*, $\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}$, and the *abstract*, $(\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$, are *necessary* and *known*. Optimisation can be done to find the *maximum likelihood estimate* of the *distribution histogram* for *known model*,

$$\{\tilde{E}_{\mathrm{o}}\} = \mathrm{maxd}(\{(E, \hat{Q}_{\mathrm{m,y},T_{\mathrm{o}},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1}\})$$

Just as in the discussion above of *classical modelled induction*, consider the case where both the *formal*, $\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{\mathrm{o}}$, and the *abstract*, $(\hat{E}_{\mathrm{h}} * T_{\mathrm{o}})^{\mathrm{X}}$, are still *necessary* but the *model*, T_{o}, is *unknown* and so both the *formal* and the *abstract* are *unknown*. Again, the *maximum likelihood estimate* for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$ can be defined as an optimisation of the *multinomial probability* conditional on the *iso-transform-independents* where both the *distribution histogram* and *transform* are treated as arguments to a likelihood function,

$$\begin{aligned} &\{(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})\} \\ &= \mathrm{maxd}(\{((E,T), \frac{Q_{\mathrm{m},U}(E, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i,y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})} Q_{\mathrm{m},U}(E, z_{\mathrm{o}})(B)}) : \\ &\qquad\qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ T \in \mathcal{T}_{U,V_{\mathrm{o}}}\}) \end{aligned}$$

It is conjectured that in *transform induction* there are some cases in which there is a unique solution for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$. This is because in *entity-like induction*, but not *law-like induction*, the denominator does not necessarily reduce to equal the numerator, so avoiding degeneracy. In the case where there is a unique solution then the maximisation can be rewritten in terms of the *transform-dependent*,

$$\{\tilde{T}_{\mathrm{o}}\} \quad = \quad \mathrm{maxd}(\{(T, \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T)}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})} Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T)}, z_{\mathrm{o}})(B)}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

The *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, is sometimes not computable because the *transform-dependent*, $A_{\mathrm{o}}^{\mathrm{Y}(\tilde{T}_{\mathrm{o}})}$, is sometimes not computable. A finite approximation to arbitrary accuracy for the *transform-dependent*, $A_k^{\mathrm{Y}(T)} \approx A^{\mathrm{Y}(T)}$, is computable. However, even an approximation is not *tractable*. The *formal-abstract* pair valued function, $Y_{U,\mathrm{i},T,z} \in \mathcal{A}_{U,\mathrm{i},V,z} :\to (\mathcal{A}_{U,W,z} \times \mathcal{A}_{U,W,z})$, is *intractable* because its computation requires the *intractable* computation of its domain of the *substrate histograms*, $\mathcal{A}_{U,\mathrm{i},V,z}$.

In *transform induction*, where the *history probability function* is *iso-transform-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}$, given some *unknown substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, in some cases the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is non-trivial,

$$\tilde{T}_{\mathrm{o}} \quad \notin \quad \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$$

Note that the optimisation is not the same as the optimisation of the *iso-transform-independent conditional generalised multinomial probability*,

$$\mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{m},\mathrm{y},T,U}(A_{\mathrm{o}}^{\mathrm{Y}(T)}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

because the normalising factor, $1/|\mathrm{ran}(Y_{U,\mathrm{i},T,z_{\mathrm{o}}})|$, implies *uniform possible induction* rather than *necessary induction*. Recall that in *classical induction* the *maximum likelihood estimate* for the *model* in the *necessary derived* case is degenerate. Only in the *uniform possible* case are there non-trivial solutions. In *abstract induction*, however, there are non-trivial solutions where the condition is *necessary abstract*. In *aligned induction*, there are non-trivial solutions where the condition is *necessary abstract* and *necessary formal*.

Consider how an approximation to the optimisation may be made more *tractable*. It is conjectured in section 'Likely histograms', above, that the *log-likelihood* with respect to the *dependent-analogue* varies with the *relative space* with respect to the *independent-analogue*,

$$\ln \frac{Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(A)}{\sum Q_{\mathrm{m},U}(A^{\mathrm{Y}(T)}, z)(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z}(A)} \quad \sim \quad \mathrm{spaceRelative}(A^{\mathrm{X}(T)})(A)$$

and conjectured further in section 'Transform alignment', above, that the *relative space* with respect to the *transform-independent* varies with the *derived alignment*,

$$\text{spaceRelative}(A^{\mathrm{X}(T)})(A) \quad \sim \quad \text{algn}(A * T)$$

This correlation was decomposed in the discussion above into three separate correlations, (i) the *dependent-independent anti-correlation*, (ii) the *underlying-lifted correlation* and (iii) the *conditional-relative correlation*. Now consider how the optimisation of the terms of these relations may form the definition of *induction* assumptions.

The *maximum likelihood estimate* for the *unknown model*, $\tilde{T}_{\mathrm{o}}$, with respect to the *dependent-analogue* is

$$\{\tilde{T}_{\mathrm{o}}\} = \text{maxd}(\{(T, \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T)}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T)}, z_{\mathrm{o}})(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

First, given the *dependent-independent anti-correlation*, assume that the *maximum likelihood estimate* of the *iso-transform-independent conditional multinomial probability* with respect to the *dependent-analogue* or *transform-dependent*, $A_{\mathrm{o}}^{\mathrm{Y}(T)}$, is also the *minimum likelihood estimate* of the *iso-transform-independent conditional multinomial probability* with respect to the *independent-analogue* or *transform-independent*, $A_{\mathrm{o}}^{\mathrm{X}(T)}$,

$$\{\tilde{T}_{\mathrm{o}}\} = \text{mind}(\{(T, \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{X}(T)}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{X}(T)}, z_{\mathrm{o}})(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

This assumption is the *iso-transform-independent dependent-independent anti-optimisation assumption*. It relies on the monotonicity of the *dependent-independent anti-correlation*.

Second, given the *underlying-lifted correlation*, assume that the *minimum likelihood estimate* of the *iso-transform-independent conditional multinomial probability* with respect to the *independent-analogue* or *transform-independent*, $A_{\mathrm{o}}^{\mathrm{X}(T)}$, is also the *minimum likelihood estimate* of the *lifted iso-transform-independent conditional multinomial probability* with respect to the *lifted independent-analogue* or *transform-independent derived*, $A_{\mathrm{o}}^{\mathrm{X}(T)} * T$,

$$\{\tilde{T}_{\mathrm{o}}\} = \text{mind}(\{(T, \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{X}(T)} * T, z_{\mathrm{o}})(A_{\mathrm{o}} * T)}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{X}(T)} * T, z_{\mathrm{o}})(B') : B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

This assumption is the *iso-transform-independent underlying-lifted optimisation assumption*. It relies on the monotonicity of the *underlying-lifted correlation*.

Third, given the *conditional-relative correlation*, assume that the *minimum likelihood estimate* of the *lifted iso-transform-independent conditional multinomial probability* with respect to the *lifted independent-analogue* or *transform-independent derived*, $A_{\mathrm{o}}^{\mathrm{X}(T)} * T$, is also the *minimum likelihood estimate* of the *relative multinomial probability* with respect to the *abstract*, $(A_{\mathrm{o}} * T)^{\mathrm{X}}$,

$$\{\tilde{T}_{\mathrm{o}}\} \quad = \quad \mathrm{mind}(\{(T, \frac{Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}}, z_{\mathrm{o}})(A_{\mathrm{o}} * T)}{Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}}, z_{\mathrm{o}})((A_{\mathrm{o}} * T)^{\mathrm{X}})}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

The negative logarithm of the *relative multinomial probability* is the *relative space* of the *derived* with respect to the *abstract*, which is the *derived alignment*,

$$\begin{aligned} -\ln \frac{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)(A * T)}{Q_{\mathrm{m},U}((A * T)^{\mathrm{X}}, z)((A * T)^{\mathrm{X}})} &= \mathrm{spaceRelative}((A * T)^{\mathrm{X}})(A * T) \\ &= \mathrm{algn}(A * T) \end{aligned}$$

So the third assumption is that the *minimum likelihood estimate* of the *lifted iso-transform-independent conditional multinomial probability* with respect to the *transform-independent derived*, $A_{\mathrm{o}}^{\mathrm{X}(T)} * T$, is also the *maximum likelihood estimate* for the *derived alignment*,

$$\{\tilde{T}_{\mathrm{o}}\} \quad = \quad \mathrm{maxd}(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

This assumption is the *iso-transform-independent conditional-relative optimisation assumption*. It relies on the monotonicity of the *conditional-relative correlation*.

A finite approximation to arbitrary accuracy of the *derived alignment*, $\mathrm{algn}(A_{\mathrm{o}} * T)$, is computable by means of an approximation to the gamma function. The computation of the *derived alignment* is *tractable* given limits on the *derived volume*, $|T^{-1}|$. So the optimisation of *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, at least for a limited subset of the *substrate transforms*, is *tractable*.

In *transform induction*, where (i) the *history probability function* is *iso-transform-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}$, given some *unknown substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *iso-transform-independent dependent-independent anti-optimisation assumption* is true, (iii) the *iso-transform-independent underlying-lifted optimisation assumption* is true, and (iv) the *iso-transform-independent* *conditional-relative optimisation assumption* is true, then the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is

$$\{\tilde{T}_{\mathrm{o}}\} \quad = \quad \mathrm{maxd}(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

It is shown in the *known transform* case above that the *maximum likelihood estimate* is more *classical* and not *formal* if (i) the *sample* is *idealised*, (ii) the *sample formal* equals the *sample abstract* and (iii) the *derived alignment* is high. This is the case for *unknown transform* too. In fact, if the three *iso-transform-independent optimisation assumptions* are true, then the *maximum likelihood estimate* for the *model*, $\tilde{T}_{\mathrm{o}}$, occurs at the maximisation of the *derived alignment*, implying that the *derived alignment* is as high as possible, $\forall T \in \mathcal{T}_{U,V_{\mathrm{o}}}\ ((A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}) \wedge (A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}) \implies \mathrm{algn}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}}) \geq \mathrm{algn}(A_{\mathrm{o}} * T))$.

Now the *independent analogue*, or *transform-independent*, equals the computable *naturalised abstract*,

$$A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}} \quad \implies \quad A^{\mathrm{X}(T)} = (A * T)^{\mathrm{X}} * T^{\dagger}$$

so the three *optimisation assumptions* are modified as follows:

The *maximum likelihood estimate* for the *unknown model*, $\tilde{T}_{\mathrm{o}}$, with respect to the *dependent-analogue* is

$$\begin{aligned}&\{\tilde{T}_{\mathrm{o}}\} = \\ &\quad \mathrm{maxd}(\{(T, \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T)}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T)}, z_{\mathrm{o}})(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})}) : \\ &\qquad\qquad T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}},\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}\})\end{aligned}$$

First, given the *dependent-independent anti-correlation*, assume that the *maximum likelihood estimate* of the *iso-transform-independent conditional multinomial probability* with respect to the *dependent-analogue* or *transform-dependent*, $A_{\mathrm{o}}^{\mathrm{Y}(T)}$, is also the *minimum likelihood estimate* of the *iso-transform-independent conditional multinomial probability* with respect to the *independent-analogue* or *naturalised abstract*, $(A_{\mathrm{o}} * T)^{\mathrm{X}} * T^{\dagger}$,

$$\begin{aligned}&\{\tilde{T}_{\mathrm{o}}\} = \\ &\quad \mathrm{mind}(\{(T, \frac{Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}} * T^{\dagger}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}} * T^{\dagger}, z_{\mathrm{o}})(B) : B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})}) : \\ &\qquad\qquad T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}},\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}\})\end{aligned}$$

This assumption is the *iso-transform-independent dependent-independent anti-optimisation assumption.* It relies on the monotonicity of the *dependent-independent anti-correlation.*

Second, given the *underlying-lifted correlation*, assume that the *minimum likelihood estimate* of the *iso-transform-independent conditional multinomial probability* with respect to the *independent-analogue* or *transform-independent*, $(A_{\mathrm{o}} * T)^{\mathrm{X}} * T^{\dagger}$, is also the *minimum likelihood estimate* of the *lifted iso-transform-independent conditional multinomial probability* with respect to the *lifted independent-analogue* or *abstract*, $(A_{\mathrm{o}} * T)^{\mathrm{X}}$,

$$\begin{aligned}
&\{\tilde{T}_{\mathrm{o}}\} = \\
&\quad \mathrm{mind}(\{(T, \frac{Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}}, z_{\mathrm{o}})(A_{\mathrm{o}} * T)}{\sum Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}}, z_{\mathrm{o}})(B') : B' \in \mathcal{A}'_{U,\mathrm{i},\mathrm{y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})}) : \\
&\qquad\qquad T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}},\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}\})
\end{aligned}$$

This assumption is the *iso-transform-independent underlying-lifted optimisation assumption.* It relies on the monotonicity of the *underlying-lifted correlation.*

Third, given the *conditional-relative correlation*, assume that the *minimum likelihood estimate* of the *lifted iso-transform-independent conditional multinomial probability* with respect to the *lifted independent-analogue* or *abstract*, $(A_{\mathrm{o}} * T)^{\mathrm{X}}$, is also the *minimum likelihood estimate* of the *relative multinomial probability* with respect to the *lifted independent-analogue* or *abstract*, $(A_{\mathrm{o}} * T)^{\mathrm{X}}$,

$$\begin{aligned}
&\{\tilde{T}_{\mathrm{o}}\} = \\
&\quad \mathrm{mind}(\{(T, \frac{Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}}, z_{\mathrm{o}})(A_{\mathrm{o}} * T)}{Q_{\mathrm{m},U}((A_{\mathrm{o}} * T)^{\mathrm{X}}, z_{\mathrm{o}})((A_{\mathrm{o}} * T)^{\mathrm{X}})}) : \\
&\qquad\qquad T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}},\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}\})
\end{aligned}$$

So the third assumption is that the *minimum likelihood estimate* of the *lifted iso-transform-independent conditional multinomial probability* with respect to the *abstract*, $(A_{\mathrm{o}} * T)^{\mathrm{X}}$, is also the *maximum likelihood estimate* for the *derived alignment*,

$$\begin{aligned}
\{\tilde{T}_{\mathrm{o}}\} \quad = \quad & \mathrm{maxd}(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) : \\
& \qquad T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}},\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}\})
\end{aligned}$$

In *transform induction*, where (i) the *history probability function* is *iso-transform-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}$, given some *unknown substrate transform* in the *sample variables* $T_{\mathrm{o}} \in \mathcal{T}_{U,V_{\mathrm{o}}}$, if it is the case that (ii) the *iso-transform-independent dependent-independent anti-optimisation assumption* is true, (iii) the *iso-transform-independent underlying-lifted optimisation assumption* is true, (iv) the *iso-transform-independent conditional-relative optimisation assumption* is true, (v) the *sample* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, (vi) the *sample formal* equals the *sample abstract*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, (vii) the *alignment* is at least intermediate, $\mathrm{algn}(A_{\mathrm{o}}) > \mathrm{algnMax}(U)(V_{\mathrm{o}}, z_{\mathrm{o}})/2$, (viii) the *distribution history size* is large with respect to the *sample size*, $z_{\mathrm{h}} \gg z_{\mathrm{o}}$, and such that (ix) the *scaled estimate distribution histogram* is *integral*, $\tilde{E}_{\mathrm{o},z_{\mathrm{h}}} \in \mathcal{A}_{\mathrm{i}}$, then (a) the *maximum likelihood estimate* of the *model*, $\tilde{T}_{\mathrm{o}}$, at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_{\mathrm{o}}$, is

$$\begin{aligned}\{\tilde{T}_{\mathrm{o}}\} \;=\; \mathrm{maxd}(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) :\\ T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}},\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}\})\end{aligned}$$

(b) the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *derived alignment*,

$$\ln \hat{Q}_{\mathrm{h,y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\; \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

(c) the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *sample*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) &\sim \ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})\\ &\sim \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})\end{aligned}$$

(d) so the *log likelihood* varies with the *size-volume* scaled *component size cardinality sum relative entropy*,

$$\begin{aligned}&\ln \hat{Q}_{\mathrm{h,y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\\ &\quad (z_{\mathrm{o}} + v_{\mathrm{o}}) \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}} + V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})\\ &\qquad\qquad - z_{\mathrm{o}} \times \mathrm{entropy}(A_{\mathrm{o}} * T_{\mathrm{o}}) \;- v_{\mathrm{o}} \times \mathrm{entropy}(V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}})\end{aligned}$$

(e) the *formal alignment* of the *maximum likelihood estimate* is zero,

$$\begin{aligned}\tilde{E}_{\mathrm{o}}^{\mathrm{X}} * \tilde{T}_{\mathrm{o}} &= \hat{A}_{\mathrm{o}}^{\mathrm{X}} * \tilde{T}_{\mathrm{o}}\\ &= (\hat{A}_{\mathrm{o}}^{\mathrm{X}} * \tilde{T}_{\mathrm{o}})^{\mathrm{X}}\\ &= (\tilde{E}_{\mathrm{o}}^{\mathrm{X}} * \tilde{T}_{\mathrm{o}})^{\mathrm{X}}\end{aligned}$$

and (f) the *derived* of the *maximum likelihood estimate* approximates to the *normalised sample derived*,

$$\tilde{E}_{\mathrm{o}} * \tilde{T}_{\mathrm{o}} \quad \approx \quad \hat{A}_{\mathrm{o}} * T_{\mathrm{o}}$$

In this case queries via the *model* of the *sample* approximate to queries via the *model* of the *distribution*,

$$(Q * \tilde{T}_{\mathrm{o}}^{\%} * \mathrm{his}(\tilde{T}_{\mathrm{o}}^{\%}) * A_{\mathrm{o}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K) \quad \approx \quad (Q * T_{\mathrm{o}}^{\%} * \mathrm{his}(T_{\mathrm{o}}^{\%}) * E_{\mathrm{h}})^{\wedge} \ \% \ (V_{\mathrm{o}} \setminus K)$$

That is, given *unknown model* where (i) the *underlying alignment* is intermediate, (ii) the *sample* is *known* to be *ideal* and (iii) the *sample formal* is *known* to be equal to the *sample abstract*, the maximisation of the *derived alignment* tends to make the properties of *aligned modelled induction* similar to those of *classical modelled induction*.

If, in addition, (x) the *component size cardinality relative entropy* of the *maximum likelihood estimate* for the *model* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, then the *sum sensitivity* varies against the *log-likelihood*,

$$\begin{aligned}\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad &\sim \quad - \ \ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \\ &\sim \quad - \ \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})\end{aligned}$$

so the *query sensitivity* to the *distribution histogram* is sometimes lower,

$$\begin{aligned}&\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}} \\ &\qquad\qquad \leq \mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})))/z_{\mathrm{o}}\end{aligned}$$

and the *model likelihood* of the *distribution histogram* is sometimes higher,

$$\hat{Q}_{\mathrm{h,y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{Q,N}) \quad \geq \quad \hat{Q}_{\mathrm{h},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z)(A_{Q,N})$$

If, further, (xi) the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, then the *sensitivity* to *model* also varies against the *log likelihood*,

$$\begin{aligned}- \ \ln |\max(\{(T, \hat{Q}_{\mathrm{h,y},T,U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}, \ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\})| \ \sim \\ - \ \ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})\end{aligned}$$

or

$$\begin{aligned}- \ \ln |\max(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}, \ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\})| \ \sim \\ - \ \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})\end{aligned}$$

So (a) by weakening the *induction* condition from *law-like necessary derived* to *entity-like necessary abstract*, (b) by strengthening the *induction* condition with *necessary formal* and (c) by strengthening the constraints on the *sample* to be *ideal* and to have *formal-abstract equivalence*, it is found that in some cases the *aligned modelled induction maximum likelihood estimate* of the *model* is non-trivial, $\tilde{T}_{\mathrm{o}} \notin \{T_{\mathrm{s}}, T_{\mathrm{u}}\}$, but retains properties of *classical modelled induction* such as allowing query via the *model*, minimising *sensitivity* to the *unknown underlying* and minimising *sensitivity* to the *model*. Furthermore, the optimisation is *tractable* depending on the limits on the searched subset of the *substrate transforms*.

Now consider the definition of *inducers* in the light of the preceding discussion. The set of *inducers* is defined in section 'Tractable alignment-bounding', and reviewed in section 'Induction'. The *inducers* are *computers* $I_z \in \mathrm{inducers}(z) \subset \mathrm{computers}$ such that (i) the domain is a set of *substrate histograms* which are at least a superset of the *integral-independent substrate histograms*, $\mathcal{A}_{z,\mathrm{xi}} \subseteq \mathrm{domain}(I_z) \subseteq \mathcal{A}_z$, (ii) the finite *time* and *space* application returns a rational-valued function of the *substrate models set*, $I_z^*(A) \in \mathcal{M}_{U_A,V_A} \to \mathbf{Q}$, and (iii) the maximum of the *inducer* application, $\mathrm{maxr} \circ I_z^*$, is positively correlated with the finite *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I_z^*) \geq 0)$$

where $\mathrm{cov}(z)(F,G) := \mathrm{covariance}(\hat{R}_z)(F,G)$ and the *renormalised geometry-weighted probability function* is $\hat{R}_z = \mathrm{normalise}(\{(A, 1/(|V_A|! \prod_{w \in V_A} |U_A(w)|!)) : A \in \mathrm{dom}(F)\})$. The set of *sized cardinal substrate histograms* $\mathcal{A}_z$ is defined,

$$\mathcal{A}_z = \{A : A \in \mathcal{A}_{\mathrm{c}} \cap \mathcal{A}_{\mathrm{i}},\ \mathrm{size}(A) = z,\ |V_A| \leq z,\ A^{\mathrm{U}} = A^{\mathrm{XF}} = A^{\mathrm{C}}\}$$

where $A^{\mathrm{CS}} = \mathrm{cartesian}(U_A)(V_A)$ and $U_A = \mathrm{implied}(\mathrm{implied}(A))$ and $V_A = \mathrm{vars}(A)$. The set of *integral-independent substrate histograms* $\mathcal{A}_{z,\mathrm{xi}}$ is defined,

$$\mathcal{A}_{z,\mathrm{xi}} = \{A : A \in \mathcal{A}_z,\ A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}\}$$

The *independent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*, also known as the *alignment-bounded iso-transform space ideal transform search set*, is defined $X_{z,\mathrm{xi,T,y,fa,j}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to$

$\ln \mathbf{Q}_{>0})$ as

$$\begin{aligned}
X_{z,\mathrm{xi,T,y,fa,j}}(A) \ = \\
\{(T, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\sum \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B) : B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)}) : \\
T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}
\end{aligned}$$

The *maximum likelihood estimate* for the *model* is,

$$\begin{aligned}
\{\tilde{T}\} &= \mathrm{maxd}(X_{z,\mathrm{xi,T,y,fa,j}}(A)) \\
&= \mathrm{mind}(\{(T, e^{-x}) : (T, x) \in X_{z,\mathrm{xi,T,y,fa,j}}(A)\}) \\
&= \mathrm{mind}(\{(T, \frac{Q_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)}{\sum Q_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(B) : B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)}) : \\
&\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\})
\end{aligned}$$

This may be compared to the case in *aligned modelled induction* where the *iso-transform-independent dependent-independent anti-optimisation assumption* is true, the *histogram* is *ideal*, $A = A * T * T^{\dagger A}$, and the *formal* equals the *abstract*, $A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}$,

$$\begin{aligned}
\{\tilde{T}\} \ = \\
\mathrm{mind}(\{(T, \frac{Q_{\mathrm{m},U_A}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(A)}{\sum Q_{\mathrm{m},U_A}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(B) : B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)}) : \\
T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\})
\end{aligned}$$

The *independent-analogue* is the *naturalised abstract* which equals the *naturalised formal*, $(A * T)^{\mathrm{X}} * T^{\dagger} = A^{\mathrm{X}} * T * T^{\dagger}$. As the *transform*, T, tends to *full functional*, $T_{\mathrm{f}} = \{\{w\}^{\mathrm{CS}\{\}V_A\mathrm{T}} : w \in V_A\}^{\mathrm{T}}$, the *naturalised abstract* tends to the *independent*, $(A * T_{\mathrm{f}})^{\mathrm{X}} * T_{\mathrm{f}}^{\dagger} = A^{\mathrm{X}} * T_{\mathrm{f}} * T_{\mathrm{f}}^{\dagger} = A^{\mathrm{X}}$. This suggests a revised definition of *inducers*. The *naturalised-abstract-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*, is defined $X_{z,\mathrm{xi,T,yx,fa,j}} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$ as

$$\begin{aligned}
X_{z,\mathrm{xi,T,yx,fa,j}}(A) \ = \\
\{(T, -\ln \frac{\hat{Q}_{\mathrm{m},U_A}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(A)}{\sum \hat{Q}_{\mathrm{m},U_A}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(B) : B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)}) : \\
T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}
\end{aligned}$$

Here the *independent*, A^{X}, is replaced by the *independent-analogue*, which is the *naturalised abstract*, $(A*T)^{\mathrm{X}}*T^{\dagger}$, so the numerator is no longer constant,

$\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}}, z)(A)$, but depends on the *model*, $\hat{Q}_{\mathrm{m},U_A}((A * T)^{\mathrm{X}} * T^{\dagger}, z)(A)$. In this revised definition of *inducers*, which is more consistent with *ideal formal-abstract aligned induction*, the maximum of the *inducer* application, $\mathrm{maxr} \circ I_z^*$, is instead constrained to be positively correlated with the finite *naturalised-abstract-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yx,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yx,fa,j}}, \mathrm{maxr} \circ I_z^*) \geq 0)$$

In section 'Substrate structures alignment', above, it is conjectured that the *alignment-bounded iso-transform space ideal transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}$, is correlated with the *derived alignment integral-independent substrate ideal formal-abstract transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}) \geq 0)$$

where the *derived alignment integral-independent substrate ideal formal-abstract transform search set* is defined,

$$\begin{aligned} X'_{z,\mathrm{xi,T,a,fa,j}}(A) &= \{(T, \mathrm{algn}(A * T)) : \\ &\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\} \end{aligned}$$

As shown above, in 'Derived alignment and conditional probability', given the *minimum alignment conjecture*, the *alignment-bounded lifted iso-transform space* is bounded

$$\begin{aligned} &\mathrm{algn}(A * T) \\ \leq\ &\Bigg(-\ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(A * T)}{\sum \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{X}} * T, z)(B') : B' \in \mathcal{A}'_{U_A,\mathrm{i,y},T,z}(A)} : \\ &\qquad\qquad A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}}\Bigg) \\ \leq\ &\mathrm{algn}(A * T) + \ln |\mathcal{A}'_{U_A,\mathrm{i,y},T,z}(A)| \end{aligned}$$

The *alignment-bounded iso-transform space* is functionally related to the *alignment-bounded lifted iso-transform space*, although it is not always the case that the *alignment-bounded iso-transform space* is bounded by the *derived alignment*, $\mathrm{algn}(A*T)$, so, strictly speaking, it is a misnomer. The correlation between the *alignment-bounded iso-transform space ideal transform* *maximum function*, $\text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa},\text{j}}$, and the *derived alignment integral-independent substrate ideal formal-abstract transform maximum function*, $\text{maxr} \circ X'_{z,\text{xi},\text{T},\text{a},\text{fa},\text{j}}$, is conjectured nonetheless.

Similarly, for the revised definition of *inducers*, conjecture that the *naturalised-abstract-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform maximum function*, $\text{maxr} \circ X_{z,\text{xi},\text{T},\text{yx},\text{fa},\text{j}}$, is also correlated with the *derived alignment integral-independent substrate ideal formal-abstract transform maximum function*, $\text{maxr} \circ X'_{z,\text{xi},\text{T},\text{a},\text{fa},\text{j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\text{cov}(z)(\text{maxr} \circ X_{z,\text{xi},\text{T},\text{yx},\text{fa},\text{j}}, \text{maxr} \circ X'_{z,\text{xi},\text{T},\text{a},\text{fa},\text{j}}) \geq 0)$$

It is not obvious, however, whether this correlation is greater than that for the existing *inducer* definition,

$$\begin{aligned}\text{cov}(z)(\text{maxr} \circ X_{z,\text{xi},\text{T},\text{yx},\text{fa},\text{j}}, \text{maxr} \circ X'_{z,\text{xi},\text{T},\text{a},\text{fa},\text{j}}) \ &\geq \\ &\text{cov}(z)(\text{maxr} \circ X_{z,\text{xi},\text{T},\text{y},\text{fa},\text{j}}, \text{maxr} \circ X'_{z,\text{xi},\text{T},\text{a},\text{fa},\text{j}})\end{aligned}$$

It is possible to go a step further and drop the *iso-transform-independent dependent-independent anti-optimisation assumption*. Now the *maximum likelihood estimate* for the *unknown model*, $\tilde{T}$, is with respect to the *dependent-analogue*, which is the *transform-dependent*, $A^{\text{Y}(T)}$,

$$\begin{aligned}\{\tilde{T}\} \ &= \\ &\text{maxd}(\{(T, \frac{Q_{\text{m},U_A}(A^{\text{Y}(T)}, z)(A)}{\sum Q_{\text{m},U_A}(A^{\text{Y}(T)}, z)(B) : B \in \mathcal{A}_{U_A,\text{i},\text{y},T,z}(A)}) : \\ &\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\text{X}} * T = (A * T)^{\text{X}},\ A = A * T * T^{\dagger A}\})\end{aligned}$$

The corresponding *transform-dependent-sample-distributed iso transform independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*, is defined $X_{z,\text{xi},\text{T},\text{yy},\text{fa},\text{j}} \in \mathcal{A}_{z,\text{xi}} \to (\mathcal{T}_{\text{f}} \to \ln \mathbf{Q}_{>0})$ as

$$\begin{aligned}X_{z,\text{xi},\text{T},\text{yy},\text{fa},\text{j}}(A) \ &= \\ &\{(T, \ln \frac{\hat{Q}_{\text{m},U_A}(A^{\text{Y}(T)}, z)(A)}{\sum \hat{Q}_{\text{m},U_A}(A^{\text{Y}(T)}, z)(B) : B \in \mathcal{A}_{U_A,\text{i},\text{y},T,z}(A)}) : \\ &\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\text{X}} * T = (A * T)^{\text{X}},\ A = A * T * T^{\dagger A}\}\end{aligned}$$

In this further revision of the definition of *inducers*, the maximum of the *inducer* application, $\text{maxr} \circ I_z^*$, is constrained to be positively correlated with the *transform-dependent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yy,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yy,fa,j}}, \mathrm{maxr} \circ I_z^*) \geq 0)$$

Again, for the further revision of the definition of *inducers*, conjecture that the *transform-dependent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal abstract transform maximum function*, $\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yy,fa,j}}$, is also correlated with the *derived alignment integral-independent substrate ideal formal-abstract transform maximum function*, $\mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yy,fa,j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}}) \geq 0)$$

Conjecture, however, that the *dependent-analogue* correlation is less than that for the *independent-analogue*,

$$\begin{aligned}\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yy,fa,j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}})\ \leq \\ \mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yx,fa,j}}, \mathrm{maxr} \circ X'_{z,\mathrm{xi,T,a,fa,j}})\end{aligned}$$

Note that the *transform-dependent*, $A^{\mathrm{Y}(T)}$, is sometimes not computable. The finite approximation to the *transform-dependent* is

$$\begin{aligned}&\{A_k^{\mathrm{Y}(T)}\} = \\ &\quad \mathrm{maxd}(\{(D/Z_k, \frac{Q_{\mathrm{m},U_A}(D,z)(A)}{\sum Q_{\mathrm{m},U_A}(D,z)(B) : B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)}) : D \in \mathcal{A}_{U_A,\mathrm{i},V_A,kz}\})\end{aligned}$$

A literal finite approximation for an *inducer* for the *transform-dependent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*, $X_{z,\mathrm{xi,T,yy,fa,j}}(A)$, requires the finite approximation to the *transform-dependent*, $A_k^{\mathrm{Y}(T)}$. Define the *finite transform-dependent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*,

$$\begin{aligned}&X_{z,\mathrm{xi,T,yy,fa,j},k}(A)\ = \\ &\quad \{(T, \ln \frac{\hat{Q}_{\mathrm{m},U_A}(A_k^{\mathrm{Y}(T)},z)(A)}{\sum \hat{Q}_{\mathrm{m},U_A}(A_k^{\mathrm{Y}(T)},z)(B) : B \in \mathcal{A}_{U_A,\mathrm{i,y},T,z}(A)}) : \\ &\qquad\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}\end{aligned}$$

Let the *literal finite transform-dependent-sample-distributed iso-transform-independent conditional dependent multinomial space ideal formal-abstract* *transform inducer* $I_{z,\mathrm{yy,l},k} \in \mathrm{inducers}(z)$ be a literal implementation of the *transform-dependent-sample-distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal abstract transform search set*, $X_{z,\mathrm{xi,T,yy,fa,j},k} \in \mathcal{A}_{z,\mathrm{xi}} \to (\mathcal{T}_{\mathrm{f}} \to \ln \mathbf{Q}_{>0})$,

$$\forall A \in \mathcal{A}_{z,\mathrm{xi}}\ (I^{*}_{z,\mathrm{yy,l},k}(A) = \{(T, I^{*}_{\approx\ln\mathbf{Q}}(y)) : (T,y) \in X_{z,\mathrm{xi,T,yy,fa,j},k}(A)\})$$

In *ideal formal-abstract transform induction* the *maximum likelihood estimate* for the *unknown model*, $\tilde{T}_{\mathrm{o}}$, with respect to the *dependent-analogue* is

$$\begin{aligned}\{\tilde{T}_{\mathrm{o}}\} = \\ &\mathrm{maxd}(\{(T, \frac{Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T)}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o}}^{\mathrm{Y}(T)}, z_{\mathrm{o}})(B) : B \in \mathcal{A}_{U,\mathrm{i,y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})}) : \\ &\qquad T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}},\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}}\})\end{aligned}$$

The computation of the *dependent-analogue*, $A^{\mathrm{Y}(T)}$, is sometimes incomputable, but even with a computable approximation, $A_k^{\mathrm{Y}(T)}$, in the *literal finite transform-dependent-sample-distributed iso-transform-independent conditional dependent multinomial space ideal formal-abstract transform inducer*, $I_{z,\mathrm{yy,l},k}$,

$$\tilde{T}_{\mathrm{o}} \ \in\ \mathrm{maxd}(I^{*}_{z_{\mathrm{o}},\mathrm{yy,l},k}(A_{\mathrm{o}}))$$

in the case where $A_{\mathrm{o}} \in \mathcal{A}_{z_{\mathrm{o}},\mathrm{xi}}$, the computation of the *maximum likelihood estimate* for the *unknown model*, $\tilde{T}_{\mathrm{o}}$, remains intractable.

The *literal derived alignment integral-independent substrate ideal formal abstract transform inducer* $I'_{z,\mathrm{a,l}} \in \mathrm{inducers}(z)$ is a literal finite approximation to the *derived alignment integral-independent substrate ideal formal-abstract transform search set*, $X'_{z,\mathrm{xi,T,a,fa,j}}(A)$,

$$\begin{aligned}I'^{*}_{z,\mathrm{a,l}}(A) &= \{(T, I^{*}_{\approx\ln\mathbf{Q}}(\mathrm{algn}(A * T))) : \\ &\qquad T \in \mathcal{T}_{U_A,V_A},\ A^{\mathrm{X}} * T = (A * T)^{\mathrm{X}},\ A = A * T * T^{\dagger A}\}\end{aligned}$$

The *induction correlation* of the *literal derived alignment inducer* is conjectured to be positive, regardless of the definition of *inducers*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,y,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}) \geq 0)$$

and

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yx,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}) \geq 0)$$

and

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yy,fa,j}}, \mathrm{maxr} \circ I_{z,\mathrm{a,l}}^{'*}) \geq 0)$$

Although the *literal derived alignment inducer*, $I_{z,\mathrm{a,l}}^{'}$, is faster than the *literal finite transform-dependent-sample-distributed iso-transform-independent conditional dependent multinomial space ideal formal-abstract transform inducer*, $I_{z,\mathrm{yy,l},k}$, it is also intractable unless some limits are imposed on the *substrate models*.

Section 'Tractable alignment-bounding' discusses the various intractabilities and the classes of limits and constraints on the structures of more tractable *inducers*. There the *tractable limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer* is defined as an *inducer*,

$$I_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}^{'} \in \mathrm{inducers}(z)$$

Given *non-independent substrate histogram* $A \in \mathcal{A}_z \setminus \{A^{\mathrm{X}}\}$, the *midising, idealising fud decomposition inducer* is defined,

$$\begin{aligned}
&I_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}^{'*}(A) = \\
&\quad \{(D, I_{\approx\mathbf{R}}^{*}(\mathrm{algnValDensSum}(U_A)(A, D^{\mathrm{D}}))) : \\
&\qquad\qquad D \in \mathcal{D}_{\mathrm{F},\infty,U_A,V_A} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})), \\
&\qquad\qquad\qquad \forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A * C * F^{\mathrm{T}}) > 0)\}
\end{aligned}$$

where (i) the *limited-models fuds*, $\mathcal{F}_{\mathrm{q}}$ is the intersection of *limited-breadth*, *limited-layer*, *limited-underlying* and *limited-derived fuds*, $\mathcal{F}_{\mathrm{q}} = \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{d}} \cap \mathcal{F}_{\mathrm{h}} \cap \mathcal{F}_{\mathrm{b}}$, (ii) $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$, (iii) $()^{\mathrm{D}} \in \mathcal{D}_{\mathrm{F}} \to \mathcal{D}$, and (iv) the *summed derived alignment valency density* $\mathrm{algnValDensSum}(U) \in \mathcal{A} \times \mathcal{D} \to \mathbf{R}$ is defined as

$$\begin{aligned}
&\mathrm{algnValDensSum}(U)(A, D) := \\
&\qquad \sum_{(C,T) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * T)/\mathrm{capacityValency}(U)((A * C * T)^{\mathrm{FS}})
\end{aligned}$$

The maximum of the *fud decomposition*, $\max(I_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}^{'*}(A))$, is obtained by searching for the *fud decomposition* $D \in \mathcal{D}_{\mathrm{F}}$ which maximises the *summed alignment valency-density*,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \quad = \sum_{(C,F) \in \mathrm{cont}(D)} \mathrm{algn}(A * C * F^{\mathrm{T}})/w_F^{1/m_F}$$

where $W_F = \mathrm{der}(F)$, $w_F = |W_F^{\mathrm{C}}|$ and $m_F = |W_F|$.

Section 'Tractable alignment-bounding' shows that the *limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is tractable in all respects. There is a summary of the removal of intractabilities in section 'Inducers'.

It is conjectured that the *summed alignment valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is positively correlated with the *literal derived alignment inducer*, $I'_{z,\mathrm{a,l}}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I'^{*}_{z,\mathrm{a,l}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}) \geq 0)$$

So the *summed alignment inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is also positively correlated with the *literal finite transform-dependent-sample-distributed iso-transform-independent conditional dependent multinomial space ideal formal-abstract transform inducer*, $I_{z,\mathrm{yy,l},k}$,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ I^{*}_{z,\mathrm{yy,l},k}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}) \geq 0)$$

and positively correlated with the *finite transform-dependent-sample distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract transform search set*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yy,fa,j},k}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}) \geq 0)$$

Conjecture that, in the case where the *model*, T_{o}, is *unknown*, the *maximum likelihood estimate* for the *model* for *transform induction*,

$$\tilde{T}_{\mathrm{o}} \ \in\ \mathrm{maxd}(I^{*}_{z_{\mathrm{o}},\mathrm{yy,l},k}(A_{\mathrm{o}}))$$

can be tractably approximated by the maximisation of the *tractable limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$,

$$\tilde{T}_{\mathrm{o}} \ \approx\ D^{\mathrm{T}}_{\mathrm{o,Sd}}$$

where

$$D_{\mathrm{o,Sd}} \ \in\ \mathrm{maxd}(I'^{*}_{z_{\mathrm{o}},\mathrm{Sd,D,F},\infty,\mathrm{n,q}}(A_{\mathrm{o}}))$$

and $A_{\mathrm{o}} \neq A^{\mathrm{X}}_{\mathrm{o}}$. The *tractable model*, $D_{\mathrm{o,Sd}}$, is defined explicitly,

$$\begin{aligned} D_{\mathrm{o,Sd}} \ \in\ \mathrm{maxd}(\{(D, I^{*}_{\approx\mathbf{R}}(\mathrm{algnValDensSum}(U)(A_{\mathrm{o}}, D^{\mathrm{D}}))) :\ & \\ D \in \mathcal{D}_{\mathrm{F},\infty,U,V_{\mathrm{o}}} \cap \mathrm{trees}(\mathcal{S} \times (\mathcal{F}_{\mathrm{n}} \cap \mathcal{F}_{\mathrm{q}})),\ & \\ \forall (C,F) \in \mathrm{cont}(D)\ (\mathrm{algn}(A_{\mathrm{o}} * C * F^{\mathrm{T}}) > 0)\}) & \end{aligned}$$

The *accuracy* of the approximation can be defined as the ratio of the *tractable model transform likelihood* to the *maximum model transform likelihood*,

$$0 \;<\; \frac{\exp(X_{z_{\rm o},\rm xi,T,yy,fa,j},k(A_{\rm o})(D^{\rm T}_{\rm o,Sd}))}{\exp(X_{z_{\rm o},\rm xi,T,yy,fa,j},k(A_{\rm o})(\tilde{T}_{\rm o}))} \;\leq\; 1$$

or, in terms of the *literal inducer*,

$$0 \;<\; \frac{I^{*}_{\rm exp}(I^{*}_{z_{\rm o},\rm yy,l},k(A_{\rm o})(D^{\rm T}_{\rm o,Sd}))}{I^{*}_{\rm exp}(I^{*}_{z_{\rm o},\rm yy,l},k(A_{\rm o})(\tilde{T}_{\rm o}))} \;\leq\; 1$$

The *accuracy* is defined explicitly,

$$\frac{Q_{{\rm m},U}(A_{{\rm o},k}^{{\rm Y}(D^{\rm T}_{\rm o,Sd})}, z_{\rm o})(A_{\rm o})}{\sum Q_{{\rm m},U}(A_{{\rm o},k}^{{\rm Y}(D^{\rm T}_{\rm o,Sd})}, z_{\rm o})(B) : B \in \mathcal{A}_{U,{\rm i},{\rm y},D^{\rm T}_{\rm o,Sd},z_{\rm o}}(A_{\rm o})}$$
$$/ \quad \frac{Q_{{\rm m},U}(A_{{\rm o},k}^{{\rm Y}(\tilde{T}_{\rm o})}, z_{\rm o})(A_{\rm o})}{\sum Q_{{\rm m},U}(A_{{\rm o},k}^{{\rm Y}(\tilde{T}_{\rm o})}, z_{\rm o})(B) : B \in \mathcal{A}_{U,{\rm i},{\rm y},\tilde{T}_{\rm o},z_{\rm o}}(A_{\rm o})}$$

The *accuracy* is computable, though not tractable and so not necessarily practicable.

In the case where (i) the *iso-transform-independent dependent-independent anti-optimisation assumption* is true, (ii) the *iso-transform-independent underlying lifted optimisation assumption* is true, and (iii) the *iso-transform-independent conditional-relative optimisation assumption* is true, then the *maximum likelihood estimate* for the *model* can be obtained from the *literal derived alignment inducer*, $I'_{z,\rm a,l}$,

$$\tilde{T}_{\rm o} \;\in\; \mathrm{maxd}(I'^{*}_{z_{\rm o},\rm a,l}(A_{\rm o}))$$

and a definition of *accuracy* can be made in terms of *derived alignment*,

$$\frac{I^{*}_{\rm exp}(I'^{*}_{z_{\rm o},\rm a,l}(A_{\rm o})(D^{\rm T}_{\rm o,Sd}))}{I^{*}_{\rm exp}(I'^{*}_{z_{\rm o},\rm a,l}(A_{\rm o})(\tilde{T}_{\rm o}))} \;=\; \frac{I^{*}_{\rm exp}(I^{*}_{\rm a}(A_{\rm o} * D^{\rm T}_{\rm o,Sd}))}{I^{*}_{\rm exp}(I^{*}_{\rm a}(A_{\rm o} * \tilde{T}_{\rm o}))}$$

In the case of *integral independent*, $A^{\rm X} \in \mathcal{A}_{\rm i}$, the exponential of the *alignment* is rational, $\exp(\mathrm{algn}(A)) \in \mathbf{Q}_{\geq 0}$, and there is no need for numeric approximation. In this case, the *derived alignment accuracy* is the exponential of the difference in *derived alignments*,

$$0 \;<\; \frac{\exp(\mathrm{algn}(A_{\rm o} * D^{\rm T}_{\rm o,Sd}))}{\exp(\mathrm{algn}(A_{\rm o} * \tilde{T}_{\rm o}))} \;\leq\; 1$$

This definition of *accuracy* is consistent with the gradient of the likelihood function at the mode, so the *derived alignment accuracy* varies against the *sensitivity* to *model*,

$$\begin{aligned}\frac{\exp(\mathrm{algn}(A_{\mathrm{o}} * D_{\mathrm{o,Sd}}^{\mathrm{T}}))}{\exp(\mathrm{algn}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}}))} &\sim \\ &-(-\ln|\max(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \\ &\qquad A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\})|)\end{aligned}$$

If the *alignment* is at least intermediate, $\mathrm{algn}(A_{\mathrm{o}}) > \mathrm{algnMax}(U)(V_{\mathrm{o}}, z_{\mathrm{o}})/2$, then the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *sample*,

$$\begin{aligned}\ln \hat{Q}_{\mathrm{h,y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) &\sim \ln \hat{Q}_{\mathrm{h},\dagger,T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \\ &\sim \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})\end{aligned}$$

If, in addition, the *component size cardinality relative entropy* of the *maximum likelihood estimate* for the *model* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, then the *sum sensitivity* varies against the *log-likelihood*,

$$\begin{aligned}\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h,y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) &\sim -\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \\ &\sim -\mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})\end{aligned}$$

If, further, the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, then the *sensitivity* to *model* also varies against the *log likelihood*,

$$\begin{aligned}-\ \ln|\max(\{(T, \mathrm{algn}(A_{\mathrm{o}} * T)) &: T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}},&\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\})| \\ &\sim -\ \ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \\ &\sim -\ \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})\end{aligned}$$

So, although the *maximum model derived alignment*, $\mathrm{algn}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}})$, appears in the denominator of the *derived alignment accuracy*, the *tractable model accuracy* in fact varies with the *derived alignment*,

$$\frac{\exp(\mathrm{algn}(A_{\mathrm{o}} * D_{\mathrm{o,Sd}}^{\mathrm{T}}))}{\exp(\mathrm{algn}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}}))} \sim \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

or

$$\mathrm{algn}(A_{\mathrm{o}} * D_{\mathrm{o,Sd}}^{\mathrm{T}}) - \mathrm{algn}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}}) \sim \mathrm{algn}(A_{\mathrm{o}} * T_{\mathrm{o}})$$

That is, although the *model* obtained from the *tractable summed alignment valency-density inducer* is merely an approximation, in the cases where the *log-likelihood* or *derived alignment* is high, and so both the *sensitivity* to *model* and the *sensitivity* to *distribution* are low, the approximation may be reasonably close nonetheless.

Consider the *practicable model* obtained by maximisation of the *summed shuffle content alignment valency-density* of the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*,

$$I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}} \in \mathrm{inducers}(z)$$

Given *substrate histogram* $A \in \mathcal{A}_z$, the *practicable fud decomposition inducer* is defined in section 'Optimisation', above, as

$$\begin{aligned}
&I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}(A) = \\
&\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A,D)))\}, \{(D_{\emptyset},0)\}) : \\
&\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D,F,d}})),\ \{D\} = Q
\end{aligned}$$

Let the *practicable fud decomposition* be

$$D_{\mathrm{o,Scsd},P} \in \mathrm{maxd}(I'^{*}_{z_{\mathrm{o}},\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}(A_{\mathrm{o}}))$$

The *practicable fud decomposition inducer* imposes a sequence on the search and other constraints that do not apply to the *tractable summed alignment valency-density decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, so conjecture that the *practicable derived alignment* is less than or equal to the *tractable derived alignment*,

$$\mathrm{algn}(A_{\mathrm{o}} * D^{\mathrm{T}}_{\mathrm{o,Scsd},P}) \ \leq\ \mathrm{algn}(A_{\mathrm{o}} * D^{\mathrm{T}}_{\mathrm{o,Sd}})$$

So conjecture, in the case where (i) the *iso-transform-independent dependent-independent anti-optimisation assumption* is true, (ii) the *iso-transform independent underlying lifted optimisation assumption* is true, and (iii) the *iso-transform-independent conditional-relative optimisation assumption* is true, that the *derived alignment accuracy* with respect to the *practicable fud decomposition inducer* is less than or equal to that of the *tractable fud decomposition inducer*,

$$\frac{\exp(\mathrm{algn}(A_{\mathrm{o}} * D^{\mathrm{T}}_{\mathrm{o,Scsd},P}))}{\exp(\mathrm{algn}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}}))} \ \leq\ \frac{\exp(\mathrm{algn}(A_{\mathrm{o}} * D^{\mathrm{T}}_{\mathrm{o,Sd}}))}{\exp(\mathrm{algn}(A_{\mathrm{o}} * \tilde{T}_{\mathrm{o}}))}$$

and, in general, the *accuracy* is such that

$$\frac{\exp(X_{z_{\mathrm{o}},\mathrm{xi,T,yy,fa,j},k}(A_{\mathrm{o}})(D^{\mathrm{T}}_{\mathrm{o,Scsd},P}))}{\exp(X_{z_{\mathrm{o}},\mathrm{xi,T,yy,fa,j},k}(A_{\mathrm{o}})(\tilde{T}_{\mathrm{o}}))} \ \leq\ \frac{\exp(X_{z_{\mathrm{o}},\mathrm{xi,T,yy,fa,j},k}(A_{\mathrm{o}})(D^{\mathrm{T}}_{\mathrm{o,Sd}}))}{\exp(X_{z_{\mathrm{o}},\mathrm{xi,T,yy,fa,j},k}(A_{\mathrm{o}})(\tilde{T}_{\mathrm{o}}))}$$

It is shown above in *classical uniform possible modelled induction*, where the *history probability function* is *uniform possible iso-derived historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}}$, that, in the case where (i) the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, but the *sample* approximates to the *naturalisation*, $A_{\mathrm{o}} \approx A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$, and (ii) the *maximum likelihood estimate relative entropy* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, (a) the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies against the *specialising derived substrate history coder space*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad -\ \mathrm{space}(C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T},\mathrm{H}}(T_{\mathrm{o}}))(H_{\mathrm{o}})$$

(b) the *sensitivity* to *distribution* varies against the *log likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad -\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

and (c) the *sensitivity* to *model* varies against the *log likelihood*,

$$\begin{aligned} -\ \ln |\max(\{(T, \hat{Q}_{\mathrm{h},\mathrm{d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})| \ \sim \\ -\ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \end{aligned}$$

It is shown above in *aligned modelled induction*, where the *history probability function* is *iso-transform-independent historically distributed*, $P = P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}}}$, that, in the case where (i) the *sample* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$, (ii) the *sample formal* equals the *sample abstract*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, (iii) the *alignment* is at least intermediate, $\mathrm{algn}(A_{\mathrm{o}}) > \mathrm{algnMax}(U)(V_{\mathrm{o}}, z_{\mathrm{o}})/2$, (iv) the *size* is less than the *volume*, $z_{\mathrm{o}} < v_{\mathrm{o}}$, and (v) the *component size cardinality relative entropy* of the *maximum likelihood estimate* for the *model* is high, $\mathrm{entropyCross}(A_{\mathrm{o}} * T_{\mathrm{o}}, V_{\mathrm{o}}^{\mathrm{C}} * T_{\mathrm{o}}) > \ln |T_{\mathrm{o}}^{-1}|$, (a) the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *sample*,

$$\ln \hat{Q}_{\mathrm{h},\mathrm{y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \quad \sim \quad \ln \hat{Q}_{\mathrm{h},\mathrm{d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

(b) the *sum sensitivity* varies against the *log-likelihood*,

$$\mathrm{sum}(\mathrm{sensitivity}(U)(\hat{Q}_{\mathrm{h},\mathrm{y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}}))) \quad \sim \quad -\ \ln \hat{Q}_{\mathrm{h},\mathrm{y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

and (c) the *sensitivity* to *model* varies against the *log likelihood*,

$$\begin{aligned} -\ \ln |\max(\{(T, \hat{Q}_{\mathrm{h},\mathrm{y},T,U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}, \\ A_{\mathrm{o}}^{\mathrm{X}} * T = (A_{\mathrm{o}} * T)^{\mathrm{X}},\ A_{\mathrm{o}} = A_{\mathrm{o}} * T * T^{\dagger A_{\mathrm{o}}}\})| \ \sim \\ -\ \ln \hat{Q}_{\mathrm{h},\mathrm{y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \end{aligned}$$

So (a) by weakening the *induction* condition from *law-like necessary derived* to *entity-like necessary abstract*, (b) by strengthening the *induction* condition with *necessary formal* and (c) by strengthening the constraints on the *sample* to be *ideal* and have *formal-abstract equivalence*, the *likelihood* and *sensitivity* properties of *aligned modelled induction* approximate to those of *classical modelled induction.*

Insofar as the *uniform possible iso-derived history probability function* approximates to the *necessary iso-transform-independent history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},\mathrm{p},T_{\mathrm{o}}} \approx P_{U,X,H_{\mathrm{h}},\mathrm{y},T_{\mathrm{o}},\mathrm{H}}$, conjecture that the *model*, $D_{\mathrm{o,Sd}}^{\mathrm{T}}$, obtained by the maximisation of the *tractable summed alignment valency-density inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is also a tractable approximation to the *maximum likelihood estimate* for the *model* for *uniform possible iso-derived induction*,

$$\tilde{T}_{\mathrm{o}} \in \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})$$

That is, in the *aligned*, *formal-abstract*, *ideal*, high *relative entropy* case, a tractable *maximum likelihood estimate* for the *model* may be obtained for *classical modelled induction* by optimisation of the *summed alignment valency-density inducer*,

$$\tilde{T}_{\mathrm{o}} \approx D_{\mathrm{o,Sd}}^{\mathrm{T}}$$

The *accuracy* of the approximation can be defined as the ratio of the *tractable model uniform possible iso-derived likelihood* to the *maximum model uniform possible iso-derived likelihood*,

$$0 < \frac{\hat{Q}_{\mathrm{h,d},D_{\mathrm{o,Sd}}^{\mathrm{T}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \leq 1$$

Just as the *tractable model iso-transform-independent accuracy* varies with the *log-likelihood*, so too does the *tractable model uniform possible iso-derived accuracy*,

$$\frac{\hat{Q}_{\mathrm{h,d},D_{\mathrm{o,Sd}}^{\mathrm{T}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \sim \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

Again, in the cases where the *log-likelihood* is high, and so the *sensitivity* to *model* is low, the tractable approximation in the *aligned*, *formal-abstract*, *ideal*, high *relative entropy* case may be reasonably close.

This positive correlation between the *tractable model uniform possible iso-derived accuracy* and the *log-likelihood*,

$$\frac{\hat{Q}_{\mathrm{h,d},D^{\mathrm{T}}_{\mathrm{o,Sd}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \;\sim\; \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

has already been established indirectly in section 'Tractable transform induction', above, by comparing the *entropy* properties of the *tractable summed alignment valency-density inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, and the *specialising fud decomposition substrate history coder*, $C_{\mathrm{G},V,\mathrm{D,F,H}}$. These properties are described in section 'Inducers and Compression' which considers the relations between the *summed alignment valency-density* and the *specialising space*. In particular, it is shown that the *summed alignment valency-density* (a) varies against the *derived entropy* of the *nullable transform*,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \;\sim\; -\ \mathrm{entropy}(A * D^{\mathrm{T}})$$

(b) varies against the *possible derived volume* $w' = |(D^{\mathrm{T}})^{-1}|$,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \;\sim\; 1/w'$$

(c) varies with the *expected component entropy*,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \;\sim\; \mathrm{entropyComponent}(A, D^{\mathrm{T}})$$

and (d) varies with the *component size cardinality relative entropy*,

$$\mathrm{algnValDensSum}(U)(A, D^{\mathrm{D}}) \;\sim\; \mathrm{entropyRelative}(A * D^{\mathrm{T}}, V^{\mathrm{C}} * D^{\mathrm{T}})$$

Given these relations it is conjectured in section 'Tractable transform induction' that the *maximum likelihood estimate* for the *model* for *specialising induction*,

$$\tilde{T}_{\mathrm{o}} \;\in\; \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

or

$$\tilde{T}_{\mathrm{o}} \;\in\; \mathrm{mind}(\{(T, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T)^{\mathrm{s}}(H_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

can be tractably approximated by the maximisation of the *tractable limited-models summed alignment valency-density substrate aligned non-overlapping infinite-layer fud decomposition inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$,

$$\tilde{T}_{\mathrm{o}} \;\approx\; D^{\mathrm{T}}_{\mathrm{o,Sd}}$$

and the *accuracy* of the *tractable model* varies with the *specialising log-likelihood*,

$$\frac{\hat{Q}_{\mathrm{G},D^{\mathrm{T}}_{\mathrm{o,Sd}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{G},\tilde{T}_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})} \sim \ln \hat{Q}_{\mathrm{G},T_{\mathrm{o}},\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})$$

Then it is conjectured that, insofar as the *uniform possible iso-derived history probability function* approximates to the *specialising history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d,p},T_{\mathrm{o}}} \approx P_{U,X,\mathrm{G},T_{\mathrm{o}},\mathrm{H}}$, the *model*, $D^{\mathrm{T}}_{\mathrm{o,Sd}}$, obtained by the maximisation of the *tractable summed alignment valency-density inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is also a tractable approximation to the *maximum likelihood estimate* for the *model* for *uniform possible iso-derived induction*,

$$\tilde{T}_{\mathrm{o}} \in \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})$$

That is, in the *near-natural*, high *relative entropy* case, a tractable *maximum likelihood estimate* for the *model* may be obtained for *classical modelled induction* by optimisation of the *summed alignment valency-density inducer*,

$$\tilde{T}_{\mathrm{o}} \approx D^{\mathrm{T}}_{\mathrm{o,Sd}}$$

Just as the *tractable model specialising accuracy* varies with the *log-likelihood*, so too does the *tractable model uniform possible iso-derived accuracy*,

$$\frac{\hat{Q}_{\mathrm{h,d},D^{\mathrm{T}}_{\mathrm{o,Sd}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \sim \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

This indirect derivation of the relation between the *tractable model uniform possible iso-derived accuracy* and the *log-likelihood* via the *entropy* properties of the *tractable inducer* and *specialising coder* in *natural classical modelled induction* provides corroboration for the more direct derivation in *formal-abstract ideal aligned modelled induction*. The *tractable summed alignment valency-density inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is directly derived from the *aligned induction* assumptions by removing the intractabilities from the *literal derived alignment integral-independent substrate ideal formal-abstract transform inducer*, $I'_{z,\mathrm{a,l}}$, while maintaining the positive correlation with the *finite transform-dependent-sample distributed iso-transform-independent conditional dependent multinomial space integral-independent substrate ideal formal abstract transform search set*,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,T,yy,fa,j},k}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}) \geq 0)$$

Then, given the *formal-abstract ideal* constraints on the *sample* in *aligned modelled induction* it is conjectured that the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *sample*,

$$\ln \hat{Q}_{\mathrm{h,y},\tilde{T}_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \sim \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

leading to the relation between the *tractable model uniform possible iso-derived accuracy* and the *log-likelihood*,

$$\frac{\hat{Q}_{\mathrm{h,d},D^{\mathrm{T}}_{\mathrm{o,Sd}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{T}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \sim \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

That is, the indirect derivation via the *entropy* properties is separate evidence for the conjecture that the *formal-abstract ideal aligned induction log-likelihood* varies with the *natural classical induction log-likelihood*.

In the discussion of *classical modelled induction*, above, consideration is given to the case where the *model* is extended from a *transform* first to a *functional definition set* and then to a *fud decomposition*. Now consider an outline for the same in *aligned modelled induction*.

In section 'Necessary derived functional definition set' the *model* is extended to a *functional definition set* from the *transform* of section 'Necessary derived'. Given some *known substrate fud*, $F_{\mathrm{o}} \in \mathcal{F}_{U,V_{\mathrm{o}}}$, such that there exists a *top transform*, $\exists T \in F_{\mathrm{o}}\ (\mathrm{der}(T) = \mathrm{der}(F_{\mathrm{o}}))$, the *derived histogram set* of the *distribution probability histogram* is $\{\hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}} : T \in F_{\mathrm{o}}\}$, where $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$. In *classical functional definition set induction*, while the *distribution probability histogram*, $\hat{E}_{\mathrm{h}}$, remains *unknown*, the *derived distribution probability histogram set*, $\{\hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}} : T \in F_{\mathrm{o}}\}$, is *known* and *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have a *derived probability histogram* equal to the *known derived distribution probability histogram* for each of the *transforms* of the *fud*, $\forall T \in F_{\mathrm{o}}\ (\hat{A}_H * T_{F_{\mathrm{o}}} = \hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}})$. The *iso-fud historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ is defined and the corresponding *iso-fud conditional stuffed historical probability distribution* is now conditional on the set of *iso-fuds*,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h,d},F,U}(E, z) \\
&:= \{(A, \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},F,z}(\{A * T_F : T \in F\})} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\
&\quad \{(A, 0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \not\leq E\}
\end{aligned}$$

where the finite set of *iso-fuds* of *derived histogram set* $\{A * T_F : T \in F\}$ is

$$D_{U,\mathrm{i},F,z}^{-1}(\{A * T_F : T \in F\}) = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ \forall T \in F\ (B * T_F = A * T_F)\}$$

Similarly, in *aligned functional definition set induction*, while the *distribution probability histogram*, $\hat{E}_{\mathrm{h}}$, remains *unknown*, the *formal-abstract-pair distribution probability histogram set*, $\{(\hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{F_{\mathrm{o}}}, (\hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}})^{\mathrm{X}}) : T \in F_{\mathrm{o}}\}$, is *known* and *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have a *formal probability histogram* equal to the *known formal distribution probability histogram* and an *abstract probability histogram* equal to the *known abstract distribution probability histogram* for each of the *transforms* of the *fud*, $\forall T \in F_{\mathrm{o}}\ (\hat{A}_H^{\mathrm{X}} * T_{F_{\mathrm{o}}} = \hat{E}_{\mathrm{h}}^{\mathrm{X}} * T_{F_{\mathrm{o}}} \wedge (\hat{A}_H * T_{F_{\mathrm{o}}})^{\mathrm{X}} = (\hat{E}_{\mathrm{h}} * T_{F_{\mathrm{o}}})^{\mathrm{X}})$. The *iso-fud-independent historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{y},F_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ can be defined analogously to the *iso-fud historically distributed history probability function*, $P_{U,X,H_{\mathrm{h}},\mathrm{d},F_{\mathrm{o}}}$, and the corresponding *iso-fud-independent conditional stuffed historical probability distribution* is now conditional on the set of *iso-fud-independents*,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{y},F,U}(E,z) \\
&\quad := \{(A, \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in Y_{U,\mathrm{i},F,z}^{-1}(Y_{U,F,z}(A))} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\
&\qquad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \not\leq E\}
\end{aligned}$$

where the finite set of *integral iso-fud-independents* of *formal-abstract-pair histogram set* $\{(A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}) : T \in F\}$ is

$$\begin{aligned}
&Y_{U,\mathrm{i},F,z}^{-1}(\{(A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}) : T \in F\}) = \\
&\quad \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ \forall T \in F\ (B^{\mathrm{X}} * T_F = A^{\mathrm{X}} * T_F \wedge (B * T_F)^{\mathrm{X}} = (A * T_F)^{\mathrm{X}})\}
\end{aligned}$$

which is the intersection of the *iso-fud-formals* and the *iso-fud-abstracts*

$$\begin{aligned}
&Y_{U,\mathrm{i},F,z}^{-1}(\{(A^{\mathrm{X}} * T_F, (A * T_F)^{\mathrm{X}}) : T \in F\}) = \\
&\qquad Y_{U,\mathrm{i},F,\mathrm{V},z}^{-1}(\{A^{\mathrm{X}} * T_F : T \in F\}) \cap Y_{U,\mathrm{i},F,\mathrm{W},z}^{-1}(\{(A * T_F)^{\mathrm{X}} : T \in F\})
\end{aligned}$$

In *classical transform induction* the special case is considered where the *sample* is constrained to be equal to the *independent analogue*, which is the *naturalisation*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger}$. In this case, the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger} \implies A_{\mathrm{o}}^{\mathrm{D}(T_{\mathrm{o}})} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

The *naturalisation* is the *likely histogram* of the *iso-derived*,

$$\{A * T * T^{\dagger}\} \;=\; \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T))) : E \in \mathcal{A}_{U,V,z}\})$$

The *fud independent analogue* corresponding to the *naturalisation* is the *fud-independent*, $A^{\mathrm{E_F}(F)} \in \mathcal{A}_{U,V,z}$, defined,

$$\{A^{\mathrm{E_F}(F)}\} \;=\; \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},F,z}(D_{U,F,z}(A)))) : E \in \mathcal{A}_{U,V,z}\})$$

The *fud-independent* approximates to the arithmetic *average* of the *naturalisations*,

$$A^{\mathrm{E_F}(F)} \;\approx\; Z_{1/|F|} * \sum_{T \in F} A * T_F * T_F^{\dagger}$$

In *classical fud induction*, it is only in the case where the *histogram* equals the *fud-independent* that the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E_F}(F_{\mathrm{o}})} \implies A_{\mathrm{o}}^{\mathrm{D_F}(F_{\mathrm{o}})} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

In *aligned transform induction*, however, the *sample* is not constrained to be equal to the *independent analogue*, $A_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})}$, but instead (a) the *sample formal* equals the *sample abstract*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, and (b) the *sample* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$. In this case the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *sample*,

$$\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \;\sim\; \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

The corresponding *likely histogram* in *aligned fud induction* is the solution to

$$\{A^{\mathrm{X_{F,fa,j}}(F)}\} \;\in\; \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},F,z}(Y_{U,F,z}(A)),\; \forall T \in F\ (B^{\mathrm{X}} * T_F = (B * T_F)^{\mathrm{X}} \;\wedge\; B * T_F * T_F^{\dagger B} = A * T_F * T_F^{\dagger A})) : E \in \mathcal{A}_{U,V,z}\})$$

Then, in section 'Unknown necessary derived', the case is considered where the *model*, T_{o}, is *unknown* and it is found that there is no singular solution to the optimisation,

$$\mathrm{maxd}(\{((E,T), \frac{Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z_{\mathrm{o}}}(A_{\mathrm{o}} * T)} Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(B)}) : E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ T \in \mathcal{T}_{U,V_{\mathrm{o}}}\}) \supseteq \mathcal{A}_{U,V_{\mathrm{o}},1} \times \{T_{\mathrm{s}}\}$$

where T_{s} is a *self transform*. The discussion of *classical induction* goes on to consider a weakening of *necessary derived* to *uniform possible derived* in section 'Uniform possible derived induction' and the corresponding extension of the *model* to *fud* in section 'Uniform possible derived functional definition set induction'. This is not required in *aligned modelled induction*, however, because it is conjectured that in *transform induction* there are some cases in which there is a unique solution for the pair $(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})$, where the optimisation is

$$\begin{aligned} &\{(\tilde{E}_{\mathrm{o}}, \tilde{T}_{\mathrm{o}})\} \\ &\quad= \mathrm{maxd}(\{((E,T), \frac{Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})} Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(B)}) : \\ &\qquad\qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ T \in \mathcal{T}_{U,V_{\mathrm{o}}}\}) \end{aligned}$$

The corresponding optimisation in *aligned fud induction* is

$$\begin{aligned} &\{(\tilde{E}_{\mathrm{o}}, \tilde{F}_{\mathrm{o}})\} \\ &\quad= \mathrm{maxd}(\{((E,F), \frac{Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in Y^{-1}_{U,\mathrm{i},F,z_{\mathrm{o}}}(Y_{U,F,z_{\mathrm{o}}}(A_{\mathrm{o}}))} Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(B)}) : \\ &\qquad\qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ F \in \mathcal{F}_{U,V_{\mathrm{o}}}\}) \end{aligned}$$

The discussion of *classical induction* goes on, in section 'Specialising induction', to consider *induction* assumptions that do not depend on the *multinomial probability* given a *distribution histogram*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})$, but instead are based on *specialising coder space*, $C_{\mathrm{G,T,H}}(T_{\mathrm{o}})^{\mathrm{s}}$. The corresponding extension of the *model* to *fud*, based on *specialising fud coder space*, $C_{\mathrm{G,F,H}}(F_{\mathrm{o}})^{\mathrm{s}}$, is discussed in section 'Specialising functional definition set induction'.

Then the discussion of *classical induction* goes on to consider *tractable induction* in section 'Tractable transform induction'. There it is conjectured that the *model*, $D^{\mathrm{T}}_{\mathrm{o,Sd}}$, obtained by the maximisation of the *tractable summed alignment valency-density inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is a tractable approximation to the *maximum likelihood estimate* for the *model* for *specialising induction*,

$$\tilde{T}_{\mathrm{o}} \in \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

or

$$\tilde{T}_{\mathrm{o}} \ \in \ \mathrm{mind}(\{(T, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T)^{\mathrm{s}}(H_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

and so a tractable approximation to the *maximum likelihood estimate* for the *model* for *uniform possible iso-derived induction*,

$$\tilde{T}_{\mathrm{o}} \ \in \ \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})$$

The discussion of the corresponding extension of the *model* to *fud*, in section 'Tractable functional definition set induction', does not depend on *tractable inducers* but rather conjectures that the *model*, $F_{\mathrm{o,gr,lsq}}$, obtained by the maximisation of the *least squares gradient descent fud search function*, $Z_{\mathrm{F,P},P,\mathrm{gr,lsq}}$, is a tractable approximation to the *maximum likelihood estimate* for the *model* for *specialising induction*,

$$\tilde{F}_{\mathrm{o}} \ \in \ \mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{G},F,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})$$

or

$$\tilde{F}_{\mathrm{o}} \ \in \ \mathrm{mind}(\{(F, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{F,H}}(F^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})) : F \in \mathcal{F}_{U,V_{\mathrm{o}}}\})$$

and so is a tractable approximation to the *maximum likelihood estimate* for the *model* for *uniform possible iso-fud induction*,

$$\begin{aligned}\tilde{F}_{\mathrm{o}} \ \in \ \mathrm{maxd}(\{(F, \hat{Q}_{\mathrm{h,d},F,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) :& \\ F \in \mathcal{F}_{U,V_{\mathrm{o}}},\ \exists T \in F\ (W_T = W_F),\ A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_F}(F)}\})&\end{aligned}$$

where the *least squares gradient descent fud search function* is defined

$$\begin{aligned}&Z_{\mathrm{F,P},P,\mathrm{gr,lsq}}(H) = \\ &\quad \{(\mathrm{fud}(\sigma)(G), -\mathrm{lsq}(\sigma)(A,G,K)) : Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{gr,lsq}})),\ \{G\} = Q\}\end{aligned}$$

and the *least squares gradient descent substrate net tree searcher*, $Z_{P,A,\mathrm{gr,lsq}}$, is, in turn, defined in terms of the *neural net substrate fud set*, $\mathcal{F}_{\infty,U,V,\sigma} = \mathcal{F}_{\infty,U,V} \cap (\mathrm{fud}(\sigma) \circ \mathrm{nets})$. The *accuracy* of the approximation is defined as the ratio of the *tractable model uniform possible iso-fud likelihood* to the *maximum model uniform possible iso-fud likelihood*,

$$0 \ < \ \frac{\hat{Q}_{\mathrm{h,d},F_{\mathrm{o,gr,lsq}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{F}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \ \leq \ 1$$

The corresponding *aligned accuracy* is defined as the ratio of the *tractable model iso-fud-independent likelihood* to the *maximum model iso-fud-independent* *likelihood*. The *accuracy*, defined with respect to the *finite fud-dependent-sample-distributed iso-fud-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract fud search set*, $X_{z,\mathrm{xi,F,yy,fa,j},k}$, may be stated explicitly,

$$\frac{Q_{\mathrm{m},U}(A_{\mathrm{o},k}^{\mathrm{Y_F}(F_{\mathrm{o,gr,lsq}})}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o},k}^{\mathrm{Y_F}(F_{\mathrm{o,gr,lsq}})}, z_{\mathrm{o}})(B) : B \in Y^{-1}_{U,\mathrm{i},F_{\mathrm{o,gr,lsq}},z_{\mathrm{o}}}(Y_{U,F_{\mathrm{o,gr,lsq}},z_{\mathrm{o}}}(A_{\mathrm{o}}))}$$
$$/ \quad \frac{Q_{\mathrm{m},U}(A_{\mathrm{o},k}^{\mathrm{Y_F}(\tilde{F}_{\mathrm{o}})}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o},k}^{\mathrm{Y_F}(\tilde{F}_{\mathrm{o}})}, z_{\mathrm{o}})(B) : B \in Y^{-1}_{U,\mathrm{i},\tilde{F}_{\mathrm{o}},z_{\mathrm{o}}}(Y_{U,\tilde{F}_{\mathrm{o}},z_{\mathrm{o}}}(A_{\mathrm{o}}))}$$

Now consider the outline of *aligned modelled induction* for the extension of the *model* to a *functional definition set*.

In section 'Necessary derived functional definition set decomposition' the *model* is extended to a *functional definition set decomposition* from the *functional definition set* of section 'Necessary derived functional definition set'. Given some non-empty *known substrate fud decomposition*, $D_{\mathrm{o}} \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\}$, such that there exists a *top transform* for all of the *fuds*, $\forall F \in \mathrm{fuds}(D_{\mathrm{o}})\ \exists T \in F\ (\mathrm{der}(T) = \mathrm{der}(F))$, the *component derived set* of the *distribution probability histogram* is $\{(C, \{\hat{E}_{\mathrm{h}} * C * T_F : T \in F\}) : (C, F) \in \mathrm{cont}(D_{\mathrm{o}})\}$, where $\mathrm{cont}(D) = \mathrm{elements}(\mathrm{contingents}(D))$ and $T_F := \mathrm{depends}(F, \mathrm{der}(T))^{\mathrm{T}}$. In *classical functional definition set decomposition induction*, while the *distribution probability histogram*, $\hat{E}_{\mathrm{h}}$, remains *unknown*, the *component derived distribution probability set*, $\{(C, \{\hat{E}_{\mathrm{h}} * C * T_F : T \in F\}) : (C, F) \in \mathrm{cont}(D_{\mathrm{o}})\}$, is *known* and *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have a *derived probability histogram* equal to the *known derived distribution probability histogram* for each of the *transforms* of the *fud* for each *slice*, $\forall (C, F) \in \mathrm{cont}(D_{\mathrm{o}})\ \forall T \in F\ (\hat{A}_H * C * T_F = \hat{E}_{\mathrm{h}} * C * T_F)$. The *iso-fud-decomposition historically distributed history probability function* $P_{U,X,H_{\mathrm{h}},\mathrm{d},D_{\mathrm{o}}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ is defined and the corresponding *iso-fud-decomposition conditional stuffed historical probability distribution* is now conditional on the set of *iso-fud-decompositions*,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h,d},D,U}(E, z) \\
&\quad := \ \{(A, \frac{Q_{\mathrm{h},U}(E, z)(A)}{\sum_{B \in D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A))} Q_{\mathrm{h},U}(E, z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^{\wedge} \cup \\
&\qquad\ \ \{(A, 0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \nleq E\}
\end{aligned}$$

where the finite *iso-fud-decompositions* of *component-derived-set* $D_{U,D,\mathrm{F},z}(A)$ is

$$D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A)) \;=\; \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ \forall (C,F) \in \mathrm{cont}(D)\ \forall T \in F\ (B * C * T_F = A * C * T_F)\}$$

Similarly, in *aligned functional definition set decomposition induction*, while the *distribution probability histogram*, $\hat{E}_\mathrm{h}$, remains *unknown*, the *component formal-abstract-pair set distribution probability histogram set*, $\{(C, \{((\hat{E}_\mathrm{h} * C)^\mathrm{X} * T_F, (\hat{E}_\mathrm{h} * C * T_F)^\mathrm{X}) : T \in F\}) : (C,F) \in \mathrm{cont}(D_\mathrm{o})\}$, is *known* and *necessary*. That is, the *history probability function*, P, is *historically distributed* but constrained such that all *drawn histories* have a *formal probability histogram* equal to the *known formal distribution probability histogram* and an *abstract probability histogram* equal to the *known abstract distribution probability histogram* for each of the *transforms* of the *fud* for each *slice*, $\forall (C,F) \in \mathrm{cont}(D_\mathrm{o})\ \forall T \in F\ ((\hat{A}_H * C)^\mathrm{X} * T_F = (\hat{E}_\mathrm{h} * C)^\mathrm{X} * T_F \wedge (\hat{A}_H * C * T_F)^\mathrm{X} = (\hat{E}_\mathrm{h} * C * T_F)^\mathrm{X})$. The *iso-fud-decomposition-independent historically distributed history probability function* $P_{U,X,H_\mathrm{h},\mathrm{y},D_\mathrm{o}} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{[0,1]}) \cap \mathcal{P}$ can be defined analogously to the *iso-fud-decomposition historically distributed history probability function*, $P_{U,X,H_\mathrm{h},\mathrm{d},D_\mathrm{o}}$, and the corresponding *iso-fud-decomposition-independent conditional stuffed historical probability distribution* is now conditional on the set of *iso-fud-decomposition-independents*,

$$\begin{aligned}
&\hat{Q}_{\mathrm{h},\mathrm{y},D,U}(E,z) \\
&\quad := \{(A, \frac{Q_{\mathrm{h},U}(E,z)(A)}{\sum_{B \in Y^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(Y_{U,D,\mathrm{F},z}(A))} Q_{\mathrm{h},U}(E,z)(B)}) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \leq E\}^\wedge \cup \\
&\qquad \{(A,0) : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A \nleq E\}
\end{aligned}$$

where the finite set of *integral iso-fud-decomposition-independents* of *component formal-abstract-pair set* $Y_{U,D,\mathrm{F},z}(A)$ is

$$\begin{aligned}
&Y^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(Y_{U,D,\mathrm{F},z}(A)) \;=\; \\
&\quad \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ \forall (C,F) \in \mathrm{cont}(D)\ \forall T \in F \\
&\qquad ((B * C)^\mathrm{X} * T_F = (A * C)^\mathrm{X} * T_F\ \wedge\ (B * C * T_F)^\mathrm{X} = (A * C * T_F)^\mathrm{X})\}
\end{aligned}$$

In *classical transform induction* the special case is considered where the *sample* is constrained to be equal to the *independent analogue*, which is the *naturalisation*, $A_\mathrm{o} = A_\mathrm{o} * T_\mathrm{o} * T_\mathrm{o}^\dagger$. In this case, the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_\mathrm{o} = A_\mathrm{o} * T_\mathrm{o} * T_\mathrm{o}^\dagger \implies A_\mathrm{o}^{\mathrm{D}(T_\mathrm{o})} = A_\mathrm{o} \implies \tilde{E}_\mathrm{o} = \hat{A}_\mathrm{o}$$

The *naturalisation* is the *likely histogram* of the *iso-derived*,

$$\{A * T * T^{\dagger}\} \ = \\ \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},T,z}(A * T))) : E \in \mathcal{A}_{U,V,z}\})$$

The *fud-decomposition-independent analogue* corresponding to the *naturalisation* is the *fud-decomposition-independent*, $A^{\mathrm{E}_{\mathrm{D,F}}(D)} \in \mathcal{A}_{U,V,z}$, defined,

$$\{A^{\mathrm{E}_{\mathrm{D,F}}(D)}\} \ = \\ \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in D^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(D_{U,D,\mathrm{F},z}(A)))) : E \in \mathcal{A}_{U,V,z}\})$$

The *fud-decomposition-independent* approximates to the *scaled sum* of the *slice* arithmetic *average* of the *naturalisations*,

$$A^{\mathrm{E}_{\mathrm{D,F}}(D)} \ \approx \ Z_z * \left(\sum_{(C,F) \in \mathrm{cont}(D)} \left(Z_{1/|F|} * \sum_{T \in F} A * C * T_F * T_F^{\dagger} \right) \right)^{\wedge}$$

In *classical fud decomposition induction*, it is only in the case where the *histogram* equals the *fud-decomposition-independent* that the *maximum likelihood estimate* is necessarily equal to the *sample probability histogram*,

$$A_{\mathrm{o}} = A_{\mathrm{o}}^{\mathrm{E}_{\mathrm{D,F}}(D_{\mathrm{o}})} \implies A_{\mathrm{o}}^{\mathrm{D}_{\mathrm{D,F}}(D_{\mathrm{o}})} = A_{\mathrm{o}} \implies \tilde{E}_{\mathrm{o}} = \hat{A}_{\mathrm{o}}$$

In *aligned transform induction*, however, the *sample* is not constrained to be equal to the *independent analogue*, $A_{\mathrm{o}}^{\mathrm{X}(T_{\mathrm{o}})}$, but instead (a) the *sample formal* equals the *sample abstract*, $A_{\mathrm{o}}^{\mathrm{X}} * T_{\mathrm{o}} = (A_{\mathrm{o}} * T_{\mathrm{o}})^{\mathrm{X}}$, and (b) the *sample* is *ideal*, $A_{\mathrm{o}} = A_{\mathrm{o}} * T_{\mathrm{o}} * T_{\mathrm{o}}^{\dagger A_{\mathrm{o}}}$. In this case the *log likelihood* of the *iso-transform-independent conditional stuffed historical probability distribution* at the *maximum likelihood estimate* varies with the *log likelihood* of the *iso-derived conditional stuffed historical probability distribution* at the *sample*,

$$\ln \hat{Q}_{\mathrm{h,y},T_{\mathrm{o}},U}(\tilde{E}_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}}) \ \sim \ \ln \hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})$$

The corresponding *likely histogram* in *aligned fud decomposition induction* is the solution to

$$\{A^{\mathrm{X}_{\mathrm{D,F,fa,j}}(D)}\} \ \in \\ \mathrm{maxd}(\{(E, \sum(Q_{\mathrm{m},U}(E,z)(B) : B \in Y^{-1}_{U,\mathrm{i},D,\mathrm{F},z}(Y_{U,D,\mathrm{F},z}(A)), \\ \forall (C,F) \in \mathrm{cont}(D) \ \forall T \in F \\ ((B * C)^{\mathrm{X}} * T_F = (B * C * T_F)^{\mathrm{X}} \ \wedge \\ B * C * T_F * T_F^{\dagger B * C} = A * C * T_F * T_F^{\dagger A * C})) : \\ E \in \mathcal{A}_{U,V,z}\})$$

Then, in section 'Unknown necessary derived', the case is considered where the *model*, T_{o}, is *unknown* and it is found that there is no singular solution to the optimisation,

$$\begin{aligned} &\mathrm{maxd}(\{((E,T), \frac{Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in D^{-1}_{U,\mathrm{i},T,z_{\mathrm{o}}}(A_{\mathrm{o}}*T)} Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(B)}) : \\ &\qquad\qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ T \in \mathcal{T}_{U,V_{\mathrm{o}}}\}) \supseteq \mathcal{A}_{U,V_{\mathrm{o}},1} \times \{T_{\mathrm{s}}\} \end{aligned}$$

where T_{s} is a *self transform*. The discussion of *classical induction* goes on to consider a weakening of *necessary derived* to *uniform possible derived* in section 'Uniform possible derived induction' and the corresponding extension of the *model* to *fud decomposition* in section 'Uniform possible derived functional definition set decomposition induction'. This is not required in *aligned modelled induction*, however, because it is conjectured that in *transform induction* there are some cases in which there is a unique solution for the pair $(\tilde{E}_{\mathrm{o}},\tilde{T}_{\mathrm{o}})$, where the optimisation is

$$\begin{aligned} &\{(\tilde{E}_{\mathrm{o}},\tilde{T}_{\mathrm{o}})\} \\ &\quad= \mathrm{maxd}(\{((E,T), \frac{Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in \mathcal{A}_{U,\mathrm{i},\mathrm{y},T,z_{\mathrm{o}}}(A_{\mathrm{o}})} Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(B)}) : \\ &\qquad\qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ T \in \mathcal{T}_{U,V_{\mathrm{o}}}\}) \end{aligned}$$

The corresponding optimisation in *aligned fud decomposition induction* is

$$\begin{aligned} &\{(\tilde{E}_{\mathrm{o}},\tilde{D}_{\mathrm{o}})\} \\ &\quad= \mathrm{maxd}(\{((E,D), \frac{Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum_{B \in Y^{-1}_{U,\mathrm{i},D,\mathrm{F},z_{\mathrm{o}}}(Y_{U,D,\mathrm{F},z_{\mathrm{o}}}(A_{\mathrm{o}}))} Q_{\mathrm{m},U}(E,z_{\mathrm{o}})(B)}) : \\ &\qquad\qquad E \in \mathcal{A}_{U,V_{\mathrm{o}},1},\ D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}}\}) \end{aligned}$$

The discussion of *classical induction* goes on, in section 'Specialising induction', to consider *induction* assumptions that do not depend on the *multinomial probability* given a *distribution histogram*, $\hat{Q}_{\mathrm{h,d},T_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}},z_{\mathrm{o}})$, but instead are based on *specialising coder space*, $C_{\mathrm{G,T,H}}(T_{\mathrm{o}})^{\mathrm{s}}$. The corresponding extension of the *model* to *fud decomposition*, based on *specialising fud decomposition coder space*, $C_{\mathrm{G,D,F,H}}(D_{\mathrm{o}})^{\mathrm{s}}$, is discussed in section 'Specialising functional definition set decomposition induction'.

Then the discussion of *classical induction* goes on to consider *tractable induction* in section 'Tractable transform induction'. There it is conjectured that the *model*, $D^{\mathrm{T}}_{\mathrm{o,Sd}}$, obtained by the maximisation of the *tractable summed* *alignment valency-density inducer*, $I'_{z,\mathrm{Sd,D,F},\infty,\mathrm{n,q}}$, is a tractable approximation to the *maximum likelihood estimate* for the *model* for *specialising induction*,

$$\tilde{T}_{\mathrm{o}} \in \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{G},T,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

or

$$\tilde{T}_{\mathrm{o}} \in \mathrm{mind}(\{(T, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{T,H}}(T)^{\mathrm{s}}(H_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}}\})$$

and so a tractable approximation to the *maximum likelihood estimate* for the *model* for *uniform possible iso-derived induction*,

$$\tilde{T}_{\mathrm{o}} \in \mathrm{maxd}(\{(T, \hat{Q}_{\mathrm{h,d},T,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) : T \in \mathcal{T}_{U,V_{\mathrm{o}}},\ A_{\mathrm{o}} \approx A_{\mathrm{o}} * T * T^{\dagger}\})$$

The discussion of the corresponding extension of the *model* to *fud decomposition*, in section 'Tractable functional definition set decomposition induction', does not depend so much on the *tractable inducer* but rather conjectures that the *model*, $D_{\mathrm{o,Scsd},P}$, obtained by the maximisation of the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}$, is a practicable approximation to the *maximum likelihood estimate* for the *model* for *specialising induction*,

$$\tilde{D}_{\mathrm{o}} \in \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{G},D,\mathrm{H},U}(z_{\mathrm{o}})(A_{\mathrm{o}})) : D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}}\})$$

or

$$\tilde{D}_{\mathrm{o}} \in \mathrm{mind}(\{(D, C_{\mathrm{G},V_{\mathrm{o}},\mathrm{D,F,H}}(D^{V_{\mathrm{o}}})^{\mathrm{s}}(H_{\mathrm{o}})) : D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}}\})$$

and so is a practicable approximation to the *maximum likelihood estimate* for the *model* for *uniform possible iso-fud-decomposition induction*,

$$\begin{aligned}\tilde{D}_{\mathrm{o}} \in \mathrm{maxd}(\{(D, \hat{Q}_{\mathrm{h,d},D,U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})) :&\\ D \in \mathcal{D}_{\mathrm{F},U,V_{\mathrm{o}}} \setminus \{\emptyset\},\ \forall F \in \mathrm{fuds}(D)\ \exists T \in F\ (W_T = W_F),&\\ A_{\mathrm{o}} \approx A_{\mathrm{o}}^{\mathrm{E_{D,F}}(D)}\})&\end{aligned}$$

where, given parameter tuple $P \in \mathcal{L}(\mathcal{X})$, the *practicable fud decomposition inducer* is defined as

$$\begin{aligned}&I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}(A) =\\ &\quad \mathrm{if}(Q \neq \emptyset, \{(D, I^{*}_{\mathrm{Scsd}}((A, D)))\}, \{(D_{\emptyset}, 0)\}) :\\ &\qquad\qquad Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A,\mathrm{D,F,d}})),\ \{D\} = Q\end{aligned}$$

and the *summed shuffle content alignment valency-density computer* $I_{\mathrm{Scsd}} \in$ computers is defined as

$$I_{\mathrm{Scsd}}^*((A,D)) = \sum (I_{\mathrm{a}}^*(A * C * F^{\mathrm{T}}) - I_{\mathrm{a}}^*((A * C)_{R(A*C)} * F^{\mathrm{T}}))/I_{\mathrm{cvl}}^*(F) : (C,F) \in \mathrm{cont}(D)$$

and $Z_{P,A,\mathrm{D,F,d}}$ is the *highest-layer limited-models infinite-layer substrate fud decompositions tree searcher*.

The *accuracy* of the approximation is defined as the ratio of the *tractable model uniform possible iso-fud-decomposition likelihood* to the *maximum model uniform possible iso-fud-decomposition likelihood*,

$$0 \ < \ \frac{\hat{Q}_{\mathrm{h,d},D_{\mathrm{o,Scsd},P},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\hat{Q}_{\mathrm{h,d},\tilde{D}_{\mathrm{o}},U}(A_{\mathrm{o},z_{\mathrm{h}}}, z_{\mathrm{o}})(A_{\mathrm{o}})} \ \leq \ 1$$

The corresponding *aligned accuracy* is defined as the ratio of the *tractable model iso-fud-decomposition-independent likelihood* to the *maximum model iso-fud-decomposition-independent likelihood*. The *accuracy*, with respect to the *finite fud-decomposition-dependent-sample-distributed iso-fud-decomposition-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract fud decomposition search set*, $X_{z,\mathrm{xi,D,F,yy,fa,j},k}$, may be stated explicitly,

$$\frac{Q_{\mathrm{m},U}(A_{\mathrm{o},k}^{\mathrm{Y_{D,F}}(D_{\mathrm{o,Scsd},P})}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o},k}^{\mathrm{Y_{D,F}}(D_{\mathrm{o,Scsd},P})}, z_{\mathrm{o}})(B) : B \in Y_{U,\mathrm{i},D_{\mathrm{o,Scsd},P},\mathrm{F},z_{\mathrm{o}}}^{-1}(Y_{U,D_{\mathrm{o,Scsd},P},\mathrm{F},z_{\mathrm{o}}}(A_{\mathrm{o}}))}$$
$$/ \quad \frac{Q_{\mathrm{m},U}(A_{\mathrm{o},k}^{\mathrm{Y_{D,F}}(\tilde{D}_{\mathrm{o}})}, z_{\mathrm{o}})(A_{\mathrm{o}})}{\sum Q_{\mathrm{m},U}(A_{\mathrm{o},k}^{\mathrm{Y_{D,F}}(\tilde{D}_{\mathrm{o}})}, z_{\mathrm{o}})(B) : B \in Y_{U,\mathrm{i},\tilde{D}_{\mathrm{o}},\mathrm{F},z_{\mathrm{o}}}^{-1}(Y_{U,\tilde{D}_{\mathrm{o}},\mathrm{F},z_{\mathrm{o}}}(A_{\mathrm{o}}))}$$

This definition of *accuracy* can be derived more directly by considering the definition of *fud decomposition inducers*. The *fud-decomposition-dependent-sample-distributed iso-fud-decomposition-independent conditional dependent multinomial space integral-independent substrate ideal formal-abstract fud decomposition search set* is defined,

$$X_{z,\mathrm{xi,D,F,yy,fa,j}}(A) \ = \ \{(D, \ln \frac{\hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{Y_{D,F}}(D)}, z)(A)}{\sum \hat{Q}_{\mathrm{m},U_A}(A^{\mathrm{Y_{D,F}}(D)}, z)(B) : B \in Y_{U_A,\mathrm{i},D,\mathrm{F},z}^{-1}(Y_{U_A,D,\mathrm{F},z}(A))}) : D \in \mathcal{D}_{\mathrm{F},U_A,V_A}, \ A = A^{\mathrm{X_{D,F,fa,j}}(D)}\}$$

The *fud decomposition induction correlation* of *inducer* I_z must be positive,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,D,F,yy,fa,j}}, \mathrm{maxr} \circ I_z^*) \geq 0)$$

Conjecture that, for some parameter tuples, P, the *fud decomposition induction correlation* of the *practicable highest-layer summed shuffle content alignment valency-density fud decomposition inducer*, $I'_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}$, is positive,

$$\forall z \in \mathbf{N}_{>0}\ (\mathrm{cov}(z)(\mathrm{maxr} \circ X_{z,\mathrm{xi,D,F,yy,fa,j}}, \mathrm{maxr} \circ I'^{*}_{z,\mathrm{Scsd,D,F},\infty,\mathrm{q},P,\mathrm{d}}) \geq 0)$$

Evidence for this conjecture may be seen by considering (a) the extension of the *aligned induction* assumptions for *fud decompositions* and (b) the approximation made by the *practicable inducer*.

In *aligned transform induction*, where (i) the *history probability function* is *iso-transform-independent historically distributed*, $P = P_{U,X,H_\mathrm{h},\mathrm{y},T_\mathrm{o}}$, given some *unknown substrate transform* in the *sample variables* $T_\mathrm{o} \in \mathcal{T}_{U,V_\mathrm{o}}$, if it is the case that (ii) the *iso-transform-independent dependent-independent anti-optimisation assumption* is true, (iii) the *iso-transform-independent underlying-lifted optimisation assumption* is true, and (iv) the *iso-transform-independent conditional-relative optimisation assumption* is true, then the *maximum likelihood estimate* of the *model*, $\tilde{T}_\mathrm{o}$, at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_\mathrm{o}$, is

$$\{\tilde{T}_\mathrm{o}\} \quad = \quad \mathrm{maxd}(\{(T, \mathrm{algn}(A_\mathrm{o} * T)) : T \in \mathcal{T}_{U,V_\mathrm{o}}\})$$

So, in *aligned functional definition set decomposition induction*, where (i) the *history probability function* is *iso-fud-decomposition-independent historically distributed*, $P = P_{U,X,H_\mathrm{h},\mathrm{y},D_\mathrm{o}}$, given some *unknown substrate fud decomposition* in the *sample variables* $D_\mathrm{o} \in \mathcal{D}_{\mathrm{F},U,V_\mathrm{o}}$, if it is the case that (ii) the *iso-fud-decomposition-independent dependent-independent anti-optimisation assumption* is true, (iii) the *iso-fud-decomposition-independent underlying-lifted optimisation assumption* is true, and (iv) the *iso-fud-decomposition-independent conditional-relative optimisation assumption* is true, then the *maximum likelihood estimate* of the *model*, $\tilde{D}_\mathrm{o}$, at the *maximum likelihood estimate* of the *distribution*, $\tilde{E}_\mathrm{o}$, depends on a maximisation of the set of *derived alignments* for each *transform* for each *fud* of the *decomposition*,

$$\{(C, \{(T, \mathrm{algn}(A_\mathrm{o} * C * T_F)) : T \in F\}) : (C, F) \in \mathrm{cont}(D_\mathrm{o})\}$$

As shown in section 'Inducers and Compression', above, in the case of the *practicable search function*, $Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}$, defined,

$$\begin{aligned}
&Z_{\mathrm{D,F,P,q,d},P,\mathrm{Scsd}}(H) = \\
&\quad \{(D, I^*_{\mathrm{Scsd}}((A_H, D))) : Q = \mathrm{leaves}(\mathrm{tree}(Z_{P,A_H,\mathrm{D,F,d}})),\ Q \neq \emptyset,\ \{D\} = Q\}\ \cup \\
&\quad \{(D_\mathrm{u}, 0)\}
\end{aligned}$$

the *fuds* of the *decomposition* are built *layer* by *layer*,

$$\forall (i,F) \in L~(\mathrm{layer}(F,\mathrm{der}(F)) = i)$$

where $\{L\} = \mathrm{paths}(\mathrm{tree}(Z_{P,B,B_R,\mathrm{L},\mathrm{d}}))$, *slice* $B = A * C$ and the *highest-layer limited-layer limited-underlying limited-breadth fud tree searcher* is

$$Z_{P,B,B_R,\mathrm{L},\mathrm{d}} = \mathrm{searchTreer}(\mathcal{F}_{\infty,U_B,V_B} \cap \mathcal{F}_{\mathrm{u}} \cap \mathcal{F}_{\mathrm{b}} \cap \mathcal{F}_{\mathrm{h}}, P_{P,B,B_R,\mathrm{L},\mathrm{d}}, \{\emptyset\})$$

So the properties of the *fuds* of the *decomposition* also depend on *layer*. In particular the *highest-layer fud tree searcher*, $Z_{P,B,B_R,\mathrm{L},\mathrm{d}}$, is constrained such that the *shuffle content alignment valency-density* of the *derived variables set* increases in each *layer*. The *shuffle content alignment valency-density* increases in each *layer*, so, in general, the *derived alignment* increases up the *layers*,

$$\forall i \in \{2 \ldots l\}~(\mathrm{algn}(B * F^{\mathrm{T}}_{\{1 \ldots i\}}) > \mathrm{algn}(B * F^{\mathrm{T}}_{\{1 \ldots i-1\}}))$$

and so

$$\forall i \in \{2 \ldots l\}~(\mathrm{algn}(A * C * F^{\mathrm{T}}_{\{1 \ldots i\}}) > \mathrm{algn}(A * C * F^{\mathrm{T}}_{\{1 \ldots i-1\}}))$$

The *limited-layer limited-underlying limited-breadth fud tree searcher* neighbourhood function is

$$\begin{aligned}
&P_{P,B,B_R,\mathrm{L}}(F) = \{G : \\
&\quad G = F \cup \{T : K \in \mathrm{topd}(\lfloor \mathrm{bmax}/\mathrm{mmax} \rfloor)(\mathrm{elements}(Z_{P,B,B_R,F,\mathrm{B}})), \\
&\qquad\qquad H \in \mathrm{topd}(\mathrm{pmax})(\mathrm{elements}(Z_{P,B,B_R,F,\mathrm{n},-,K})), \\
&\qquad\qquad w \in \mathrm{der}(H),~I = \mathrm{depends}(\mathrm{explode}(H),\{w\}),~T = I^{\mathrm{TPT}}\}, \\
&\quad \mathrm{layer}(G,\mathrm{der}(G)) \leq \mathrm{lmax}\}
\end{aligned}$$

where $Z_{P,B,B_R,F,\mathrm{n},-,K}$ is the *contracted decrementing linear non-overlapping fuds list maximiser*. The *decrementing maximiser* optimiser function is

$$\begin{aligned}
X_{P,B,B_R,F,\mathrm{n},-,K} = \{&(H, I^{*}_{\mathrm{csd}}((B,B_R,G'))) : \\
&H \in \mathcal{F}_{U_B,\mathrm{n},-,K,\overline{\mathrm{b}},\mathrm{mmax},\overline{2}},~G' = \mathrm{depends}(F \cup H, \mathrm{der}(H))\}
\end{aligned}$$

The *shuffle content alignment valency-density* increases in each *layer* because the *decrementing maximiser*, $Z_{P,B,B_R,F,\mathrm{n},-,K}$, maximises the *shuffle content alignment valency-density*, $I^{*}_{\mathrm{csd}}((B,B_R,G'))$, by *value rolling* the *underlying tuple*, K. Conjecture, therefore that the *derived alignment* for the *transform* for the *fud*, $\mathrm{algn}(A * C * G'^{\mathrm{T}})$, also tends to be maximised. In fact, the *fud tree searcher*, $Z_{P,B,B_R,\mathrm{L},\mathrm{d}}$, *explodes* the resultant *fud*, $\mathrm{explode}(H)$, so the cardinality of the set of *iso-fud-independents*, $Y^{-1}_{U,\mathrm{i},G,z_B}(Y_{U,G,z_B}(B))$, is larger than it otherwise would be. This, however, is merely a detail of the implementation, and a *practicable inducer* can easily be defined with the stricter *iso-set*, $Y^{-1}_{U,\mathrm{i},F\cup\{H^{\mathrm{T}}\},z_B}(Y_{U,F\cup\{H^{\mathrm{T}}\},z_B}(B))$.

To conclude, conjecture that for all *model* types there is a definition of *aligned modelled induction* corresponding to every definition of *classical modelled induction*. That is, both *classical induction* and *aligned induction* may be regarded as special cases of *induction* in general.

6 References

[1] D. Berend and T. Tassa, Improved bounds on Bell numbers and on moments of sums of random variables. Probability and Mathematical Statistics Vol. 30 (2010), pp. 185-205.

[2] Bengio, Y. (2009). Learning Deep Architectures for AI. Foundations and Trends in Machine Learning, V2(1). Now Publishers.

A Miscellaneous

A.1 Algebra of history

The multiply $(*)$ and reduce(V) functions are more obviously dual, in the sense that they increase or decrease the cardinality of the set of *variables*, when defined for *history*. These functions do not require equality to be determined for the *event identifier*. For *multiplication* we can simply put the *event identifiers* in a pair to preserve uniqueness, $(*) \in \mathcal{H} \times \mathcal{H} \to \mathcal{H}$,

$$H * G := \{((X,Y), S \cup T) : (X,S) \in H, (Y,T) \in G, |\mathrm{vars}(S \cup T)| = |S \cup T|\}$$

In the case of *reduction* the *event identifiers* are left entirely alone, reduce $\in \mathrm{P}(\mathcal{V}) \to (\mathcal{H} \to \mathcal{H})$

$$\mathrm{reduce}(V)(H) := \{(X, \mathrm{filter}(V,S)) : (X,S) \in H\}$$

By contrast, *addition* and particularly *subtraction* are more awkward for *histories* because we must concern ourselves with the equality of the *event identifiers*. The natural definition for *addition* would be $H + G := H \cup G$, but only where $\mathrm{dom}(H) \cap \mathrm{dom}(G) = \emptyset$. If this condition does not hold, then we must prefix the *event identifiers* with positional identifiers, for example,

$$H + G := \{((l, X), S) : (X, S) \in H\} \cup \{((r, Y), T) : (Y, T) \in G\}$$

where $\mathrm{vars}(H) = \mathrm{vars}(G)$. Similarly, the natural definition for *subtraction* would be $H - G := H \setminus G$, but only where $G \subseteq H$. This would be the case if G is drawn from H, thus appearing in the *support* of the *historical distribution*. If not, however, we must map the *event identifiers* with an outer dot product and then do the set minus, for example,

$$H - G := H \setminus \{(D_X, S) : (X, S) \in G\}$$

where $D \in H \ \cdot = \ G$ and $\forall ((X, S), (Y, T)) \in D \ (S = T)$.

The *reciprocal* $(1/)$ is not easy to implement for *history* at all. Division of *histories* may be best left as quotient pair $\mathcal{H} \times \mathcal{H}$, similar to the definition of the rational numbers $\mathbf{Q}$.

A.2 Histogram expressions

The set of *operators* $\mathcal{O}_U$ on *histograms* in *system* U is defined

$$\mathcal{O}_U = \{(+),\ (-),\ (*),\ (1/)\} \cup \{\mathrm{reduce}(V) : V \in \mathrm{P}(\mathrm{vars}(U))\}$$

A *histogram expression* in *system* U is an ordered tree where the nodes are (i) *histograms*, $\mathcal{A}_U$, or (ii) free variable identifiers which are pairs of (a) a natural number and (b) a set of *variables*, $\mathbf{N} \times \mathrm{P}(\mathcal{V})$, or (iii) pairs of (a) an *operator* and (b) a non-empty recursive list of *histogram expressions*, $\mathcal{O}_U \times \mathcal{L}(\mathcal{E}_U)$. The *arity* of the *operator-expression list* pair $(\mathrm{op}, L) \in \mathcal{O}_U \times \mathcal{L}(\mathcal{E}_U)$ is the length of the list, $|L|$. The *reduction operators* are unary. There are $2^{|U|}$ *reduction operators*. The *reciprocal operator* is unary. The *subtraction operator* is binary. The *addition* and *multiplication operators* are n-ary, with notation $\sum$ and $\prod$ respectively, where they are not specifically binary.

The set of all *histogram expressions* $\mathcal{E}_U$ in *system* U is recursively defined as

$$\mathcal{E}_U = \mathcal{A}_U \cup (\mathbf{N} \times \mathrm{P}(\mathcal{V})) \cup (\mathcal{O}_U \times \mathcal{L}(\mathcal{E}_U))$$

where $\mathcal{L}$ is the generic list. The set of *histogram expressions* is a special case of a lambda calculus.

The function vars $\in \mathcal{E}_U \to \mathrm{P}(\mathcal{V})$ returns the *variables* of the root of a *histogram expression*. *Histogram expressions* are constrained such that the free variable identifier pairs, $\mathbf{N} \times \mathrm{P}(\mathcal{V})$, in the *expression* tree form a list. The function free $\in \mathcal{E}_U \to \mathcal{L}(\mathrm{P}(\mathcal{V}))$ returns the list of free variable identifiers. The function paths $\in \mathcal{E}_U \to \mathrm{P}(\mathcal{L}(\mathcal{E}_U))$ returns all paths in the tree between a free variable identifier and the root.

The function substitute $\in \mathcal{L}(\mathcal{A}) \times \mathcal{E}_U \to \mathcal{E}_U$ substitutes the free variables in the *expression* tree. The substitution substitute(L, N) is defined if $|L| = |\mathrm{free}(N)|$ and $\forall i \in \{1 \ldots |L|\}\ (\mathrm{vars}(L_i) = \mathrm{free}(N)(i))$. The substituted *expression* has no free variables, free(substitute$(L, N)) = \emptyset$.

The function evaluate $\in \mathcal{E}_U \to \mathcal{A}_U$ evaluates an *expression*. It is defined only for *expressions* without free variables free$(N) = \emptyset$.

Define a notation of a *histogram expression* $N \in \mathcal{E}_U$ followed by an argument list of *histograms* in round brackets, $N(A)$, $N(A, B)$, etc, for the substitution of the argument list and evaluation of the substituted *histogram expression*. That is, the *histogram expression application* is

$$N(A) = \mathrm{evaluate}(\mathrm{substitute}(\{(1, A)\}, N))$$

and

$$N(A, B) = \mathrm{evaluate}(\mathrm{substitute}(\{(1, A), (2, B)\}, N))$$

and so on.

The implementation of the substitution and evaluation steps are not described here, except to note that there are a number of ways in which an *expression* can be simplified before substitution. *Operators* are left associative, but *multiplications* are commutative and *additions* are commutative. For example, a sequence of *multiplications* on *histograms* in an *expression* can be reduced to a single *histogram*, for example, $A * B * C * D = A * E$ where A is the free variable. A sequence of *reductions* is equivalent to the *reduction* of the intersection of the *variables* $A \;\%\; W \;\%\; V = A \;\%\; (W \cap V)$. An *addition* followed by a *reduction* can be split into the *addition* of *reductions* $(A+B) \;\%\; W = (A \;\%\; W)+(B \;\%\; W)$ where A and B are free variables.

The set of *models* $\mathcal{M}_U$ in *system* U are a subset of *histogram expressions*, $\mathcal{M}_U \subset \mathcal{E}_U$. A *model* $M \in \mathcal{M}_U$ in *variables* V, (i) is a *histogram expression* which has a single free variable *histogram*, free$(M) = \{(1, V)\}$, (ii) *evaluates* to a *histogram* in the same *variables*, vars(M) = vars$(M(A)) = V$ where

$V = \text{vars}(A)$, and (iii) is constrained such that all paths at some point are *reduced* to exclude the argument *variables*

$$\forall P \in \text{paths}(M)\ \exists E \in \text{set}(P)\ (\text{vars}(E) \cap V = \emptyset)$$

Models are *frames* between $M(A)$ and A such that at some point there is an internal representation in *variables* exclusive of the *variables* of start and end nodes of the graph of the *histogram expression*.

The *application* notation can be extended to lists of *histogram expressions*. A *histogram expression list* $R \in \mathcal{L}(\mathcal{E}_U)$ may be applied to an argument *histogram* $A \in \mathcal{A}_U$ in sequence to produce a list of *histograms*. Define *histogram expression list application* $R(A) \in \mathcal{L}(\mathcal{A}_U)$ such that $R(A)_1 = A$, $|R(A)| = |R| + 1$ and $\forall i \in \{1 \ldots |R|\}\ (R(A)_{i+1} = R_i(R(A)_i))$. Each of the *expressions* in the *expression list* has exactly one free variable, $\forall N \in \text{set}(R)\ (|\text{free}(N)| = 1)$, which has the same *variables* as the previous *expression*, $\forall i \in \{1 \ldots |R| - 1\}\ (\text{free}(R_{i+1}) = \{(1, \text{vars}(R_i))\})$. The *histogram expression list*, R, and its *application*, $R(A)$, together resemble a Markov chain if the *histogram expression list* is independent of the *application*. The *application*, $R(A)$, is *size conserving* if the *histograms* have the same *size* as the argument, $\forall B \in \text{set}(R(A))\ (\text{size}(B) = \text{size}(A))$. The *application*, $R(A)$, is *variables conserving* if the *histograms* have the same *variables* as the argument, $\forall B \in \text{set}(R(A))\ (\text{vars}(B) = \text{vars}(A))$. The *application*, $R(A)$, is *congruent* if the *histograms* are *congruent* to the argument, $\forall B \in \text{set}(R(A))\ (\text{congruent}(B, A))$.

Similarly, the *application* notation can be extended to trees of *histogram expressions*. A *histogram expression tree* $R \in \text{trees}(\mathcal{E}_U)$ may be applied to an argument *histogram* $A \in \mathcal{A}_U$ recursively to produce a tree of *histograms*. Define *histogram expression tree application* $R(A) \in \text{trees}(\mathcal{A}_U)$ such that $\text{dom}(\text{roots}(R(A))) = \{A\}$ and $R(A) = \{(A, R'(N(A))) : (N, R') \in R\}$ where $\emptyset(A) = \emptyset$. Each of the *expressions* in the *expression tree* has exactly one free variable, $\forall N \in \text{elements}(R)\ (|\text{free}(N)| = 1)$.

Let $N_{(D,I)} \in \mathcal{E}_U$ be a *histogram expression* of *delta* $(D, I) \in \mathcal{A} \times \mathcal{A}$ having *variables* $\text{vars}(D) = \text{vars}(I) = V$. The *expression application* to *histogram* A in *variables* V is such that $N_{(D,I)}(A) = A - D + I \in \mathcal{A}$. Define

$$N_{(D,I)} = ((+), \{(1, ((-), \{(1, (1, V)), (2, D)\})), (2, I)\})$$

Let $N_R \in \mathcal{E}_U$ be a *histogram expression* of *roll* $R \in \text{rolls}$ having *variables* V. The *expression application* to *histogram* A in *variables* V is such that

$N_R(A) = A * R \in \mathcal{A}$. The *roll histogram expression* can be defined in terms of a *delta histogram expression*, $N_R = N_{(N_D,N_I)}$. The *delta* is defined as

$$\left(\sum_{S \in A^{\mathrm{S}} \cap \mathrm{dom}(R)} \{(S, A_S)\}, \sum_{S \in A^{\mathrm{S}} \cap \mathrm{dom}(R)} \{(R_S, A_S)\} \right)$$

which equals

$$\left(\sum_{S \in A^{\mathrm{S}} \cap \mathrm{dom}(R)} A * \{S\}^{\mathrm{U}}, \sum_{S \in A^{\mathrm{S}} \cap \mathrm{dom}(R)} A * \{S\}^{\mathrm{U}} \ \%\emptyset * \{R_S\}^{\mathrm{U}} \right)$$

so define

$$\begin{aligned} N_D &= ((\Sigma), \{(Q_S, N_S) : S \in \mathrm{dom}(Q)\}) \\ N_I &= ((\Sigma), \{(Q_S, N_{S,R}) : S \in \mathrm{dom}(Q)\}) \end{aligned}$$

where

$$N_S = ((*), \{(1, (1, V)), (2, \{S\}^{\mathrm{U}})\})$$

and

$$N_{S,R} = ((*), \{(1, (\mathrm{reduce}(\emptyset), \{(1, N_S)\})), (2, \{R_S\}^{\mathrm{U}})\})$$

and $Q \in \mathrm{enums}(A^{\mathrm{S}} \cap \mathrm{dom}(R))$.

Let $N_T \in \mathcal{E}_U$ be a *histogram expression* of *transform* $T \in \mathcal{T}$ having *underlying variables* $\mathrm{und}(T) = V$. The *expression application* to *histogram* A in *variables* V is such that $N_T(A) = A * T = A * X \ \% \ W$ where $(X, W) = T$. Define

$$N_T = (\mathrm{reduce}(W), \{(1, ((*), \{(1, (1, V)), (2, X)\}))\})$$

A.3 Cardinality of the power functional definition set

We can calculate an upper bound on the cardinality of the *power functional definition set* on *variables* V in *system* U. Let the *dimension* of the zeroth *layer* $n = |V|$ and the maximum *valency* $d = \mathrm{maxr}(\{|U_v| : v \in V\})$. The number of *layers* $l = \mathrm{bell}(d^n)$. Let $N, D, C \in \mathcal{L}(\mathrm{N})$ and $|N| = |D| = |C| = l + 1$. Let $N_1 = n$ and $D_1 = d$. Let $C_i = \sum_{j \in \{1 \dots i\}} N_j$.

Let D be such that

$$\forall i \in \{1 \dots l\} \ (D_{i+1} = D_i^{C_i})$$

Let N be such that

$$\forall i \in \{1 \dots l\}\ (N_{i+1} = 2^{C_i}\ \mathrm{bell}(D_{i+1}))$$

Then

$$|\mathrm{power}(U)(V)| < C_{l+1} - n$$

A.4 Constructing states order from variables and values orders

Let D_{V} be an order on the *variables* in *system* U, $D_{\mathrm{V}} \in \mathrm{enums}(\mathrm{vars}(U))$. Let $D_{\mathrm{W}} \in \mathcal{V} \to (\mathcal{W} \leftrightarrow \mathbf{N})$ be a set of orders on the *values* on each of the *variables*, such that

$$\forall (v, Y) \in D_{\mathrm{W}}\ (Y \in \mathrm{enums}(U_v))$$

then we can construct an order D_{S} on the *states*, $D_{\mathrm{S}} \in \mathrm{enums}(\mathcal{S}_U)$

$$D_{\mathrm{S}} = \mathrm{order}(\{(S, \sum(t^j i : (v, w) \in S,\ j = D_{\mathrm{V}}(v),\ i = D_{\mathrm{W}}(v)(w))) : S \in \mathcal{S}_U\}, \mathcal{S}_U)$$

where $t = \mathrm{maxr}(\{(v, |W|) : (v, W) \in U\}) + 1$.

A.5 Coders

A *code* is an algorithm or type which defines a *code domain*, an *encode* method and a *decode* method such that there is a bijection between the *code domain* and the natural numbers. Define the set of *codes* as codes. Define $\mathcal{X}$ as the universal set. Define domain $\in$ codes $\to \mathrm{P}(\mathcal{X})$. Define encode $\in$ codes $\to (\mathcal{X} \to \mathbf{N})$ such that $\mathrm{encode}(C) \in \mathrm{domain}(C) :\to \mathbf{N}$. Define decode $\in$ codes $\to (\mathbf{N} \to \mathcal{X})$ such that $\mathrm{decode}(C) \in \mathbf{N} \to: \mathrm{domain}(C)$. Finally constrain *codes* such that

$$\forall C \in \mathrm{codes}\ \forall x \in \mathrm{domain}(C)\ (\mathrm{decode}(C)(\mathrm{encode}(C)(x)) = x)$$

A subset of the *codes* are *list codes*, codeLists $\subset$ codes, for which the *code domains* are lists of objects in a non-empty *listable domain*. Define listable $\in$ codeLists $\to \mathrm{P}(\mathcal{X})$.

$$\forall C \in \mathrm{codeLists}\ \Diamond Y = \mathrm{listable}(C)\ \forall i \in \mathbf{N}\ \forall Q \in Y^i\ (\mathrm{list}(Q) \in \mathrm{domain}(C))$$

An example of a *list code* is $C \in$ codeLists which is parameterised by some bijection $K_C \subset \mathcal{L}(\mathcal{X}) \leftrightarrow \mathbf{N}$. This bijection maps lists of objects to natural numbers. Here the *code domain* $\mathrm{domain}(C) = \mathrm{dom}(K_C)$ and the *listable domain* $Y = \text{listable}(C)$ is such that $\text{dom}(K_C) = \{\text{list}(Q) : i \in \mathbf{N},\ Q \in Y^i\}$. Define *encode* for C as $\text{encode}(C)(L) := K_C(L)$ and *decode* as $\text{decode}(C)(n) := \text{flip}(K_C)(n)$. Thus C *encodes* any list of objects $L \in \mathcal{L}(Y)$ into a natural number $K_C(L) \in \mathbf{N}$ and *decodes* the natural number n back to the original list $(L, n) \in K_C$. Here the *encode* and *decode* functions of this *list code* are easily defined in terms of the relation K_C.

Now consider a subset of *list codes*, called *coders*, coders $\subset$ codeLists, which are not parameterised by straightforward relations, but instead are defined by the *code* and *space* of each object of the *list domain*. In this context *list domains* are called *coder domains*. *Coders* also define the *encode* and *decode* methods in terms of *code* and *space* rather than by indexing a relation with a list of objects. *Coders* redefine the *decode* method to add an extra argument which specifies the length of the list to be decoded. Consider a *coder* $C \in$ coders which is parameterised by a tuple $Q_C \in (\mathcal{X} \leftrightarrow \mathbf{N}) \times (\mathcal{X} \to \mathbf{N}_{>0}) \times (\mathbf{N} \to \mathcal{X})$. *Coder definition* Q_C is a tuple of (i) a *code* represented by a bijection between objects and natural numbers, (ii) a space function of the objects and (iii) a function which decodes the head object of an encoded list. Define the parameterisation of *coders*

$$\text{definition} \in \text{coders} \to (\mathcal{X} \leftrightarrow \mathbf{N}) \times (\mathcal{X} \to \mathbf{N}_{>0}) \times (\mathbf{N} \to \mathcal{X})$$

such that

$$\forall C \in \text{coders}\ \Diamond(E, S, D) = \text{def}(C)\ (\text{dom}(E) = \text{dom}(S) = \text{ran}(D) = \text{listable}(C))$$

and

$$\forall C \in \text{coders}\ \Diamond(E, S, D) = \text{def}(C)\ \forall x \in \text{listable}(C)\ (E_x < S_x)$$

and

$$\forall C \in \text{coders}\ \Diamond(E, S, D) = \text{def}(C)\ (\text{flip}(E) \subset D)$$

Define encode $\in$ coders $\to (\mathcal{L}(\mathcal{X}) \to \mathbf{N})$ and $\text{encode}(C) \in \mathcal{L}(Y) :\to \mathbf{N}$ as

$$\text{encode}(C)(L) := \text{encode}(C)(\text{sequence}(L))$$

where $Y = \text{listable}(C)$.

Define $\text{encode}(C) \in \mathcal{K}(Y) :\to \mathbf{N}$

$$\begin{aligned} \text{encode}(C)((x, K)) &:= \text{encode}(C)(K) \times S_x + E_x \\ \text{encode}(C)(\emptyset) &:= 0 \end{aligned}$$

where $(E, S, D) = \text{definition}(C)$.

Define decode $\in$ coders $\to (\mathbf{N} \times \mathbf{N} \to \mathcal{L}(\mathcal{X}))$ and $\text{decode}(C) \in \mathbf{N} \times \mathbf{N} \to: \mathcal{L}(Y)$ as

$$\text{decode}(C)(l, n) := \text{list}(\text{decode}(C)(l, n))$$

Define $\text{decode}(C) \in \mathbf{N} \times \mathbf{N} \to: \mathcal{K}(Y)$

$$\begin{aligned}\text{decode}(C)(l, n) &:= (D_n, \text{decode}(C)(l - 1, n/S(D_n)))\\ \text{decode}(C)(0, n) &:= \emptyset\end{aligned}$$

The divide operator is the natural number operator.

Now we constrain *coders* to be *list codes*

$$\begin{aligned}&\forall C \in \text{coders} \ \Diamond Y = \text{listable}(C)\\ &\quad \forall i \in \mathbf{N} \ \forall Q \in Y^i \ \Diamond L = \text{list}(Q) \ (\text{decode}(C)(|L|, \text{encode}(C)(L)) = L)\end{aligned}$$

If the *coder domain* is finite then E and S are finite. However D is always infinite because its domain is the encoding of any list of the *coder domain*.

This definition of *coder* allows us to define *space* for each of the elements of the *coder domain*, $\text{space}(C) \in Y :\to \ln \mathbf{N}_{>0}$, where $Y = \text{listable}(C)$,

$$\text{space}(C)(x) := \ln S_x$$

where $(E, S, D) = \text{definition}(C)$.

The *space* required to encode a list $L \in \mathcal{L}(Y)$ is the sum of the *spaces* of the list elements,

$$\sum_{i \in \{1 \ldots |L|\}} \text{space}(C)(L_i) \ \geq \ \ln(\text{encode}(C)(L) + 1)$$

In the case where the *space* is constant, $|\text{ran}(S)| = 1$, the *decode* method can be implemented generically. This is called the *fixed-width* case. All other cases are *variable-width coders*. In the *fixed-width* case, where $\{s\} = \text{ran}(S)$, we can define a *decode* function in terms of the *code* parameter E and the *space* parameter $Y \times \{s\}$ of the *coder definition*

$$\begin{aligned}\text{decode}(C)(l, n) &:= (\text{flip}(E)(n\%s), \text{decode}(C)(l - 1, n/s))\\ \text{decode}(C)(0, n) &:= \emptyset\end{aligned}$$

Modulus and divide are the natural number operators.

The *decode* parameter D, where $(E, S, D) = \text{definition}(C)$, for *variable-width coders* must have an algorithm that is explicitly defined for the *coder*.

Define the subset of *coders* having some particular *listable domain* Y with a convenient construct, $\text{coders}(Y) \subset \text{coders}$

$$\text{coders}(Y) = \{C : C \in \text{coders},\ \text{listable}(C) = Y\}$$

The *total space* of a *coder* of a non-empty finite *coder domain*, Y, is the *space* of any of the lists of Y, $\{\text{flip}(N) : N \in \text{enums}(Y)\}$. By Gibbs' inequality,

$$\sum_{x \in Y} \text{space}(C)(x) \geq |Y| \ln |Y|$$

where $C \in \text{coders}(Y)$. A *minimal coder* is a member of the subset of the *coders* which are such that the *total space* is equal to $|Y| \ln |Y|$. Contrast this to an enumeration of Y which cannot itself be the *space* in a valid *coder* $\forall S \in \text{enums}(Y)\ ((E, S, D) \neq \text{definition}(C))$, because

$$\sum_{x \in Y} \ln S_x = \sum_{i \in \{1 \ldots |Y|\}} \ln i = \ln |Y|! < |Y| \ln |Y|$$

where $S \in \text{enums}(Y)$. We can see that such a *coder* which had an enumeration of Y as the definition of the *space* would have *total space* equal to that required for a permutation of Y, but not a list, and hence Y would not be *listable* by the *coder*.

Consider a non-empty list $L \in \mathcal{L}(Y)$ of a *coder domain* Y. The list, L, implies a *probability function* $P \in (Y :\to \mathbf{Q}_{>0}) \cap \mathcal{P}$ which is such that $\text{dom}(P) = Y$ and $\text{sum}(P) = 1$. Let $Q = \text{count}(\text{flip}(L))$ in

$$P = \{(x, f/\text{sum}(Q)) : (x, f) \in Q\} \cup ((Y \setminus \text{dom}(Q)) \times \{0\})$$

Let *coder* C have *domain* Y so that $\text{decode}(|L|, \text{encode}(C)(L)) = L$. The *expected space* of *coder* C of an element in Y in *probability function* P is

$$\text{expected}(P)(\text{space}(C)) = \sum_{x \in Y} P_x \times \text{space}(C)(x)$$

The *scaled expected space* in the uniform *probability function* $P = Y \times \{1/|Y|\}$ where $\text{flip}(L) \in \text{enums}(Y)$ equals the *total space*

$$|Y| \times \text{expected}(Y \times \{1/|Y|\})(\text{space}(C)) = \text{sum}(\text{space}(C))$$

See the discussion 'Coders and entropy', below, for the case where the *expected space* equals the *entropy* of the *probability function*.

One can go on to consider *coders* of *coders*. That is, *coders* that encode and decode by means of nested *coders*. An example is of the *variable coder* of *histories* which has an *entropy coder* of *states* for each *history*. See appendix 'Entropy encoding of states', below. Another example is where the *coder domain* consists of lists or trees. See appendix 'List and tree coders', below.

A.5.1 List and tree coders

Consider the subset $\mathrm{coders}(\mathcal{L}(\mathcal{Z}))$ of *coders* where the *coder domains* are lists of some underlying type, $\mathcal{Z}$, which itself has a *coder* in $\mathrm{coders}(\mathcal{Z})$. The underlying *coder* is specified in the parameters of the *coders* of lists. Here we shall define two kinds of *coder* where the *coder domain* consists of lists. The first kind, called the *limited coder* of lists has a maximum list length also specified in the parameters, so that the *limited coder domain* is finite if the *underlying coder domain* is also finite. The second kind, called the *unlimited coder* of lists, needs no such parameter, using a termination flag instead. The *coder domain* of the *unlimited coder* of lists is by definition infinite, regardless of whether the *underlying coder domain* is infinite.

Define the *limited coder* of lists

$$\mathrm{coderListLimited} \in \mathrm{coders}(\mathcal{Z}) \times \mathbf{N} \to \mathrm{coders}(\mathcal{L}(\mathcal{Z}))$$

Let $C_{\mathrm{Z}} \in \mathrm{coders}(\mathcal{Z})$ in

$$C_{\mathrm{L}} = \mathrm{coderListLimited}(C_{\mathrm{Z}}, y) \in \mathrm{coders}(\mathcal{L}_y(\mathcal{Z}))$$

where $\mathcal{L}_y(\mathcal{Z})$ is the set of all lists of the *underlying coder domain* of length less than or equal to y, $\mathcal{L}_y(\mathcal{Z}) = \{L : L \in \mathcal{L}(\mathcal{Z}),\ |L| \leq y\}$. $\mathcal{L}_y(\mathcal{Z})$ is a finite set if $\mathcal{Z}$ is finite.

The *code* E_{L} of the *coder definition* of C_{L}, $(E_{\mathrm{L}}, S_{\mathrm{L}}, D_{\mathrm{L}}) = \mathrm{definition}(C_{\mathrm{L}})$, is defined such that a list $L \in \mathcal{L}_y(\mathcal{Z})$ is encoded as a pair of its length and the list itself $(|L|, L)$

$$\begin{aligned} E_{\mathrm{L}}(L) &:= \mathrm{encode}(C_{\mathrm{Z}})(L) \times (y+1) + |L| \\ E_{\mathrm{L}}(\emptyset) &:= 0 \end{aligned}$$

The *space* is defined

$$\begin{aligned} S_{\mathrm{L}}(L) &:= (y+1) \prod_{(i,x) \in L} S_{\mathrm{Z}}(x) \\ S_{\mathrm{L}}(\emptyset) &:= (y+1) \end{aligned}$$

where $(E_{\mathrm{Z}}, S_{\mathrm{Z}}, D_{\mathrm{Z}}) = \mathrm{definition}(C_{\mathrm{Z}})$. Define the *decode* parameter

$$\begin{aligned} D_{\mathrm{L}}(n) &:= \mathrm{decode}(C_{\mathrm{Z}})(n\%(y+1), n/(y+1)) \\ D_{\mathrm{L}}(0) &:= \emptyset \end{aligned}$$

If we take the core *space* of a list L as $\sum_{(i,x) \in L} \mathrm{space}(C_{\mathrm{Z}})(x)$, then the additional *overhead space* to encode the list in *coder* C_{L} is $\ln(y+1)$.

In the case of a *fixed-width underlying coder*, $S_{\mathrm{Z}} = \mathcal{Z} \times \{s\}$, then we can simplify. Define $\mathrm{decLim}(E_{\mathrm{Z}}, s) \in \mathbf{N} \times \mathbf{N} \to \mathcal{K}(\mathcal{Z})$.

$$\begin{aligned} D_{\mathrm{L}}(n) &:= \mathrm{list}(\mathrm{decLim}(E_{\mathrm{Z}}, s)(n\%(y+1), n/(y+1))) \\ D_{\mathrm{L}}(0) &:= \emptyset \end{aligned}$$

and

$$\begin{aligned} \mathrm{decLim}(E_{\mathrm{Z}}, s)(i, n) &:= (\mathrm{flip}(E_{\mathrm{Z}})(n\%s), \mathrm{decLim}(E_{\mathrm{Z}}, s)(i-1, n/s)) \\ \mathrm{decLim}(E_{\mathrm{Z}}, s)(0, n) &:= \emptyset \end{aligned}$$

The *space* parameter of a list is $S_{\mathrm{L}}(L) := (y+1)s^{|L|}$ and so $\mathrm{space}(C_{\mathrm{L}})(L) = \ln(y+1) + |L| \ln s$.

A closely related *limited list coder* $C_{\mathrm{L,N}}$ allows only non-empty lists in the *coder domain*

$$C_{\mathrm{L,N}} = \mathrm{coderListNonEmptyLimited}(C_{\mathrm{Z}}, y) \in \mathrm{coders}(\mathcal{L}_y(\mathcal{Z}) \setminus \{\emptyset\})$$

The *code* $E_{\mathrm{L,N}}$ is defined $E_{\mathrm{L,N}}(L) := \mathrm{encode}(C_{\mathrm{Z}})(L) \times y + |L| - 1$, the *space* $S_{\mathrm{L,N}}$ is defined $S_{\mathrm{L,N}}(L) := y \prod_{(i,x) \in L} S_{\mathrm{Z}}(x)$ and the *decode* $D_{\mathrm{L,N}}$ parameter is defined $D_{\mathrm{L,N}}(n) := \mathrm{list}(\mathrm{decLim}(S_{\mathrm{Z}}, D_{\mathrm{Z}})(n\%y+1, n/y))$. The difference between C_{L} and $C_{\mathrm{L,N}}$ is that the *space* of a non-empty list is smaller in $C_{\mathrm{L,N}}$. The *overhead space* to encode the list in *coder* $C_{\mathrm{L,N}}$ is $\ln y$.

Also related to *limited list coders* are *set coders* where the *coder domain* is not a list of some *underlying coder domain*, $\mathcal{L}(\mathcal{Z})$ but instead a set, $\mathrm{P}(\mathcal{Z})$.

Set coder domains are finite and their *underlying coder domains* are also finite.

$$C_{\mathrm{S}} = \mathrm{coderSet}(C_{\mathrm{Z}}) \in \mathrm{coders}(\mathrm{P}(\mathcal{Z}))$$

C_{S} encodes a subset $Q \subset \mathcal{Z}$ by first encoding the subset cardinality $|Q|$, then the combination. The combination maps to a list of the underlying objects ordered by the *encode* parameter of the *underlying coder definition*, $\mathrm{order}(E_{\mathrm{Z}}, Q)$. The *space* is $\mathrm{space}(C_{\mathrm{S}})(Q) = \ln(y+1)+\ln(y!/(z!(y-z)!)$ where $y = |\mathcal{Z}|$ and $z = |Q|$. Note that there is no need to encode the underlying directly. An alternative method is to encode a list of bits which signify set membership. In this method the *space* comes to $y \ln 2$.

The second kind of *coder*, which is the *unlimited coder* of lists, uses a termination flag to stop the decode iteration. The *coder domain* is infinite, whether the *underlying coder domain* is finite or not. Define the *unlimited coder* of lists

$$\mathrm{coderListTerminating} \in \mathrm{coders}(\mathcal{Z}) \to \mathrm{coders}(\mathcal{L}(\mathcal{Z}))$$

Let $C_{\mathrm{Z}} \in \mathrm{coders}(\mathcal{Z})$ in

$$C_{\mathrm{U}} = \mathrm{coderListTerminating}(C_{\mathrm{Z}}) \in \mathrm{coders}(\mathcal{L}(\mathcal{Z}))$$

The *code* E_{U} of the *coder definition* of C_{U}, $(E_{\mathrm{U}}, S_{\mathrm{U}}, D_{\mathrm{U}}) = \mathrm{definition}(C_{\mathrm{U}})$, is defined such that a list $L \in \mathcal{L}(\mathcal{Z})$ is encoded as a list of pairs of a continuation code $0 \in \mathrm{bits}$ and the element of the list itself, or the termination code $1 \in \mathrm{bits}$, $\mathcal{L}((\{0\} \times \mathcal{Z}) \cup \{1\})$

$$E_{\mathrm{U}}(L) := \mathrm{encTerm}(E_{\mathrm{Z}}, S_{\mathrm{Z}})(\mathrm{sequence}(L))$$

where $(E_{\mathrm{Z}}, S_{\mathrm{Z}}, D_{\mathrm{Z}}) = \mathrm{definition}(C_{\mathrm{Z}})$. Define $\mathrm{encTerm}(E_{\mathrm{Z}}, S_{\mathrm{Z}}) \in \mathcal{K}(\mathcal{Z}) \to \mathbf{N}$ as

$$\begin{aligned}
\mathrm{encTerm}(E_{\mathrm{Z}}, S_{\mathrm{Z}})((x, K)) &:= (\mathrm{encTerm}(E_{\mathrm{Z}}, S_{\mathrm{Z}})(K) \times S_{\mathrm{Z}}(x) + E_{\mathrm{Z}}(x)) \times 2 \\
\mathrm{encTerm}(E_{\mathrm{Z}}, S_{\mathrm{Z}})(\emptyset) &:= 1
\end{aligned}$$

The *space* is defined

$$\begin{aligned}
S_{\mathrm{U}}(L) &:= 2 \prod_{(i,x) \in L} 2 S_{\mathrm{Z}}(x) = 2^{(|L|+1)} \prod_{(i,x) \in L} S_{\mathrm{Z}}(x) \\
S_{\mathrm{U}}(\emptyset) &:= 2
\end{aligned}$$

Define the *decode* parameter

$$\begin{aligned} D_{\mathrm{U}}(n) &:= \mathrm{list}(\mathrm{decTerm}(S_{\mathrm{Z}}, D_{\mathrm{Z}})(n)) \\ D_{\mathrm{U}}(1) &:= \emptyset \\ D_{\mathrm{U}}(0) &:= \emptyset \end{aligned}$$

Define $\mathrm{decTerm}(S_{\mathrm{Z}}, D_{\mathrm{Z}}) \in \mathbf{N} \to \mathcal{K}(\mathcal{Z})$ as

$$\mathrm{decTerm}(S_{\mathrm{Z}}, D_{\mathrm{Z}})(n) := (D_{\mathrm{Z}}(n/2), \mathrm{decTerm}(S_{\mathrm{Z}}, D_{\mathrm{Z}})(n/(2S_{\mathrm{Z}}(D_{\mathrm{Z}}(n/2)))))$$

where $n\%2 = 0$, otherwise $\mathrm{decTerm}(S_{\mathrm{Z}}, D_{\mathrm{Z}})(n) := \emptyset$.

The *overhead space* to encode the list in *coder* C_{U} is $(|L| + 1) \ln 2$.

In the case of a *fixed-width underlying coder*, $S_{\mathrm{Z}} = \mathcal{Z} \times \{s\}$, that by definition must have a finite *coder domain* $\mathcal{Z}$, we can replace the termination flag with a terminating code number s, $\mathcal{L}(\mathcal{Z} \cup \{s\})$

$$C_{\mathrm{U,F}} = \mathrm{coderListTerminatingFixed}(C_{\mathrm{Z}}) \in \mathrm{coders}(\mathcal{L}(\mathcal{Z}))$$

Then

$$E_{\mathrm{U,F}}(L) := \mathrm{encTerm}(E_{\mathrm{Z}}, s)(\mathrm{sequence}(L))$$

Define $\mathrm{encTerm}(E_{\mathrm{Z}}, s) \in \mathcal{K}(\mathcal{Z}) \to \mathbf{N}$ as

$$\begin{aligned} \mathrm{encTerm}(E_{\mathrm{Z}}, s)((x, K)) &:= \mathrm{encTerm}(E_{\mathrm{Z}}, s)(K) \times (s + 1) + E_{\mathrm{Z}}(x) \\ \mathrm{encTerm}(E_{\mathrm{Z}}, s)(\emptyset) &:= s \end{aligned}$$

The *space* is defined

$$S_{\mathrm{U,F}}(L) := (s + 1)^{|L|+1}$$

Define the *decode* parameter

$$\begin{aligned} D_{\mathrm{U,F}}(n) &:= \mathrm{list}(\mathrm{decTerm}(E_{\mathrm{Z}}, s)(n)) \\ D_{\mathrm{U,F}}(s) &:= \emptyset \end{aligned}$$

Define $\mathrm{decTerm}(E_{\mathrm{Z}}, s) \in \mathbf{N} \to \mathcal{K}(\mathcal{Z})$ as

$$\mathrm{decTerm}(E_{\mathrm{Z}}, s)(n) := (\mathrm{flip}(E_{\mathrm{Z}})(n\%(s + 1)), \mathrm{decTerm}(E_{\mathrm{Z}}, s)(n/(s + 1)))$$

where $n\%(s + 1) \neq s$, otherwise $\mathrm{decTerm}(E_{\mathrm{Z}}, s)(n) := \emptyset$.

The *overhead space* to encode the list in *coder* $C_{\mathrm{U,F}}$ is $(|L| + 1) \ln(s + 1) - |L| \ln s$.

Another *unlimited list coder*, $C_{\mathrm{U,N}}$, allows only non-empty lists in the *coder domain*

$$C_{\mathrm{U,N}} = \mathrm{coderListNonEmptyTerminating}(C_{\mathrm{Z}}) \in \mathrm{coders}(\mathcal{L}(\mathcal{Z}) \setminus \{\emptyset\})$$

The *code* $E_{\mathrm{U,N}}$ of the *coder definition* of $C_{\mathrm{U,N}}$ is defined such that a non-empty list $L \in \mathcal{L}(\mathcal{Z} \setminus \{\emptyset\})$ is encoded as a list of pairs of a termination flag and the element of the list itself, $\mathcal{L}(\mathrm{bits} \times \mathcal{Z})$. The *overhead space* to encode the list in *coder* $C_{\mathrm{U,N}}$ is $|L| \ln 2$.

In some cases the *underlying coder* depends on the list so far. In order to implement this requirement we define a function of the list which returns the *underlying coder*. These *list coders* are called *lookback list coders*. For example the *lookback unlimited list coder*

$$\mathrm{coderListTerminatingLookback} \in (\mathcal{L}(\mathcal{Z}) \to \mathrm{coders}(\mathcal{Z})) \to \mathrm{coders}(\mathcal{L}(\mathcal{Z}))$$

Let $B_{\mathrm{Z}} \in \mathcal{L}(\mathcal{Z}) \to \mathrm{coders}(\mathcal{Z})$ in

$$C_{\mathrm{U,B}} = \mathrm{coderListTerminatingLookback}(B_{\mathrm{Z}}) \in \mathrm{coders}(\mathcal{L}(\mathcal{Z}))$$

The *coder* $C_{\mathrm{U,B}}$ is defined exactly as C_{U} above except that we replace $(E_{\mathrm{Z}}, S_{\mathrm{Z}}, D_{\mathrm{Z}}) = \mathrm{definition}(C_{\mathrm{Z}})$ with $(E_{\mathrm{Z}}, S_{\mathrm{Z}}, D_{\mathrm{Z}}) = \mathrm{definition}(B_{\mathrm{Z}}(M))$ where $M \in \mathcal{L}(\mathcal{Z})$ is the list already encoded or decoded. For example

$$E_{\mathrm{U,B}}(L) := \mathrm{encTerm}(B_{\mathrm{Z}})(\emptyset, \mathrm{sequence}(L))$$

Define $\mathrm{encTerm}(B_{\mathrm{Z}}) \in \mathcal{K}(\mathcal{Z}) \times \mathcal{K}(\mathcal{Z}) \to \mathbf{N}$ as

$$\begin{aligned} \mathrm{encTerm}(B_{\mathrm{Z}})(M, (x, K)) &:= (\mathrm{encTerm}(B_{\mathrm{Z}})((x, M), K) \times S_{\mathrm{Z}}(x) + E_{\mathrm{Z}}(x)) \times 2 \\ \mathrm{encTerm}(B_{\mathrm{Z}})(M, \emptyset) &:= 1 \end{aligned}$$

where $(E_{\mathrm{Z}}, S_{\mathrm{Z}}, D_{\mathrm{Z}}) = \mathrm{definition}(B_{\mathrm{Z}}(\mathrm{reverse}(\mathrm{list}(M))))$.

The *underlying coder function* B_{Z} is infinite, but the *space* of $C_{\mathrm{U,B}}$ is well defined given a finite argument list.

Consider the *coder* of trees, $\mathrm{trees}(\mathcal{Z})$, of some underlying type $\mathcal{Z}$. Trees are functions and hence are sets, so it is possible to define a *limited coder*, as above, say by defining the maximum depth and cardinality of tree. There are a number of different ways in which limits could be imposed. However, here we shall consider unlimited trees with the constraint that circularities are excluded. Sets can only be encoded as sets if they are finite, so to encode an unlimited tree $T \in \mathrm{trees}(\mathcal{Z})$ we must first convert it to a list tree $\mathrm{listTrees}(\mathcal{Z})$ with an order $D \in \mathcal{Z} \leftrightarrow \mathbf{N}$, $\mathrm{listTree}(D, T) \in \mathrm{listTrees}(\mathcal{Z})$, and then use an *unlimited list tree coder*

$$\mathrm{coderListTree} \in \mathrm{coders}(\mathcal{Z}) \to \mathrm{coders}(\mathrm{listTrees}(\mathcal{Z}))$$

Let $C_{\mathrm{Z}} \in \mathrm{coders}(\mathcal{Z})$ in

$$C_{\mathrm{U,T}} = \mathrm{coderListTree}(C_{\mathrm{Z}}) \in \mathrm{coders}(\mathrm{listTrees}(\mathcal{Z}))$$

The *code* $E_{\mathrm{U,T}}$ of the *coder definition* of $C_{\mathrm{U,T}}$, where $(E_{\mathrm{U,T}}, S_{\mathrm{U,T}}, D_{\mathrm{U,T}}) = \mathrm{definition}(C_{\mathrm{U,T}})$, is defined similarly to the encode of an *unlimited list coder* C_{U}, except that recursion excludes the current list tree to prevent circularities

$$E_{\mathrm{U,T}}(L) := \mathrm{encListTree}(C'_{\mathrm{U,T}}, E_{\mathrm{Z}}, S_{\mathrm{Z}})(\mathrm{sequence}(L))$$

where $C'_{\mathrm{U,T}} \in \mathrm{coders}(\mathrm{listTrees}(\mathcal{Z}) \setminus \{L\})$ and $(E_{\mathrm{Z}}, S_{\mathrm{Z}}, D_{\mathrm{Z}}) = \mathrm{definition}(C_{\mathrm{Z}})$. $C'_{\mathrm{U,T}}$ is equal to $C_{\mathrm{U,T}}$ except that it is undefined for list tree L, $L \notin \mathrm{dom}(E'_{\mathrm{U,T}})$. Define $\mathrm{encListTree}(C_{\mathrm{U,T}}, E_{\mathrm{Z}}, S_{\mathrm{Z}}) \in \mathcal{K}(\mathcal{Z}) \to \mathbf{N}$ as

$$\begin{aligned}&\mathrm{encListTree}(C_{\mathrm{U,T}}, E_{\mathrm{Z}}, S_{\mathrm{Z}})(((x, M), K)) := \\ &\quad((\mathrm{encListTree}(C'_{\mathrm{U,T}}, E_{\mathrm{Z}}, S_{\mathrm{Z}})(K) \times S_{\mathrm{Z}}(x) + E_{\mathrm{Z}}(x)) \times S_{\mathrm{U,T}}(M) + E_{\mathrm{U,T}}(M)) \times 2\end{aligned}$$

where $\mathrm{encListTree}(C_{\mathrm{U,T}}, E_{\mathrm{Z}}, S_{\mathrm{Z}})(\emptyset) := 1$ and $(E_{\mathrm{U,T}}, S_{\mathrm{U,T}}, D_{\mathrm{U,T}}) = \mathrm{definition}(C_{\mathrm{U,T}})$ and $\mathrm{listable}(C'_{\mathrm{U,T}}) = \mathrm{listable}(C_{\mathrm{U,T}}) \setminus \{\mathrm{list}(((x, M), K))\}$.

The *space* is defined

$$\begin{aligned}S_{\mathrm{U,T}}(L) &:= 2 \prod_{(i,(x,M)) \in L} 2 \times S_{\mathrm{Z}}(x) \times S'_{\mathrm{U,T}}(M) \\ S_{\mathrm{U,T}}(\emptyset) &:= 2\end{aligned}$$

Define the *decode* parameter

$$\begin{aligned}D_{\mathrm{U,T}}(n) &:= \mathrm{list}(\mathrm{decListTree}(C'_{\mathrm{U,T}}, S_{\mathrm{Z}}, D_{\mathrm{Z}})(n)) \\ D_{\mathrm{U,T}}(1) &:= \emptyset \\ D_{\mathrm{U,T}}(0) &:= \emptyset\end{aligned}$$

$C'_{\mathrm{U,T}}$ is equal to $C'_{\mathrm{U,T}}$ except that $n \notin \mathrm{dom}(D'_{\mathrm{U,T}})$. Define $\mathrm{decListTree}(C_{\mathrm{U,T}}, S_{\mathrm{Z}}, D_{\mathrm{Z}}) \in \mathbf{N} \to \mathcal{K}(\mathcal{Z})$ as

$$\mathrm{decListTree}(C_{\mathrm{U,T}}, S_{\mathrm{Z}}, D_{\mathrm{Z}})(n) := ((x, M), \mathrm{decListTree}(C'_{\mathrm{U,T}}, S_{\mathrm{Z}}, D_{\mathrm{Z}})(m/S_{\mathrm{Z}}(x)))$$

where $M = D_{\mathrm{U,T}}(n/2)$, $m = n/(2S_{\mathrm{U,T}}(M))$, $x = D_{\mathrm{Z}}(m)$ if $n\%2 = 0$ otherwise $\mathrm{decListTree}(C_{\mathrm{U,T}}, S_{\mathrm{Z}}, D_{\mathrm{Z}})(n) := \emptyset$.

If we take the core *space* of a list tree L as $\sum_{(i,x) \in Q} \mathrm{space}(C_{\mathrm{Z}})(x)$, where $Q = \mathrm{concat}(L)$ is the depth-first traversal concatenation, then the additional *overhead space* to encode the list tree in *coder* $C_{\mathrm{U,T}}$ is $(2|Q| + 1) \ln 2$.

We can define a *lookback unlimited list tree coder* in a similar manner to the *lookback unlimited list coder* above. Again we supply a function returning *underlying coders* but here it has two arguments. The first argument is the previous sibling list tree and the second is the sub list tree for this node

$$\begin{aligned}&\mathrm{coderListTreeLookback} \in \\ &\quad (\mathrm{listTrees}(\mathcal{Z}) \times \mathrm{listTrees}(\mathcal{Z}) \to \mathrm{coders}(\mathcal{Z})) \to \mathrm{coders}(\mathrm{listTrees}(\mathcal{Z}))\end{aligned}$$

Let $B_{\mathrm{Z}} \in \mathrm{listTrees}(\mathcal{Z}) \times \mathrm{listTrees}(\mathcal{Z}) \to \mathrm{coders}(\mathcal{Z})$ in

$$C_{\mathrm{U,T,B}} = \mathrm{coderListTreeLookback}(B_{\mathrm{Z}}) \in \mathrm{coders}(\mathrm{listTrees}(\mathcal{Z}))$$

Then the encode and decode methods are as above except that $(E_{\mathrm{Z}}, S_{\mathrm{Z}}, D_{\mathrm{Z}}) = \mathrm{definition}(B_{\mathrm{Z}}(L, M))$ in the definition of $\mathrm{encListTree}(C_{\mathrm{U,T,B}}, E_{\mathrm{Z}}, S_{\mathrm{Z}})(L, ((x, M), K))$ where L is the previous sibling list tree and M is the child list tree.

If a list L contains the same object in more than place, $|\mathrm{ran}(L)| < |L|$, we can possibly reduce the *space* of a *list coder* by using *references* to the objects. A *reference* is a position in the list, $\mathrm{dom}(L) \subset \mathbf{N}_{>0}$. A *referencing list coder*, having *coder domain* of lists, seeks to use *references* rather than encoding objects in order to reduce *space*. Encoding a *referencing coder* requires a check on each object in the list to see if it already exists elsewhere in the list and if so creating a *reference* pointing to the existing object. Decoding a *reference* is *dereferencing*. Although the *overhead space* of a *referencing list coder* is larger than for a non-referencing *list coder*, requiring *space* for a flag to indicate whether the head object is a *reference* or a literal object and also *space* for the position in a *reference*, the *space* of highly redundant lists having large *space* of objects may be lower.

Referencing list tree coders can be constructed too, for example by defining a *reference* as a position in the concatenated tree. The position is in $\mathrm{dom}(\mathrm{concat}(\mathrm{listTree}(D, T))) \subset \mathbf{N}_{>0}$ for some tree T and order D. Another method would be to define the *reference* as a tuple in $\bigcup\{\mathbf{N}_{>0}^i : i \in \mathbf{N}_{>0}\}$ of the node position in the list tree. Note that these *references* refer to the objects of the nodes. That is, the first of the pair.

A graph is a tree that contains the same node at more than one position in the tree, $|\text{nodes}(T)| < |\text{concat}(\text{listTree}(D,T))|$ for some tree T and any order D. If a duplicate node has children and the recursive set of descendants contains the node, then there is a circularity. We can encode graphs by using *node references*. A *node reference* is similar to a *reference* except that it refers to the entire node. That is, the pair of the object and children. If the children of a node is empty then a *node reference* is equivalent to a *reference*.

A.5.2 Coders and entropy

If a finite *coder domain* Y has some *probability function* $P \in (Y :\to \mathbf{Q}_{>0}) \cap \mathcal{P}$, which is such that $\text{dom}(P) = Y$ and $\text{sum}(P) = 1$, associated with it, then we may be able to construct an *entropy coder* $C \in \text{coders}(Y)$ such that

$$\forall x \in Y\ (\text{space}(C)(x) = \ln \frac{1}{P_x})$$

That is, the *space* of an element of the *coder domain* is the logarithm of the surprisal. In most cases of probability functions an *entropy coder* cannot be constructed. The *coder* requires that

$$\forall x \in Y\ (\frac{1}{P_x} \in \mathbf{N})$$

at least. Also there are constraints on the parameters of the *coder definition* $(E, S, D) = \text{definition}(C)$.

If the *entropy coder* exists, the *expected space* of an element in the *coder domain* is

$$\begin{aligned}&\text{expected}(P)(\text{space}(C)) = \\ &\quad \sum_{x \in Y} P_x \times \text{space}(C)(x) = \sum_{x \in Y} P_x \ln \frac{1}{P_x} = -\sum_{x \in Y} P_x \ln P_x = \text{entropy}(P)\end{aligned}$$

The *scaled expected space* $|Y| \times \text{entropy}(P)$ is greater than or equal to the *minimal space* of the *coder domain* $|Y| \ln |Y|$, with equality only for the uniform *probability function* $P = Y \times \{1/|Y|\}$.

An *entropy coder* C_e has the smallest expected *space* of all *coders* given the *probability function*, P, because the relative entropy of any other *coder* is positive by Gibbs' Inequality,

$$\forall C \in \text{coders}(\text{dom}(P))\ (\text{expected}(P)(C^\text{s}) \geq \text{expected}(P)(C_\text{e}^\text{s}))$$

where expected$(P)(C_{\mathrm{e}}^{\mathrm{s}}) = \mathrm{entropy}(P)$.

A.5.3 Binary coders

An example of a *fixed-width coder* is

$$\mathrm{coderBitstringShortest} \in \mathrm{coders}(\mathrm{bits})$$

Let $C_{\mathbf{B}} = \mathrm{coderBitstringShortest}$ and $(E, S, D) = \mathrm{definition}(C_{\mathbf{B}})$. Let $E = \{(0, 0), (1, 1)\}$, $S = \mathrm{bits} \times \{2\}$ and $D(n) := n\%2$. A *bits coder* implies a bijection $\mathbf{N} \leftrightarrow \mathcal{L}(\mathrm{bits})$. In this case it provides a means to map any number to the shortest *bitstring* that can contain it. Obviously this is useful in a physical implementation in computer memory.

We can apply *bits coder* $C_{\mathbf{B}}$ to the *encode* method of *coder* $C \in \mathrm{coders}(Y)$ to produce a *bitstring*, $\mathrm{decode}(C_{\mathbf{B}})(\mathrm{encode}(C)(L)) \in \mathcal{L}(\mathrm{bits})$ where $L \in \mathcal{L}(Y)$. If it is also the case that the *space* parameter S, where $(E, S, D) = \mathrm{definition}(C)$, is always a multiple of two, $\forall x \in Y\ (S_x \in \{2^n : n \in \mathrm{N}\})$ then C is a *binary coder* and the maximum length of any *bitstring* is constrained

$$\forall L \in \mathcal{L}(Y)\ (|\mathrm{decode}(C_{\mathbf{B}})(|L|, \mathrm{encode}(C)(L))| \leq \sum_{(i,x) \in L} \log_2(S_x))$$

This means that, for *binary coders*, the *space* required for implementation in a *bitstring* is no greater than the *space* required for the *coder* itself. (Note that the function $\mathrm{space}(C) \in Y \to \ln \mathbf{N}_{>0}$ is defined as the natural logarithm rather than the base 2 logarithm.)

A *prefix-free coder* C is a *binary coder* constrained such that

$$\sum(\frac{1}{S_x} : (x, i) \in E) \leq 1$$

where $(E, S, D) = \mathrm{definition}(C)$. If it is the case that the sum is 1 and there is a probability mass function on the *coder domain*, $P \in Y \to \mathbf{Q}_{>0}$, such that $P = \{(x, 1/S_x) : x \in Y\}$, then C is also an *entropy coder* for P. However, no *prefix-free coder* can be a *minimal coder* because the sum of a *fixed-width prefix-free coder*, required for a uniform probability mass function, must be less than 1.

A.6 Entropy encoding of states

The *index coder* C_{H} which encodes a *history's events space* into a list of *states*, $\mathcal{L}(\mathcal{S})$, and thence to a list of natural numbers, $\mathcal{L}(\mathbf{N})$, above, is a *fixed-width coder* of the *states* because the *space* required is proportional to the *size*, $\text{spaceEvents}(U)(H) := z \ln v$. However, a *variable-width coder* of the *states* may require less *space*. Define a theoretical *variable-width coder* C_{E} of *histories*

$$C_{\mathrm{E}} = \text{coderHistoryVariable}(U, X, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \text{coders}(\mathcal{H}_{U,X})$$

This *coder* contains a nested *entropy coder* of *states* which is constructed separately for the *histogram* A of each given *history* of the *listable domain* $H \in \mathcal{H}_{U,X}$, where $A = \text{histogram}(A)$. The *entropy coder* of *states* enables us to *encode* a *variable-width* list of *states* which is ordered in the same way as the *event identifiers* of the given *history* H. Let $C \in \text{coders}(\text{states}(A))$, having *definition* $(E, S, D) = \text{definition}(C)$, be defined such that $S = (A\%\emptyset)/A$, where $S \in \text{states}(A) \to \mathbf{N}_{>0}$. Here we assume that the *histogram* A happens to define an *entropy coder*

$$\forall R \in \text{states}(A)\ (\text{space}(C)(R) = \ln \frac{1}{P_R} = \ln \frac{z}{A_R})$$

where $P = A/(A\%\emptyset)$ and $z = \text{size}(A)$. Of course, this is rarely the case, but we shall assume that it is true in order to determine the minimum *space* of *variable coder* C_{E}. The *space* parameter, $S \in \text{states}(A) \to \mathbf{N}_{>0}$, of the *entropy coder* of *states*, C, is defined by A, so the *space* to *encode* S itself is the same as the *coder* of *histograms*, C_{A} previously defined as the sum of the *variables space*, *size space* and *counts space*

$$\text{space}(C_{\mathrm{A}})(A) = \text{spVar}(U)(|\text{vars}(A)|) + \text{space}(|X| + 1) + \text{spCt}(U)(A)$$

where spVar = spaceVariables, and spCt = spaceCounts. We shall not define the *encode* parameter $E \in \text{states}(A) \to \mathbf{N}$, but merely determine the *space* required for a definition. Define $\text{spaceCodeEntropy} \in \mathcal{A}_{\mathrm{i}} \to \ln \mathbf{N}_{>0}$ so that

$$\text{spaceCodeEntropy}(A) = \sum_{R \in A^{\mathrm{S}}} \ln S_R$$

as

$$\text{spaceCodeEntropy}(A) := \sum_{R \in A^{\mathrm{S}}} \ln \frac{z}{A_R}$$

where $z = \text{size}(A)$, $z > 0$ and $A = \text{trim}(A)$. The *space* of E is $\text{spaceEntropy}(A)$.

Not only shall we not define the infinite relation $D \in \mathcal{L}(\text{states}(A)) \to \mathbf{N}$, we shall completely ignore the *space* required to define it. In practice we can define *coders* of *states* using algorithms, such as the Huffman *binary coder*, which can be *entropy coders*. For these the *decode* algorithm can be considered part of the definition of the *coder*.

Having defined the minimum *space* of the *entropy coder* of *states*, C, for the *history* H we can define the *variable events space* spaceEventsVariable $\in \mathcal{A}_{\mathrm{i}} \to \ln \mathbf{N}_{>0}$ so that

$$\mathrm{spaceEventsVariable}(A) = \sum_{R \in A^{\mathrm{S}}} A_R \ln S_R$$

as

$$\mathrm{spaceEventsVariable}(A) := \sum_{R \in A^{\mathrm{S}}} A_R \ln \frac{z}{A_R}$$

where $z = \mathrm{size}(A)$, $z > 0$ and $A = \mathrm{trim}(A)$. The *variable events space* is the *sized entropy*, $\mathrm{spaceEventsVariable}(A) = z \times \mathrm{entropy}(A)$.

The total minimum *space* of a theoretical *variable coder* of a *history* H is the sum of the *variables space*, *ids space*, *histogram counts space*, *entropy code space* and *variable events space*

$$\begin{aligned}
&\mathrm{space}(C_{\mathrm{E}})(H) = \\
&\quad \mathrm{spVar}(U)(|\mathrm{vars}(H)|) + \mathrm{spId}(|X|, |H|) + \mathrm{spCt}(U)(A) + \mathrm{spEnt}(A) + \mathrm{spEvVar}(A)
\end{aligned}$$

where $A = \mathrm{histogram}(H)$, $\mathrm{spVar} = \mathrm{spaceVariables}$, $\mathrm{spId} = \mathrm{spaceIds}$, $\mathrm{spCt} = \mathrm{spaceCounts}$, $\mathrm{spEnt} = \mathrm{spaceCodeEntropy}$ and $\mathrm{spEvVar} = \mathrm{spaceEventsVariable}$.

If we compare the *space* of the *variable coder* to that of the *classification coder* of *histories* we find that the former is always greater than or equal to the latter

$$\begin{aligned}
&\mathrm{space}(C_{\mathrm{E}})(H) - \mathrm{space}(C_{\mathrm{G}})(H) = \\
&\qquad \mathrm{spEnt}(A) + \mathrm{spEvVar}(A) - \mathrm{spCl}(A) \\
&= \sum_{R \in A^{\mathrm{S}}} (A_R + 1) \ln \frac{z}{A_R} - \left(\ln z! - \sum_{R \in A^{\mathrm{S}}} \ln A_R! \right) \\
&= (z + |A|) \ln z - \ln z! - \sum_{R \in A^{\mathrm{S}}} ((A_R + 1) \ln A_R - \ln A_R!) \\
&\geq 0
\end{aligned}$$

where $\mathrm{spCl} = \mathrm{spaceClassification}$. Even if we ignore $\mathrm{spaceCodeEntropy}(A)$, for example in the case of a *variable-width history coder* of the subset of *histories* having a constant parameter *histogram*, $C_{\mathrm{E}} \in \mathrm{coders}(\{H : H \in \mathcal{H}_{U,X},\ \mathrm{histogram}(H) = A\})$, the *space* of the *variable coder* is still greater than or equal to that of the *classification coder* because the log unit-translated gamma function, $\ln \Gamma_! x$, is convex with respect to the log-linear function, $x \ln x$.

A.7 Independent histogram space

The *integral congruent support* of the *multinomial distribution* in *variables* V and *size* z in *system* U is defined above as

$$\mathcal{A}_{U,\mathrm{i},V,z} = \{A : A \in \mathcal{A}_{U,\mathrm{i}},\ A^{\mathrm{U}} = V^{\mathrm{C}},\ \mathrm{size}(A) = z\}$$

The cardinality of which is that of the weak compositions $|\mathrm{C}'(V^{\mathrm{C}}, z)|$

$$|\mathcal{A}_{U,\mathrm{i},V,z}| = \frac{(z+v-1)!}{z!\ (v-1)!}$$

where $v = \mathrm{volume}(U)(V)$.

Define the subset of the *integral congruent support* that consists of *independent histograms* as

$$\mathcal{A}_{U,\mathrm{i},V,z,\mathrm{x}} = \{A : A \in \mathcal{A}_{U,\mathrm{i},V,z},\ A = A^{\mathrm{X}}\}$$

So

$$|\mathcal{A}_{U,\mathrm{i},V,z,\mathrm{x}}| \leq \frac{(z+v-1)!}{z!\ (v-1)!}$$

As defined above, the subset of the *independent* function, $Y_{U,\mathrm{i},V,z} = \{(A, A^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},V,z}\} \subset \mathrm{independent}$, partitions the *integral congruent support*, $\mathrm{ran}(\mathrm{inverse}(Y_{U,\mathrm{i},V,z})) \in \mathrm{B}(\mathcal{A}_{U,\mathrm{i},V,z})$. The cardinality of its range is

$$|\mathrm{ran}(Y_{U,\mathrm{i},V,z})| = \prod_{u \in V} \frac{(z+|U_u|-1)!}{z!\ (|U_u|-1)!}$$

Now $\mathcal{A}_{U,\mathrm{i},V,z,\mathrm{x}} \subseteq \mathrm{ran}(Y_{U,\mathrm{i},V,z})$ and so

$$|\mathcal{A}_{U,\mathrm{i},V,z,\mathrm{x}}| \leq \prod_{u \in V} \frac{(z+|U_u|-1)!}{z!\ (|U_u|-1)!}$$

Consider an *independent coder* of *histograms* $C_{\mathrm{A,x}}$

$$C_{\mathrm{A,x}} = \mathrm{coderHistogramIndependent}(U, y, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{A}_{U,\mathrm{i},\leq y})$$

where $y = |X|$ and $X \subset \mathcal{X}$ is a finite subset of *event identifiers*. The *coder domain* is the set of *trimmed histograms* of *size* less than or equal to y in *system* U, $\mathcal{A}_{U,\mathrm{i},\leq y}$. So the *independent coder* can encode all *histograms* whether *independent* or not. However, it is defined to require less *space* for *independent histograms*, $A = A^{\mathrm{X}}$. The *coder* $C_{\mathrm{A,x}}$ adds a flag $b \in \mathrm{bits} = \{0, 1\}$

to indicate whether the *histogram* is *independent* or not. The *coder* $C_{\mathrm{A,x}}$ has intermediate tuple $((n, N_V), z, b, R_A) \in \mathbf{N}^2 \times \mathbf{N} \times \mathrm{bits} \times \mathbf{N}$. If *independent*, then the *space* of the encoding of the *histogram* in R_A is possibly smaller depending on the *variables' valencies*

$$R_A \in \{1 \ldots |\mathcal{A}_{U,\mathrm{i},V,z,\mathrm{x}}|\}$$

instead of

$$R_A \in \{1 \ldots \frac{(z+v-1)!}{z!\ (v-1)!}\}$$

If not *independent*, then the *space* of the encoding of the *histogram* in R_A is possibly smaller too

$$R_A \in \{1 \ldots |\mathcal{A}_{U,\mathrm{i},V,z} \setminus \mathcal{A}_{U,\mathrm{i},V,z,\mathrm{x}}|\}$$

The addition of the flag reduces the *counts space* of *independent histograms*, but the *total space* of this modification, $C_{\mathrm{A,x}}$, of the *histogram coder*, C_{A}, increases because of the cost of the additional *bit* for each *histogram* in the *coder domain*, $|\mathcal{A}_{U,\mathrm{i},\leq y}| \ln 2$. The additional *total space* is

$$\begin{aligned}
&|\mathcal{A}_{U,\mathrm{i},\leq y}| \ln 2 + \\
&\quad \sum(\ln |\mathcal{A}_{U,\mathrm{i},V_A,z_A,\mathrm{x}}| : A \in \mathcal{A}_{U,\mathrm{i},\leq y},\ A = A^{\mathrm{X}}) + \\
&\quad \sum(\ln |\mathcal{A}_{U,\mathrm{i},V_A,z_A} \setminus \mathcal{A}_{U,\mathrm{i},V_A,z_A,\mathrm{x}}| : A \in \mathcal{A}_{U,\mathrm{i},\leq y},\ A \neq A^{\mathrm{X}}) - \\
&\sum(\ln |\mathcal{A}_{U,\mathrm{i},V_A,z_A}| : A \in \mathcal{A}_{U,\mathrm{i},\leq y})
\end{aligned}$$

where $z = \mathrm{size}$ and $V = \mathrm{vars}$.

The remaining terms of the intermediate tuple, (n, N_V) and z, are encoded in exactly the same way as for the *histogram coder*, C_{A}.

The *total space* of the *independent coder*, $C_{\mathrm{A,x}}$, is conjectured to be greater than or equal to the *total space* of the *histogram coder*, C_{A}, $\mathrm{sum}(\mathrm{space}(C_{\mathrm{A,x}})) \geq \mathrm{sum}(\mathrm{space}(C_{\mathrm{A}}))$, and so $C_{\mathrm{A,x}}$ is not a *minimal coder*. The difference in *total space* between the two *coders* depends on the *system*, U, which defines the *coder domain*, $\mathcal{A}_{U,\mathrm{i},\leq y}$. For example, if $|\mathrm{vars}(U)| = 1$ then all of the *histograms* of the *support* must be *independent* because they are *mono-variate*. In this case the flag is pure overhead.

The *generic independent classification coder* of *histories* $C_{\mathrm{G,A,x}}$ takes the *independent coder*, $C_{\mathrm{A,x}}$, as the underlying *coder* of *histograms*

$$C_{\mathrm{G,A,x}} = \mathrm{coderClassificationGeneric}(C_{\mathrm{A,x}}, X, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

The *coder domain* is $\mathcal{H}_{U,X}$, so the *generic independent classification coder* can encode all *histories* in a *system* U and *identifier set* X whether *independent* or not.

Now consider instead a *coder* $C_{\mathrm{A,p}}$, similar to the *independent coder* $C_{\mathrm{A,x}}$, that encodes the *perimeter histogram expression* that is *equivalent* to the *independent histogram*. If *histogram* A is *independent* $A = A^{\mathrm{X}}$ then by definition

$$A^{\mathrm{X}} = Z_A * \prod\{\frac{A}{Z_A} \% \{u\} : u \in V\}$$

where $Z_A = A\%\emptyset$ and $V = \mathrm{vars}(A)$. This is *equivalent* to

$$\mathrm{scalar}(z^{-(n-1)}) * \prod\{A\%\{u\} : u \in V\}$$

where $z = \mathrm{size}(A)$ and $n = |V|$. This *histogram expression* is a scaled product of the *reduced histograms* for each *variable* in V.

Construct the *perimeter coder* of *histograms* $C_{\mathrm{A,p}}$

$$C_{\mathrm{A,p}} = \mathrm{coderHistogramPerimeter}(U, y, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{A}_{U,\mathrm{i},\leq y})$$

The intermediate tuple $((n, N_V), z, b, L) \in \mathbf{N}^2 \times \mathbf{N} \times \mathrm{bits} \times \mathcal{L}(\mathbf{N})$ adds a flag b to indicate whether the *histogram* is *independent* or not. If it is not *independent*, $A \neq A^{\mathrm{X}}$, then the flag b is reset (or false), 0, and the next argument L will be a list containing only the encoding of the *histogram* $L = \{(1, R_A)\}$ where R_A is defined as in the *histogram coder*, C_{A}

$$R_A \in \{1 \ldots \frac{(z+v-1)!}{z!\ (v-1)!}\}$$

where $v = |V^{\mathrm{C}}|$. If the *histogram* is *independent*, $A = A^{\mathrm{X}}$, then the flag b is set (or true), 1, and the list consists of encodings of the *perimeter* of *reduced histograms* for each *variable*. The *perimeter histograms* are *integral* because the *histogram* is *integral*, $A = A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$. Given order D_{A} choose enumerations R

$$\forall u \in V\ (R_u \in \mathrm{enums}(\{\mathrm{trim}(B) : B \in \mathcal{A}_{U,\mathrm{i},\{u\},z}\}))$$

which are such that

$$\forall u \in V\ (R_u(A^{\mathrm{X}}\%\{u\}) \in \{1 \ldots \frac{(z+|U_u|-1)!}{z!\ (|U_u|-1)!}\})$$

Given order D_{V} let $W = \mathrm{order}(D_{\mathrm{V}}, V)$ and so

$$L = \{(W_u, R_u(A^{\mathrm{X}}\%\{u\})) : u \in V\}$$

The decoding can rely on the fact that all the *histograms* in the list have the same *size* z. The list itself is a limited list and so its *space* is the sum of the *counts space* for each *reduction* of the *histogram*

$$\sum_{u \in V} \text{spaceCounts}(U)(A^{\mathrm{X}}\%\{u\}) = \sum_{u \in V} \ln \frac{(z+|U_u|-1)!}{z!\ (|U_u|-1)!}$$

Define the *space* of the encoding of the *perimeter* as $\text{spacePerimeter}(U) \in \mathcal{A}_{U,\mathrm{i}} \to \ln \mathbf{N}_{>0}$

$$\text{spacePerimeter}(U)(A) := \sum_{u \in V} \ln \frac{(z+|U_u|-1)!}{z!\ (|U_u|-1)!}$$

When compared to the *histogram coder*, C_{A}, the *counts space* of the *histograms* in the *histogram expression* of an *independent history* decreases by

$$\ln |\mathcal{A}_{U,\mathrm{i},V,z}| - \sum_{u \in V} \ln |\mathcal{A}_{U,\mathrm{i},\{u\},z}|$$

or

$$\ln \frac{(z+v-1)!}{z!\ (v-1)!} - \sum_{u \in V} \ln \frac{(z+|U_u|-1)!}{z!\ (|U_u|-1)!}$$

That is, $\text{spaceCounts}(U)(A) - \text{spacePerimeter}(U)(A)$. As shown in the section 'Iso-independents' above, it is the case that

$$\frac{(z+v-1)!}{z!\ (v-1)!} \geq \prod_{u \in V} \frac{(z+|U_u|-1)!}{z!\ (|U_u|-1)!}$$

and so the difference is always positive

$$\text{spaceCounts}(U)(A) - \text{spacePerimeter}(U)(A) \geq 0$$

The *space* of an *independent histogram*, $A = A^{\mathrm{X}}$, in the *perimeter coder*, $C_{\mathrm{A,p}}$, is greater than or equal to the *space* of the same *histogram* in the *independent coder*, $C_{\mathrm{A,x}}$, that calculates the *congruent independent histograms*, $\mathcal{A}_{U,\mathrm{i},V,z,\mathrm{x}} = \{B : B \in \mathcal{A}_{U,\mathrm{i},V,z},\ B = B^{\mathrm{X}}\}$, explicitly because

$$|\mathcal{A}_{U,\mathrm{i},V,z,\mathrm{x}}| \leq \prod_{u \in V} |\mathcal{A}_{U,\mathrm{i},\{u\},z}| = \prod_{u \in V} \frac{(z+|U_u|-1)!}{z!\ (|U_u|-1)!}$$

Also, unlike $C_{\mathrm{A,x}}$, there is not a symmetrical reduction in the *space* of *non-independent histograms* that corresponds to the *histogram expression* encapsulation of the *space* of *independent histograms*. Hence the *total space* of $C_{\mathrm{A,p}}$

is greater than that of $C_{\mathrm{A,x}}$, $\mathrm{sum}(\mathrm{space}(C_{\mathrm{A,p}})) \geq \mathrm{sum}(\mathrm{space}(C_{\mathrm{A,x}}))$, which is greater than the *total space* of a *minimal coder*. However, depending on implementation, $C_{\mathrm{A,p}}$ may be more practicable than $C_{\mathrm{A,x}}$. In exchange for the larger *space* of both *independent* and *non-independent histograms*, an implementation of $C_{\mathrm{A,p}}$ may only require a little more computation time than C_{A}, less than that required by $C_{\mathrm{A,x}}$, depending on the calculation of $\mathcal{A}_{U,\mathrm{i},V,z,\mathrm{x}}$.

If the *independent histogram* $A = A^{\mathrm{X}}$ is also *regular* such that $\exists d \in \mathbf{N}\ \forall u \in V\ (|U_u| = d)$ then the decrease in *counts space* between C_{A} and $C_{\mathrm{A,p}}$ is

$$\ln \frac{(z+d^n-1)!}{z!\ (d^n-1)!} - n \ln \frac{(z+d-1)!}{z!\ (d-1)!}$$

where $z = \mathrm{size}(A)$ and $n = |V|$. If $z > d$, then the *counts space* in $C_{\mathrm{A,p}}$ is at most $nd \ln z$. If $z \geq v^2$, where $v = d^n$, the *counts space* in C_{A} is at least $(d^n-1)\ln d^n$ and so the *space* decrease is at least $(v-1)\ln v - nd\ln z$. If the decrease is greater than the cost of the *independent* flag, $\mathrm{space}(|\mathrm{bits}|) = \ln 2$, then there is an overall decrease in *space*. One can think of the *perimeter coder* as encoding the *perimeter* rather than the *volume* of *histograms* when they are *independent*.

The *generic perimeter classification coder* of *histories* $C_{\mathrm{G,A,p}}$ takes the *perimeter coder*, $C_{\mathrm{A,p}}$, as the underlying *histogram coder*

$$C_{\mathrm{G,A,p}} = \mathrm{coderClassificationGeneric}(C_{\mathrm{A,p}}, X, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

The *coder domain* is $\mathcal{H}_{U,X}$, so the *generic perimeter classification coder* can encode all *histories* in a *system* U and *identifier set* X whether *independent* or not.

Now consider the *history* H, in *variables* $V = \mathrm{vars}(H)$ and *size* $z = |H|$ in *system* U, which is such that its *histogram* $A = \mathrm{histogram}(H)$ is not necessarily *independent*. The *dimensional classification coder reduces* the *history* to a set of *histories*, one for each *variable*, regardless of whether the *histogram* is *independent* or not,

$$\{(u, \{(x, S\%\{u\}) : (x,S) \in H\}) : u \in V\} \in V \to \mathcal{H}_{U,X}$$

Define the constructor of the *dimensional classification coder* of *histories*

$$C_{\mathrm{G,n}} = \mathrm{coderClassificationDimensional}(U, X, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

The intermediate tuple is $((n, N_V), (z, Z_I), L, M) \in \mathbf{N}^2 \times \mathbf{N}^2 \times \mathcal{L}(\mathbf{N}) \times \mathcal{L}(\mathbf{N})$. The first pair encodes the set of *variables* in the same way as the *classification coder*, C_{G}, above. The *space* is $\mathrm{spaceVariables}(U)(n)$. The second pair encodes the set of *event identifiers* also in the same way as the *classification coder*. The *space* is $\mathrm{spaceIds}(y, z)$ where $y = |X|$.

The third element of the intermediate tuple, L, encodes the list of *reduced perimeter histograms*, $\{(u, A\%\{u\}) : u \in V\} \in V \to \mathcal{A}_{U,\mathrm{i}}$. The method used is the same as that for the *perimeter coder*, $C_{\mathrm{A,p}}$, above, in the special case of *independent histogram*. Here, however, the *histogram*, A, is *reduced* in all cases. Given order D_{A} choose enumerations

$$\forall u \in V \ (R_u \in \mathrm{enums}(\{\mathrm{trim}(B) : B \in \mathcal{A}_{U,\mathrm{i},\{u\},z}\}))$$

which is such that

$$\forall u \in V \ (R_u(A\%\{u\}) \in \{1 \ldots \frac{(z + |U_u| - 1)!}{z! \ (|U_u| - 1)!}\})$$

Given order D_{V} let $W = \mathrm{order}(D_{\mathrm{V}}, V)$ and so

$$L = \{(W_u, R_u(A\%\{u\})) : u \in V\}$$

The list itself is a limited list and so its *space* is the sum of the *counts space* for each *reduction* of the *histogram*, $\mathrm{spacePerimeter}(U)(A)$.

The last element of the intermediate tuple, M, encodes the list of *reduced classifications*, $\{(u, \mathrm{classification}(\{(x, S\%\{u\}) : (x, S) \in H\})) : u \in V\} \in V \to \mathcal{G}_{U,X}$. The method is the same as for the *classification coder*, C_{G}, above, applied to each *reduced classification* in sequence. That is

$$\forall u \in V \ (F_u(Q_u) \in \{1 \ldots \frac{z!}{\prod_{S \in \mathrm{dom}(G_u)} |G_u(S)|!}\})$$

where $G_u = \mathrm{classification}(\{(x, S\%\{u\}) : (x, S) \in H\})$ and $Q_u = \mathrm{ran}(G_u)$. So

$$M = \{(W_u, F_u(Q_u)) : u \in V\}$$

Again, the list itself is a limited list and so its *space* is the sum of the *classification space* for each *reduction* of the *histogram*

$$\sum_{u \in V} \mathrm{spaceClassification}(A\%\{u\}) = n \ln z! - \sum_{u \in V} \sum_{S \in (A\%\{u\})^{\mathrm{S}}} \ln (A\%\{u\})(S)!$$

where $n = |V|$. The total *classification space* can be approximated by Stirling's approximation

$$nz \ln z - \sum_{u \in V} \sum_{S \in (A\%\{u\})^{\mathrm{S}}} A\%\{u\}(S) \ln A\%\{u\}(S)$$

The *dimensional classification coder* decodes by joining the *reduced histories*, $\{(x, \bigcup\{N_i(x) : i \in \{1 \ldots n\}\}) : x \in \mathrm{dom}(H)\} = H$ where $N \in \mathcal{L}(\mathcal{H}_{U,X})$.

In the case of a *regular history* of *dimension* $n = |V|$ and *valency* $\{d\} = \{|U_u| : u \in V\}$ the *counts space* may be less than that of the *classification coder*, C_{G}, depending on the *size* z and the *volume* d^n as can be seen above for the *perimeter classification coder* when encoding an *independent histogram*. The total *classification space* of the *dimensional classification coder*, $C_{\mathrm{G,n}}$, in the case of *uniform history*, $A = \mathrm{resize}(z, V^{\mathrm{C}})$, approximates to

$$nz \ln z - nd\frac{z}{d} \ln \frac{z}{d} = nz \ln d$$

which may be compared to that of the *classification coder*, C_{G}

$$z \ln z - d^n \frac{z}{d^n} \ln \frac{z}{d^n} = z \ln d^n = nz \ln d$$

Conjecture that in some high *entropy* cases the *space* of a *history* in the *dimensional classification coder*, $C_{\mathrm{G,n}}$, is less than that of the *classification coder*, C_{G}. Conversely, conjecture that in some low *entropy* cases the *space* in the *dimensional classification coder*, $C_{\mathrm{G,n}}$, is greater than that of the *classification coder*, C_{G}.

Unlike the *generic independent classification coder*, $C_{\mathrm{G,A,x}}$, and the *generic perimeter classification coder*, $C_{\mathrm{G,A,p}}$, the *dimensional classification coder*, $C_{\mathrm{G,n}}$, does modify the calculation of the *events classification space*. Therefore there is no underlying *coder* of *histograms* which encapsulates all the differences in encoding logic between the *dimensional classification coder*, $C_{\mathrm{G,n}}$, and the *classification coder*, C_{G}.

Now consider a *coder* of *histograms* based on the *iso-independent* set. Each *complete histogram*, $A^{\mathrm{U}} = V^{\mathrm{C}}$, has an associated finite set of *integral iso-independents*. This set is the equivalence class component of the partition of the *trimmed integral histograms*, $\mathcal{A}_{U,\mathrm{i},\leq y}$, implied by the subset of the *independent* function, $Y_{U,\mathrm{i},\leq y} = \{(A, A^{\mathrm{X}}) : A \in \mathcal{A}_{U,\mathrm{i},\leq y}\} \subset \mathrm{independent}$. That is, the partition $\mathrm{ran}(Y^{-1}_{U,\mathrm{i},\leq y}) \in \mathrm{B}(\mathcal{A}_{U,\mathrm{i},\leq y})$. The set of *iso-independents* of *histogram* A is $Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}}) \subseteq \mathcal{A}_{U,\mathrm{i},\leq y}$.

Define the constructor of the *iso-independent coder* of *histograms*

$$C_{\mathrm{A,y}} = \mathrm{coderHistogramIsoIndependent}(U, y, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{A}_{U,\mathrm{i},\leq y})$$

The intermediate tuple is $((n, N_V), z, L, J_A) \in \mathbf{N}^2 \times \mathbf{N} \times \mathcal{L}(\mathbf{N}) \times \mathbf{N}$. The first pair encodes the set of *variables* in the same way as the *histogram coder*, C_{A}, above. The *space* is $\mathrm{spaceVariables}(U)(n)$. The second element encodes the *size*, also in the same way as the *histogram coder*. The *space* is $\mathrm{spaceSize}(y)$ where $y = |X|$. The third element of the intermediate tuple, L, encodes the list of *reduced perimeter histograms*, $\{(u, A\%\{u\}) : u \in V\} \in V \to \mathcal{A}_{U,\mathrm{i}}$. The method used is the same as that for the *dimensional classification coder*, $C_{\mathrm{G,n}}$, above. The *histogram* is encoded as a *perimeter* regardless of whether it is *independent* or not. The *space* of the list L is the sum of the *counts space* for each *reduction* of the *histogram*, $\mathrm{spacePerimeter}(U)(A)$.

The fourth element, J_A, is the position of the *histogram*, A, in an ordering of the *trimmed iso-independents*. When the *perimeter* has been decoded from L the *trimmed independent histogram*, A^{X}, is known, where $A^{\mathrm{X}} = \mathrm{scalar}(z^{-(n-1)}) * \prod_{u \in V} L(W_u)$ and $W = \mathrm{order}(D_{\mathrm{V}}, V)$. Given order D_{A} choose $J \in \mathrm{enums}(Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}}))$. Thus

$$J_A \in \{1 \dots |Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})|\}$$

So the *space* of the *iso-independent position*, J_A, is $\ln |Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})|$. Define the *space* of the encoding of the *iso-independent position*

$$\mathrm{spaceIsoIndependentPosition}(U) \in \mathcal{A}_U \to \ln \mathbf{N}_{>0}$$

as

$$\mathrm{spaceIsoIndependentPosition}(U)(A) := \ln |Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})|$$

which is defined if $A^{\mathrm{X}} \in \mathrm{ran}(Y_{U,\mathrm{i},\leq y})$.

The *space* of the *iso-independent coder* $C_{\mathrm{A,y}}$ of a *trimmed histogram* $A \in \mathcal{A}_{U,\mathrm{i},\leq y}$ is the sum of the *variables space*, *size space*, *perimeter space* and *iso-independent position space*

$$\begin{aligned} \mathrm{space}(C_{\mathrm{A,y}})(A) \;=\;& \mathrm{spaceVariables}(U)(|\mathrm{vars}(A)|) + \\ & \mathrm{spaceSize}(y) + \\ & \mathrm{spacePerimeter}(U)(A) + \\ & \mathrm{spaceIsoIndependentPosition}(U)(A) \end{aligned}$$

As conjectured above, the cardinality of the *iso-independents* corresponding to A varies with the *entropy* of A^{X}

$$\ln |Y_{U,\mathrm{i},V,z}^{-1}(A^{\mathrm{X}})| \quad \sim \quad z \times \mathrm{entropy}(A^{\mathrm{X}})$$

Therefore, those *histograms* that have low *entropy independent histogram* will require less *space* to encode the *iso-independent position*, J_A, in the *iso-independent coder*, $C_{\mathrm{A,y}}$. The average *iso-independent position space* of the subset of the *coder domain* having *variables* V and *size* z, $\{A : A \in \mathcal{A}_{U,\mathrm{i},\leq y},\ \mathrm{vars}(A) = V,\ \mathrm{size}(A) = z\}$, is

$$\ln \frac{|\mathcal{A}_{U,\mathrm{i},V,z}|}{|\mathrm{ran}(Y_{U,\mathrm{i},V,z})|} = \ln \frac{(z+v-1)!}{z!\ (v-1)!} \prod_{u \in V} \frac{z!\ (|U_u|-1)!}{(z+|U_u|-1)!}$$

The *generic iso-independent classification coder* of *histories* $C_{\mathrm{G,A,y}}$ takes the *iso-independent coder*, $C_{\mathrm{A,y}}$, as the underlying *histogram coder*

$$C_{\mathrm{G,A,y}} = \mathrm{coderClassificationGeneric}(C_{\mathrm{A,y}}, X, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

The *space* of the *generic iso-independent classification coder* $C_{\mathrm{G,A,y}}$ of a *history* $H \in \mathcal{H}_{U,X}$ is the sum of the *variables space*, *ids space*, *perimeter space*, *iso-independent position space* and *classification space*

$$\begin{aligned}
\mathrm{space}(C_{\mathrm{G,A,y}})(H) \quad = \quad & \mathrm{spaceVariables}(U)(|\mathrm{vars}(H)|) + \\
& \mathrm{spaceIds}(|X|, |H|) + \\
& \mathrm{spacePerimeter}(U)(A) + \\
& \mathrm{spaceIsoIndependentPosition}(U)(A) + \\
& \mathrm{spaceClassification}(A)
\end{aligned}$$

where $A = \mathrm{histogram}(H)$.

There are *non-generic classification coders* of *histories* that are based on the *iso-independent* set. Define the constructor of the *superposed iso-independent classification coder* of *histories*

$$C_{\mathrm{G,y,u}} = \mathrm{coderClassificationIsoSuperposed}(U, X, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

The intermediate tuple is $((n, N_V), (z, Z_I), L, J_A, F_Q) \in \mathbf{N}^2 \times \mathbf{N}^2 \times \mathcal{L}(\mathbf{N}) \times \mathbf{N} \times \mathbf{N}$. The tuple is calculated exactly as for the *generic iso-independent classification coder*, $C_{\mathrm{G,A,y}}$. In particular, the last element, F_Q, of the intermediate tuple encodes the *classification* of the *events* of *history* $H \in \mathcal{H}_{U,X}$.

Given D_{X}, choose enumeration F of the enumerations of the partitions of the *event identifiers* corresponding to the *classification*

$$F \in \mathrm{enums}(\{P : P \in \mathrm{B}(\mathrm{ids}(G)),\ \exists X \in P :\leftrightarrow: Q\ \forall (Y,Z) \in X\ (|Y| = |Z|)\})$$

where $G = \mathrm{classification}(H)$ and $Q = \mathrm{ran}(G) \in \mathrm{B}(\mathrm{ids}(G))$ is the partition of the *event identifiers*. The *classification* is such that

$$F_Q \in \{1 \ldots \frac{z!}{\prod_{S \in \mathrm{dom}(G)} |G_S|!}\}$$

In the *generic iso-independent classification coder*, $C_{\mathrm{G,A,y}}$, the *events classification space* of *histogram* $A = \mathrm{histogram}(H)$ is $\mathrm{spaceClassification}(A)$. Here, however, the *space* is fixed to the largest *classification space* of the *iso-independents*. That is,

$$\begin{aligned} & \ln \mathrm{maxr}(\{(B, \frac{z!}{\prod_{S \in B^{\mathrm{S}}} B_S!}) : B \in Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})\}) \\ = \quad & \mathrm{maxr}(\{(B, \mathrm{spaceClassification}(B)) : B \in Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})\}) \end{aligned}$$

Define the *space* of the encoding of the *iso-independent superposed classification* as $\mathrm{spaceClassificationIsoSuperposed}(U) \in \mathcal{A}_U \to \ln \mathbf{N}_{>0}$

$$\begin{aligned} & \mathrm{spaceClassificationIsoSuperposed}(U)(A) := \\ & \quad \mathrm{maxr}(\{(B, \mathrm{spaceClassification}(B)) : B \in Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})\}) \end{aligned}$$

which is defined if $A^{\mathrm{X}} \in \mathrm{ran}(Y_{U,\mathrm{i},\leq y})$. This *space* is always large enough to encode F_Q

$$\mathrm{spaceClassificationIsoSuperposed}(U)(A) \geq \mathrm{spaceClassification}(A)$$

The *space* of the *superposed iso-independent classification coder* $C_{\mathrm{G,y,u}}$ of a *history* $H \in \mathcal{H}_{U,X}$ is the sum of the *variables space*, *ids space*, *perimeter space*, *iso-independent position space* and *iso-independent superposed classification space*

$$\begin{aligned} \mathrm{space}(C_{\mathrm{G,y,u}})(H) \quad = \quad & \mathrm{spaceVariables}(U)(|\mathrm{vars}(H)|) + \\ & \mathrm{spaceIds}(|X|, |H|) + \\ & \mathrm{spacePerimeter}(U)(A) + \\ & \mathrm{spaceIsoIndependentPosition}(U)(A) + \\ & \mathrm{spaceClassificationIsoSuperposed}(U)(A) \end{aligned}$$

where $A = \text{histogram}(H)$.

The *space* of a *history* H in the *superposed iso-independent classification coder*, $C_{\text{G,y,u}}$, is always greater than or equal to the *space* of the *history* in the *generic iso-independent classification coder*, $C_{\text{G,A,y}}$

$$\forall H \in \mathcal{H}_{U,X}\ (\text{space}(C_{\text{G,y,u}})(H) \geq \text{space}(C_{\text{G,A,y}})(H))$$

So the *total space* is larger, $\text{sum}(\text{space}(C_{\text{G,y,u}})) \geq \text{sum}(\text{space}(C_{\text{G,A,y}}))$. However, the *coder* $C_{\text{G,y,u}}$ has the property that all of the *histories* in any given *iso-independent* set parameterised by V and z have equal *space*

$$\begin{aligned}&\forall A^{\text{X}} \in \text{ran}(Y_{U,\text{i},\leq y})\\ &\quad(|\{\text{space}(C_{\text{G,y,u}})(H) : H \in \mathcal{H}_{U,X},\ B = \text{histogram}(H),\ B^{\text{X}} = A^{\text{X}}\}| = 1)\end{aligned}$$

In addition, if the *independent histogram* is *integral*, $A^{\text{X}} \in \mathcal{A}_{\text{i}}$, and hence an *iso-independent*, $A^{\text{X}} \in Y^{-1}_{U,\text{i},\leq y}(A^{\text{X}})$, then it is conjectured that its *classification space* is greater than or equal to the *classification space* of the *iso-independents*

$$\forall B \in Y^{-1}_{U,\text{i},\leq y}(A^{\text{X}})\ (\text{spaceClassification}(B) \leq \text{spaceClassification}(A^{\text{X}}))$$

See the discussion of *alignment* where the *independent histogram* is *integral* in the section 'Minimum alignment', below. In this case the *iso-independent superposed classification space* equals the *classification space* of the *independent histogram*

$$\text{spaceClassificationIsoSuperposed}(U)(A) = \text{spaceClassification}(A^{\text{X}})$$

Note that the *classification space* of the *independent histogram* does not depend on *system* U.

Next consider another *non-generic classification coder* of *histories*. Define the constructor of the *parallel iso-independent classification coder* of *histories*

$$C_{\text{G,y,p}} = \text{coderClassificationIsoParallel}(U, X, D_{\text{V}}, D_{\text{S}}, D_{\text{X}}) \in \text{coders}(\mathcal{H}_{U,X})$$

The intermediate tuple is $((n, N_V), (z, Z_I), L, K_Q) \in \mathbf{N}^2 \times \mathbf{N}^2 \times \mathcal{L}(\mathbf{N}) \times \mathbf{N}$. The first three elements of the tuple are calculated as for the *generic iso-independent classification coder*, $C_{\text{G,A,y}}$. The last element, K_Q, encodes the position of the partition of the *event identifiers*, $Q = \text{ran}(G)$ where $G =$

classification(H), in an enumeration of the partitions corresponding to the *iso-independent histograms* of A^{X}. Given D_{X}, choose enumeration K

$$K \in \mathrm{enums}(\{P : P \in \mathrm{B}(\mathrm{ids}(G)),\ B \in Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}}), \\ \exists X \in P :\leftrightarrow: B^{\mathrm{S}}\ \forall (Y,S) \in X\ (|Y| = B_S)\})$$

The enumeration is such that

$$K_Q \in \{1 \ldots \sum_{B \in Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})} \frac{z!}{\prod_{S \in B^{\mathrm{S}}} B_S!}\}$$

Define the *space* of the encoding of the *parallel iso-independent classification* as spaceClassificationIsoParallel$(U) \in \mathcal{A}_U \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceClassificationIsoParallel}(U)(A) := \ln \sum_{B \in Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})} \frac{z!}{\prod_{S \in B^{\mathrm{S}}} B_S!}$$

which is defined if $A^{\mathrm{X}} \in \mathrm{ran}(Y_{U,\mathrm{i},\leq y})$. The *space* of the *parallel iso-independent classification coder* $C_{\mathrm{G,y,p}}$ of a *history* $H \in \mathcal{H}_{U,X}$ is the sum of the *variables space*, *ids space*, *perimeter space*, and *parallel iso-independent classification space*

$$\begin{aligned} \mathrm{space}(C_{\mathrm{G,y,u}})(H) \quad = \quad & \mathrm{spaceVariables}(U)(|\mathrm{vars}(H)|) + \\ & \mathrm{spaceIds}(|X|,|H|) + \\ & \mathrm{spacePerimeter}(U)(A) + \\ & \mathrm{spaceClassificationIsoParallel}(U)(A) \end{aligned}$$

where $A = \mathrm{histogram}(H)$.

Although the *parallel iso-independent classification coder*, $C_{\mathrm{G,y,p}}$, does not encode the *iso-independent position space*, spaceIsoIndependentPosition$(U)(A)$, the *space* of a *history* in the *parallel iso-independent classification coder*, $C_{\mathrm{G,y,p}}$, is always greater than or equal to the *space* of the *history* in the *superposed iso-independent classification coder*, $C_{\mathrm{G,y,u}}$

$$\forall H \in \mathcal{H}_{U,X}\ (\mathrm{space}(C_{\mathrm{G,y,p}})(H) \geq \mathrm{space}(C_{\mathrm{G,y,u}})(H))$$

The *coder* $C_{\mathrm{G,y,p}}$ also has the property that all of the *histories* in any given *iso-independent* set parameterised by V and z have equal *space*

$$\forall A^{\mathrm{X}} \in \mathrm{ran}(Y_{U,\mathrm{i},\leq y}) \\ (|\{\mathrm{space}(C_{\mathrm{G,y,p}})(H) : H \in \mathcal{H}_{U,X},\ B = \mathrm{histogram}(H),\ B^{\mathrm{X}} = A^{\mathrm{X}}\}| = 1)$$

Following the *parallel iso-independent classification coder*, define the constructor of the *sequential iso-independent classification coder* of *histories*

$$C_{\mathrm{G,y,s}} = \mathrm{coderClassificationIsoSequential}(U, X, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$$

The intermediate tuple is $((n, N_V), (z, Z_I), L, M) \in \mathbf{N}^2 \times \mathbf{N}^2 \times \mathcal{L}(\mathbf{N}) \times \mathcal{L}(\mathbf{N})$. The first three elements of the tuple are calculated as for the *generic iso-independent classification coder*, $C_{\mathrm{G,A,y}}$. The last element, M, encodes a list whose elements correspond to the *iso-independent histograms* of A^{X}, $|M| = |Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})|$. Each element of M encodes the partitions of the *event identifiers* corresponding to the *iso-independent histogram*. Given order D_{A} choose $J \in \mathrm{enums}(Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}}))$. Thus

$$J_A \in \{1 \ldots |Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})|\}$$

Then

$$M_{J_A}(Q) \in \{1 \ldots \frac{z!}{\prod_{S \in \mathrm{dom}(G)} |G_S|!}\}$$

where $A = \mathrm{histogram}(H)$, $G = \mathrm{classification}(H)$ and $Q = \mathrm{ran}(G)$. Define the *space* of the encoding of the *sequential iso-independent classification* in the enumeration, M, as $\mathrm{spaceClassificationIsoSequential}(U) \in \mathcal{A}_U \to \ln \mathbf{N}_{>0}$

$$\mathrm{spaceClassificationIsoSequential}(U)(A) := \sum_{B \in Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})} \ln \frac{z!}{\prod_{S \in B^{\mathrm{S}}} B_S!}$$

which is defined if $A^{\mathrm{X}} \in \mathrm{ran}(Y_{U,\mathrm{i},\leq y})$. Thus

$$\mathrm{spaceClassificationIsoSequential}(U)(A) = \sum_{B \in Y^{-1}_{U,\mathrm{i},\leq y}(A^{\mathrm{X}})} \mathrm{spaceClassification}(B)$$

The *space* of the *sequential iso-independent classification coder* $C_{\mathrm{G,y,s}}$ of a *history* $H \in \mathcal{H}_{U,X}$ is the sum of the *variables space*, *ids space*, *perimeter space*, and *sequential iso-independent classification space*

$$\begin{aligned}\mathrm{space}(C_{\mathrm{G,y,s}})(H) \;=\; & \mathrm{spaceVariables}(U)(|\mathrm{vars}(H)|) + \\ & \mathrm{spaceIds}(|X|, |H|) + \\ & \mathrm{spacePerimeter}(U)(A) + \\ & \mathrm{spaceClassificationIsoSequential}(U)(A)\end{aligned}$$

where $A = \mathrm{histogram}(H)$.

The *space* of a *history* H in the *sequential iso-independent classification coder*, $C_{\mathrm{G,y,s}}$, is always greater than or equal to the *space* of the *history* in the *parallel iso-independent classification coder*, $C_{\mathrm{G,A,p}}$

$$\forall H \in \mathcal{H}_{U,X}\ (\mathrm{space}(C_{\mathrm{G,y,s}})(H) \geq \mathrm{space}(C_{\mathrm{G,A,p}})(H))$$

Thus

$$\mathrm{sum}(C^{\mathrm{s}}_{\mathrm{G,y,s}}) \geq \mathrm{sum}(C^{\mathrm{s}}_{\mathrm{G,y,p}}) \geq \mathrm{sum}(C^{\mathrm{s}}_{\mathrm{G,y,u}}) \geq \mathrm{sum}(C^{\mathrm{s}}_{\mathrm{G,A,y}})$$

where $C^{\mathrm{s}} := \mathrm{space}(C)$. The *coder* $C_{\mathrm{G,y,s}}$ also has the property that all of the *histories* in any given *iso-independent* set parameterised by V and z have equal *space*

$$\begin{aligned}&\forall A^{\mathrm{X}} \in \mathrm{ran}(Y_{U,\mathrm{i},\leq y})\\ &\quad (|\{\mathrm{space}(C_{\mathrm{G,y,s}})(H) : H \in \mathcal{H}_{U,X},\ B = \mathrm{histogram}(H),\ B^{\mathrm{X}} = A^{\mathrm{X}}\}| = 1)\end{aligned}$$

There are other examples of *history coders*, $C \in \mathrm{coders}(\mathcal{H}_{U,X})$, that use the *independent histogram*, A^{X}, or related concepts to modify the *space* function, $\mathrm{space}(C) \in \mathcal{H}_{U,X} \to \ln \mathbf{N}_{>0}$, in order to minimise the *space* of *expected histories*, $\mathrm{expected}(P)(\mathrm{space}(C)) \in \mathbf{Q}_{\geq 0} \ln \mathbf{N}_{>0}$, which depend on the *probability function*, $P \in (\mathcal{H}_{U,X} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$.

For example, if there exists an *integral independent histogram* B^{X} which is such that $B^{\mathrm{X}} \leq A$ and the difference in *size* between the *integral independent histogram* and the *history's histogram*, $\mathrm{size}(A - B^{\mathrm{X}})$, is small, then the *space* of the *size* of the *perimeter* plus *space* of a *perimeter* encoding of B^{X} plus the *counts space* of the encoding of the difference, $A - B^{\mathrm{X}}$, may be less than the *counts space* of the *histogram*, A

$$\begin{aligned}&\mathrm{space}(z) + \mathrm{spacePerimeter}(U)(B^{\mathrm{X}}) + \mathrm{spaceCounts}(U)(A - B^{\mathrm{X}})\\ \leq\ &\mathrm{spaceCounts}(U)(A)\end{aligned}$$

Another example is the *ideal transform coder*, described in the section 'Inducers and Compression' below, which searches for an *ideal transform* T, which is such that $A * T * T^{\dagger A} = A$. The resultant *transform partitions* the *volume*, $T^{\mathrm{P}} \in \mathrm{B}(V^{\mathrm{CS}})$, into *independent components*, $\forall C \in T^{\mathrm{P}}\ (A * C^{\mathrm{U}} = (A * C^{\mathrm{U}})^{\mathrm{X}})$, which can be individually encoded as *perimeters* given the *derived histogram*, $A * T$, that encodes their *sizes*.

A.8 Distribution space

The various *history coders*, above, have *coder domain* $\mathcal{H}_{U,X}$. The *expected space* of *coder* C, $\mathrm{expected}(P)(\mathrm{space}(C))$, depends on the *probability function*

P of the *coder domain*, $P \in (\mathcal{H}_{U,X} \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$. As can be seen above, the *classification coder*, C_{G}, requires least *space* for low *entropy histories*. In contrast, the *index history coder*, C_{H}, uses less *space* than the *classification coder*, C_{G}, where the *entropy* is high. Thus the *coder* can be chosen such that the probability in P of low *space* encodings is high and the probability of high *space* encodings is low. That is, such that the *space* varies inversely with the probability, $\mathrm{space}(C) \sim \{(H, 1/P_H) : H \in \mathcal{H}_{U,X}\}$. For example, an *entropy coder* is defined such that $\mathrm{space}(C)(H) = -\ln P_H$. See appendix 'Coders and entropy'. *Minimal coders*, where the *probability function* is uniform, $P = \mathcal{H}_{U,X} \times \{1/|\mathcal{H}_{U,X}|\}$, imply an absolute standard of the *expected space* of $\ln |\mathcal{H}_{U,X}|$, by which *coders* may be compared. However, the calculation of the cardinality of $\mathcal{H}_{U,X}$ depends on the *variables* of the *histories*.

Consider a special case of *history coders* where the *variables* V of the *histories* are fixed. Parameterise the *coders* with some non-empty *distribution history* $H_E \in \mathcal{H}_U \setminus \{\emptyset\}$, of *size* $z_E = |H_E|$ in *variables* V, from which a subset $H \subseteq H_E$ is *drawn without replacement*. The *coder domain* is the powerset of the *distribution history*, $\mathrm{P}(H_E)$. The cardinality of the *distribution coder domain* is $|\mathrm{P}(H_E)| = 2^{z_E}$ which implies a *minimal space* per *history* of $z_E \ln 2$.

An example of a *minimal distribution coder* $C \in \mathrm{coders}(\mathrm{P}(H_E))$ is such that the *code* is in $\mathrm{enums}(\mathrm{P}(H_E))$. The *space* of $H \subseteq H_E$ is $\mathrm{space}(C)(H) = z_E \ln 2$.

Consider the *fixed width* analogue for the *index coder* of *histories*, C_{H}, where the *distribution history*, H_E, is known. The *index distribution coder*, $C_{\mathrm{Q,1}}$, is constructed

$$C_{\mathrm{Q,1}} = \mathrm{coderDistributionIndex}(H_E, D_{\mathrm{X}}) \in \mathrm{coders}(\mathrm{P}(H_E))$$

The *index distribution coder* is parameterised by (i) the *distribution history*, H_E, and (ii) the order on the *event identifiers*, $D_{\mathrm{X}} \in \mathrm{enums}(X)$, where X is some set of *event identifiers* such that $\mathrm{ids}(H_E) \subseteq X$.

The intermediate tuple is $(z, L) \in \mathbf{N} \times \mathcal{L}(\mathbf{N})$. The first element of the tuple encodes the *size* of the *drawn history*, $z = |H|$. The *space* is $\ln(z_E + 1)$.

The second element of the tuple encodes a list of the subset of the *events* of H_E indexed by the enumeration $\mathrm{order}(D_{\mathrm{X}}, \mathrm{ids}(H_E)) \in \mathrm{enums}(\mathrm{ids}(H_E))$. That is, $\mathrm{set}(L) \subseteq \{1 \ldots z_E\}$ and $|\mathrm{set}(L)| = |L| = z$, where $z_E = |H_E|$. The *space* of the element L is $z \ln z_E$.

The *space* of the *index distribution coder* is

$$\mathrm{space}(C_{\mathrm{Q,l}})(H) := \mathrm{spaceSize}(z_E) + z \times \mathrm{space}(z_E)$$

A similar *coder* of *histories* where the *distribution history*, H_E, is known, is the *subset distribution coder*, $C_{\mathrm{Q,p}}$. It is constructed from the same parameters as $C_{\mathrm{Q,l}}$. The intermediate tuple is $(z, N_H) \in \mathbf{N} \times \mathbf{N}$. The first element of the tuple encodes the *size* of the *drawn history*, $z = |H|$. The *space* is $\ln(z_E + 1)$. The second element of the tuple encodes the subset of the *events* of H_E. Given order D_{X} choose an enumeration $N \in \mathrm{enums}(\{R : R \in \mathrm{P}(H_E),\ |R| = z\})$. The *space* of the element N_H is

$$\mathrm{spaceSubset}(z_E, z) = \ln \frac{z_E!}{z!\ (z_E - z)!} = \underline{z} \ln z_E - \underline{z} \ln z$$

where $x^{\underline{y}}$ is falling factorial. The notation is abused such that $\underline{y} \ln x = \ln x^{\underline{y}}$.

The *space* of the *subset distribution coder* is

$$\mathrm{space}(C_{\mathrm{Q,p}})(H) := \mathrm{spaceSize}(z_E) + \mathrm{spaceSubset}(z_E, z)$$

The *space* of the *subset distribution coder* is less than the *space* of the *index distribution coder* because H is *drawn without replacement* from H_E. That is, H is a subset, $H \subseteq H_E$,

$$\mathrm{space}(C_{\mathrm{Q,l}})(H) - \mathrm{space}(C_{\mathrm{Q,p}})(H) = z \ln z_E - (\underline{z} \ln z_E - \underline{z} \ln z) > 0$$

Neither the *index distribution coder*, $C_{\mathrm{Q,l}}$, nor the *subset distribution coder*, $C_{\mathrm{Q,p}}$, are defined in terms of *distributions* of *histograms*, $\mathcal{Q} \subset \mathcal{A}_{\mathrm{i}} \to \mathbf{Q}_{\geq 0}$. The *historical distribution* of *histograms*, $Q_{\mathrm{h}} \in \mathcal{A}_{\mathrm{i}} \times \mathbf{N} \to \mathcal{Q}$, forms the basis for the *historical distribution coder* of *histories* $C_{\mathrm{Q,h}}$ constructed

$$C_{\mathrm{Q,h}} = \mathrm{coderDistributionHistorical}(U, H_E, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathrm{P}(H_E))$$

The *historical distribution coder* is parameterised by (i) the *system*, U, (ii) the non-empty *distribution history*, H_E, (iii) the order on the *variables*, D_{V}, (iv) the order on the *states*, D_{S}, and (v) the order on the *event identifiers*, D_{X}.

The intermediate tuple is $(z, R_A, J_H) \in \mathbf{N} \times \mathbf{N} \times \mathbf{N}$. The first element of the tuple encodes the *size* of the *drawn history*, $z = |H|$. The *space* is $\ln(z_E + 1)$.

The second element of the tuple encodes the set of *counts* of the *histogram* of the *drawn history*, $A = \text{histogram}(H)$. The *space* of the element R_A is $\text{spaceCounts}(U)(A)$.

The third element of the intermediate tuple, J_H, encodes the position in a list of the *histories* which are subsets of H_E having *histogram* A. Given order D_{A}, constructed from D_{V} and D_{S}, and order D_{X} choose the enumeration

$$J \in \text{enums}(\{G : G \subseteq H_E,\ \text{histogram}(G) = A\})$$

which is such that $J_H \in \{1 \ldots Q_{\text{h}}(E, z)(A)\}$ where $E = \text{histogram}(H_E)$ and the set of *historical distributions* $Q_{\text{h}} \in \mathcal{A}_{\text{i}} \times \mathbf{N} \to \mathcal{Q}$ is defined

$$Q_{\text{h}}(E, z)(A) = \prod_{S \in A^{\text{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!} \in \mathbf{N}_{>0}$$

Define *historical distribution space* $\text{spaceDistributionHistorical} \in \mathcal{A}_{\text{i}} \times \mathbf{N} \to (\mathcal{A}_{\text{i}} \to \ln \mathbf{N}_{>0})$ as

$$\text{spaceDistributionHistorical}(E, z)(A) := \ln Q_{\text{h}}(E, z)(A)$$

The *space* of the *historical distribution coder* is

$$\begin{aligned} \text{space}(C_{\text{Q,h}})(H) &= \text{spaceSize}(z_E) + \\ &\quad \text{spaceCounts}(U)(A) + \\ &\quad \text{spaceDistributionHistorical}(E, z)(A) \end{aligned}$$

where $z_E = |H_E|$, $E = \text{histogram}(H_E)$, $A = \text{histogram}(H)$ and $z = |H|$. Here the *variables* V are known from the parameters of the constructor and so there is no need to encode them. The *space* required to encode the *event identifiers* is incorporated into the *historical distribution space*, $\ln Q_{\text{h}}(E, z)(A)$. The *counts space* only depends on the *volume* and *size* of the *history*,

$$\text{spaceCounts}(U)(A) = \text{spaceCompositionWeak}(v, z)$$

where $v = \text{volume}(U)(H_E)$. The *space* of the *empty history* is defined $\text{space}(C_{\text{Q,h}})(\emptyset) := \text{spaceSize}(z_E)$.

The *space* of the *historical distribution coder* may be compared to a *minimal coder*. In the case of a high *entropy uniform cartesian histogram*,

$A = \text{scalar}(z/v) * V^{\text{C}} \in \mathcal{A}_{\text{i}}$ where $v = |V^{\text{C}}|$, and a *uniform cartesian distribution histogram*, $E = \text{scalar}(z_E/v) * V^{\text{C}} \in \mathcal{A}_{\text{i}}$, the *space* is

$$\begin{aligned}
&\text{space}(C_{\text{Q,h}})(H) \\
&= \ln(z_E+1) + \ln\frac{(z+v-1)!}{z!\ (v-1)!} + \ln\prod_{S \in A^{\text{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!} \\
&= \ln(z_E+1) + \overline{v}\ln z - \underline{v}\ln v - \ln(z/v) - \sum_{S \in A^{\text{S}}} \underline{A_S}\ln A_S + \sum_{S \in A^{\text{S}}} \underline{A_S}\ln E_S \\
&= \ln(z_E+1) + \overline{v}\ln z - \underline{v}\ln v - \ln(z/v) - v\underline{(z/v)}\ln(z/v) + v\underline{(z/v)}\ln(z_E/v) \\
&< v\underline{(z/v)}\ln(z_E/v) + \overline{v}\ln z + \ln(z_E+1)
\end{aligned}$$

where $x^{\overline{y}}$ is rising factorial and $x^{\underline{y}}$ is falling factorial. The notation is abused such that $\overline{y}\ln x = \ln x^{\overline{y}}$ and $\underline{y}\ln x = \ln x^{\underline{y}}$. It can be seen that in some cases, if $v \leq z \ll z_E$, the *space* is less than that for a *minimal coder*, $\text{space}(C_{\text{Q,h}})(H) < z_E \ln 2$.

The *space* of a *history* in the *historical distribution coder* in the low *entropy* case of a *singleton histogram* $A = \{(S, z)\}$ and *distribution histogram* such that $E_S = z$ is

$$\begin{aligned}
&\text{space}(C_{\text{Q,h}})(H) \\
&= \ln(z_E+1) + \overline{v}\ln z - \underline{v}\ln v - \ln(z/v) - \underline{z}\ln z + -\underline{z}\ln z \\
&< \overline{v}\ln z + \ln(z_E+1)
\end{aligned}$$

For $v < z < z_E$ the *space* is less than that for an *index distribution coder*, $\text{space}(C_{\text{Q,h}})(H) < \text{space}(C_{\text{Q,l}})(H) = z\ln z_E + \ln(z_E+1)$.

As shown above, the *historical distribution* can be rewritten in terms of the *multinomial coefficient*

$$Q_{\text{h}}(E,z)(A) = \prod_{S \in A^{\text{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!} = \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!}\frac{1}{z!}\prod_{S \in A^{\text{S}}} E_S^{\underline{A_S}}$$

Compare this to a *multinomial distribution* in $Q_{\text{m}} \in \mathcal{A}_{\text{i}} \times \mathbf{N} \to \mathcal{Q}$ defined

$$Q_{\text{m}}(E,z)(A) = \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!}\prod_{S \in A^{\text{S}}} E_S^{A_S} \in \mathbf{N}_{>0}$$

The *coder domain* of the powerset of the *distribution histogram*, $\text{P}(H_E)$, can be encoded by means of the *multinomial distribution*, albeit requiring more *space* than the *historical distribution coder*, $C_{\mathrm{Q,h}}$. The *multinomial distribution coder* of *histories* $C_{\mathrm{Q,m}}$ is defined

$$C_{\mathrm{Q,m}} = \mathrm{coderDistributionMultinomial}(U, H_E, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathrm{P}(H_E))$$

The *multinomial distribution coder*, $C_{\mathrm{Q,m}}$, is constructed with the same parameters as the *historical distribution coder*, $C_{\mathrm{Q,h}}$.

The intermediate tuple is $(z, R_A, K_H) \in \mathbf{N} \times \mathbf{N} \times \mathbf{N}$. The first and second elements of the tuple encodes the *size* $z = |H|$ and the *histogram* $A = \mathrm{histogram}(H)$ in the same way as the *historical distribution coder*, $C_{\mathrm{Q,h}}$.

The third element of the intermediate tuple, K_H, encodes the position in a list of the *histories* which have *histogram* A. The *histories* are constructed from subsets of $H_E^z = \{L : L \in \mathcal{L}(H_E),\ |L| = z\}$ by modifying the *event identifiers*. Given order D_{A}, constructed from D_{V} and D_{S}, and order D_{X} choose the enumeration

$$K \in \mathrm{enums}(\{G : L \in \mathcal{L}(H_E),\ G = \{((i,x),S) : (i,(x,S)) \in L\},\ \mathrm{his}(G) = A\})$$

where his = histogram. The enumeration of modified *histories* $K \in \mathcal{H} \to \mathbf{N}$ is such that

$$\exists G \in \mathrm{dom}(K)\ ((\{(x,S) : ((i,x),S) \in G\} = H) \wedge (K_G \in \{1 \ldots Q_{\mathrm{m}}(E,z)(A)\}))$$

where $E = \mathrm{histogram}(H_E)$. Any of the modified *histories*, $G \in \mathrm{dom}(K)$, for which $\{(x,S) : ((i,x),S) \in G\} = H$, can be chosen to encode H. Given K_H the decode is $\{(x,S) : ((i,x),S) \in \mathrm{flip}(K)(K_H)\} = H$.

Define *multinomial distribution space* spaceDistributionMultinomial $\in \mathcal{A}_{\mathrm{i}} \times \mathbf{N} \to (\mathcal{A}_{\mathrm{i}} \to \ln \mathbf{N}_{>0})$ as

$$\mathrm{spaceDistributionMultinomial}(E,z)(A) := \ln Q_{\mathrm{m}}(E,z)(A)$$

The *space* of the *multinomial distribution coder* is

$$\begin{aligned}\mathrm{space}(C_{\mathrm{Q,m}})(H) \quad = \quad & \mathrm{spaceSize}(z_E) + \\ & \mathrm{spaceCounts}(U)(A) + \\ & \mathrm{spaceDistributionMultinomial}(E,z)(A)\end{aligned}$$

where $z_E = |H_E|$, $E = \mathrm{histogram}(H_E)$, $A = \mathrm{histogram}(H)$ and $z = |H|$. The *space* of the *empty history* is defined $\mathrm{space}(C_{\mathrm{Q,m}})(\emptyset) := \mathrm{spaceSize}(z_E)$.

The *space* of a *history* $H \subseteq H_E$ in the *multinomial distribution coder* is greater than or equal to the *space* in the *historical distribution coder*

$$\text{space}(C_{\text{Q,m}})(H) \geq \text{space}(C_{\text{Q,h}})(H)$$

The *space* is sometimes larger because

$$|H_E^z| = z_E^z \geq |\{H : H \subseteq H_E,\ |H| = z\}| = z_E^{\underline{z}}/z!$$

There are modified *histories* in the *multinomial distribution coder* that do not correspond to elements of the *coder domain*, $\mathrm{P}(H_E)$, because some contain *replaced events* and because there are up to $z!$ permutations depending on the number of *replaced events*. See the discussion, above, comparing *historical distributions* and *multinomial distributions*.

In fact, the *multinomial distribution* would be a more efficient method if the *coder domain* was $\{L : L \in \mathcal{L}(H_E),\ |L| \leq z_E\}$ rather than $\mathrm{P}(H_E)$. In this case, each $K_G \in \{1 \ldots Q_{\text{m}}(E,z)(A)\}$ would encode a different element of the *coder domain* and there would be no unused *space*.

Also the *multinomial distribution* could be used to encode lists of cardinality greater than $|H_E|$. In this case the *coder domain* would be $\{L : L \in \mathcal{L}(H_E),\ |L| \leq y\}$ where $y \in \mathbf{N}$. Here y would be a parameter of the constructor of a *multinomial distribution coder*, $(U, H_E, y, D_{\text{V}}, D_{\text{S}})$, and additional *space* would be required for (i) the length of a limited list, $\ln(y+1)$, or (ii) the termination flag *space* of an unlimited list.

In order to compare the *distribution coders* to the *classification coders*, let $C_{\text{G},V}$ be a special case of the *classification coder* of *histories*, C_{G}, such that the *variables*, V, are fixed. The *coder domain* is $X \to V^{\text{CS}} = \{H : H \in \mathcal{H}_{U,X},\ \text{vars}(H) = V\} \supset \mathrm{P}(H_E)$ where $X = \text{dom}(H_E)$. Thus

$$\text{space}(C_{\text{G},V})(H) = \text{space}(C_{\text{G}})(H) - \text{spaceVariables}(U)(V)$$

where *history* $H \subseteq H_E$. The *space* of the *substrate classification coder*, $C_{\text{G},V}$, is

$$\begin{aligned}\text{space}(C_{\text{G},V})(H) \quad = \quad & \text{spaceIds}(z_E, z)\, + \\ & \text{spaceCounts}(U)(A)\, + \\ & \text{spaceClassification}(A)\end{aligned}$$

where $z_E = |H_E|$, $A = \text{histogram}(H)$ and $z = |H|$. The *events classification space* of the *histogram*, $\text{spaceClassification}(A)$, is defined as the logarithm of the *multinomial coefficient*

$$\text{spaceClassification}(A) := \ln \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!}$$

As shown above, the *events classification space* approximates to the *sized entropy* of the *histogram*

$$\text{spaceClassification}(A) \approx z \times \text{entropy}(A)$$

The *space* of a *history* $H \subseteq H_E$ in the *multinomial distribution coder*, $\text{space}(C_{\mathrm{Q,m}})(H)$, varies as the *frequency* $Q_{\mathrm{m}}(E, z)(A)$. This in turn varies as the *multinomial coefficient*, $z!/\prod_{S \in A^{\mathrm{S}}} A_S!$. Thus the *space* of both the *classification coder* and the *multinomial distribution coder* varies with the *entropy* of the *histogram*, $\text{entropy}(A)$. Low *entropy histories* tends to require less *space* than high *entropy*.

However, the *frequencies* of the *multinomial distribution*, $Q_{\mathrm{m}}(E, z)$, depend on the *permutorial* part, $\prod_{S \in A^{\mathrm{S}}} E_S^{A_S}$, as well as the *multinomial coefficient*, $z!/\prod_{S \in A^{\mathrm{S}}} A_S!$. As shown above, the logarithm of the *multinomial probability distribution* approximates to the negative *sized relative entropy* between the *complete sample histogram*, $A + V^{\mathrm{CZ}}$, and the *distribution histogram*, E,

$$\begin{aligned}
\ln \hat{Q}_{\mathrm{m},U}(E, z)(A + V^{\mathrm{CZ}}) &\approx \sum_{S \in A^{\mathrm{FS}},\ P_S > 0} A_S \ln \frac{P_S}{N_S} \\
&= -z \sum_{S \in A^{\mathrm{FS}},\ P_S > 0} N_S \ln \frac{N_S}{P_S} \\
&= -z \times \text{entropyRelative}(N, P)
\end{aligned}$$

where $P = \text{resize}(1, E)$ and $N = \text{resize}(1, A)$. So the *space* of the *multinomial distribution coder*, $\text{space}(C_{\mathrm{Q,m}})(H)$, varies inversely with the *relative entropy*. In the case of the *classification coder*, $C_{\mathrm{G},V}$, there is no *distribution histogram*. It is equivalent to the *multinomial distribution* where the *distribution histogram* is *uniform cartesian*, $E = V^{\mathrm{C}}$

$$\begin{aligned}
\text{spaceClassification}(A) &= \ln Q_{\mathrm{m}}(V^{\mathrm{C}}, z)(A) \\
&= \ln \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!} \prod_{S \in A^{\mathrm{S}}} 1^{A_S} \\
&= \ln \frac{z!}{\prod_{S \in A^{\mathrm{S}}} A_S!}
\end{aligned}$$

The *space* of a *history* in the *multinomial distribution coder*, $\text{space}(C_{\text{Q,m}})(H)$, is maximised when the *relative entropy* is minimised. That occurs when the *sample histogram* equals the *scaled distribution histogram*, $A = \text{scalar}(z/z_E) * E$. In probabilistic terms, this is when the *sample histogram* equals the *mean*, $A = \text{mean}(\hat{Q}_{\text{m},U}(E + V^{\text{CZ}}, z))$. *Sample histograms* that are far from the *mean* with respect to the *variance*, $\text{var}(U)(\hat{Q}_{\text{m},U}(E + V^{\text{CZ}}, z))$, have lower *space*. Note that, as shown above, *variance* varies as the *entropy* of the *distribution histogram*, $\text{entropy}(E)$.

The *space* of a *history* in the *multinomial distribution coder* in the case of a high *entropy uniform cartesian histogram*, $A = \text{scalar}(z/v) * V^{\text{C}} \in \mathcal{A}_{\text{i}}$ where $v = |V^{\text{C}}|$, and a *uniform cartesian distribution histogram*, $E = \text{scalar}(z_E/v) * V^{\text{C}} \in \mathcal{A}_{\text{i}}$ is

$$\begin{aligned}
&\text{space}(C_{\text{Q,m}})(H) \\
&= \ln(z_E + 1) + \ln \frac{(z + v - 1)!}{z!\ (v-1)!} + \ln \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!} \prod_{S \in A^{\text{S}}} E_S^{A_S} \\
&= \ln(z_E + 1) + \overline{v} \ln z - \underline{v} \ln v - \ln(z/v) + \underline{z} \ln z - \sum_{S \in A^{\text{S}}} \underline{A_S} \ln A_S + \sum_{S \in A^{\text{S}}} A_S \ln E_S \\
&= \ln(z_E + 1) + \overline{v} \ln z - \underline{v} \ln v - \ln(z/v) + \underline{z} \ln z - v\underline{(z/v)} \ln(z/v) + z \ln(z_E/v) \\
&< z \ln(z_E/v) + \underline{z} \ln z + \overline{v} \ln z + \ln(z_E + 1)
\end{aligned}$$

It can be seen that even in some high *entropy* cases the *space* is less than that for a *minimal coder*, $\text{space}(C_{\text{Q,m}})(H) < z_E \ln 2$, if $v \leq z \ll z_E$.

The *space* of a *history* in the *multinomial distribution coder* in the low *entropy* case of a *singleton histogram* $A = \{(S, z)\}$ and *distribution histogram* such that $E_S = z$ is

$$\begin{aligned}
&\text{space}(C_{\text{Q,m}})(H) \\
&= \ln(z_E + 1) + \overline{v} \ln z - \underline{v} \ln v - \ln(z/v) + \underline{z} \ln z - \underline{z} \ln z + z \ln z \\
&< z \ln z + \overline{v} \ln z + \ln(z_E + 1)
\end{aligned}$$

For some $v < z$ the *space* is less than that for an *index distribution coder*, $\text{space}(C_{\text{Q,m}})(H) < \text{space}(C_{\text{Q,l}})(H) = z \ln z_E + \ln(z_E + 1)$.

The *space* of a *history* in the *classification coder* is

$$\begin{aligned}
&\text{space}(C_{\text{G},V})(H) \\
&= \ln(z_E + 1) + \ln \frac{z_E!}{z!\ (z_E - z)!} + \ln \frac{(z + v - 1)!}{z!\ (v-1)!} + \ln \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!}
\end{aligned}$$

The difference between the two *coders* is

$$\text{space}(C_{\text{G},V})(H) - \text{space}(C_{\text{Q,m}})(H) = \ln \frac{z_E!}{z!\ (z_E - z)!} - \ln \prod_{S \in A^{\text{S}}} E_S^{A_S}$$

In the case of *uniform cartesian histogram*, $A = \text{scalar}(z/v) * V^{\text{C}}$, and *uniform cartesian distribution histogram*, $E = \text{scalar}(z_E/v) * V^{\text{C}} \in \mathcal{A}_{\text{i}}$, the difference is

$$\text{space}(C_{\text{G},V})(H) - \text{space}(C_{\text{Q,m}})(H) = \underline{z} \ln z_E + z \ln v - \underline{z} \ln z - z \ln z_E$$

Thus the *space* of the *multinomial distribution coder* is greater than that of the *classification coder*, $\text{space}(C_{\text{Q,m}})(H) > \text{space}(C_{\text{G},V})(H)$, in the high *entropy* case.

Compare the low *entropy* case of a *singleton histogram* $A = \{(S, z)\}$ and *distribution histogram* such that $E_S = z$

$$\text{space}(C_{\text{G},V})(H) - \text{space}(C_{\text{Q,m}})(H) = \underline{z} \ln z_E - \underline{z} \ln z - z \ln z$$

In the low *entropy* case the *space* of the *multinomial distribution coder* is less than that of the *classification coder*, $\text{space}(C_{\text{Q,m}})(H) < \text{space}(C_{\text{G},V})(H)$, if $z_E \gg z^2$. Thus the parameterisation of the *distribution coders*, $C_{\text{Q,h}}$ and $C_{\text{Q,m}}$, by the *distribution history*, H_E, reduces the *space* in some cases.

The *historical distribution* and the *multinomial distribution* of the *distribution coders* naturally imply a *probability function*. For example, let $P = \{(H, \hat{Q}_{\text{m},U}(E, |H|)(A)) : H \subseteq H_E,\ A = \text{his}(H) + V^{\text{CZ}}\}$ where $E = \text{his}(H_E) + V^{\text{CZ}}$ then $\hat{P} \in (\text{P}(H_E) \to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is the conditional *multinomial probability* given the *size*. Of course, the expected *space* of the *multinomial distribution coder* in this implied *probability function*, $\text{expected}(\hat{P})(\text{space}(C_{\text{Q,m}})) \in \mathbf{Q}_{\geq 0} \ln \mathbf{N}_{>0}$, is high because the *space* varies with the implied probability, $\text{space}(C_{\text{Q,m}}) \sim \hat{P}$, rather than inversely as would be the case in an *entropy coder* of $\hat{P}$.

In the discussion above the *alignment* has been derived in terms of *relative probability*. It may equally be derived in terms of *relative space*. Let H_E be some non-empty *distribution history*, $H_E \in \mathcal{H}_{U,X} \setminus \{\emptyset\}$, of *size* $z_E = |H_E|$ and *variables* $V = \text{vars}(H_E)$ in *system* U and *event identifiers* set X. The subset *sample history* $H \subset H_E$, of *size* $z = |H|$, is *drawn* from the *distribution* *history without replacement*. Consider the *distribution coder domain* $\mathrm{P}(H_E)$. The *multinomial distribution coder* of *histories* $C_{\mathrm{Q,m}}$ is defined above, as

$$C_{\mathrm{Q,m}} = \mathrm{coderDistributionMultinomial}(U, H_E, D_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathrm{P}(H_E))$$

where D_{V}, D_{S} and D_{X} are orders on the *variables*, *states* and *event identifiers*. The *space* of the *multinomial distribution coder* is

$$\begin{aligned}\mathrm{space}(C_{\mathrm{Q,m}})(H) \;=\;& \mathrm{spaceSize}(z_E) + \\ & \mathrm{spaceCounts}(U)(A) + \\ & \mathrm{spaceDistributionMultinomial}(E, z)(A)\end{aligned}$$

where $E = \mathrm{histogram}(H_E)$, $A = \mathrm{histogram}(H)$ and *multinomial distribution space* is

$$\mathrm{spaceDistributionMultinomial}(E, z)(A) := \ln Q_{\mathrm{m}}(E, z)(A)$$

Here the *sample histogram* is *trimmed*, $A \in \{\mathrm{trim}(B) : B \in \mathcal{A}_{U,\mathrm{i},V,z}\}$.

In the case where the *independent sample histogram* is in the *histograms* of the *coder domain*, $A^{\mathrm{X}} \in \{\mathrm{histogram}(G) : G \subseteq H_E\}$, and is therefore *integral*, $A^{\mathrm{X}} \in \mathcal{A}_{\mathrm{i}}$, then the *multinomial distribution coder space* of *history* H may be decomposed into (i) the *independent multinomial distribution coder space* and (ii) *relative dependent multinomial distribution coder space*

$$\begin{aligned}\mathrm{space}(C_{\mathrm{Q,m}})(H) \;=\;& \mathrm{spaceSize}(z_E) + \\ & \mathrm{spaceCounts}(U)(A^{\mathrm{X}}) + \\ & \mathrm{spaceDistributionMultinomial}(E, z)(A^{\mathrm{X}}) + \\ & (\mathrm{spaceDistributionMultinomial}(E, z)(A) - \\ & \mathrm{spaceDistributionMultinomial}(E, z)(A^{\mathrm{X}}))\end{aligned}$$

The *counts space* of the *independent sample* equals the *counts space* of the *sample histogram*, $\mathrm{spaceCounts}(U)(A^{\mathrm{X}}) = \mathrm{spaceCounts}(U)(A)$, because the *counts space* only depends on the *size*, z, and *volume*, $v = |V^{\mathrm{C}}|$, and the *independent* is *congruent*, $\mathrm{congruent}(A, A^{\mathrm{X}})$.

Let H_{X} be any *history drawn* from the *distribution history* such that it's *histogram* equals the *independent sample histogram*,

$$H_{\mathrm{X}} \in \{G : G \subseteq H_E,\ \mathrm{histogram}(G) = A^{\mathrm{X}}\}$$

The *independent multinomial distribution coder space* equals the *space* of H_X

$$\begin{aligned}\text{space}(C_{Q,m})(H_X) &= \text{spaceSize}(z_E) + \\ &\quad \text{spaceCounts}(U)(A^X) + \\ &\quad \text{spaceDistributionMultinomial}(E, z)(A^X)\end{aligned}$$

The *relative dependent multinomial distribution coder space* of *history* H is

$$\begin{aligned}&\text{spaceDistMult}(E, z)(A) - \text{spaceDistMult}(E, z)(A^X) \\ &= \ln Q_m(E, z)(A) - \ln Q_m(E, z)(A^X) \\ &= \ln \frac{z!}{\prod_{S \in A^S} A_S!} \prod_{S \in A^S} E_S^{A_S} - \ln \frac{z!}{\prod_{S \in A^{XS}} A_S!} \prod_{S \in A^{XS}} E_S^{A_S^X}\end{aligned}$$

where spaceDistMult = spaceDistributionMultinomial. Thus the negative *relative dependent multinomial distribution coder space* equals the *alignment* minus the *mis-alignment*

$$\begin{aligned}&-(\text{spaceDistMult}(E, z)(A) - \text{spaceDistMult}(E, z)(A^X)) \\ &= \sum_{S \in A^S} \ln A_S! - \sum_{S \in A^{XS}} \ln A_S^X! - \sum_{S \in A^{XS}} (A_S - A_S^X) \ln E_S \\ &= \text{alignment}(A) - \sum_{S \in A^{XS}} (A_S - A_S^X) \ln E_S\end{aligned}$$

In the case where the *sample histogram* is *independent*, $A = A^X$, the negative *relative dependent multinomial distribution coder space* is 0

$$-(\text{spaceDistMult}(E, z)(A^X) - \text{spaceDistMult}(E, z)(A^X)) = 0$$

In the case where the *distribution history* is *independent*, $E = E^X$, the *mis-alignment* is zero and the negative *independently-distributed relative dependent multinomial distribution coder space* equals the *alignment*

$$-(\text{spaceDistMult}(E^X, z)(A) - \text{spaceDistMult}(E^X, z)(A^X)) = \text{alignment}(A)$$

Similarly, the *historical distribution coder* of *histories* $C_{Q,h}$ is constructed

$$C_{Q,h} = \text{coderDistributionHistorical}(U, H_E, D_V, D_S, D_X) \in \text{coders}(P(H_E))$$

The *space* of the *multinomial distribution coder* is

$$\begin{aligned}\text{space}(C_{Q,h})(H) &= \text{spaceSize}(z_E) + \\ &\quad \text{spaceCounts}(U)(A) + \\ &\quad \text{spaceDistributionHistorical}(E, z)(A)\end{aligned}$$

where $E = \text{histogram}(H_E)$, $A = \text{histogram}(H)$ and *historical distribution space* is

$$\text{spaceDistributionHistorical}(E,z)(A) := \ln Q_{\text{h}}(E,z)(A)$$

In the case where the *independent sample histogram* is in the *histograms* of the *coder domain*, $A^{\text{X}} \in \{\text{histogram}(G) : G \subseteq H_E\}$, then the *historical distribution coder space* of *history* H may be decomposed into (i) the *independent historical distribution coder space* and (ii) *relative dependent historical distribution coder space*

$$\begin{aligned}\text{space}(C_{\text{Q,h}})(H) &= \text{spaceSize}(z_E) + \\ &\quad \text{spaceCounts}(U)(A^{\text{X}}) + \\ &\quad \text{spaceDistributionHistorical}(E,z)(A^{\text{X}}) + \\ &\quad (\text{spaceDistributionHistorical}(E,z)(A) - \\ &\quad \text{spaceDistributionHistorical}(E,z)(A^{\text{X}}))\end{aligned}$$

The *relative dependent historical distribution coder space* of *history* H is

$$\begin{aligned}&\text{spaceDistHist}(E,z)(A) - \text{spaceDistHist}(E,z)(A^{\text{X}}) \\ &\quad= \ln Q_{\text{h}}(E,z)(A) - \ln Q_{\text{h}}(E,z)(A^{\text{X}}) \\ &\quad= \ln \prod_{S \in A^{\text{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!} - \ln \prod_{S \in A^{\text{XS}}} \frac{E_S!}{A^{\text{X}}_S!\ (E_S - A^{\text{X}}_S)!} \\ &\quad= \ln \frac{z!}{\prod_{S \in A^{\text{S}}} A_S!} \frac{1}{z!} \prod_{S \in A^{\text{S}}} E_S^{\underline{A_S}} - \ln \frac{z!}{\prod_{S \in A^{\text{XS}}} A^{\text{X}}_S!} \frac{1}{z!} \prod_{S \in A^{\text{XS}}} E_S^{\underline{A^{\text{X}}_S}}\end{aligned}$$

where spaceDistHist = spaceDistributionHistorical. Thus the negative *relative dependent historical distribution coder space* equals the *alignment* minus the *historical mis-alignment*

$$\begin{aligned}&-(\text{spaceDistHist}(E,z)(A) - \text{spaceDistHist}(E,z)(A^{\text{X}})) \\ &\quad= \sum_{S \in A^{\text{S}}} \ln A_S! - \sum_{S \in A^{\text{XS}}} \ln A^{\text{X}}_S! - \sum_{S \in A^{\text{XS}}} (\underline{A_S} - \underline{A^{\text{X}}_S}) \ln E_S \\ &\quad= \text{alignment}(A) - \sum_{S \in A^{\text{XS}}} (\underline{A_S} - \underline{A^{\text{X}}_S}) \ln E_S\end{aligned}$$

where the falling factorial notation is abused such that $\underline{y} \ln x = \ln x^{\underline{y}}$. The *historical mis-alignment* depends on the *distribution histogram*, E, and its *size*, z_E. In the case where the *distribution history* is *independent*, $E = E^{\text{X}}$, conjecture that the *historical mis-alignment* tends to zero as the *distribution history size*, z_E, tends to infinity

$$\lim_{z_E \to \infty} \sum_{S \in A^{\mathrm{XS}}} (\underline{A_S} - \underline{A^{\mathrm{X}}_S}) \ln E^{\mathrm{X}}_S = 0$$

Thus in the limit the negative *independently-distributed relative dependent historical distribution coder space* equals the *alignment*

$$\lim_{z_E \to \infty} -(\mathrm{spaceDistHist}(E^{\mathrm{X}}, z)(A) - \mathrm{spaceDistHist}(E^{\mathrm{X}}, z)(A^{\mathrm{X}})) = \mathrm{alignment}(A)$$

As shown above the *space* in the *multinomial distribution coder* is greater than or equal to that of the *space* in the *historical distribution coder* because of the unused codes in the *multinomial distribution coder* that are not needed to represent the *coder domain*, $\mathrm{P}(H_E)$. However, for *independent distribution histogram*, $E = E^{\mathrm{X}}$, in the limit as the *size* z_E tends to infinity, the *independently-distributed relative dependent multinomial distribution coder space* equals the *independently-distributed relative dependent historical distribution coder space* which equals the *alignment*.

A.9 Transform and partition space

Let the *coder domain* $\mathcal{T}_{U,\mathrm{i},\leq y} \subset \mathcal{A}_{U,\mathrm{i},\leq y} \times \mathrm{P}(\mathcal{V}_U)$ be the set of *transforms* in *system* U having *trimmed integral histograms* of maximum *size* $y \in \mathbf{N}$

$$\mathcal{T}_{U,\mathrm{i},\leq y} = \{(X, W) : (X, W) \in \mathcal{T}_U,\ X \in \mathcal{A}_{U,\mathrm{i},\leq y}\}$$

where the finite set of *trimmed integral histograms* having *size* less than or equal to y is

$$\mathcal{A}_{U,\mathrm{i},\leq y} = \{\mathrm{trim}(A) : A \in \mathcal{A}_{U,\mathrm{i}},\ \mathrm{size}(A) \leq y\}$$

In the case of a *regular system* U, having *dimension* $n = |U|$ and such that all the *variables* have the same *valency* d, $\forall u \in \mathrm{vars}(U)\ (|U_u| = d)$, the cardinality of the *coder domain* $\mathcal{T}_{U,\mathrm{i},\leq y}$ is such that $|\mathcal{T}_{U,\mathrm{i},\leq y}| < 2^n |\mathcal{A}_{U,\mathrm{i},\leq y}|$. Hence

$$|\mathcal{T}_{U,\mathrm{i},\leq y}| < y 2^{2n} \frac{(y + d^n - 1)!}{y!\ (d^n - 1)!}$$

The *coder domain* $\mathcal{T}_{U,\mathrm{i},\leq y}$ is finite so a *minimal coder* $C_{\mathrm{T,m}} \in \mathrm{coders}(\mathcal{T}_{U,\mathrm{i},\leq y})$ can be constructed by enumeration of the *coder domain* in a similar fashion to the *minimal histogram coder*, $C_{\mathrm{A,m}}$, above. The *space* of the *minimal coder* is $\mathrm{space}(C_{\mathrm{T,m}})(T) = \ln |\mathcal{T}_{U,\mathrm{i},\leq y}|$.

Consider the *transform coder*

$$C_{\mathrm{T}} = \mathrm{coderTransform}(U, y, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{T}_{U,\mathrm{i},\leq y})$$

where $U \in \mathcal{U}$ is a *system*, $y \in \mathbf{N}$ is the cardinality of the *identifier set*, D_{V} and D_{S} are orders on the *variables* and *states* in U. The *space* of the *transform coder* C_{T} is that of the *trimmed integral histogram coder*, C_{A}, plus *space* required to define the subset of the *transform's variables* which are the *derived variables*. This extra *space* is that of a pair $\mathbf{N} \times \mathbf{N}$, the first of which defines the cardinality of the subset *derived variables*, $|W|$, and the second of which defines the combination

$$\mathrm{space}(C_{\mathrm{T}})((X, W)) = \mathrm{space}(C_{\mathrm{A}})(X) + \mathrm{space}(n + 1) + \mathrm{spaceSubset}(n, |W|)$$

where $n = |\mathrm{vars}(X)|$. An alternative method would be to *encode* the definition of W in a list $\mathcal{L}(\mathrm{bits})$ which would have *space* $n \ln 2$ instead. A third method is to define all the *variables* together in a weak composition for the three sets of *variables*, $V = \mathrm{und}(T)$, $W = \mathrm{der}(T)$ and $\mathrm{dom}(U)$, plus a classification, $\mathrm{spCom}(3, |U|) + \mathrm{spCl}(\{(V, |V|), (W, |W|), (U, |U| - |V| - |W|)\})$, where $\mathrm{spCom} = \mathrm{spaceCompositionWeak}$ and $\mathrm{spCl} = \mathrm{spaceClassification}$.

Now consider a *coder* having a subset *coder domain* which is the *unit transforms* $C_{\mathrm{T,U}} \in \mathrm{coders}(\mathcal{T}_{U,\mathrm{i},\leq y} \cap \mathcal{T}_{\mathrm{U}})$. In this special case the *space* of the weak composition that defines the *counts* of *transform* T, $\mathrm{spaceCounts}(U)(X)$ where $(X, W) = T$, is not needed. Instead the *histogram* X is simply defined by means of a subset of the *volume*

$$\begin{aligned}\mathrm{space}(C_{\mathrm{T,U}})((X, W)) \quad = \quad & \mathrm{spaceVariables}(U)(n) + \\ & \mathrm{space}(n + 1) + \mathrm{spaceSubset}(n, |W|) + \\ & \mathrm{space}(|X^{\mathrm{C}}| + 1) + \mathrm{spaceSubset}(|X^{\mathrm{C}}|, |X^{\mathrm{F}}|)\end{aligned}$$

where $n = |\mathrm{vars}(X)|$. Again, the subset of the *volume* could be encoded instead in a list of *space* $|X^{\mathrm{C}}| \ln 2$.

To move on to *functional transforms*, first consider the *coder* of *partitions*. The *partition coder* is constructed

$$C_{\mathrm{P}} = \mathrm{coderPartition}(U, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{R}_U)$$

where $U \in \mathcal{U}$ is a *system*, D_{V} and D_{S} are orders on the *variables* and *states* in U, and the *coder domain* $\mathcal{R}_U$ is the finite set of all *partitions* of the *variables* of the *system* U

$$\mathcal{R}_U = \bigcup\{\mathrm{B}(W^{\mathrm{CS}}) : W \in \mathrm{P}(\mathrm{vars}(U))\}$$

where B is the partition function. A *partition* $P \in \mathcal{R}_U$ can be encoded in an intermediate tuple $((n, N_V), R_P) \in \mathbf{N}^2 \times \mathbf{N}$. The first element, (n, N_V), encodes the *variables* in the same way as the *history coder* C_{H}, above. The *space* is $\mathrm{spaceVariables}(U)(n)$ where $V = \mathrm{vars}(P)$ and $n = |V|$. The last element of the tuple, R_P, chooses one of the *partitions*. Given order D_{S} choose $R \in \mathrm{enums}(\mathrm{B}(V^{\mathrm{CS}}))$. Then

$$R_P \in \{1 \ldots |\mathrm{B}(V^{\mathrm{CS}})|\} = \{1 \ldots \mathrm{bell}(v)\}$$

where *volume* $v = |V^{\mathrm{C}}|$. Define $\mathrm{spacePartition}(U) \in \mathrm{P}(\mathcal{V}_U) \to \ln \mathbf{N}_{>0}$ as

$$\mathrm{spacePartition}(U)(V) := \ln \mathrm{bell}(v)$$

Define $\mathrm{spacePartition}(U)(\emptyset) := 0$. The *space* of the *partition* is

$$\mathrm{space}(C_{\mathrm{P}})(P) = \mathrm{spaceVariables}(U)(|V|) + \mathrm{spacePartition}(U)(V)$$

where $V = \mathrm{vars}(P)$.

The set of *binary partitions* is $\mathcal{R}_{U,\mathrm{b}} = \{\{C,\ W^{\mathrm{CS}} \setminus C\} : W \in \mathrm{P}(\mathrm{vars}(U)),\ C \in \mathrm{P}(W^{\mathrm{CS}})\} = \{P : P \in \mathcal{R}_U,\ |P| = 2\}$. The *binary partition coder* is constructed

$$C_{\mathrm{P,b}} = \mathrm{coderPartitionBinary}(U, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{R}_{U,\mathrm{b}})$$

Define $\mathrm{spacePartitionBinary}(U) \in \mathrm{P}(\mathcal{V}_U) \to \ln \mathbf{N}_{>0}$ as

$$\mathrm{spacePartitionBinary}(U)(V) := v \ln 2$$

Define $\mathrm{spacePartitionBinary}(U)(\emptyset) := 0$. The *space* of the *binary partition* is

$$\mathrm{space}(C_{\mathrm{P,b}})(P) = \mathrm{spaceVariables}(U)(|V|) + \mathrm{spacePartitionBinary}(U)(V)$$

where $V = \mathrm{vars}(P)$.

The *partition transform* $T \in \mathcal{T}_{U,\mathrm{P}} = \{P^{\mathrm{T}} \in \mathcal{R}_U\} \subset \mathcal{T}_{U,\mathrm{f},1}$ has a single *derived partition variable* P from which it is constructed, $T = P^{\mathrm{T}} = (\{S \cup \{(P, C)\} : C \in P,\ S \in C\}^{\mathrm{U}}, \{P\})$. The *partition variable*, $P = T^{\mathrm{P}}$, is not in the *system*, $P \notin \mathrm{vars}(U)$, but is in the infinite *implied system*, $P \in \mathrm{vars}(\mathrm{implied}(U))$. The *coder* of *partition transforms*

$$C_{\mathrm{T,P}} = \mathrm{coderTransformPartition}(U, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{T}_{U,\mathrm{P}})$$

is exactly equal to *partition coder*, C_{P}, after mapping bijectively between the *coder domains*, $\mathcal{T}_{U,\mathrm{P}} :\leftrightarrow: \mathcal{R}_U$. The *space* is equal, $\mathrm{space}(C_{\mathrm{T,P}})(T) = \mathrm{space}(C_{\mathrm{P}})(T^{\mathrm{P}})$.

Similarly, the *space* of the *binary partition transform coder* $C_{\mathrm{T,P,b}}$ equals the *space* of the *binary partition coder*, $C^{\mathrm{s}}_{\mathrm{T,P,b}}(T) = C^{\mathrm{s}}_{\mathrm{P,b}}(T^{\mathrm{P}})$.

To *encode* a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$, where all of the *transform's variables* are in the *system*, $\mathrm{vars}(T) \in \mathrm{dom}(U)$, map the *transform's* subset of the *derived variables' cartesian states*, $(X\%W)^{\mathrm{S}} \subseteq W^{\mathrm{CS}}$ where $(X, W) = T$, to the components of the *partition* $P = T^{\mathrm{P}} \in \mathrm{B}(V^{\mathrm{CS}})$ of the *underlying variables* $V = \mathrm{und}(T)$, such that $|X\%W| = |P|$. The *space* to *encode* the map between the *derived* and *underlying* is $\mathrm{spaceSubset}(|W^{\mathrm{C}}|, |P|)$ plus the *space* of the permutation $\ln |P|!$, or, equivalently, the *space* of the falling factorial $\ln |W^{\mathrm{C}}|^{\underline{|P|}}$. The *space* to define the *partition* is smaller than for the *partition coder* because the number of *components* is fixed by the definition of the *derived variables* W in the *system* U and by the functional mapping, $X\%W :\leftrightarrow: P$, of the *components* to the *derived states*. Instead use the Stirling number of the second kind $\mathrm{stir} \in \mathbf{N}_{>0} \times \mathbf{N} \to \mathbf{N}_{>0}$ rather than the Bell number. The *partition space* is $\ln(\mathrm{stir}(|V^{\mathrm{C}}|, |P|))$. Let

$$C_{\mathrm{T,f,1}} = \mathrm{coderTransformOneFunc}(U, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{T}_{U,\mathrm{f},1})$$

in

$$\begin{aligned}
\mathrm{space}(C_{\mathrm{T,f,1}})(T) \quad = \quad & \mathrm{spaceVariables}(U)(n) + \\
& \mathrm{space}(n+1) + \mathrm{spaceSubset}(n, |W|) + \\
& \mathrm{space}(|W^{\mathrm{C}}|) + \mathrm{spaceSubset}(|W^{\mathrm{C}}|, |T^{\mathrm{P}}|) + \mathrm{space}(|T^{\mathrm{P}}|!) + \\
& \mathrm{space}(\mathrm{stir}(|V^{\mathrm{C}}|, |T^{\mathrm{P}}|))
\end{aligned}$$

where $n = |\mathrm{vars}(T)|$, $W = \mathrm{der}(T)$ and $V = \mathrm{und}(T)$.

The *coder* $C_{\mathrm{T,f,U}} \in \mathrm{coders}(\mathcal{T}_{U,\mathrm{f,U}})$ is intermediate between *coders* $C_{\mathrm{T,U}} \in \mathrm{coders}(\mathcal{T}_{U,\mathrm{i},\leq y} \cap \mathcal{T}_{\mathrm{U}})$ and $C_{\mathrm{T,f,1}} \in \mathrm{coders}(\mathcal{T}_{U,\mathrm{f},1})$. It requires *space* to define both the *partition* and the subset of the *volume* and so is larger than the *one functional coder*.

Conjecture that the *space* of a *one functional transform* $T \in \mathcal{T}_{U,\mathrm{f},1}$ having at least one *derived variable*, $|\mathrm{der}(T)| \geq 1$, must be such that

$$\mathrm{space}(C_{\mathrm{T,P}})(T^{\mathrm{PT}}) \leq \mathrm{space}(C_{\mathrm{T,f,1}})(T) \leq \mathrm{space}(C_{\mathrm{T,f,U}})(T) \leq \mathrm{space}(C_{\mathrm{T,U}})(T)$$

Of course, this conjecture depends on the exact definition of $C_{\mathrm{T,f,U}}$ which is not made explicit here.

A.10 Functional definition set space and Decomposition space

If *system* U is finite then the set of *functional definition sets*, $\mathcal{F}_U \subseteq \mathrm{P}(\mathcal{T}_{U,\mathrm{f,U}})$, in that *system* is also finite. Therefore there exists a *minimal coder*

$C_{\mathrm{F,m}} \in \mathrm{coders}(\mathcal{F}_U)$ such that the *space* of a member of the *coder domain* $F \in \mathcal{F}_U$ is constant, $\mathrm{space}(C_{\mathrm{F,m}})(F) = \ln |\mathcal{F}_U|$. Similarly there can be constructed a *minimal coder* of the set of *one functional definition sets*, $\mathcal{F}_{U,1} \subseteq \mathrm{P}(\mathcal{T}_{U,\mathrm{f},1})$, $C_{\mathrm{F,1,m}} \in \mathrm{coders}(\mathcal{F}_{U,1})$, such that $\mathrm{space}(C_{\mathrm{F,1,m}})(F) = \ln |\mathcal{F}_{U,1}|$ where $F \in \mathcal{F}_{U,1}$.

The set of *functional definition sets* is a subset of the powerset of the set of *unit functional transforms* in the finite *system* U, $\mathcal{F}_U \subseteq \mathrm{P}(\mathcal{T}_{U,\mathrm{f,U}})$, so a similar method of *coder* would be to encode the subset of *transforms*. Let $C_{\mathrm{F,S}} \in \mathrm{coders}(\mathcal{F}_U)$ and $F \in \mathcal{F}_U$, and define the encoding such that $\mathrm{space}(C_{\mathrm{F,S}})(F) = \mathrm{space}(|\mathcal{T}_{U,\mathrm{f,U}}| + 1) + \mathrm{spaceSubset}(|\mathcal{T}_{U,\mathrm{f,U}}|, |F|)$. This *coder* is not *minimal* in non-trivial *systems* because there are members of the powerset of the set of *one functional transforms*, $\mathrm{P}(\mathcal{T}_{U,\mathrm{f,U}})$, which are not *functional definition sets*.

It is possible to avoid the necessity of calculating $\mathcal{F}_U$ or $\mathcal{T}_{U,\mathrm{f,U}}$ (or, in the case of *one functional fuds* and *transforms*, $\mathcal{F}_{U,1}$ or $\mathcal{T}_{U,\mathrm{f},1}$), required by the methods above, by means of *list coders* (see appendix 'List and tree coders'). *List coders* encode a list in a generic manner, leaving the details of the encoding of the elements of the list to an *underlying coder*. Convert the *fud* $F \in \mathcal{F}_U$ to a list of *transforms* $L \in \mathcal{L}(\mathcal{T}_{U,\mathrm{f,U}})$ and then encode the list in one of the *list coders* $\mathrm{coders}(\mathcal{L}(\mathcal{T}_{U,\mathrm{f,U}}))$. The list L is the inverse of one of the enumerations of F, $L = \mathrm{flip}(M)$ where $M \in \mathrm{enums}(F)$. Any ordering of the list may be chosen. In particular, there is no need to order by *dependency*. If we choose to limit the cardinality of the *fud*, $|F| \leq y$, we can use a *limited list coder*

$$C_{\mathrm{F,L}} = \mathrm{coderListLimited}(C_{\mathrm{T,f,U}}, y) \in \mathrm{coders}(\mathcal{L}_y(\mathcal{T}_{U,\mathrm{f,U}}))$$

where $\mathcal{L}_y(\mathcal{T}_{U,\mathrm{f,U}})$ is the set of lists of *unit functional transforms* in *system* U of length less that or equal to y. The *underlying coder* $C_{\mathrm{T,f,U}}$ is described above in the section 'Transform and partition space'. The parameter y need not be greater than the maximum possible cardinality of a *fud* in the *system* U, $y \leq |\mathcal{T}_{U,\mathrm{f,U}}|$. *Fuds* which exclude *null transforms* have at least one *derived variable* per *transform*. Non-empty *fuds* which exclude *disjoint transforms* have at least one *underlying variable*. Let $\mathcal{F}_{\mathrm{X}} \subset \mathcal{F}$ be the set of *fuds* which exclude both special cases, $\forall F \in \mathcal{F}_{\mathrm{X}}\ \forall T \in F\ (\mathrm{der}(T) \neq \emptyset \wedge \mathrm{und}(T) \neq \emptyset)$. To encode only *fuds* in $\mathcal{F}_{\mathrm{X}}$ places a maximum on the list length, $y \leq r - 1$ where $r = |\mathrm{vars}(U)|$.

If we do not wish to explicitly limit the cardinality of the *fud* we can use the *unlimited list coder*

$$C_{\mathrm{F,U}} = \mathrm{coderListTerminating}(C_{\mathrm{T,f,U}}) \in \mathrm{coders}(\mathcal{L}(\mathcal{T}_{U,\mathrm{f,U}}))$$

Neither the *limited list coder* of *functional definition sets*, $C_{\mathrm{F,L}}$, nor the *unlimited list coder*, $C_{\mathrm{F,U}}$, are *minimal coders*, because of the cost of the list *overhead space*, and because the set of *fuds* is a proper subset of the powerset of *unit functional transforms* in non-trivial *systems*, $\mathcal{F}_U \subset \mathrm{P}(\mathcal{T}_{U,\mathrm{f,U}})$.

The *space* of a *transform list coder* of a *fud* is the sum the *spaces* of the *transforms* $\sum_{T \in F} \mathrm{space}(C_{\mathrm{T,f,U}})(T)$ plus the *overhead space*. The *overhead space* is a constant, $\ln(y+1)$, in $C_{\mathrm{F,L}}$, and a linear function of $|F|$, $(|F|+1)\ln 2$, in $C_{\mathrm{F,U}}$.

Consider *partition functional definition set coders*, $\mathrm{coders}(\mathcal{F}_{U,\mathrm{P}})$. *Partition fuds* are a subset of *one fuds*, $\mathcal{F}_{U,\mathrm{P}} \subset \mathcal{F}_{U,1}$, and so are a subset of the set of sets of *unit functional transforms*, $\mathcal{F}_{U,\mathrm{P}} \subset \mathrm{P}(\mathcal{T}_{U,\mathrm{f,U}})$. So a *fud* $F \in \mathcal{F}_{U,\mathrm{P}}$ could be encoded in *list coder* such as $C_{\mathrm{F,L}}$ or $C_{\mathrm{F,U}}$ with *underlying coder* $C_{\mathrm{T,f,U}}$. However if the *system* U were to contain all of the *partition variables* of the *partition fuds* in $\mathcal{F}_{U,\mathrm{P}}$ then both the *system* and the set of *partition fuds* would be necessarily infinite. So an *underlying coder* cannot be constructed from *system* U or $\mathrm{implied}(U)$. A finite *system* can be constructed from the *partition fuds* in finite *system* U, $U' = \{(P,P) : F \in \mathcal{F}_{U,\mathrm{P}},\ P \in \mathrm{dom}(\mathrm{def}(F))\} \cup U$. The *list coders* $C_{\mathrm{F,U}}, C_{\mathrm{F,L}} \in \mathrm{coders}(\mathcal{L}(\mathcal{T}_{U',\mathrm{f,U}}))$ and *underlying coder* $C_{\mathrm{T,f,U}}$ are then parameterised by the finite *system* U'.

A more efficient *partition functional definition set coder* implementation in a *list coder* is to use the *partition transform coder*, $C_{\mathrm{T,P}}$, as the *underlying coder* and to construct the *system* incrementally by looking back at the list of *partition transforms* so far. In this way, the *underlying coder* need not choose the *underlying variables* of the *partition* from all of the *variables* in the *fuds* of $\mathcal{F}_{U,\mathrm{P}}$. That is, instead of using *system* U', the *partition transform coder*, $C_{\mathrm{T,P}}$, can use the much smaller *system* U plus the *partition variables* previously defined in the list of the *fud's transforms*. Of course, the list must be ordered by *fud dependency* so that the references exist. Let

$$C_{\mathrm{F,U,P}} = \mathrm{coderFudPartition}(U, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{F}_{U,\mathrm{P}})$$

In the *non-minimal coders* above, the *fud* has been treated as a unordered collection of *transforms*, ignoring the *definitions* constraint on the *fud*. See appendix 'Functional definition set coders' for the details of the implementation of $C_{\mathrm{F,U,P}}$ and a discussion of other *fud coders* that use the *fud* constraints to reduce *space*.

At this stage note that the *space* of the *partition fud coder*, $C_{\mathrm{F,U,P}}$, is at least the sum of the *space* of the set of *partitions*

$$\begin{aligned}\text{space}(C_{\text{F,U,P}})(F) &> \sum_{T \in F} \text{spacePartition}(U')(\text{und}(T)) \\ &= \sum_{T \in F} \ln \text{bell}(\text{volume}(U')(\text{vars}(T^{\text{P}}))\end{aligned}$$

Consider the *partition space* of $F = \{P_K^{\text{T}}, P_{V \setminus K}^{\text{T}}\} \in \mathcal{F}_{U,\text{P}}$ where $V \subset \text{vars}(U)$, $K \subset V$ and $P_X \in \text{B}(X^{\text{CS}})$. Let V be *regular* having even *dimension* $n = |V|$, $n/2 \in \mathbf{N}$ and *valency* $\{d\} = \{|U_u| : u \in V\}$. Let the subset K be such that $|K| = n/2$. Then the *partition space* of F is

$$\text{spacePartition}(U)(K) + \text{spacePartition}(U)(V \setminus K) = 2 \ln \text{bell}(d^{n/2})$$

The *partition space* of the *equivalent transform* $F^{\text{TPT}} \in \mathcal{T}_{U,\text{P}}$ is

$$\text{spacePartition}(U)(V) = \ln \text{bell}(d^n)$$

Conjecture that $\forall i \in \mathbf{N}\ ((\text{bell}(i))^2 \leq \text{bell}(i^2))$ because the Bell number is log convex. Therefore conjecture that the *partition space* of the *fud* F is less than or equal to the *partition space* of its *equivalent transform*, F^{T}, $2 \ln \text{bell}(d^{n/2}) \leq \ln \text{bell}(d^n)$.

A *coder* of *multi-partition transforms*, $C_{\text{T,P*}}$, may be implemented as a special case of the *partition fud coder*, $C_{\text{F,U,P}}$, by *exploding* the *contracted transform*, $\text{explode}(T^{\%}) \in \mathcal{F}_{U,\text{P}}$ where $T \in \mathcal{T}_{U,\text{P*}}$.

Now consider a *coder* of *partition fud decompositions*. The set of *partition fud decompositions* is defined

$$\mathcal{D}_{\text{F},U,\text{P}} = \mathcal{D}_{\text{F}} \cap \text{trees}(\mathcal{S}_{U'} \times \mathcal{F}_{U,\text{P}})$$

where the finite *system* U' is defined $U' = \{(P,P) : F \in \mathcal{F}_{U,\text{P}},\ P \in \text{dom}(\text{def}(F))\} \cup U$. The *unlimited partition fud decomposition coder* is constructed

$$C_{\text{D,F,U,P}} = \text{coderDecompFudPartition}(U, D_{\text{V}}, D_{\text{S}}) \in \text{coders}(\mathcal{D}_{\text{F},U,\text{P}})$$

The *fud decomposition coder* implements the encoding of the *decomposition* tree of *state-fud* pairs by means of an *unlimited list tree coder*, $C_{\text{U,T}}$, defined in appendix 'List and tree coders'. The *fuds* of the *state-fud* pairs are encoded by an *unlimited partition fud coder*, $C_{\text{F,U,P}}$. Let $D \in \mathcal{D}_{\text{F},U,\text{P}}$. The *state* S of the pair is in the *derived states* of the parent *fud* G in the tree,

$S \in W^{\mathrm{CS}}$, where $((\cdot, G), (S, \cdot)) \in \mathrm{steps}(D)$ and $W = \mathrm{der}(G)$. So the encoding and decoding of the *state*, S, is preceded by the encoding and decoding of the parent *fud*, G. The *state* may be encoded by indexing the *cartesian* of the *derived variables* of the parent *fud*. That is, by choosing an enumeration from $\mathrm{enums}(W^{\mathrm{CS}})$. The *space* of the *state* encoding is $\ln |W^{\mathrm{C}}|$.

The *fud decomposition coder space* is greater than the *space* of the *partition fud* encodings, and so is greater than the total *space* of the *partition transforms* of the *fuds*,

$$\begin{aligned}
C^{\mathrm{s}}_{\mathrm{D,F,U,P}}(D) &> \sum_{F \in \mathrm{fuds}(D)} C^{\mathrm{s}}_{\mathrm{F,U,P}}(F) \\
&> \sum_{F \in \mathrm{fuds}(D)} \sum_{T \in F} \mathrm{spacePartition}(U')(\mathrm{und}(T)) \\
&= \sum_{F \in \mathrm{fuds}(D)} \sum_{T \in F} \ln \mathrm{bell}(\mathrm{volume}(U')(\mathrm{vars}(T^{\mathrm{P}})))
\end{aligned}$$

An *unlimited partition transform decomposition coder*, $C_{\mathrm{D,U,P}}$, is a special case of the *unlimited partition fud decomposition coder*, $C_{\mathrm{D,F,U,P}}$, constructed

$$C_{\mathrm{D,U,P}} = \mathrm{coderDecompPartition}(U, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{D}_{U,\mathrm{P}})$$

where $\mathcal{D}_{U,\mathrm{P}} = \mathcal{D} \cap \mathrm{trees}(\mathcal{S}_{U'} \times \mathcal{T}_{U,\mathrm{P}})$.

A.11 Functional definition set coders

Consider the finite *coder domain* of *functional definition sets*, $\mathcal{F}_U \subseteq \mathrm{P}(\mathcal{T}_{U,\mathrm{f,U}})$, in the finite *system* U. In the section 'Functional definition set space and Decomposition space', above, the *non-minimal fud coders* treated a *fud* $F \in \mathcal{F}_U$ as a collection of *transforms*, ignoring the *definitions* constraint on the *fud*. That is, that no *derived variable* of a *transform* in the *fud* can be a *derived variable* in another *transform*, $\mathrm{ran}(F) \setminus \emptyset \in \mathrm{B}(\mathrm{dom}(\mathrm{def}(F)))$ where $\mathrm{def}(F) \neq \emptyset$ and def = definitions. The *space* of the *fud coder* can be reduced by making use of this constraint. The *classification definition fud coders* $C_{\mathrm{F,L,C}} \in \mathrm{coders}(\mathcal{F}_U)$ and $C_{\mathrm{F,L,C,1}} \in \mathrm{coders}(\mathcal{F}_{U,1})$ encode part of the *fud* in initial *space*, followed by a *limited list coder* of the remaining parts of the *transforms*. Treat the *fud* as a classification of the *derived variables* of the *transforms*. In order to do this first encode the cardinality of the *fud*, $|F|$. Do this by using a *limited list coder* of maximum length y, in which case the *space* of the cardinality is the up-front *overhead space* $\ln(y + 1)$. Then the *space* required to specify the cardinalities of *derived variables* in the *transforms* of $F \in \mathcal{F}_U$ is weak composition *space* spaceCompositionWeak$(|F|+1, |U|)$. If the *coder domain* is constrained to *fuds* which exclude *null transforms* and *disjoint transforms*, $\mathcal{F}_U \cap \mathcal{F}_{\mathrm{X}}$, then $y \leq |U|-1$ and the *space* of the cardinality of the *fud* is at most $\ln |U|$. In this case, the *space* to specify the cardinalities of *derived variables* is only strong composition *space* spaceComposition$(|F|+1, |U|)$, where spaceComposition $\in \mathbf{N}_{>0} \times \mathbf{N}_{>0} \to \ln \mathbf{N}_{>0}$ is defined as spaceComposition$(k,n) := \ln |\mathrm{C}(\{1 \ldots k\}, n)|$. In both cases, $\mathcal{F}_U$ and $\mathcal{F}_U \cap \mathcal{F}_{\mathrm{X}}$, the *space* required to specify the classification of *derived variables* in the *transforms* is spaceClassification$(\{(T, |\mathrm{der}(T)|) : T \in F\})$. The *space* of the *underlying variables* in the *limited list coder* is $\sum$(spaceVariables$(U)(|\mathrm{und}(T)|) : T \in F)$. The remaining *space* of the *states* of the *histograms* of the *transforms* is the same as defined in the respective *underlying coders*, $C_{\mathrm{T,f,U}}$ for *fud coder* $C_{\mathrm{F,L,C}}$ and $C_{\mathrm{T,f,1}}$ for *one fud coder* $C_{\mathrm{F,L,C,1}}$. In the case of *one functional definition sets*, $F \in \mathcal{F}_{U,1}$, this *space* for the *limited list coder* of each $T \in F$ is space$(|W^{\mathrm{C}}|)$ + spaceSubset$(|W^{\mathrm{C}}|, p)$ + space$(p!)$ + space(stir$(|V^{\mathrm{C}}|, p)$) where $p = |\mathrm{inverse}(T)|$, $W = \mathrm{der}(T)$ and $V = \mathrm{und}(T)$. The *overhead space* of the *limited list coder* of $\ln(y+1)$ performs the dual purpose of defining the length of the list, $|L|$ where $L \in \mathcal{L}_y(\mathcal{T}_{U,\mathrm{f,U}})$ or $L \in \mathcal{L}_y(\mathcal{T}_{U,\mathrm{f},1})$, and encoding the cardinality of the *fud*, $|F| = |L|$, for the *definition classification* above.

The method of *definition classification* does not address the *space* that can be saved in the *underlying variables* by utilising the *depends* relations between the *transforms* of the *fud*. The *underlying variables* of a *transform* $T \in F$ need not be chosen as a subset of the entire *system* U, but only need to be chosen from the *underlying variables* of the *fud* and the *definition* of the underlying *depends fud* of the *transform*, $\mathrm{und}(F) \cup \mathrm{vars}(\mathrm{depends}(F, \mathrm{der}(T)) \setminus \{T\}) \subseteq \mathrm{vars}(U)$. In this method the *space* of the *underlying variables* of *transform* T in *fud* F is not spVar$(U)(|\mathrm{und}(T)|)$, but rather spVar$(Q)(|\mathrm{und}(T)|)$ where *system* $Q \subseteq U$ is defined

$$Q = \{(v, U_v) : v \in \mathrm{und}(F) \cup \mathrm{vars}(\mathrm{depends}(F, \mathrm{der}(T)) \setminus \{T\})\}$$

and spVar = spaceVariables. Define system$(U) \in \mathcal{F}_U \to \mathcal{U}$ as

$$\mathrm{system}(U)(F) := \{(v, U_v) : v \in \mathrm{vars}(F)\}$$

First encode the *underlying variables* of the *fud* in *space* spVar$(U)(|\mathrm{und}(F)|)$. These *underlying variables* imply a subset of the *system* $R \subseteq U$ such that $R = \{(v, U_v) : v \in \mathrm{und}(F)\}$. Then choose a list ordered such that the *system* Q so far can be implied from a *lookback list coder*. A finite *system* cannot be implied where there are *fud circularities*, so consider only *one* *functional definition sets* $\mathcal{F}_{U,1}$. Choose a list $L = \text{flip}(M) \in \mathcal{L}(\mathcal{T}_{U,\text{f},1})$ where $M \in \text{enums}(F)$ ordered such that all of the underlying *depends* set of any *transform* precedes the *transform*

$$\forall (i, T) \in L \ (\text{depends}(F, \text{der}(T)) \subseteq \{S : (j, S) \in L,\ j \leq i\})$$

Define a variation on the *one functional transform coder* that takes the *underlying variables* from *system* $Q \subseteq U$ and the *derived variables* from *system* $U \setminus Q$. Let

$$C_{\text{T,F,1,S}} = \text{coderTransformOneFuncSplitSystem}(U, Q, D_{\text{V}}, D_{\text{S}}) \in \text{coders}(\mathcal{T}_{U,\text{f},1})$$

in

$$\begin{aligned}\text{space}(C_{\text{T,F,1,S}})(T) \ = \ & \text{space}(|U| + 1) + \text{space}(n) + \\ & \text{spaceSubset}(|U \setminus Q|, |W|) + \text{spaceSubset}(|Q|, |V|) + \\ & \text{space}(|W^{\text{C}}|) + \text{spaceSubset}(|W^{\text{C}}|, p) + \text{space}(p!) + \\ & \text{space}(\text{stir}(|V^{\text{C}}|, p))\end{aligned}$$

where $n = |\text{vars}(T)|$, $p = |\text{inverse}(T)|$, $W = \text{der}(T)$ and $V = \text{und}(T)$. By splitting the *system* into Q and $U \setminus Q$ the *coder* partly addresses the *space* saved by *definitions constraint* on the *fud*. By contrast, the *definition classification coders*, which also use the *definitions constraint* to reduce *space*, do not need to order the underlying list by *dependency*.

Let $B_{\text{T,F,1,S}} \in \mathcal{L}(\mathcal{T}_{U,\text{f},1}) \to \text{coders}(\mathcal{T}_{U,\text{f},1})$ in

$$C_{\text{F,U,S,B}} = \text{coderListTerminatingLookback}(B_{\text{T,F,1,S}}) \in \text{coders}(\mathcal{L}(\mathcal{T}_{U,\text{f},1}))$$

be defined as

$$B_{\text{T,F,1,S}}(L) := \text{enc}(U, \text{system}(U)(\text{ran}(L)) \cup R, D_{\text{V}}, D_{\text{S}})$$

where enc = coderTransformOneFuncSplitSystem.

The overall *space* for this *coder* $C_{\text{F,U,S}} \in \text{coders}(\mathcal{F}_{U,1})$, including the up-front encoding of the *underlying variables* of the *fud* F and the *list coder* with an underlying *split system one functional transform coder* is

$$\text{space}(C_{\text{F,U,S}})(F) := \text{spVar}(U)(|\text{und}(F)|) + \text{space}(C_{\text{F,U,S,B}})(L)$$

However, the order of the list in the *list coder* does not necessarily minimise the cardinality of the *system* Q given to the *underlying coder* to encode the *transform's underlying variables*. Typically there are more *underlying variables* than *derived variables*, $|V| > |W|$, in a *transform*, so we wish to minimise $|Q|$ in order to use the least *space*. Ideally $|\mathrm{system}(U)(\mathrm{ran}(L))| = |\mathrm{vars}(\mathrm{depends}(F, \mathrm{der}(T)) \setminus \{T\})|$, but sometimes it is larger. A method of minimising $|Q|$ is to use a list tree having the same layout as the *fud's variables tree*, $\mathrm{fudsTreeVariable}(F) \in \mathrm{trees}(\mathcal{V})$. The list tree is encoded using a *lookback list tree coder* which can construct the *underlying variables system* Q from the child list tree. If the tree has duplicate nodes then it is a graph and we could use a *referencing coder* to avoid duplicating the encoding of *transforms*. Of course, *list tree* and *list graph coders* would have additional *overhead space*.

Now consider *partition functional definition set coders*, $\mathrm{coders}(\mathcal{F}_{U,\mathrm{P}})$. *Partition fuds* are a subset of *one fuds*, $\mathcal{F}_{U,\mathrm{P}} \subset \mathcal{F}_{U,1}$, so we could use a *coder* such as $C_{\mathrm{F,L,C,1}}$ or $C_{\mathrm{F,U,S}}$ above. However, the additional constraint imposed on the *transforms* of *partition fuds* that the set of *derived variables* is a singleton set of the *partition variable*, $\forall T \in F\ (\mathrm{der}(T) = \{\mathrm{partition}(T)\})$, allows us to both simplify and reduce the *space* of the *coder* of a *partition fud* $F \in \mathcal{F}_{U,\mathrm{P}}$. First, the *defined variables* of the *fud* $\mathrm{dom}(\mathrm{def}(F))$ can imply a *system* disjoint of the *system* R where R contains the *underlying variables* of the *fud*, $\mathrm{und}(F) \subseteq \mathrm{vars}(R)$. Define $\mathrm{system} \in \mathcal{F} \to \mathcal{U}$ as

$$\mathrm{system}(F) := \{(P, P) : P \in \mathrm{dom}(\mathrm{def}(F))\}$$

Note that the *system* function here is defined for all *fuds*, $F \in \mathcal{F}$, but we are only interested in *partition fuds*, $F \in \mathcal{F}_{U,\mathrm{P}}$ where $\mathrm{system}(F) \cup R \subseteq U$. Second, while encoding the list in the *lookback unlimited list coder* we do not need to encode the *derived variable* explicitly because it is encoded in the *space* defined by the Bell number of the *underlying volume* rather than the lesser *space* defined by the Stirling number of the second kind. Let

$$C_{\mathrm{F,U,P}} = \mathrm{coderFudPartition}(R, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{F}_{U,\mathrm{P}})$$

Again, the subset of the *underlying variables* of the *fud* are encoded first before the list encoding takes place, $\mathrm{spVar}(R)(|\mathrm{und}(F)|)$. Then choose a list $L = \mathrm{flip}(M) \in \mathcal{L}(\mathcal{T}_{U,\mathrm{P}})$ where $M \in \mathrm{enums}(F)$ ordered such that all of the underlying *depends* set of any *transform* precedes the *transform* in the same way as the *one fud coder* $C_{\mathrm{F,U,S}}$ above. Let $B_{\mathrm{T,F,P}} \in \mathcal{L}(\mathcal{T}_{U,\mathrm{P}}) \to \mathrm{coders}(\mathcal{T}_{U,Q,\mathrm{P}})$ in

$$C_{\mathrm{F,U,P,B}} = \mathrm{coderListTerminatingLookback}(B_{\mathrm{T,F,P}}) \in \mathrm{coders}(\mathcal{L}(\mathcal{T}_{U,\mathrm{P}}))$$

be defined as

$$B_{\mathrm{T,F,P}}(L) := \mathrm{coderTransformPartition}(\mathrm{system}(\mathrm{ran}(L)) \cup R, D_{\mathrm{V}}, D_{\mathrm{S}})$$

Here the restricted *coder domain*, $\mathcal{T}_{U,Q,\mathrm{P}}$ is defined by $Q = \mathrm{system}(\mathrm{ran}(L)) \cup R$. So, strictly speaking, $C_{\mathrm{F,U,P,B}}$ has a *coder domain* which is a proper subset of $\mathcal{L}(\mathcal{T}_{U,\mathrm{P}})$. That is, only the *dependency* ordered lists of *partition transforms*.

The overall *space* of the *fud* F in this *coder* $C_{\mathrm{F,U,P}} \in \mathrm{coders}(\mathcal{F}_{U,P})$, including the up-front encoding of the *underlying variables*, $\mathrm{und}(F)$, and the *unlimited list coder* $C_{\mathrm{F,U,P,B}}$, with underlying *partition transform coder* $C_{\mathrm{T,P}}$, of the list of *transforms*, $L \in \mathcal{L}(\mathcal{T}_{U,\mathrm{P}})$, is

$$\mathrm{space}(C_{\mathrm{F,U,P}})(F) := \mathrm{spVar}(R)(|\mathrm{und}(F)|) + \mathrm{space}(C_{\mathrm{F,U,P,B}})(L)$$

The *fud space* is subject to an inequality which does not depend on L

$$\begin{aligned}&\mathrm{space}(C_{\mathrm{F,U,P}})(F) \leq \\ &\quad \mathrm{spVar}(R)(|\mathrm{und}(F)|) + \sum_{T \in F} \mathrm{space}(C_{\mathrm{T,P}})(T) + (|F|+1)\ln 2\end{aligned}$$

where

$$C_{\mathrm{T,P}} = \mathrm{coderTransformPartition}(U, D_{\mathrm{V}}, D_{\mathrm{S}}) \in \mathrm{coders}(\mathcal{T}_{U,\mathrm{P}})$$

and $\mathrm{system}(F) \cup R \subseteq U$. Now $\mathrm{space}(C_{\mathrm{T,P}})(T) = \mathrm{space}(C_{\mathrm{P}})(\mathrm{partition}(T))$ so

$$\begin{aligned}&\mathrm{space}(C_{\mathrm{F,U,P}})(F) \leq \\ &\quad \mathrm{spVar}(R)(|V_F|) + \sum_{T \in F}(\mathrm{spVar}(U)(|V_T|) + \mathrm{spPart}(U)(P_T)) + (|F|+1)\ln 2\end{aligned}$$

where $V = \mathrm{und}$, $P = \mathrm{partition}$, $\mathrm{spVar} = \mathrm{spVar}$ and $\mathrm{spPart} = \mathrm{spacePartititon}$.

The *variables space* is minimised in the inequality when $U = \mathrm{system}(F) \cup R$. An *unlimited list coder* minimises the *variables space* of a *fud* in the *partition fud coder*. For example, if a *power partition set*, $X = \mathrm{power}(U)(V) \in \mathcal{F}_{U,\mathrm{P}}$, has large cardinality and hence a large *system*, $|U| > |X|$, even small subsets $F \subset X$, where $\mathrm{und}(F) \subseteq V$ and $|F| \ll |X|$, would require large *variables space*. Conversely, the use of an *unlimited list coder* also avoids setting a maximum effective limit on the cardinality of the *system* implied by *limited list coders*, $y \leq |U| - 1$.

A.12 Derived history coders and search

The *derived history coders* discussed in section 'Derived history space', above, such as the *specialising derived substrate history coder*, $C_{\mathrm{G},V,\mathrm{T,H}}$, are *substrate history coders* in that they are given a *model* such as a *transform* T as a constructor parameter. The *substrate history coder domain* is then restricted to the *histories* in the *underlying variables* of the *transform*, $\mathcal{H}_{U,V,X}$, where $V = \mathrm{und}(T)$.

Now consider how the *derived substrate history coders* may be generalised to the unrestricted *history coder domain* where the *histories* may be in any of the *system variables*, $\mathcal{H}_{U,X} = \bigcup\{X \to V^{\mathrm{CS}} : V \subseteq \mathrm{vars}(U)\}$. The generalisation may be accomplished by removing the *transform* from the parameters altogether and (i) computing a *transform*, T_H, at the beginning of the encoding, (ii) encoding the *transform*, T_H, in a *transform coder*, and (iii) encoding the *history*, H, by means of a newly instantiated *derived substrate history coder* parameterised by the constructed *transform*, T_H. A subsequent decoding of the *history*, H, then proceeds by (i) decoding the *transform* T_H in the *transform coder*, and (ii) decoding the *history*, H, by means of a newly instantiated *derived substrate history coder* parameterised by the decoded *transform*, T_H.

The *derived history coders* need not be restricted to *one functional transforms* of a *substrate*, but can be generalised to any *model* (i) that can be a *model* parameter of the *history coder* and (ii) for which a *model coder* exists.

In particular, of the *substrate models* (i) the *partition transforms* of the *base fud*, $F_{U,V} = \{P^{\mathrm{T}} : P \in \mathrm{B}(V^{\mathrm{CS}})\} \subset \mathcal{F}_{U,V} \subset \mathcal{F}_{U,\mathrm{P}}$, can be encoded in a *partition transform coder*, $C_{\mathrm{T,P}}$, (ii) the *substrate fuds*, $\mathcal{F}_{U,V} \subset \mathcal{F}_{U,\mathrm{P}}$, can be encoded in an *unlimited partition fud coder*, $C_{\mathrm{F,U,P}}$, and (iii) the *substrate fud decompositions*, $\mathcal{D}_{\mathrm{F},U,V} \subset \mathcal{D}_{\mathrm{F},U,\mathrm{P}}$, can be encoded in an *unlimited partition fud decomposition coder*, $C_{\mathrm{D,F,U,P}}$.

The process of computing the *model* for each encoded *history* can be generalised also. This is done by defining the computation as a maximisation of a logarithm extended rational valued function of the *models* for the *history*. Let $Z \in \mathcal{H} \to (\mathcal{M} \to (\mathbf{Q} \cup \mathbf{Q} \ln \mathbf{Q}_{>0}))$ be a search function parameterised by *history*, where $\mathcal{M}$ is the set of *models*. The search function, Z, is constrained such that the *underlying variables* of the *models* equal the *history variables*, $\forall H \in \mathrm{dom}(Z)\ \forall M \in \mathrm{dom}(Z(H))\ (\mathrm{und}(M) = \mathrm{vars}(H))$. Then for *history* $H \in \mathcal{H}_{U,X}$ the *model* $M \in \mathcal{M}$ chosen for encoding is such that $M \in \mathrm{maxd}(Z(H))$. If there is more than one *model* in the maximum function's domain, $|\mathrm{maxd}(Z(H))| > 1$, then a *model* is chosen arbitrarily. For example, $\{M\} = \mathrm{maxd}(\mathrm{order}(D_{\mathcal{X}}, \mathrm{maxd}(Z(H))))$, where $D_{\mathcal{X}} \in \mathrm{enums}(\mathcal{X})$ is an arbitrary ordering.

In the particular case where a *derived history coder* is implemented by encoding some subset of the *substrate models*, $\mathcal{M}_{U,V}$, the search function is constrained $Z(H) \in \mathcal{M}_{U,V} \to (\mathbf{Q} \cup \mathbf{Q} \ln \mathbf{Q}_{>0})$.

An instance of the class of *derived history coders* may then be constructed by specifying (i) the *model coder* $C_{\mathrm{M}} \in \mathrm{coders}(\mathcal{M})$, (ii) the *model search function* $Z \in \mathcal{H} \to (\mathcal{M} \to (\mathbf{Q} \cup \mathbf{Q} \ln \mathbf{Q}_{>0}))$, and (iii) the *derived substrate history coder* function parameterised by *model* $C_{\mathrm{V}} \in \mathcal{M} \to \mathrm{coders}(\mathcal{H})$. The parameters of the constructor are constrained $\forall H \in \mathcal{H}_{U,X}\ \forall M \in \mathrm{dom}(Z(H))\ (M \in \mathrm{domain}(C_{\mathrm{M}}) \wedge M \in \mathrm{dom}(C_{\mathrm{V}}))$.

The *generic derived history coder* is constructed

$$\begin{aligned}&C(C_{\mathrm{M}}, Z, C_{\mathrm{V}}) = \\ &\quad \mathrm{coderHistoryDerivedGeneric}(U, X, C_{\mathrm{M}}, Z, C_{\mathrm{V}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \in \mathrm{coders}(\mathcal{H}_{U,X})\end{aligned}$$

A *history* $H \in \mathcal{H}_{U,X}$ is encoded as a tuple $(E_{\mathrm{M}}(M), E_{\mathrm{V}}(H)) \in \mathbf{N} \times \mathbf{N}$ where $M \in \mathrm{maxd}(Z(H))$, $(E_{\mathrm{M}}, \cdot, \cdot) = \mathrm{def}(C_{\mathrm{M}})$ and $(E_{\mathrm{V}}, \cdot, \cdot) = \mathrm{def}(C_{\mathrm{V}}(M))$.

The *space* of the *history* in the *generic derived history coder*, $C(C_{\mathrm{M}}, Z, C_{\mathrm{V}})$, is the *space* of the *model*, M, in the given *model coder*, C_{M}, plus the *space* of the *history*, H, in the given *derived substrate history coder*, $C_{\mathrm{V}}(M)$,

$$C(C_{\mathrm{M}}, Z, C_{\mathrm{V}})^{\mathrm{s}}(H) = C_{\mathrm{M}}^{\mathrm{s}}(M) + C_{\mathrm{V}}(M)^{\mathrm{s}}(H)$$

Derived history coders require more *space* to encode a *history* than their underlying *derived substrate history coders* do to encode the same *history*, $C(C_{\mathrm{M}}, Z, C_{\mathrm{V}})^{\mathrm{s}}(H) > C_{\mathrm{V}}(M)^{\mathrm{s}}(H)$, because of the necessity of the additional *space* to encode the *model*, $C_{\mathrm{M}}^{\mathrm{s}}(M)$. However, *derived history coders* are more flexible. Not only do they allow *histories* with arbitrary *substrates* within the *system*, but they also allow different *models* for *histories* in the same *substrate*. So $\mathrm{vars}(H_1) = \mathrm{vars}(H_2)$, where $H_1, H_2 \in \mathcal{H}_{U,X}$, does not necessarily imply $\mathrm{maxd}(Z(H_1)) = \mathrm{maxd}(Z(H_2))$. That is, a *derived substrate history coder* of domain $\mathcal{H}_{U,V,X}$ must encode all of the *histories* of the domain with the same *model* M, where $\mathrm{und}(M) = V$, defined as a parameter at *coder* instantiation. In contrast, a *derived history coder* can have different instantiations of its *derived substrate history coder* depending on the *history*, possibly with different *space*.

In addition, the *model space* typically varies with the *volume* rather than the *size*, whereas the *substrate history space* varies with *volume* and *size*. So the proportion of *derived history coder space* corresponding to the *model* decreases as the *size* of the *history* increases. For example, consider a *derived history coder* $C(C_{\mathrm{T,P}}, Z, C_{\mathrm{H},V,\mathrm{T,H}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$ constructed with a *partition transform coder*, $C_{\mathrm{T,P}}$, and *index derived substrate history coder* $C_{\mathrm{H},V,\mathrm{T,H}}$. Let $H \in \mathcal{H}_{U,X}$ and $T \in \mathrm{maxd}(Z(H))$. The *partition transform coder space* is

$$\mathrm{space}(C_{\mathrm{T,P}})(T) = \mathrm{spaceVariables}(U)(|V|) + \mathrm{spacePartition}(U)(V)$$

where $V = \mathrm{und}(T)$. The *index derived substrate history coder space* is

$$\begin{aligned} \mathrm{space}(C_{\mathrm{H},V,\mathrm{T,H}})(H) \;=\; & \mathrm{spaceIds}(|X|, |H|) + \\ & \mathrm{spaceEventsDerived}(U)(H, T) + \\ & \mathrm{spaceEventsPartition}(A, T) \end{aligned}$$

where $A = \mathrm{histogram}(H)$. So the *space* is

$$\begin{aligned} C(C_{\mathrm{T,P}}, Z, C_{\mathrm{H},V,\mathrm{T,H}})^{\mathrm{s}}(H) \;=\; & \mathrm{spaceVariables}(U)(|V|) + \\ & \mathrm{spaceIds}(|X|, |H|) + \\ & \mathrm{spacePartition}(U)(V) \\ & \mathrm{spaceEventsDerived}(U)(H, T) + \\ & \mathrm{spaceEventsPartition}(A, T) \end{aligned}$$

Partition space does not depend on *size*, but varies as $\ln \mathrm{bell}(v) < v \ln v$ where $v = |V^{\mathrm{C}}|$. The *events space* is $\mathrm{spaceEventsDerived}(U)(H, T) := z \ln w'$, where $w' = |(V^{\mathrm{C}} * T)^{\mathrm{F}}| = |T^{-1}|$, and the *partitioned events space* is $\mathrm{spaceEventsPartition}(A, T) := \sum_{(R,C) \in T^{-1}} (A * T)_R \ln |C|$. Both the *events space* and the *partitioned events space* scale with *size*, z. In this example the fraction of *space* which is *model space* varies as v/z.

A special case of a *model search function* is the *minimum space model search function* $Z_{\mathrm{m}}(C_{\mathrm{M}}, C_{\mathrm{V}}) \in \mathcal{H} \to (\mathcal{M} \to (\mathbf{Q} \cup \mathbf{Q} \ln \mathbf{Q}_{>0}))$, which is parameterised by a *model coder*, C_{M}, and a *derived substrate history coder* function, C_{V}. Here the *model* is chosen to minimise the *space* of the *history* encoding,

$$\begin{aligned} & Z_{\mathrm{m}}(C_{\mathrm{M}}, C_{\mathrm{V}})(H) = \\ & \quad \{(M, -(C^{\mathrm{s}}_{\mathrm{M}}(M) + C_{\mathrm{V}}(M)^{\mathrm{s}}(H))) : M \in \mathrm{domain}(C_{\mathrm{M}}),\ \mathrm{und}(M) = \mathrm{vars}(H)\} \end{aligned}$$

The *minimum space model search function* is such that for any *search function* $Z \in \mathcal{H} \to (\mathcal{M} \to (\mathbf{Q} \cup \mathbf{Q} \ln \mathbf{Q}_{>0}))$ the *minimum space model search function* always has least *space*

$$\forall H \in \mathcal{H}_{U,X}\ (C(C_{\mathrm{M}}, Z, C_{\mathrm{V}})^{\mathrm{s}}(H) \geq C(C_{\mathrm{M}}, Z_{\mathrm{m}}(C_{\mathrm{M}}, C_{\mathrm{V}}), C_{\mathrm{V}})^{\mathrm{s}}(H))$$

For any *generic derived history coder* $C(C_{\mathrm{M}}, Z, C_{\mathrm{V}})$ there exists a *minimum space derived history coder* $C(C_{\mathrm{M}}, Z_{\mathrm{m}}(C_{\mathrm{M}}, C_{\mathrm{V}}), C_{\mathrm{V}})$. That is, the search of the *models* can always be performed by a complete traversal of the finite *model* set such that the *space* of the encoding is minimised.

Consider the subset of *histories* for which a *state* is uniquely associated with an *event identifier*. Define the *unit histories*

$$\mathcal{H}_{U,X,\mathrm{U}} = \bigcup\{X \leftrightarrow V^{\mathrm{CS}} : V \subseteq \mathrm{vars}(U)\}$$

These *histories* are such that the *histograms* are *unit*, $\forall H \in \mathcal{H}_{U,X,\mathrm{U}}\ \Diamond A = \mathrm{histogram}(H)\ (A = A^{\mathrm{F}})$. The *size* z equals the *effective volume*, $z = |A^{\mathrm{F}}| \leq v$, where $z = |H|$, $V = \mathrm{vars}(H)$ and $v = |V^{\mathrm{C}}|$.

A *minimum space derived history coder* of the *unit histories* may be implemented with the *binary partition transform coder*, $C_{\mathrm{T,P,b}}$, and the *specialising derived substrate history coder* $C_{\mathrm{G},V,\mathrm{T,H}}$. Define the *binary partition minimum space specialising derived history coder*

$$C_{\mathrm{P,b,m,G,T,H}} = \mathrm{coderHistoryDerivedGeneric}(U, X, C_{\mathrm{T,P,b}}, Z_{\mathrm{m}}(C_{\mathrm{T,P,b}}, C_{\mathrm{G},V,\mathrm{T,H}}), C_{\mathrm{G},V,\mathrm{T,H}}, D_{\mathrm{S}}, D_{\mathrm{X}})$$

The *space* is

$$\begin{aligned} C^{\mathrm{s}}_{\mathrm{P,b,m,G,T,H}}(H) \ = \ & \mathrm{spaceVariables}(U)(|V|) + \\ & \mathrm{spaceIds}(|X|, |H|) + \\ & \mathrm{spacePartitionBinary}(U)(V) \\ & \mathrm{spaceCountsDerived}(U)(A, T) + \\ & \mathrm{spaceClassification}(A * T) + \\ & \mathrm{spaceEventsPartition}(A, T) \end{aligned}$$

where $V = \mathrm{vars}(H)$ and $A = \mathrm{histogram}(H)$. The *minimum space binary partition* is always $\{A^{\mathrm{FS}}, V^{\mathrm{CS}} \setminus A^{\mathrm{FS}}\}$. So the *space* is $C^{\mathrm{s}}_{\mathrm{P,b,m,G,T,H}}(H) = \mathrm{spVar}(U)(|V|) + \mathrm{spIds}(|X|, |H|) + v \ln 2 + \ln(z + 1) + z \ln z$. Compare the *space* to the *canonical history coders*. The difference in *space* for the *index history coder*, C_{H}, is

$$C^{\mathrm{s}}_{\mathrm{P,b,m,G,T,H}}(H) - C^{\mathrm{s}}_{\mathrm{H}}(H) \quad = \quad v \ln 2 + \ln(z+1) + z \ln z - z \ln v$$

The difference is necessarily positive. Similarly, the difference in *space* for the *classification history coder*, C_{G}, is

$$C^{\mathrm{s}}_{\mathrm{P,b,m,G,T,H}}(H) - C^{\mathrm{s}}_{\mathrm{G}}(H) \quad = \quad v \ln 2 + \ln(z+1) + z \ln z - \overline{z} \ln v$$

which is also always positive. That is, the *binary partition minimum space specialising derived history coder*, $C_{\mathrm{P,b,m,G,T,H}}$, always requires more *space* to encode a *unit history* than both of the *canonical coders*, C_{H} and C_{G}, even though the search is the *minimum space* search. This is because the *model space*, $v \ln 2$, varies with the *volume*, as is typical, while the *canonical coder space*, $z \ln v$ and $\overline{z} \ln v$, varies with the *size*.

For fixed *volume* the *coder space*, $C^{\mathrm{s}}_{\mathrm{P,b,m,G,T,H}}(H)$, varies with the *size*, z, which equals the *component cardinality*, $|A^{\mathrm{F}}|$. So the *space* varies against the *component size cardinality relative entropy*.

In order to construct a *minimum space specialising derived history coder* that requires a smaller *space* to encode a *unit history* than the *canonical coders*, a *model* M is needed such that its encoding *space* in *coder* C_{M} is approximately constrained

$$C^{\mathrm{s}}_{\mathrm{M}}(M) < z \ln \frac{v+z}{2z}$$

Note that the discussion above of the process of encoding in *derived history coders* does not address the computational tractability or practicability. Especially in the case of the *minimum space* search, where the entire set of the *model coder* domain is traversed, the computation may be infeasible. The discussion in later sections below considers *search functions* that (i) traverse only subsets of the *model* set, and (ii) have different metrics, not necessarily an encoding *space* valued function of the *models*.

Given a *system* U and *event identifiers* X, a *history coder domain probability function* $P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is defined as *entropic* with respect to *history coder* $C \in \mathrm{coders}(\mathcal{H}_{U,X})$ if the *coder* is an *entropy coder*. See appendix 'Coders and entropy' for the definition of the *entropy coder*. The *coder* is an *entropy history coder* if and only if the *space* of a *history* equals the negative logarithm of the non-zero probability, $\forall H \in \mathcal{H}_{U,X}\ (P_H > 0 \implies C^{\mathrm{s}}(H) = -\ln P_H)$. Then the expected *space* of the *coder* equals the *entropy* of the *history probability function*.

$$\begin{aligned} \mathrm{expected}(P)(C^{\mathrm{s}}) &= \sum_{H \in \mathcal{H}_{U,X}} P_H \times C^{\mathrm{s}}(H) \\ &= -\sum(P_H \ln P_H : H \in \mathcal{H}_{U,X},\ P_H > 0) \\ &= \mathrm{entropy}(P) \end{aligned}$$

An *entropy coder* has the smallest expected *space* of all *coders* given the *probability function*.

Similar to the definition of *entropic history probability functions*, a *history coder domain probability function* $P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is defined as *structured* with respect to *derived history coder* $C(C_{\mathrm{M}}, Z, C_{\mathrm{V}}) \in \mathrm{coders}(\mathcal{H}_{U,X})$, if the expected *space* of the *derived history coder* is less than the expected lesser *space* of the *canonical history coders*, (i) *index history coder*, C_{H}, and (ii) *classification history coder*, C_{G},

$$\mathrm{expected}(P)(C(C_{\mathrm{M}}, Z, C_{\mathrm{V}})^{\mathrm{s}}) < \mathrm{expected}(P)(\mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}, C^{\mathrm{s}}_{\mathrm{G}}))$$

where $\mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}, C^{\mathrm{s}}_{\mathrm{G}}) \in \mathcal{H}_{U,X} \to \ln \mathbf{N}_{>0}$. That is, *history* $H \in \mathcal{H}_{U,X}$ in *structured history probability function* P with respect to *derived history coder* $C(C_{\mathrm{M}}, Z, C_{\mathrm{V}})$ has a *model* $M \in \mathrm{mind}(Z(H))$ for which it is expected that $C^{\mathrm{s}}_{\mathrm{M}}(M) + C_{\mathrm{V}}(M)^{\mathrm{s}}(H) < \mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}(H), C^{\mathrm{s}}_{\mathrm{G}}(H))$, where $\mathrm{minimum}(x, y) = \mathrm{if}(x < y, x, y)$.

The *degree of structure* is defined $\mathrm{structure}(U, X) \in ((\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}) \times \mathrm{coders}(\mathcal{H}_{U,X}) \to \mathbf{Q} \ln \mathbf{Q}_{>0} / \ln \mathbf{Q}_{>0}$ as

$$\mathrm{structure}(U, X)(P, C) := \frac{\mathrm{canonical}(U, X)(P) - \mathrm{expected}(P)(C^{\mathrm{s}})}{\mathrm{canonical}(U, X)(P) - \mathrm{entropy}(P)}$$

where $\mathrm{canonical}(U, X) \in ((\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}) \to \mathbf{Q}_{\geq 0} \ln \mathbf{N}_{>0}$ is defined

$$\mathrm{canonical}(U, X)(P) := \mathrm{expected}(P)(\mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}, C^{\mathrm{s}}_{\mathrm{G}}))$$

The *degree of structure* is undefined if the *canonical coders* are already *entropic*, $\mathrm{canonical}(U, X)(P) = \mathrm{entropy}(P)$. The *degree of structure* is defined for all *history coders*, not just *derived history coders*.

Define the *compression* of *coder* C with respect to *probability function* P

as a synonym for the *degree of structure* of *probability function* P with respect to the *coder* C.

The *degree of structure* is always less than or equal to one,

$$\forall P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}\ (\mathrm{structure}(U,X)(P,C) \leq 1)$$

If the *degree of structure* equals one, $\mathrm{structure}(U,X)(P,C) = 1$, the *coder*, C, is an *entropy coder* of the *probability function*, P, $\mathrm{expected}(P)(C^{\mathrm{s}}) = \mathrm{entropy}(P)$.

If the *degree of structure* less than or equal to zero, $\mathrm{structure}(U,X)(P,C) \leq 0$, the *probability function*, P, is *structureless* with respect to the *coder*, C, or, equivalently, the *coder*, C, is *non-compressing* with respect to the *probability function*, P. For example, the theoretical *variable-width history coder*, C_{E}, is *non-compressing* with respect to all *probability functions* for which it can be defined, because the *space* is always greater than or equal to the *space* of the *classification coder*, $C^{\mathrm{s}}_{\mathrm{E}}(H) \geq C^{\mathrm{s}}_{\mathrm{G}}(H)$.

Structured history probability functions are less strongly constrained than *entropic history probability functions* because *entropy coders* have least expected *space*, $0 < \mathrm{structure}(U,X)(P,C) \leq 1$.

Histories that are *structured* with respect to *derived history coders*,

$$\mathrm{structure}(U,X)(P,C(C_{\mathrm{M}},Z,C_{\mathrm{V}})) > 0$$

are expected to be lawlike in that the structures can be encapsulated in a *model* such that encoding *space* of the *history* plus the additional *space* of the *model* is less than the cost of encoding the *history* in the structureless *canonical coders*. If a *history probability function* is *structured* with respect to some *derived history coder* $C(C_{\mathrm{M}},Z,C_{\mathrm{V}})$, then it is also at least as *structured* with respect to the *minimum space derived history coder* $C(C_{\mathrm{M}},Z_{\mathrm{m}}(C_{\mathrm{M}},C_{\mathrm{V}}),C_{\mathrm{V}})$,

$$\mathrm{structure}(U,X)(P,C(C_{\mathrm{M}},Z_{\mathrm{m}}(C_{\mathrm{M}},C_{\mathrm{V}}),C_{\mathrm{V}})) \geq \mathrm{structure}(U,X)(P,C(C_{\mathrm{M}},Z,C_{\mathrm{V}}))$$

Structured histories are not necessarily assumed to be each encoded with the same *model*, $\exists M \in \mathrm{dom}(Z)\ \forall H \in \mathcal{H}_{U,X}\ (\mathrm{maxd}(Z(H)) = \{M\})$, only that there exists some *model* for some *histories* such that the expected *space* is smaller than the *canonical space*, $\exists H \in \mathcal{H}_{U,X}\ \exists M \in \mathrm{maxd}(Z(H))\ (C^{\mathrm{s}}_{\mathrm{M}}(M) + C_{\mathrm{V}}(M)^{\mathrm{s}}(H) < \mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}(H), C^{\mathrm{s}}_{\mathrm{G}}(H)))$.

There is no *structured probability function* of *unit histories*, $\mathcal{H}_{U,X,\mathrm{U}}$, with respect to the *binary partition minimum space specialising derived history coder*, $C_{\mathrm{P,b,m,G,T,H}}$,

$$\forall P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}\ (\mathrm{structure}(U,X)(P, C_{\mathrm{P,b,m,G,T,H}}) < 0)$$

because the *space* of any non-empty *history* is always greater than in both *canonical coders*

$$\forall H \in \mathcal{H}_{U,X,\mathrm{U}} \setminus \{\emptyset\}\ (C^{\mathrm{s}}_{\mathrm{P,b,m,G,T,H}}(H) > \mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}(H), C^{\mathrm{s}}_{\mathrm{G}}(H)))$$

A *history coder* $C_{\mathrm{min(H,G)}}$ of the lesser *space* of the *canonical history coders* can be implemented with a flag to indicate which of the *canonical coders* was chosen. The *space* is $C^{\mathrm{s}}_{\mathrm{min(H,G)}}(H) = \mathrm{minimum}(C^{\mathrm{s}}_{\mathrm{H}}(H), C^{\mathrm{s}}_{\mathrm{G}}(H)) + \ln 2$. The *lesser canonical history coder*, $C_{\mathrm{min(H,G)}}$, is necessarily *structureless*,

$$\forall P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}\ (\mathrm{structure}(U,X)(P, C_{\mathrm{min(H,G)}}) < 0)$$

because of the additional *space* of the flag.

Conjecture that there is no *coder* such that the uniform *history probability function*, $\hat{\mathcal{H}}_{U,X} = \mathcal{H}_{U,X} \times \{1/|\mathcal{H}_{U,X}|\} \in \mathcal{P}$, has *structure*,

$$\forall C \in \mathrm{coders}(\mathcal{H}_{U,X})\ (\mathrm{structure}(U,X)(\hat{\mathcal{H}}_{U,X}, C) < 0)$$

where $\mathrm{canonical}(U,X)(\hat{\mathcal{H}}_{U,X}) \neq \mathrm{entropy}(\hat{\mathcal{H}}_{U,X})$.

The *degree of structure* has two arguments, (i) the *probability function* $P \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$, and (ii) the *coder* $C \in \mathrm{coders}(\mathcal{H}_{U,X})$. The function can be viewed as (i) a measure of the *structure* of the *histories* of the *probability function*, P, with respect to a fixed *coder*, C, or (ii) a measure of the *compression*, or *canonical-entropic* relative *space*, of the *coder*, C, given a *probability function*, P.

In the first case, *probability functions* may be compared by *structure* given the *partition minimum space specialising derived history coder*,

$$C_{\mathrm{P,m,G,T,H}} = \mathrm{coderHistoryDerivedGeneric}(U, X, C_{\mathrm{T,P}}, Z_{\mathrm{m}}(C_{\mathrm{T,P}}, C_{\mathrm{G},V,\mathrm{T,H}}), C_{\mathrm{G},V,\mathrm{T,H}}, D_{\mathrm{S}}, D_{\mathrm{X}})$$

The *speciality* is the *degree of structure* of *probability function* P with respect to the *partition minimum space specialising derived history coder*. Define $\mathrm{speciality}(U,X) \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P} \to \mathbf{Q} \ln \mathbf{Q}_{>0} / \ln \mathbf{Q}_{>0}$ as

$$\mathrm{speciality}(U,X)(P) := \mathrm{structure}(U,X)(P, C_{\mathrm{P,m,G,T,H}})$$

For an example of comparison by *structure*, let $P_1, P_2 \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$. Then P_1 is more *specially structured* than P_2 if $\mathrm{speciality}(U,X)(P_1) > \mathrm{speciality}(U,X)(P_2)$. Here the given *coder*, $C_{\mathrm{P,m,G,T,H}}$, has a simply defined *model space*, requiring only that the *partition* be encoded. The *minimum search function*, $Z_{\mathrm{m}}(C_{\mathrm{T,P}}, C_{\mathrm{G},V,\mathrm{T,H}})$, does a brute force search over the *partition transforms*, $\mathrm{domain}(C_{\mathrm{T,P}}) \subset \mathcal{T}_{U,\mathrm{P}}$, choosing the *transform* with the least encoding *space* of the *history*, so the *speciality structure* is maximised with respect to *search function*.

Similarly, the *generality* is the *degree of structure* of *probability function* P with respect to the *partition minimum space generalising derived history coder*. Define $\mathrm{generality}(U,X) \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P} \to \mathbf{Q} \ln \mathbf{Q}_{>0} / \ln \mathbf{Q}_{>0}$ as

$$\mathrm{generality}(U,X)(P) := \mathrm{structure}(U,X)(P, C_{\mathrm{P,m,H,T,G}})$$

In the second case, *coders* may be compared by *compression* given the *probability function*, P. For example, let $C_1, C_2 \in \mathrm{coders}(\mathcal{H}_{U,X})$. Then C_1 is more *compressing* than C_2 if $\mathrm{structure}(U,X)(P, C_1) > \mathrm{structure}(U,X)(P, C_2)$.

A *probability function* P can be characterised by the *relative compression* of *coders* that differ only in *model*. The *generic minimum space specialising derived history coder* is parameterised by *model* C_{M},

$$\begin{aligned} &C_{\mathrm{m,G,T,H}}(C_{\mathrm{M}}) = \\ &\quad \mathrm{coderHistoryDerivedGeneric}(U, X, C_{\mathrm{M}}, Z_{\mathrm{m}}(C_{\mathrm{M}}, C_{\mathrm{G},V,\mathrm{T,H}}), C_{\mathrm{G},V,\mathrm{T,H}}, D_{\mathrm{S}}, D_{\mathrm{X}}) \end{aligned}$$

Similarly, define (i) the *generic minimum space specialising fud history coder* $C_{\mathrm{m,G,F,H}}(C_{\mathrm{M}})$, (ii) the *generic minimum space specialising decomposition history coder* $C_{\mathrm{m,G,D,H}}(C_{\mathrm{M}})$ and (iii) the *generic minimum space specialising fud decomposition history coder* $C_{\mathrm{m,G,D,F,H}}(C_{\mathrm{M}})$.

The *relative redundant speciality* of the *probability function* P is the *relative compression* between the *multi-partition transform coder*, $C_{\mathrm{T,P^*}}$, and the *partition transform coder*, $C_{\mathrm{T,P}}$,

$$\mathrm{structure}(U,X)(P, C_{\mathrm{m,G,T,H}}(C_{\mathrm{T,P^*}})) - \mathrm{structure}(U,X)(P, C_{\mathrm{m,G,T,H}}(C_{\mathrm{T,P}}))$$

The *relative layered redundant speciality* of the *probability function* P is the *relative compression* between the *unlimited partition fud coder*, $C_{\mathrm{F,U,P}}$, and the *multi-partition transform coder*, $C_{\mathrm{T,P^*}}$,

$$\mathrm{structure}(U,X)(P, C_{\mathrm{m,G,F,H}}(C_{\mathrm{F,U,P}})) - \mathrm{structure}(U,X)(P, C_{\mathrm{m,G,T,H}}(C_{\mathrm{T,P^*}}))$$

The *relative contingent speciality* of the *probability function* P is the *relative compression* between the *unlimited partition transform decomposition coder*, $C_{\mathrm{D,U,P}}$, and the *partition transform coder*, $C_{\mathrm{T,P}}$,

$$\mathrm{structure}(U,X)(P,C_{\mathrm{m,G,D,H}}(C_{\mathrm{D,U,P}})) - \mathrm{structure}(U,X)(P,C_{\mathrm{m,G,T,H}}(C_{\mathrm{T,P}}))$$

The *relative layered redundant contingent speciality* of the *probability function* P is the *relative compression* between the *unlimited partition fud decomposition coder*, $C_{\mathrm{D,F,U,P}}$, and the *unlimited partition transform decomposition coder*, $C_{\mathrm{D,U,P}}$,

$$\mathrm{structure}(U,X)(P,C_{\mathrm{m,G,D,F,H}}(C_{\mathrm{D,F,U,P}})) - \mathrm{structure}(U,X)(P,C_{\mathrm{m,G,D,H}}(C_{\mathrm{D,U,P}}))$$

Note that, while the *partition transform coder*, $C_{\mathrm{T,P}}$, is straightforwardly implemented by encoding the *partition* T^{P} in *space* $\mathrm{spacePartition}(U)(T^{\mathrm{P}}) := \ln \mathrm{bell}(v)$, where $v = |V^{\mathrm{C}}|$ and $V = \mathrm{und}(T)$, the other *model coders* require the encoding of the structure in lists and trees, and so there are various implementations. This includes the *multi-partition transform coder*, $C_{\mathrm{T,P^*}}$, which is implemented by encoding the *exploded contracted transform*, $\mathrm{explode}(T^{\%}) \in \mathcal{F}_{U,\mathrm{P}}$ in a *partition fud coder* such as the *unlimited partition fud coder*, $C_{\mathrm{F,U,P}}$.

A *probability function* P can also be characterised by the *relative compression* of *coders* that differ only in *derived substrate history coder*. The *generic partition minimum space history coder* is parameterised by *substrate history coder* C_{V},

$$\begin{aligned}&C_{\mathrm{P,m}}(C_{\mathrm{V}}) = \\ &\quad \mathrm{coderHistoryDerivedGeneric}(U,X,C_{\mathrm{T,P}},Z_{\mathrm{m}}(C_{\mathrm{T,P}},C_{\mathrm{V}}),C_{\mathrm{V}},D_{\mathrm{S}},D_{\mathrm{X}})\end{aligned}$$

The *midity* of the *probability function* P is the difference in *compression* between (a) the sum of (i) the *index derived substrate history coder* $C_{\mathrm{H},V,\mathrm{T,H}}$, and (ii) the *classification derived substrate history coder* $C_{\mathrm{G},V,\mathrm{T,G}}$, and (b) the sum of (iii) the *specialising derived substrate history coder* $C_{\mathrm{G},V,\mathrm{T,H}}$, and (iv) the *generalising derived substrate history coder* $C_{\mathrm{H},V,\mathrm{T,G}}$. Define $\mathrm{midity}(U,X) \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P} \to \mathbf{Q} \ln \mathbf{Q}_{>0} / \ln \mathbf{Q}_{>0}$ as

$$\begin{aligned}
&\mathrm{midity}(U,X)(P) := \\
&\qquad \big(\mathrm{structure}(U,X)(P,C_{\mathrm{P,m}}(C_{\mathrm{H},V,\mathrm{T,H}})) + \mathrm{structure}(U,X)(P,C_{\mathrm{P,m}}(C_{\mathrm{G},V,\mathrm{T,G}}))\big) \\
&\quad - \big(\mathrm{structure}(U,X)(P,C_{\mathrm{P,m}}(C_{\mathrm{G},V,\mathrm{T,H}})) + \mathrm{structure}(U,X)(P,C_{\mathrm{P,m}}(C_{\mathrm{H},V,\mathrm{T,G}}))\big) \\
&\quad = \big(\mathrm{structure}(U,X)(P,C_{\mathrm{P,m}}(C_{\mathrm{H},V,\mathrm{T,H}})) + \mathrm{structure}(U,X)(P,C_{\mathrm{P,m}}(C_{\mathrm{G},V,\mathrm{T,G}}))\big) \\
&\quad - \big(\mathrm{speciality}(U,X)(P) + \mathrm{generality}(U,X)(P)\big)
\end{aligned}$$

Consider a *history probability function* which is defined in terms of a *historical distribution*, Q_{h}. The *historical distribution* is defined for *sample histogram* A of *size* z *drawn* from *distribution histogram* E as

$$Q_{\mathrm{h}}(E,z)(A) = \prod_{S \in A^{\mathrm{S}}} \binom{E_S}{A_S} = \prod_{S \in A^{\mathrm{S}}} \frac{E_S!}{A_S!\ (E_S - A_S)!} \in \mathbf{N}_{>0}$$

The *historical probability distribution* is normalised,

$$\hat{Q}_{\mathrm{h}}(E,z)(A) = Q_{\mathrm{h}}(E,z)(A)/\tbinom{z_E}{z}$$

where $z_E = \mathrm{size}(E)$.

Let $H_E \subseteq \mathcal{H}_{U,X}$ be a *distribution history* and E be its *distribution histogram*, $E = \mathrm{histogram}(H_E)$. The *distribution history*, H_E, has *substrate* V_E equal to the *system variables*, $V_E = \mathrm{vars}(E) = \mathrm{vars}(U)$. Its *volume* is $v_E = |V_E^{\mathrm{C}}|$. Its domain is the entire set of *event identifiers*, $\mathrm{ids}(H_E) = X$, so that the *distribution history* is a left total function, $H_E \in X :\to V_E^{\mathrm{CS}}$, and its *size* z_E equals the cardinality of the *event identifiers*, $z_E = \mathrm{size}(E) = |X|$. The *historically distributed history probability function* $P_{U,X,H_E} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is defined

$$\begin{aligned} P_{U,X,H_E} \ :=\ & \{(H, 1/(z_E 2^{v_E} \tbinom{z_E}{z_H}))) : H \in \mathcal{H}_{U,X},\ H \subseteq H_E \% V_H,\ H \neq \emptyset\} \cup \\ & \{(H,0) : H \in \mathcal{H}_{U,X},\ H \not\subseteq H_E \% V_H\} \cup \{(\emptyset, 0)\} \end{aligned}$$

where $V_H = \mathrm{vars}(H)$ and $z_H = |H|$. The *historically distributed history probability*, $P_{U,X,H_E}(H)$, is independent of the *event identifiers*, $\mathrm{ids}(H) \subseteq X$, because the *probability* depends only on the *history size*, z_H. The *historically distributed history probability function*, P_{U,X,H_E}, is uniform with respect to the *substrate* subset of the *system variables*, $V_H \subseteq V_E$, and the *draw size*, $z_H \leq z_E$. That is,

$$\begin{aligned} P_{U,X,H_E} \ =\ & \{(H, 1/\tbinom{z_E}{z}) : V \subseteq V_E,\ z \in \{1 \ldots z_E\},\ H \subseteq H_E \% V,\ |H| = z\}^{\wedge} \cup \\ & \{(H,0) : H \in \mathcal{H}_{U,X},\ H \not\subseteq H_E \% V_H\} \cup \{(\emptyset, 0)\} \end{aligned}$$

where $\hat{Y} = (Y)^{\wedge} = \mathrm{normalise}(Y)$. Or

$$\begin{aligned} & \forall V \subseteq V_E\ \forall z \in \{1 \ldots z_E\} \\ & \quad (\sum (P_{U,X,H_E}(H) : H \in \mathcal{H}_{U,X},\ V_H = V,\ |H| = z) = 1/(z_E 2^{v_E})) \end{aligned}$$

The *historically distributed history probability function*, P_{U,X,H_E}, is the *probability function* of the first *drawn history* $H \subseteq H_E \% V_H$ of arbitrary *variables*

$V_H \subseteq V_E$ and *size* $z_H \in \{1 \dots z_E\}$ from *distribution history* $H_E \subseteq \mathcal{H}_{U,X}$.

Now for arbitrary *drawn history* $H \subseteq H_E\%V_H$, the *historical probability* of *drawing without replacement* its *histogram* $A_H = \text{histogram}(H)$ from the *distribution histogram*, $E = \text{histogram}(H_E)$, is the expected *historically distributed history probability* of the *histogram*, A_H, times the normalising factor,

$$\hat{Q}_{\text{h}}(E\%V_H, z_H)(A_H) = z_E 2^{v_E} \sum(P_{U,X,H_E}(G) : G \in \mathcal{H}_{U,X},\ A_G = A_H)$$

The set of *sized cardinal substrate histograms* $\mathcal{A}_z$, defined above in section 'Distinct geometry sized cardinal substrate histograms', is the set of *complete integral cardinal substrate histograms* of *size* z and *dimension* less than or equal to the *size* such that the *independent* is *completely effective*

$$\mathcal{A}_z = \{A : A \in \mathcal{A}_{\text{c}} \cap \mathcal{A}_{\text{i}},\ \text{size}(A) = z,\ |V_A| \leq z,\ A^{\text{U}} = A^{\text{XF}} = A^{\text{C}}\}$$

Each *substrate histogram* $A \in \mathcal{A}_z$ has $|V_A|! \prod_{w \in V_A} |U_A(w)|!$ *cardinal substrate permutations*. These *frame* mappings partition the *substrate histograms* into equivalence classes having the same *geometry*. Let P_z be the partition, $P_z \in \text{B}(\mathcal{A}_z)$, such that the components of P_z are the equivalence classes by *cardinal substrate permutation*, $\forall C \in P_z\ \forall A \in C\ (|C| = |V_A|! \prod_{w \in V_A} |U_A(w)|)$.

Each of the *substrate histograms* in a component of P_z, that are equivalent by *cardinal substrate permutation*, have the same *entropy*, $\forall C \in P_z\ \forall A, B \in C\ (\text{entropy}(A) = \text{entropy}(B))$.

The expected function of the *renormalised geometry-weighted probability function*, $\hat{R}_z$, that operates on real-valued functions of the *sized cardinal substrate histograms*, $\mathcal{A}_z \to \mathbf{R}$, is defined $\text{ex}(z)(F) := \text{expected}(\hat{R}_z)(F)$ where $\hat{R}_z = \text{normalise}(\{(A, 1/(|V_A|! \prod_{w \in V_A} |U_A(w)|!)) : A \in \text{dom}(F)\})$.

The set of *non-trivially ideal sized cardinal substrate histograms* $\mathcal{A}_{z,\dagger}$ is a subset defined

$$\begin{aligned}\mathcal{A}_{z,\dagger} = \{A : A \in \mathcal{A}_z,\ z > |V_A^{\text{C}}|,& \\ \exists T \in \mathcal{T}_{U_A,V_A}\ (T^{\text{P}} \notin \{V_A^{\text{CS}\{\}}, \{V_A^{\text{CS}}\}\} \wedge A = A * T * T^{\dagger A})\}&\end{aligned}$$

where $V_A = \text{vars}(A)$ and $U_A = \text{implied}(A)$. Each *non-trivially ideal histogram* $A \in \mathcal{A}_{z,\dagger}$ has an *ideal transform* such that the *transform's partition* is neither (i) the *self partition*, $V_A^{\text{CS}\{\}}$, nor (ii) the *unary partition*, $\{V_A^{\text{CS}}\}$. Thus (i) the *partition* cannot be a *reframe*, and (ii) the *histogram* is not necessarily *independent*. The *size* is constrained to be greater than the *volume*, $z > v$, so at least one *component* of the *partition* has a *size* greater than one, $\exists(\cdot, C) \in T^{-1}\ (\text{size}(A * C) > 1)$. Therefore a *trimmed non-trivially ideal histogram* cannot be the *histogram* of a *unit history* $H \notin \mathcal{H}_{U,X,\text{U}}$ where $A = \text{histogram}(H) \in \text{trim}(\mathcal{A}_{z,\dagger})$.

The *historically distributed history probability function* $P_{U,X,H_E} \in (\mathcal{H}_{U,X} :\to \mathbf{Q}_{\geq 0}) \cap \mathcal{P}$ is a function only of the *distribution histogram* E if (i) the *system* is the *implied system*, $U = \text{implied}(E)$, (ii) the *distribution history* is constructed from the *distribution histogram*, $H_E = \text{history}(E)$, and (iii) the set of *event identifiers* is the domain of the constructed *history*, $X = \text{ids}(H_E) \in \mathcal{S} \times \mathbf{N}_{>0}$. So the *degree of structure* of the *historically non-trivially ideal distributed history probability function* with respect to the *partition minimum space specialising derived history coder* is the *speciality*, $\text{speciality}(U, X)(P_{U,X,H_E})$ where $E \in \text{trim}(\mathcal{A}_{z_E,\dagger})$.

Conjecture that the expected *geometry-permutation-weighted speciality* varies with *distribution history size*

$$\text{ex}(z_E)(\{(E', \text{speciality}(U_E, X_E)(P_{U,X,H_E})) : E' \in \mathcal{A}_{z_E,\dagger},\ E = \text{trim}(E')\}) \sim z_E$$

where $U_E = \text{implied}(E)$ and $X_E = \text{ids}(\text{history}(E))$.

That is, conjecture that as the *size* of the *non-trivially ideal distribution histogram*, E, increases, the corresponding *historically distributed history probability function*, P_{U,X,H_E}, tends to have increasing *degree of structure* with respect to the *specialising coder*, $C_{\text{P,m,G,T,H}}$.

A.13 Computers

The set of *computers*, computers, is a type class that formalises computation *time* and representation *space*. Define the application of a *computer*, $\text{apply} \in \text{computers} \to (\mathcal{X} \to \mathcal{Y})$, where $\mathcal{X}$ and $\mathcal{Y}$ are universal sets. Define the domain of the application of a *computer*, $\text{domain} \in \text{computers} \to \text{P}(\mathcal{X})$, and the range of the application of a *computer*, $\text{range} \in \text{computers} \to \text{P}(\mathcal{Y})$, such that

$$\forall I \in \text{computers}\ (\text{apply}(I) \in \text{domain}(I) \to \text{range}(I))$$

and

$$\forall I \in \text{computers}\ (\text{dom}(\text{apply}(I)) = \text{domain}(I))$$

Define the shorthand $I^* := \text{apply}(I)$. The definition here of the application is left total, but not necessarily right total, so the application of a subsequent *computer*, $J^*(I^*(x))$, requires the domain of J to be a superset of the range of I, $\text{domain}(J) \supseteq \text{range}(I)$. In the cases where the range of I is a subset of its own domain, $\text{range}(I) \subseteq \text{domain}(I)$, the *computer* I may be applied to itself recursively, $I^* \circ I^*$, $I^* \circ I^* \circ I^*$, and so on.

An *inverse computer* $J \in \text{computer}$ of a *computer* I is such that $\text{domain}(J) = \text{range}(I)$, $\text{range}(J) = \text{domain}(I)$ and $\forall x \in \text{domain}(I)\ (J^*(I^*(x)) = x)$. Note that if an *inverse computer* J exists, the *computer* I must be right total.

The computation or application *time* is defined as $\text{time} \in \text{computers} \to (\mathcal{X} \to \mathbf{N}_{>0})$. Define the shorthand $I^{\text{t}} := \text{time}(I)$. If the *time* of some given argument x is finite, $I^{\text{t}}(x) < \infty$, then x is *computable*.

Note that the range is available to the compute *time* calculation via apply. If it is necessary to perform the application in order to calculate the computation *time* then the computation is non-deterministic and the *time* of the *time* calculation itself must be greater than the application *time*. For example, let $I, J \in \text{computers}$, $\text{domain}(J) = \{I\} \times \text{domain}(I)$, $\text{range}(J) = \mathbf{N}_{>0}$ and $\forall x \in \text{domain}(I)\ (\text{apply}(J)((I, x)) = \text{time}(I)(x))$. Let I and J be such that $\text{time}(J)$ depends on $\text{apply}(I)$, then $\forall x \in \text{domain}(I)\ (\text{time}(J)((I, x)) > \text{time}(I)(x))$.

The representation *space* is defined as $\text{space} \in \text{computers} \to (\mathcal{X} \to \ln \mathbf{N}_{>0})$. Define the shorthand $I^{\text{s}} := \text{space}(I)$. If the *space* of some given argument x is finite, $I^{\text{s}}(x) < \infty$, then x is *representable*.

A *computer* is *tractable* if both the *time* complexity and the *space* complexity is no worse than polynomial with respect to any underlying variable.

A possible definition of a *computer* could be in terms of a Turing Machine. For example, the tuple $(\text{in}, \text{out}, \text{step}, \text{term})$ of (i) an input function, $\text{in} \in \mathcal{X} \to \mathbf{N}$, (ii) an output function, $\text{out} \in \mathbf{N} \to \mathcal{X}$, (iii) a transition function, $\text{step} \in \mathbf{N} \to \mathbf{N}$, and (iv) a set of terminating states, $\text{term} \subset \mathbf{N}$. Let $L \in \mathcal{L}(\mathbf{N}) \setminus \{\emptyset\}$ be the list of the states such that $L_1 = \text{in}(x)$, $\forall i \in \{1 \ldots t-1\}\ (L_i \notin \text{term} \wedge L_{i+1} = \text{step}(L_i))$ and $L_t \in \text{term}$ where $t = |L|$. Then $I^*(x) = \text{out}(L_t)$ and the *time* to compute x is the number of states in the sequence, $I^{\text{t}}(x) = t$. The *time* cannot be zero, because the conversion from and to the representation, $\mathbf{N}$, is treated as a step. The *space* to represent x is $\ln(\text{maxr}(L) + 1)$. The *space* is zero if $L = \{(1, 0)\}$.

Define $I_+ = \text{adder} \in \text{computers}$ such that $\text{domain}(I_+) = \mathbf{Q} \times \mathbf{Q}$ and $\text{range}(I_+) = \mathbf{Q}$, such that

$$\forall (a, b) \in \text{domain}(I_+)\ (\text{apply}(I_+)((a, b)) = a + b)$$

The *adder* is also constrained such that the self addition of zero has least *time*, $\forall q_1, q_2 \in \text{domain}(I_+)\ (I_+^{\text{t}}((q_1, q_2)) \geq I_+^{\text{t}}((0, 0))$.

Define $I_\times = \text{multiplier} \in \text{computers}$ such that $\text{domain}(I_\times) = \mathbf{Q} \times \mathbf{Q}$ and $\text{range}(I_\times) = \mathbf{Q}$, such that

$$\forall (a, b) \in \text{domain}(I_\times)\ (\text{apply}(I_\times)((a, b)) = ab)$$

The *multiplier* is also constrained such that the self multiplication of one has least *time*, $\forall q_1, q_2 \in \text{domain}(I_\times)\ (I_\times^{\text{t}}((q_1, q_2)) \geq I_\times^{\text{t}}((1, 1))$.

Define $I_0 = \text{resetter} \in \text{computers}$ such that $\text{domain}(I_0) = \text{range}(I_0) = \mathbf{Q}$, such that

$$\forall a \in \text{domain}(I_0)\ (\text{apply}(I_0)(a) = 0)$$

Define $I_{\text{L,s}} = \text{listSetter}(X) \in \text{computers}$ such that $\text{domain}(I_{\text{L,s}}) = \{(L, (i, x)) : L \in \mathcal{L}(X),\ i \in \{1 \ldots |L|\},\ x \in X\}$, $\text{range}(I_{\text{L,s}}) = \mathcal{L}(X)$, such that

$$\forall (L, (i, x)) \in \text{domain}(I_{\text{L,s}})\ (\text{apply}(I_{\text{L,s}})((L, (i, x))) = L \setminus \{(i, L_i)\} \cup \{(i, x)\})$$

Define $I_{\text{L,g}} = \text{listGetter}(X) \in \text{computers}$ such that $\text{domain}(I_{\text{L,g}}) = \{(L, i) : L \in \mathcal{L}(X),\ i \in \{1 \ldots |L|\}\}$, $\text{range}(I_{\text{L,g}}) = X$, such that

$$\forall (L, i) \in \text{domain}(I_{\text{L,g}})\ (\text{apply}(I_{\text{L,g}})((L, i)) = L_i)$$

The *time* complexity of the list get and set operations is constant

$$\exists m \in \mathbf{N}_{>0}\ (I_{\text{L,s}}^{\text{t}} \in \text{O}(\text{domain}(I_{\text{L,s}}) \times \{1\}, m))$$

and

$$\exists m \in \mathbf{N}_{>0}\ (I_{\text{L,g}}^{\text{t}} \in \text{O}(\text{domain}(I_{\text{L,g}}) \times \{1\}, m))$$

The *space* complexity of the list get and set operations varies as the length of the list, assuming the *space* of the elements of the list is constant

$$\exists m \in \mathbf{N}_{>0}\ (I_{\text{L,s}}^{\text{s}} \in \text{O}(\{((L, (i, x)), n) : (L, (i, x)) \in \text{domain}(I_{\text{L,s}}),\ n = |L|\}, m))$$

and

$$\exists m \in \mathbf{N}_{>0}\ (I_{\text{L,g}}^{\text{s}} \in \text{O}((\{((L, i), n)) : (L, i) \in \text{domain}(I_{\text{L,g}}),\ n = |L|\}, m))$$

Define $I_{\mathrm{B,s}} = \mathrm{mapBinarySetter}(X) \in \mathrm{computers}$ such that $\mathrm{domain}(I_{\mathrm{B,s}}) = \{(B,(i,x)) : B \in \mathcal{B}(X),\ i \in \mathrm{domain}(B),\ x \in X\}$, $\mathrm{range}(I_{\mathrm{B,s}}) = \mathcal{B}(X)$, such that

$$\forall (B,(i,x)) \in \mathrm{domain}(I_{\mathrm{B,s}})\ ((i,x) \in \mathrm{function}(\mathrm{apply}(I_{\mathrm{B,s}})((B,(i,x)))))$$

Define $I_{\mathrm{B,g}} = \mathrm{mapBinaryGetter}(X) \in \mathrm{computers}$ such that $\mathrm{domain}(I_{\mathrm{B,g}}) = \{(B,i) : B \in \mathcal{B}(X),\ i \in \mathrm{domain}(B)\}$, $\mathrm{range}(I_{\mathrm{B,g}}) = X$, such that

$$\forall (B,i) \in \mathrm{domain}(I_{\mathrm{B,g}})\ (\mathrm{apply}(I_{\mathrm{B,g}})((B,i)) = \mathrm{find}(B,i))$$

The *time* complexity of the *binary map accessors* is that of the find operation which is $\ln n$ where $n = |\mathrm{function}(B)|$

$$\exists m \in \mathbf{N}_{>0}$$
$$(I^{\mathrm{t}}_{\mathrm{B,s}} \in \mathrm{O}(\{((B,(i,x)), \ln n) : (B,(i,x)) \in \mathrm{domain}(I_{\mathrm{B,s}}),\ n = |\mathrm{f}(B)|\}, m))$$

and

$$\exists m \in \mathbf{N}_{>0}$$
$$(I^{\mathrm{t}}_{\mathrm{B,g}} \in \mathrm{O}(\{((B,i), \ln n) : (B,i) \in \mathrm{domain}(I_{\mathrm{B,g}}),\ n = |\mathrm{f}(B)|\}, m))$$

where $\mathrm{f} = \mathrm{function}$.

The *space* complexity of the *binary map accessors* is $n \ln n$ where $n = |\mathrm{function}(B)|$

$$\exists m \in \mathbf{N}_{>0}$$
$$(I^{\mathrm{t}}_{\mathrm{B,s}} \in \mathrm{O}(\{((B,(i,x)), n \ln n) : (B,(i,x)) \in \mathrm{domain}(I_{\mathrm{B,s}}),\ n = |\mathrm{f}(B)|\}, m))$$

and

$$\exists m \in \mathbf{N}_{>0}$$
$$(I^{\mathrm{t}}_{\mathrm{B,g}} \in \mathrm{O}(\{((B,i), n \ln n) : (B,i) \in \mathrm{domain}(I_{\mathrm{B,g}}),\ n = |\mathrm{f}(B)|\}, m))$$

Set operations on a domain $X \subset \mathcal{X}$ can be implemented in a binary map given some *coder* $C \in \mathrm{coders}$ such that $\mathrm{domain}(C) = X$. For example, define $I_{\mathrm{B,S,i}} = \mathrm{setBinaryInserter}(C) \in \mathrm{computers}$ such that $\mathrm{range}(I_{\mathrm{B,S,i}}) = \{B : B \in \mathcal{B}(X),\ \mathrm{flip}(\mathrm{function}(B)) \subseteq E\}$ and $\mathrm{domain}(I_{\mathrm{B,S,i}}) = \mathrm{range}(I_{\mathrm{B,S,i}}) \times X$ where $(E,\cdot,\cdot) = \mathrm{def}(C)$

$$\forall (B,x) \in \mathrm{domain}(I_{\mathrm{B,S,i}})\ ((E_x, x) \in \mathrm{function}(\mathrm{apply}(I_{\mathrm{B,S,i}})((B,x))))$$

Again, the *time* complexity is that of the find function, $\ln n$, assuming that the encode opertion has constant *time* complexity. The *space* complexity is also that of the *binary map accessors*, $n \ln n$.

Given a *coder* $C \in \mathrm{coders}$ such that $\mathrm{domain}(C) = X$ nested in *coder* $C' \in \mathrm{coders}$ having a domain which is the powerset, $\mathrm{domain}(C') = \mathrm{P}(X)$, the *powersetter* can be defined $I_{\mathrm{B,S,P}} = \mathrm{setBinaryPowersetter}(C', C) \in \mathrm{computers}$ such that $\mathrm{domain}(I_{\mathrm{B,S,P}}) = \{B : B \in \mathcal{B}(X),\ \mathrm{flip}(\mathrm{function}(B)) \subseteq E\}$ and $\mathrm{range}(I_{\mathrm{B,S,P}}) = \{B : B \in \mathcal{B}(X),\ \mathrm{flip}(\mathrm{function}(B)) \subseteq E'\}$ where $(E, \cdot, \cdot) = \mathrm{def}(C)$ and $(E', \cdot, \cdot) = \mathrm{def}(C')$

$$\begin{aligned}&\forall B \in \mathrm{domain}(I_{\mathrm{B,S,P}})\\ &\qquad (\bigcup\{\mathrm{range}(Q) : Q \in \mathrm{range}(\mathrm{apply}(I_{\mathrm{B,S,P}})(B))\} = \mathrm{P}(\mathrm{range}(B)))\end{aligned}$$

In other words, the application $I^*_{\mathrm{B,S,P}}(B)$ on a binary set B of set $Y \subseteq X$, $\mathrm{range}(B) = Y$, is equivalent to computing $\mathrm{P}(Y)$. Implementation of the *powersetter* can be done by folding over the given binary set with the current accumulated powerset, creating a singleton and adding the current element to each of the sets in the current powerset. The list of cardinalities of the powersets is such that $L_{i+1} = 2L_i + 1$ where $L \in \mathcal{L}(\mathbf{N})$. Thus $\mathrm{sum}(L) < n2^n$ where $n = |L| = |Y|$. The *time* complexity of the *powersetter* is $n2^n \ln n$.

Similarly, given $C'' \in \mathrm{coders}$ having a domain which is the powerset of the powerset, $\mathrm{domain}(C'') = \mathrm{P}(\mathrm{P}(X))$, the *partitioner* can be defined $I_{\mathrm{B,S,B}} = \mathrm{setBinaryPartitioner}(C'', C', C) \in \mathrm{computers}$ which is such that the application $I^*_{\mathrm{B,S,B}}(B)$ on a binary set B of set $Y \subseteq X$, $\mathrm{range}(B) = Y$, is equivalent to computing $\mathrm{B}(Y)$. Implementation of the *partitioner* can be done by folding over the given binary set with the current accumulated set of partitions. The list of cardinalities of the sets of partitions is such that $L_{i+1} = (i+1)L_i$ where $L \in \mathcal{L}(\mathbf{N})$. Thus $\mathrm{sum}(L) < nn!$ where $n = |L| = |Y|$. The *time* complexity of the *partitioner* is $nn! \ln n$.

Contrast the computation of the *natural number binary map setter* to the *poset binary map setter*. Define $I_{\mathrm{B,P,s}} = \mathrm{mapBinaryPosetSetter}(I_{\pm}, Y, X) \in \mathrm{computers}$ such that

$$\mathrm{domain}(I_{\mathrm{B,P,s}}) = \{(B, (y, x)) : B \in \mathrm{mapBinaryPosets}(Y, X),\ y \in Y,\ x \in X\}$$

and $\mathrm{range}(I_{\mathrm{B,P,s}}) = \mathrm{mapBinaryPosets}(Y, X)$, such that

$$\forall (B, (y, x)) \in \mathrm{domain}(I_{\mathrm{B,P,s}})\ ((y, x) \in \mathrm{function}(\mathrm{apply}(I_{\mathrm{B,P,s}})((B, (y, x)))))$$

The find computation of the *poset binary map setter* is implemented with a nested *computer*, *comparator* $I_{\pm} \in$ computers, which is such that domain$(I_{\pm}) = Y \times Y$, range$(I_{\pm}) = \{-1, 0, 1\}$, and $\forall y_1, y_2 \in Y\ ((y_1 < y_2 \implies I^*_{\pm}((y_1, y_2)) = -1) \wedge (y_1 = y_2 \implies I^*_{\pm}((y_1, y_2)) = 0) \wedge (y_1 > y_2 \implies I^*_{\pm}((y_1, y_2)) = 1))$.

Define $I_{\mathrm{B,P,g}} = \mathrm{mapBinaryPosetGetter}(I_{\pm}, Y, X) \in$ computers such that domain$(I_{\mathrm{B,P,g}}) = \{(B, y) : B \in \mathrm{mapBinaryPosets}(Y, X),\ y \in \mathrm{domain}(B)\}$, range$(I_{\mathrm{B,P,g}}) = X$, such that

$$\forall (B, y) \in \mathrm{domain}(I_{\mathrm{B,P,g}})\ (\mathrm{apply}(I_{\mathrm{B,P,g}})((B, y)) = \mathrm{find}(B, y))$$

The *time* of the *poset binary map getter* is at least that of the *time* of the self comparison in the *comparator* of the given index, $I^{\mathrm{t}}_{\mathrm{B,P,g}}((B, y)) > I^{\mathrm{t}}_{\pm}((y, y))$. If it is the case that self comparison has the greatest *time*, $\forall (y_1, y_2) \in Y\ (I^{\mathrm{t}}_{\pm}((y_1, y_1)) \geq I^{\mathrm{t}}_{\pm}((y_1, y_2)))$, then the *time* complexity of the *poset binary map accessors* is that of the find operation, which is $\ln n$ where $n = |\mathrm{function}(B)|$, times the self comparison *time* of the *comparator*

$$\exists m \in \mathbf{N}_{>0}\ (I^{\mathrm{t}}_{\mathrm{B,P,s}} \in \mathrm{O}(\{((B, (y, x)), I^{\mathrm{t}}_{\pm}((y, y)) \times \ln n) : \\ (B, (y, x)) \in \mathrm{domain}(I_{\mathrm{B,P,s}}),\ n = |\mathrm{function}(B)|\}, m))$$

and

$$\exists m \in \mathbf{N}_{>0}\ (I^{\mathrm{t}}_{\mathrm{B,P,g}} \in \mathrm{O}(\{((B, y), I^{\mathrm{t}}_{\pm}((y, y)) \times \ln n) : \\ (B, y) \in \mathrm{domain}(I_{\mathrm{B,P,g}}),\ n = |\mathrm{function}(B)|\}, m))$$

A.14 Search and optimisation

Searchers, searchers$(\mathcal{X})$, encapsulate partial or complete traversal of some search set by neighbourhood. There are two sub-types of *searcher*, (i) *tree searchers*

$$\mathrm{searchTreers}(\mathcal{X}) \subset \mathrm{searchers}(\mathcal{X})$$

and (ii) *list searchers*

$$\mathrm{searchListers}(\mathcal{X}) \subset \mathrm{searchers}(\mathcal{X})$$

Both types have (i) some set $X \subset \mathcal{X}$ to search, (ii) an initial subset of the search set $R \subseteq X$, and (iii) a total neighbourhood function on the search set. *Tree searchers* have a total neighbourhood function having a domain equal to the search set and a range of subsets of the search set, $N \in X :\to \mathrm{P}(X)$ where $\mathrm{dom}(N) = X$. *List searchers* have a neighbourhood function having a domain equal to the powerset of the search set and a range of subsets of the search set, $P \in \mathrm{P}(X) :\to \mathrm{P}(X)$ where $\mathrm{dom}(P) = \mathrm{P}(X)$. Define the constructor of a *tree searcher* as

$$\begin{aligned}&\mathrm{searchTreer} \in \\ &\quad \mathrm{P}(\mathcal{X}) \times (\mathcal{X} \to \mathrm{P}(\mathcal{X})) \times \mathrm{P}(\mathcal{X}) \to \mathrm{searchTreers}(\mathcal{X})\end{aligned}$$

such that $\mathrm{searchTreer}(X, N, R) \in \mathrm{searchTreers}(X) \subset \mathrm{searchers}(X)$.

Define the constructor of a *list searcher* as

$$\begin{aligned}&\mathrm{searchLister} \in \\ &\quad \mathrm{P}(\mathcal{X}) \times (\mathrm{P}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})) \times \mathrm{P}(\mathcal{X}) \to \mathrm{searchListers}(\mathcal{X})\end{aligned}$$

such that $\mathrm{searchLister}(X, P, R) \in \mathrm{searchListers}(X) \subset \mathrm{searchers}(X)$.

Given a *tree searcher* $Z = \mathrm{searchTreer}(X, N, R)$ the function tree returns a tree of elements of the search set, $\mathrm{tree}(Z) \in \mathrm{trees}(X)$, of all possible non-circular partial or complete traverses of X from R via successive neighbouring elements of X. Define

$$\mathrm{tree} \in \mathrm{searchTreers}(\mathcal{X}) \to \mathrm{trees}(\mathcal{X})$$

as $\mathrm{tree}(Z) := \mathrm{ts}(R, \emptyset)$ where $\mathrm{searchTreer}(X, N, R) = Z$ and $\mathrm{ts} \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{X}) \to \mathrm{trees}(\mathcal{X})$ is defined

$$\mathrm{ts}(Y, J) := \{(y, \mathrm{ts}(N(y) \setminus (J \cup \{y\}), J \cup \{y\})) : y \in Y\}$$

Define $\mathrm{elements} \in \mathrm{searchTreers}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})$ as

$$\mathrm{elements}(Z) := \mathrm{elements}(\mathrm{tree}(Z))$$

The elements of the search tree $T = \mathrm{tree}(Z)$ form a subset of the search set, $\mathrm{elements}(Z) = \mathrm{elements}(T) \subseteq X$. If the elements equals the search set, $\mathrm{elements}(Z) = X$, then the search is said to be a complete traversal. Otherwise the search is a partial traversal, $\mathrm{elements}(Z) \neq X$. The roots of the search tree equal the initial subset, $\mathrm{roots}(T) = R$. If the search set, X, is finite then the neighbourhood function, N, and the initial subset, R, are finite, and hence the search tree is finite, $|\mathrm{leaves}(T)| < \infty$, $\mathrm{depth}(T) < \infty$ and $|\mathrm{elements}(T)| < \infty$.

The paths of the trees exclude circularities and thus contain each element no more than once, $\forall L \in \mathrm{paths}(T)\ (|\mathrm{set}(L)| = |L|)$. The first element of each path is in the subset R, $\forall L \in \mathrm{paths}(T)\ (L_1 \in R)$. The successive element of each element in a path must be in the neighbourhood of the element excluding the elements of the path up to that point, $\forall L \in \mathrm{paths}(T)\ \forall i \in \{1 \ldots |L|-1\}\ (L_{i+1} \in N(L_i) \setminus \mathrm{set}(L_{\{1 \ldots i\}}))$.

In the case where the tree neighbourhood function returns X for all elements and the initial set is X, $Z = \mathrm{searchTreer}(X, X \times \{X\}, X)$, then the paths of the search form the set of all permutations, $|\mathrm{paths}(\mathrm{tree}(Z))| = |X|!$. In the case where the neighbourhood function returns the empty set, $Z = \mathrm{searchTreer}(X, X \times \{\emptyset\}, X)$, then the search tree nodes equals the roots, $\mathrm{tree}(Z) = X \times \{\emptyset\}$, and the cardinality of the paths equals that of the initial set, $|\mathrm{paths}(\mathrm{tree}(Z))| = |X|$.

An example of a tree neighbourhood function returns decremented cardinality parent partitions given a partition of some set J, $X = \mathrm{B}(J)$, and $N = \{(Q, \{P : P \in X,\ \mathrm{parent}(P,Q),\ |P| = |Q|-1\}) : Q \in X\} \in X :\to \mathrm{P}(X)$.

Given a *list searcher* $Z = \mathrm{searchLister}(X, P, R)$ the function list returns a list of subsets of the search set, $\mathrm{list}(Z) \in \mathcal{L}(\mathrm{P}(X))$. The list, $\mathrm{list}(Z)$, is a non-circular partial or complete traverse of X from R via successive neighbouring subsets of X. Define

$$\mathrm{list} \in \mathrm{searchListers}(\mathcal{X}) \to \mathcal{L}(\mathrm{P}(X))$$

as $\mathrm{list}(Z) := \mathrm{list}(\mathrm{ls}(R, \emptyset))$ where $\mathrm{searchLister}(X, P, R) = Z$ and $\mathrm{ls} \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{X}) \to \mathcal{K}(\mathrm{P}(\mathcal{X}))$ is defined

$$\begin{aligned} \mathrm{ls}(Y, K) &:= (Y, \mathrm{ls}(P(Y) \setminus (K \cup Y), K \cup Y)) \\ \mathrm{ls}(\emptyset, \cdot) &:= \emptyset \end{aligned}$$

Define $\mathrm{elements} \in \mathrm{searchListers}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})$ as

$$\mathrm{elements}(Z) := \bigcup \mathrm{set}(\mathrm{list}(Z))$$

The elements of the search list $L = \mathrm{list}(Z)$ form a subset of the search set, $\mathrm{elements}(Z) = \bigcup \mathrm{set}(L) \subseteq X$. If the elements equals the search set, $\mathrm{elements}(Z) = X$, then the search is said to be a complete traversal. Otherwise the search is a partial traversal, $\mathrm{elements}(Z) \neq X$. The first element of the list is the initial subset, $L_1 = R$. If the search set, X, is finite then the neighbourhood function, P, and the initial subset, R, are finite, and hence the search list is finite, $|L| < \infty$ and $|\mathrm{set}(L)| < \infty$.

The list excludes circularities and thus contains each element no more than once, $\forall i \in \{1 \ldots |L|-1\}\ (\bigcup \mathrm{set}(L_{\{1\ldots i\}}) \cap L_{i+1} = \emptyset)$.

In the case where the initial set is X, $Z = \mathrm{searchLister}(X, P, X)$, then the list is a singleton, $\mathrm{list}(Z) = \{(1, X)\}$, whatever the neighbourhood function, P. In the case where the list neighbourhood function returns X for all subsets of the search set, $Z = \mathrm{searchLister}(X, \mathrm{P}(X) \times \{X\}, R)$, and the initial set is a proper subset of the search set, $R \neq X$, then the list has a length of two, $\mathrm{list}(Z) = \{(1, R), (2, X \setminus R)\}$. In the case where the neighbourhood function returns the empty set, $Z = \mathrm{searchLister}(X, \mathrm{P}(X) \times \{\emptyset\}, R)$, then the list is a singleton, $\mathrm{list}(Z) = \{(1, R)\}$.

The set of *optimisers* $\mathrm{optimisers}(\mathcal{X})$ is a subset of *searchers*,

$$\mathrm{optimisers}(\mathcal{X}) \subset \mathrm{searchers}(\mathcal{X})$$

which constrain the application of the neighbourhood function by post-applying an inclusion function. On a search set X the inclusion function $I \in \mathrm{P}(X) :\to \mathrm{P}(X)$ returns a subset of the argument, $\forall Y \subseteq X\ (I(Y) \subseteq Y)$. Given an extensive tree neighbourhood function $N \in X :\to \mathrm{P}(X)$ the application of the inclusion function results in a new neighbourhood function $M = \{(x, I(N(x))) : x \in X\} \in X :\to \mathrm{P}(X)$, so that $\forall x \in X\ (M(x) \subseteq N(x))$. Given an extensive list neighbourhood function $P \in \mathrm{P}(X) :\to \mathrm{P}(X)$ the application of the inclusion function results in a new neighbourhood function $Q = \{(Y, I(P(Y))) : Y \subseteq X\} \in \mathrm{P}(X) :\to \mathrm{P}(X)$, so that $\forall Y \subseteq X\ (Q(Y) \subseteq P(Y))$.

Consider two sub-types of *optimisers*, (i) *tree optimisers*

$$\mathrm{optimiseTreer}(\mathcal{X}) \subset \mathrm{optimisers}(\mathcal{X})$$

and (ii) *list optimisers*

$$\mathrm{optimiseListers}(\mathcal{X}) \subset \mathrm{optimisers}(\mathcal{X})$$

Define the constructor of a *tree optimiser*

$$\begin{aligned}&\mathrm{optimiseTreer} \in \\ &\quad \mathrm{P}(\mathcal{X}) \times (\mathcal{X} \to \mathrm{P}(\mathcal{X})) \times (\mathrm{P}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})) \times \mathrm{P}(\mathcal{X}) \to \mathrm{optimiseTreers}(\mathcal{X})\end{aligned}$$

such that $\mathrm{optimiseTreer}(X, N, I, R) \in \mathrm{optimiseTreers}(X) \subset \mathrm{optimisers}(X)$.

Define the constructor of a *list optimiser*

$$\text{optimiseLister} \in \\ \mathrm{P}(\mathcal{X}) \times (\mathrm{P}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})) \times (\mathrm{P}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})) \times \mathrm{P}(\mathcal{X}) \to \text{optimiseListers}(\mathcal{X})$$

such that $\text{optimiseLister}(X, P, I, R) \in \text{optimiseListers}(X) \subset \text{optimisers}(X)$.

Re-define the *tree optimiser* tree function

$$\text{tree} \in \text{optimiseTreers}(\mathcal{X}) \to \text{trees}(\mathcal{X})$$

as $\text{tree}(Z) := \text{tree}(\text{searchTreer}(X, M, I(R)))$ where $\text{optimiseTreer}(X, N, I, R) = Z$ and $M = \{(x, I(N(x))) : x \in X\}$.

Re-define the *list optimiser* list function

$$\text{list} \in \text{optimiseListers}(\mathcal{X}) \to \mathcal{L}(\mathrm{P}(X))$$

as $\text{list}(Z) := \text{list}(\text{searchLister}(X, Q, I(R)))$ where $\text{optimiseLister}(X, P, I, R) = Z$ and $Q = \{(Y, I(P(Y))) : Y \subseteq X\}$.

Define the *tree optimiser* searched set as the union of the extensive neighbourhoods $\text{searched} \in \text{optimiseTreers}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})$ as

$$\text{searched}(Z) := \bigcup\{N(x) : x \in \text{elements}(Z)\} \cup R$$

Define the *list optimiser* searched set as the union of the extensive neighbourhoods $\text{searched} \in \text{optimiseListers}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})$ as

$$\text{searched}(Z) := \bigcup\{P(Y) : Y \in \text{set}(\text{list}(Z))\} \cup R$$

Define the traversable set as the elements of the search of the extensive neighbourhoods $\text{traversable} \in \text{optimiseTreers}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})$ as

$$\text{traversable}(Z) := \text{elements}(\text{searchTreer}(X, N, R))$$

and $\text{traversable} \in \text{optimiseListers}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})$ as

$$\text{traversable}(Z) := \text{elements}(\text{searchLister}(X, P, R))$$

An *optimiser* is potentially traversable if $\text{traversable}(Z) = X$.

Thus, for all $Z \in \text{optimisers}(X)$

$$I(\text{elements}(Z)) \subseteq \text{elements}(Z) \subseteq \text{searched}(Z) \subseteq \text{traversable}(Z) \subseteq X$$

Consider a variation in which the *optimisers* do not apply the inclusion function to the initial set. Consider two sub-types of *optimisers*, (i) *tree tail optimisers*

$$\mathrm{optimiseTailTreers}(\mathcal{X}) \subset \mathrm{optimisers}(\mathcal{X})$$

and (ii) *list tail optimisers*

$$\mathrm{optimiseTailListers}(\mathcal{X}) \subset \mathrm{optimisers}(\mathcal{X})$$

Define the constructor of a *tree tail optimiser*

$$\begin{aligned}&\mathrm{optimiseTailTreer} \in \\ &\quad \mathrm{P}(\mathcal{X}) \times (\mathcal{X} \to \mathrm{P}(\mathcal{X})) \times (\mathrm{P}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})) \times \mathrm{P}(\mathcal{X}) \to \mathrm{optimiseTailTreers}(\mathcal{X})\end{aligned}$$

Define the constructor of a *list tail optimiser*

$$\begin{aligned}&\mathrm{optimiseTailLister} \in \\ &\quad \mathrm{P}(\mathcal{X}) \times (\mathrm{P}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})) \times (\mathrm{P}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})) \times \mathrm{P}(\mathcal{X}) \to \mathrm{optimiseTailListers}(\mathcal{X})\end{aligned}$$

Re-define the *tree tail optimiser* tree function

$$\mathrm{tree} \in \mathrm{optimiseTailTreers}(\mathcal{X}) \to \mathrm{trees}(\mathcal{X})$$

as $\mathrm{tree}(Z) := \mathrm{tree}(\mathrm{searchTreer}(X, M, R))$ where $\mathrm{optimiseTailTreer}(X, N, I, R) = Z$ and $M = \{(x, I(N(x))) : x \in X\}$.

Re-define the *list tail optimiser* list function

$$\mathrm{list} \in \mathrm{optimiseTailListers}(\mathcal{X}) \to \mathcal{L}(\mathrm{P}(X))$$

as $\mathrm{list}(Z) := \mathrm{list}(\mathrm{searchLister}(X, Q, R))$ where $\mathrm{optimiseTailLister}(X, P, I, R) = Z$ and $Q = \{(Y, I(P(Y))) : Y \subseteq X\}$.

Consider the optimisation of some real-valued total function on the search set $F \in X :\to \mathbf{R}$. In particular, consider the maximisation inclusion function, $I = \{(Y, \mathrm{maxd}(\{(y, F(y)) : y \in Y\})) : Y \subseteq X\}$, applied to some tree neighbourhood function $N \in X :\to \mathrm{P}(X)$. The true maximum value is $\mathrm{maxr}(\{(y, F(y)) : y \in \mathrm{dom}(F)\})$ and the optimised maximum value from initial set R of *optimiser* $Z = \mathrm{optimiseTreer}(X, N, I, R)$ is $\mathrm{maxr}(\{(y, F(y)) : y \in \mathrm{elements}(Z)\})$.

Computationally it is inefficient to apply the function F to each element of the searched neighbourhoods, $\mathrm{searched}(Z)$, and then to re-apply it to the maximum subsets of these, $\mathrm{elements}(Z)$. If the function F is the search set rather than X then the values, $\mathrm{ran}(F)$, are carried around in the search tree. Let $Z' = \mathrm{optimiseTreer}(F, N', \max, R')$ where $N' := \{((x,r), \{(y, F(y)) : y \in N(x)\}) : (x,r) \in F\} = \{((x,r), \mathrm{filter}(N(x), F)) : (x,r) \in F\} \in F \to \mathrm{P}(F)$ and $R' = \mathrm{filter}(R, F) \subseteq F$. The application of function F need only be done once for the set $\mathrm{dom}(\mathrm{searched}(Z'))$. The true maximum value is $\mathrm{maxr}(F)$ and the optimised maximum value is $\mathrm{maxr}(\mathrm{elements}(Z'))$.

The *maximisers* $\mathrm{maximisers}(\mathcal{X})$ is a subset of the *optimisers*,

$$\mathrm{maximisers}(\mathcal{X}) \subset \mathrm{optimisers}(\mathcal{X} \times \mathbf{R})$$

which constrain the search set to real-valued functions.

Consider two sub-types of *maximisers*, (i) *tree maximisers*

$$\mathrm{maximiseTreers}(\mathcal{X}) \subset \mathrm{maximisers}(\mathcal{X})$$

and (ii) *list maximisers*

$$\mathrm{maximiseListers}(\mathcal{X}) \subset \mathrm{maximisers}(\mathcal{X})$$

Define the constructor of a *tree maximiser*

$$\mathrm{maximiseTreer} \in (\mathcal{X} \to \mathbf{R}) \times ((\mathcal{X} \times \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})) \times ((\mathcal{X} \to \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})) \times (\mathcal{X} \to \mathbf{R}) \to \mathrm{maximiseTreers}(\mathcal{X})$$

such that $\mathrm{maximiseTreer}(X, N, I, R) = \mathrm{optimiseTreer}(X, N, I, R)$.

Define the constructor of a *list maximiser*

$$\mathrm{maximiseLister} \in (\mathcal{X} \to \mathbf{R}) \times ((\mathcal{X} \to \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})) \times ((\mathcal{X} \to \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})) \times (\mathcal{X} \to \mathbf{R}) \to \mathrm{maximiseListers}(\mathcal{X})$$

such that $\mathrm{maximiseLister}(X, P, I, R) = \mathrm{optimiseLister}(X, P, I, R)$.

Consider two sub-types of *maximisers*, (i) *tree tail maximisers*

$$\mathrm{maximiseTailTreers}(\mathcal{X}) \subset \mathrm{maximisers}(\mathcal{X})$$

and (ii) *list tail maximisers*

$$\mathrm{maximiseTailListers}(\mathcal{X}) \subset \mathrm{maximisers}(\mathcal{X})$$

Define the constructor of a *tree tail maximiser*

maximiseTailTreer $\in$

$(\mathcal{X} \to \mathbf{R}) \times ((\mathcal{X} \times \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})) \times ((\mathcal{X} \to \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})) \times (\mathcal{X} \to \mathbf{R}) \to$
$\text{maximiseTailTreers}(\mathcal{X})$

such that $\text{maximiseTailTreer}(X, N, I, R) = \text{optimiseTailTreer}(X, N, I, R)$.

Define the constructor of a *list tail maximiser*

maximiseTailLister $\in$

$(\mathcal{X} \to \mathbf{R}) \times ((\mathcal{X} \to \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})) \times ((\mathcal{X} \to \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})) \times (\mathcal{X} \to \mathbf{R}) \to$
$\text{maximiseTailListers}(\mathcal{X})$

such that $\text{maximiseTailLister}(X, P, I, R) = \text{optimiseTailLister}(X, P, I, R)$.

Define the constructor of a *single maximiser* which is a special case of a *list maximiser* with an empty neighbourhood function

$$\begin{aligned}&\text{maximiseSingler} \in \\ &\quad (\mathcal{X} \to \mathbf{R}) \times ((\mathcal{X} \to \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})) \times (\mathcal{X} \to \mathbf{R}) \to \\ &\qquad \text{maximiseListers}(\mathcal{X})\end{aligned}$$

such that $\text{maximiseSingler}(X, I, R) = \text{optimiseLister}(X, \mathrm{P}(X) \times \{\emptyset\}, I, R)$.

Note that although the *maximiser* functions are defined in terms of the max aggregation function below, the inclusion function need not be equal to the max function. For example, $I = \text{top}(n)$. The inclusion function could also terminate a search by returning the empty set.

The true value is the maximum value of the search set. Define $\text{true} \in \text{maximisers}(\mathcal{X}) \to \mathbf{R}$ as

$$\text{true}(Z) := \text{maxr}(X)$$

where $\text{maximiseTreer}(X, N, I, R) = Z$ or $\text{maximiseLister}(X, P, I, R) = Z$.

The optimum is the maximum value of the elements. Define $\text{optimum} \in \text{maximisers}(\mathcal{X}) \to \mathbf{R}$ as

$$\text{optimum}(Z) := \text{maxr}(\text{elements}(Z))$$

The optimum value is not necessarily a leaf value in the case of *tree maximisers*, $\text{maxr}(\text{leaves}(\text{tree}(Z))) \leq \text{optimum}(Z)$. Nor is the optimum value necessarily in the last element in the case of *list maximisers*, $\text{maxr}(L_{|L|}) \leq \text{optimum}(Z)$ where $L = \text{list}(Z)$.

The error value is the difference between the true value and the optimum value. Define $\text{error} \in \text{maximisers}(\mathcal{X}) \to \mathbf{R}$

$$\text{error}(Z) := \text{true}(Z) - \text{optimum}(Z)$$

The arbitrary value is the expected maximum of arbitrarily chosen subsets of the search set, $Y \subseteq X$, having cardinality equal to that of the searched set, $|Y| = |\text{searched}(Z)|$. Define $\text{arbitrary} \in \text{maximisers}(\mathcal{X}) \to \mathbf{R}$ as

$$\text{arbitrary}(Z) := \text{average}(\{(Y, \text{maxr}(Y)) : Y \subseteq X,\ |Y| = |\text{searched}(Z)|\})$$

The difference between the optimum value and the arbitrary value is the gain. Define $\text{gain} \in \text{maximisers}(\mathcal{X}) \to \mathbf{R}$ as

$$\text{gain}(Z) := \text{optimum}(Z) - \text{arbitrary}(Z)$$

Note that the gain value is not necessarily positive. If the gain is zero then the *maximiser* is equivalent to a brute force search. The gain rate is the gain per cardinality of the searched set. Define $\text{rate} \in \text{maximisers}(\mathcal{X}) \to \mathbf{R}$ as

$$\text{rate}(Z) := \text{gain}(Z)/|\text{searched}(Z)|$$

The rate is undefined if the search set is empty, $\text{searched}(Z) = \emptyset$.

The gain may be conjectured to depend on several factors. First, the gain is conjectured to increase with the difference between (i) the expected maximum of a neighbourhood and (ii) the expected maximum of an arbitrarily chosen neighbourhood. For *tree maximisers* that is

$$\begin{aligned}&\text{average}(\{(x, \text{maxr}(N(x))) : x \in X\}) - \\ &\quad \text{average}(\{(x, \text{average}(\{(Y, \text{maxr}(Y)) : Y \subseteq X,\ |Y| = |N(x)|\})) : x \in X\})\end{aligned}$$

where $\text{maximiseTreer}(X, N, I, R) = Z$. Secondly, the optimisation gain is conjectured to increase with the difference between (i) the expected maximum of a neighbourhood per cardinality of the neighborhood and (ii) the expected maximum of an arbitrarily chosen neighbourhood per cardinality

$$\begin{aligned}&\text{average}(\{(x, \text{maxr}(N(x))/|N(x)|) : x \in X\}) - \\ &\quad \text{average}(\{(x, \text{average}(\{(Y, \text{maxr}(Y)/|Y|) : Y \subseteq X,\ |Y| = |N(x)|\})) : x \in X\})\end{aligned}$$

Conjecture that smaller neighbourhoods require longer paths and deeper trees, and therefore more elements to traverse the searched set, $\text{searched}(Z)$. Similarly for *list maximisers* smaller neighbourhoods require a longer list, and therefore more elements to traverse the searched set.

Thirdly, the gain is conjectured to increase with the cardinality of the initial set, $|R|$. Larger initial sets allow more paths to be searched in a *tree maximiser* or larger subsets of the search set in a *list maximiser*, increasing the cardinality of the searched set, $|\text{searched}(Z)|$.

A.15 Likelihood functions and Fisher information

Let the m-parameter n-dimensional *parameterised probability density functions*

$$\text{ppdfs}(m, n) \subset \mathbf{R}^m \to (\mathbf{R}^n :\to \mathbf{R}_{[0,1]})$$

be such that for all $P \in \text{ppdfs}(m, n)$ and for all $\theta \in \text{dom}(P)$, the *probability density function*, $P(\theta) \in \mathbf{R}^n :\to \mathbf{R}_{[0,1]}$ is continuous and

$$\int_{X \in \mathbf{R}^n} P(\theta)(X)\ dX = 1$$

The corresponding set of n-dimensional m-parameter *likelihood functions*

$$\text{lfs}(n, m) \subset \mathbf{R}^n :\to (\mathbf{R}^m \to \mathbf{R}_{[0,1]})$$

is such that for all $L \in \text{lfs}(n, m)$ and for all $X \in \mathbf{R}^n$, the *likelihood function*, $L(X) \in \text{dom}(L(X)) \to \mathbf{R}_{[0,1]}$ is continuous and

$$\text{lfs}(n, m) := \bigcup \left\{ \left\{ (X, \{(\theta, P(\theta)(X)) : \theta \in \text{dom}(P)\}) : X \in \mathbf{R}^n \right\} : P \in \text{ppdfs}(m, n) \right\}$$

So the *likelihood functions* are such that $\forall P \in \text{ppdfs}(m, n)\ \forall \theta \in \text{dom}(P)\ \exists L \in \text{lfs}(n, m)\ \forall X \in \mathbf{R}^n\ (L(X)(\theta) = P(\theta)(X))$.

Given a *parameterised probability density function* $P \in \text{ppdfs}(m, n)$ and its corresponding *likelihood function* $L \in \text{lfs}(n, m)$, the *maximum likelihood estimate* of the parameters $\tilde{\theta} \in \mathbf{R}^m$, under certain regularity conditions, at observation coordinate $X_{\text{o}} \in \mathbf{R}^n$ is the mode of the *likelihood function*,

$$\{\tilde{\theta}\} = \text{maxd}(L(X_{\text{o}}))$$

At $\tilde{\theta}$ the gradient of the *likelihood function* is zero,

$$\forall j \in \{1 \ldots m\}(\partial_j(L(X_{\mathrm{o}}))(\tilde{\theta}) = 0)$$

where $\partial_j \in (\mathcal{L}(\mathbf{R}) \to \mathbf{R}) \to (\mathcal{L}(\mathbf{R}) \to \mathbf{R})$ is defined $\partial_i(F) := \{(Z, \partial F(Z)/\partial Z_i) : Z \in \mathrm{dom}(F)\}$ and F is a continuous function.

At $\tilde{\theta}$ the curvature of the *likelihood function* is negative,

$$\forall j \in \{1 \ldots m\}(\partial_j^2(L(X_{\mathrm{o}}))(\tilde{\theta}) < 0)$$

so the set of modes of the *likelihood function* at X is a singleton, $|\mathrm{maxd}(L(X_{\mathrm{o}}))| = 1$.

The *score* of the j-th parameter at coordinate $X \in \mathbf{R}^n$ is the gradient of the *likelihood function* per *likelihood* or *probability density*, which equals the gradient of the *log-likelihood*,

$$\{(\theta, \frac{\partial_j(L(X))(\theta)}{L(X)(\theta)}) : \theta \in \mathrm{dom}(L(X))\} = \partial_j(\ln \circ L(X))$$

which is undefined where $L(X)(\theta) = 0$.

The expected value of the *score* is always zero,

$$\begin{aligned}
&\forall \theta \in \mathrm{dom}(P)\ \forall j \in \{1 \ldots m\} \\
&\quad \int (\partial_j(\ln \circ L(X))(\theta) \times P(\theta)(X)\ dX : X \in \mathbf{R}^n,\ P(\theta)(X) > 0) \\
&= \int (\frac{\partial_j(L(X))(\theta)}{L(X)(\theta)} \times P(\theta)(X)\ dX : X \in \mathbf{R}^n,\ P(\theta)(X) > 0) \\
&= \int_{X \in \mathbf{R}^n} \partial_j(L(X))(\theta)\ dX \\
&= \partial_j(\{(\theta', \int_{X \in \mathbf{R}^n} L(X)(\theta')\ dX) : \theta' \in \mathrm{dom}(P)\})(\theta) \\
&= \partial_j(\{(\theta', 1) : \theta' \in \mathrm{dom}(P)\})(\theta) = 0
\end{aligned}$$

That is, the expected first order sensitivity of the *probability density function*, P, to the parameters is zero.

The *Fisher information* of the j-th parameter on the diagonal $I_{P,j} \in \mathrm{dom}(P) \to \mathbf{R}_{\geq 0}$ is defined as the second moment,

$$I_{P,j}(\theta) := \int (\partial_j(\ln \circ L(X))(\theta))^2 \times P(\theta)(X)\ dX : X \in \mathbf{R}^n,\ P(\theta)(X) > 0$$

Under the regularity conditions, the *Fisher information* is always greater than zero, $I_{P,j}(\theta) > 0$. Under certain further conditions the *Fisher information* is the negative of the second derivative,

$$I_{P,j}(\theta) = -\int \partial_j^2(\ln \circ L(X))(\theta) \times P(\theta)(X)\ dX : X \in \mathbf{R}^n,\ P(\theta)(X) > 0$$

Given some observation coordinate $X_\text{o} \in \mathbf{R}^n$ the *maximum likelihood estimate* is also the mode of *log-likelihood function*, $\{\tilde{\theta}\} = \text{maxd}(\ln \circ L(X_\text{o})) = \text{maxd}(L(X_\text{o}))$, because the natural logarithm function, ln, is monotonic, $\forall j \in \{1 \ldots m\}(\partial_j(\ln \circ L(X_\text{o}))(\tilde{\theta}) = 0)$. Thus the second derivative of the j-th parameter at the *maximum likelihood estimate*, $\tilde{\theta}$, is the curvature of the *log-likelihood function*, $\partial_j^2(\ln \circ L(X_\text{o}))(\tilde{\theta})$, which is negative under the conditions.

In the case where (i) the modal *probability density* parameterised by the *maximum likelihood estimate* occurs at the observation coordinate, $X_\text{o} \in \text{maxd}(P(\tilde{\theta}))$, and (ii) the sum negative curvature of the *probability density function* at the *maximum likelihood estimate*, $-\sum_{i \in \{1 \ldots n\}} \partial_i^2(P(\tilde{\theta}))(X_\text{o})$, is high, the *Fisher information* of the j-th parameter at the *maximum likelihood estimate* of the parameters, $I_{P,j}(\tilde{\theta})$, approximates to the negative curvature of the *log-likelihood* at the *maximum likelihood estimate* times the modal *likelihood*,

$$\begin{aligned}
I_{P,j}(\tilde{\theta}) &= -\int \partial_j^2(\ln \circ L(X))(\tilde{\theta}) \times P(\tilde{\theta})(X)\ dX : X \in \mathbf{R}^n,\ P(\tilde{\theta})(X) > 0 \\
&\approx -\ \partial_j^2(\ln \circ L(X_\text{o}))(\tilde{\theta}) \times \text{maxr}(P(\tilde{\theta})) \\
&= -\ \partial_j^2(\ln \circ L(X_\text{o}))(\tilde{\theta}) \times L(X_\text{o})(\tilde{\theta})
\end{aligned}$$

Therefore the *Fisher information* for arbitrary *probability density function* would be expected to vary with the *log-likelihood*, $I_{P,j}(\tilde{\theta}) \sim \ln L(X_\text{o})(\tilde{\theta})$. That is, in some cases, the sensitivity of the *probability density function* to parameter at the *maximum likelihood estimate* varies with the *log-likelihood*.

The modal *probability density*, $P(\tilde{\theta})(X_\text{o})$, varies with the sum negative curvature of the *probability density function*, so the *Fisher information* varies with the sum negative curvature,

$$I_{P,j}(\tilde{\theta}) \ \sim \ -\sum_{i \in \{1 \ldots n\}} \partial_i^2(P(\tilde{\theta}))(X_\text{o})$$

For those centrally organised *probability density functions* which have a definition of variance $\text{var}(n) \in (\mathbf{R}^n :\to \mathbf{R}) \to \mathbf{R}$, the variance varies against the sum negative curvature at the *maximum likelihood estimate*, so the *Fisher information* would be expected to vary against the variance of the *probability density function*, at least in the case of low variance,

$$I_{P,j}(\tilde{\theta}) \quad \sim \quad - \operatorname{var}(n)(P(\tilde{\theta}))$$

That is, in some cases, the sensitivity of the *probability density function* to parameter at the *maximum likelihood estimate* varies against the variance of the *probability density function*.

The *binomial distribution* is the discrete *probability function* defined

$$\begin{aligned}
\operatorname{bpmf} &\in \mathbf{N}_{>0} \to (\mathbf{Q}_{(0,1)} \to ((\mathbf{N} \to \mathbf{Q}_{(0,1)}) \cap \mathcal{P})) \\
\operatorname{bpmf}(n) &\in \mathbf{Q}_{(0,1)} \to ((\{0 \ldots n\} :\to \mathbf{Q}_{(0,1)}) \cap \mathcal{P}) \\
\operatorname{bpmf}(n)(p)(k) &:= \binom{n}{k} p^k (1-p)^{n-k} \in \mathbf{Q}_{(0,1)}
\end{aligned}$$

The *binomial distribution* can be generalised to a *parameterised probability density function* defined in terms of the unit-translated gamma function, $\Gamma_! x = \Gamma(x+1)$,

$$\begin{aligned}
\operatorname{bppdf} &\in \mathbf{N}_{>0} \to \operatorname{ppdfs}(1,1) \\
\operatorname{bppdf}(n) &\in \mathbf{R}_{(0,1)} \to (\mathbf{R} :\to \mathbf{R}_{[0,1)}) \\
\operatorname{bppdf}(n)(p)(k) &:= \frac{n!}{\Gamma_! k \ \Gamma_!(n-k)} p^k (1-p)^{n-k} \in \mathbf{R}_{(0,1)}
\end{aligned}$$

where $0 \leq k \leq n$ and $0 < p < 1$, otherwise if $0 < p < 1$, $\operatorname{bppdf}(n)(p)(k) := 0$, otherwise $\operatorname{bppdf}(n)(p)(k)$ is undefined. Note that, strictly speaking, the *binomial parameterised probability density function*, $\operatorname{bppdf}(n)(p)$, is only coninuous in the limit as n tends to infinity.

The corresponding *likelihood function* $\operatorname{blf}(n) \in \operatorname{lfs}(1,1)$ is defined $\operatorname{blf}(n)(k)(p) := \operatorname{bppdf}(n)(p)(k)$. The *likelihood*, $\operatorname{blf}(n)(k)(p)$, and *log-likelihood*, $\ln(\operatorname{blf}(n)(k)(p))$, are defined if and only if $0 < p < 1$.

Given observation coordinate $k_{\mathrm{o}} \in \mathbf{R}_{(0,n)}$ the *maximum likelihood estimate* for the parameter of the *probability density function* is $\tilde{p} = k_{\mathrm{o}}/n$, where $\{\tilde{p}\} = \operatorname{maxd}(\operatorname{blf}(n)(k_{\mathrm{o}}))$. The *Fisher information* of the parameter $p \in \mathbf{R}_{(0,1)}$ is

$$I_{\operatorname{bppdf}(n)}(p) = \frac{n}{p(1-p)}$$

which is minimised where $p = 0.5$. That is, the *Fisher information* of the *maximum likelihood estimate* of the parameter, $I_{\mathrm{bppdf}(n)}(\tilde{p})$, is minimised if $k_{\mathrm{o}} = n/2$, where $\mathrm{blf}(n)(n/2)$ is expected to be least sensitive to $\tilde{p}$. The *Fisher information* of the *maximum likelihood estimate* varies against the variance, $n\tilde{p}(1-\tilde{p})$.

The *multiple binomial parameterised probability density function* $\mathrm{mbppdf}(n) \in \mathrm{ppdfs}(v, v)$, where $v \in \mathbf{N}_{>0}$, is defined

$$\begin{aligned}
&\mathrm{mbppdf}(n)(P) := \\
&\quad \{(K, \prod_{i \in \{1 \ldots v\}} \frac{n!}{\Gamma_! K_i \ \Gamma_!(n - K_i)} P_i^{K_i}(1-P_i)^{n-K_i}) : K \in \mathbf{R}^v_{[0,n]}\} \cup \\
&\quad (\mathbf{R}^v \setminus \mathbf{R}^v_{[0,n]}) \times \{0\}
\end{aligned}$$

where $n \in \mathbf{N}_{>0}$ and $P \in \mathbf{R}^v_{(0,1)}$, otherwise $\mathrm{mbppdf}(n)(P)$ is undefined.

The *multiple binomial likelihood function* $\mathrm{mblf}(z) \in \mathrm{lfs}(v, v)$ is defined

$$\mathrm{mblf}(n)(K) \ := \ \{(P, \mathrm{mbppdf}(n)(P)(K)) : P \in \mathbf{R}^v_{(0,1)}\}$$

where $K \in \mathbf{R}^v$.

Given observation coordinate $K_{\mathrm{o}} \in \mathbf{R}^v_{[0,n]}$ the *maximum likelihood estimate* for the parameter of the *probability density function* is $\tilde{P} = \{(i, K_{\mathrm{o}}(i)/n) : i \in \{1 \ldots v\}\}$, where $\{\tilde{P}\} = \mathrm{maxd}(\mathrm{mblf}(n)(K_{\mathrm{o}}))$. The *Fisher information* of the parameter $P_j \in \mathbf{R}_{(0,1)}$ is

$$I_{\mathrm{mbppdf}(n),j}(P_j) = \frac{n}{P_j(1-P_j)}$$

The *multinomial parameterised probability density function* $\mathrm{mppdf}(n) \in \mathrm{ppdfs}(v, v)$, where $v \in \mathbf{N}_{>0}$, is defined

$$\begin{aligned}
&\mathrm{mppdf}(n)(P) := \\
&\quad \{(K, \frac{n!}{\prod_{i \in \{1 \ldots v\}} \Gamma_! K_i} \prod_{i \in \{1 \ldots v\}} P_i^{K_i}) : K \in \mathbf{R}^v_{[0,n]}, \ \sum_{i \in \{1 \ldots v\}} K_i = n\} \cup \\
&\quad \{(K, 0) : K \in \mathbf{R}^v_{[0,n]}, \ \sum_{i \in \{1 \ldots v\}} K_i \neq n\} \cup \\
&\quad (\mathbf{R}^v \setminus \mathbf{R}^v_{[0,n]}) \times \{0\}
\end{aligned}$$

where $n \in \mathbf{N}_{>0}$, $P \in \mathbf{R}^v_{(0,1)}$ and $\sum_{i \in \{1 \dots v\}} P_i = 1$, otherwise $\text{mppdf}(n)(P)$ is undefined.

The *multinomial likelihood function* $\text{mlf}(z) \in \text{lfs}(v, v)$ is defined

$$\text{mlf}(n)(K) \ := \ \{(P, \text{mppdf}(n)(P)(K)) : P \in \mathbf{R}^v_{(0,1)}\}$$

where $K \in \mathbf{R}^v$. Note that the *multinomial likelihood function* only requires that each parameter is in the open set between zero and one, $P \in \mathbf{R}^v_{(0,1)} = \{r : r \in \mathbf{R},\ 0 < r < 1\}^v$, so P is not necessarily a *probability function*. That is, in some cases $P \notin \mathcal{P}$ or $\sum_{i \in \{1 \dots v\}} P_i \neq 1$. This is to allow well defined partial derivatives in free parameters. So $\partial_i(\text{mlf}(n)(K))(P)$ is the sensitivity of the *likelihood* to the i-th parameter at P, where $\partial_j \in (\mathcal{L}(\mathbf{R}) \to \mathbf{R}) \to (\mathcal{L}(\mathbf{R}) \to \mathbf{R})$ is defined $\partial_j(F) := \{(Z, \partial F(Z)/\partial Z_j) : Z \in \text{dom}(F)\}$ and F is a continuous function.

Given observation coordinate $K_{\text{o}} \in \mathbf{R}^v_{[0,n]}$, where $\sum_{i \in \{1 \dots v\}} K_{\text{o}}(i) = n$, the *maximum likelihood estimate* for the parameter of the *probability density function* is $\tilde{P} = \{(i, K_{\text{o}}(i)/n) : i \in \{1 \dots v\}\}$. That is, although the *multinomial parameterised probability density function* is constrained in the sum of the coordinates, $\sum_{i \in \{1 \dots v\}} K_i = n$, a Lagrangian multiplier can be used to prove that the *maximum likelihood estimate* is equal to that for the parameter of the *multiple binomial parameterised probability density function*, $\{\tilde{P}\} = \text{maxd}(\text{mlf}(n)(K_{\text{o}})) = \text{maxd}(\text{mblf}(n)(K_{\text{o}}))$. Similarly, along the diagonal the *Fisher information* of the parameter $P_j \in \mathbf{R}_{(0,1)}$ is also equal to that for the parameter of the *multiple binomial parameterised probability density function*

$$I_{\text{mppdf}(n),j}(P_j) = I_{\text{mbppdf}(n),j}(P_j) = \frac{n}{P_j(1 - P_j)}$$

B Useful functions

B.1 Entropy and Gibbs' inequality

Define $\text{entropy} \in (\mathcal{X} \to \mathbf{Q}_{\geq 0}) \to \mathbf{Q}_{\geq 0} \ln \mathbf{Q}_{>0}$

$$\text{entropy}(N) := -\sum(\hat{N}_x \ln \hat{N}_x : x \in \text{dom}(N),\ N_x > 0)$$

where $\text{sum}(N) > 0$ and normalised $\hat{N} = \{(x, q/\text{sum}(N)) : (x, q) \in N\}$. Define $\text{entropy}(\emptyset) := 0$. Here entropy is defined such that it is independent of $\text{sum}(N)$.

Define entropyCross $\in (\mathcal{X} \to \mathbf{Q}_{\geq 0}) \times (\mathcal{X} \to \mathbf{Q}_{\geq 0}) \to \mathbf{Q}_{\geq 0} \ln \mathbf{Q}_{>0}$

$$\text{entropyCross}(N, M) := -\sum(\hat{N}_x \ln \hat{M}_x : x \in \text{dom}(N),\ M_x > 0)$$

where $\text{sum}(N) > 0$, $\text{sum}(M) > 0$ and $\text{dom}(N) \subseteq \text{dom}(M)$.

Gibbs' inequality states that

$$-\sum(P_x \ln P_x : x \in \text{dom}(P),\ P_x > 0) \leq -\sum(P_x \ln Q_x : x \in \text{dom}(P),\ Q_x > 0)$$

where $P, Q \in \mathcal{X} \to \mathbf{Q}_{\geq 0}$, $\text{sum}(P) = 1$, $\text{dom}(Q) \supseteq \text{dom}(P)$ and $\text{sum}(Q) \leq 1$. P is a *probability function*, $P \in \mathcal{P}$, but Q is not necessarily a *probability function*.

Define entropyRelative $\in (\mathcal{X} \to \mathbf{Q}_{\geq 0}) \times (\mathcal{X} \to \mathbf{Q}_{\geq 0}) \to \mathbf{Q}_{\geq 0} \ln \mathbf{Q}_{>0}$

$$\text{entropyRelative}(N, M) := \sum(\hat{N}_x \ln \frac{\hat{N}_x}{\hat{M}_x} : x \in \text{dom}(N),\ N_x > 0,\ M_x > 0)$$

where $\text{sum}(N) > 0$, $\text{sum}(M) > 0$ and $\text{dom}(N) \subseteq \text{dom}(M)$. The relative entropy equals the cross entropy minus the entropy,

$$\text{entropyRelative}(N, M) \quad = \quad \text{entropyCross}(N, M) - \text{entropy}(N)$$

By Gibbs' inequality the relative entropy is positive, $\text{entropyRelative}(N, M) \geq 0$. So the cross entropy is greater than or equal to the entropy,

$$\text{entropyCross}(N, M) \geq \text{entropy}(N)$$

B.2 Probability functions

The set of *probability functions* $\mathcal{P}$ is the set of rational valued functions such that the values are bounded $[0, 1]$ and sum to 1,

$$\mathcal{P} \subset \mathcal{X} \to \{q : q \in \mathbf{Q},\ 0 \leq q \leq 1\}$$

and

$$\forall P \in \mathcal{P}\ (\text{sum}(P) = 1)$$

A *probability function* cannot be empty, $\emptyset \notin \mathcal{P}$. Note that the events of a *probability definition*, $\text{dom}(P)$ where $P \in \mathcal{P}$, are defined here as elementary events or outcomes. That is, the events are exclusive and do not form a

σ-field.

Any non-empty, finite $\mathcal{X}$-valued function of $\mathcal{Y}$ implies a distribution of $\mathcal{Y}$ over $\mathcal{X}$ and hence a *probability function* by normalisation,

$$\forall R \in \mathcal{Y} \to \mathcal{X}\ (0 < |R| < \infty \implies \{(x, |C|) : (x, C) \in R^{-1}\}^\wedge \in \mathcal{P})$$

where $()^{-1}$:= inverse and $()^\wedge$:= normalise. Conversely, any *probability function* implies the existence of at least one non-empty, finite $\mathcal{X}$-valued function of integer,

$$\forall P \in \mathcal{P}\ \exists R \in \mathbf{N} \to \mathcal{X}$$
$$(0 < |R| < \infty\ \wedge\ \{(x, |C|) : (x, C) \in R^{-1}\}^\wedge = \{(x, p) : (x, p) \in P,\ p > 0\})$$

The set of *weak probability functions* $\mathcal{P}'$ is superset of *probability functions*, $\mathcal{P}' \supset \mathcal{P}$, that weakens the summation constraint such that the sum is less than or equal to 1,

$$\mathcal{P}' \subset \mathcal{X} \to \{q : q \in \mathbf{Q},\ 0 \leq q \leq 1\}$$

and

$$\forall P' \in \mathcal{P}'\ (\text{sum}(P') \leq 1)$$

The empty function is a *weak probability function*, $\emptyset \in \mathcal{P}'$.

The *expected* or *mean* of a *probability function* $P \in \mathcal{P}$ applied to a given function $F \in \mathcal{X} \to \mathbf{R}$ is defined expected $\in \mathcal{P} \to ((\mathcal{X} \to \mathbf{R}) \to \mathbf{R})$, and expected$(P) \in (\text{dom}(P) \to \mathbf{R}) \to \mathbf{R}$ as

$$\text{expected}(P)(F) := \sum_{x \in \text{dom}(F)} P_x F_x$$

Define expected$(P)(\emptyset) := 0$. If $F \in \mathcal{X} \to \mathbf{Q}$ then expected$(P)(F) \in \mathbf{Q}$. If $F \in \mathcal{X} \to \mathbf{Q}_{\geq 0}$ then expected$(P)(F) \in \mathbf{Q}_{\geq 0}$.

The covariance of given functions $F, G \in \mathcal{X} \to \mathbf{R}$ is defined covariance $\in \mathcal{P} \to ((\mathcal{X} \to \mathbf{R}) \times (\mathcal{X} \to \mathbf{R}) \to \mathbf{R})$, and covariance$(P) \in (\text{dom}(P) \to \mathbf{R}) \times (\text{dom}(P) \to \mathbf{R}) \to \mathbf{R}$ as

$$\begin{aligned}&\text{covariance}(P)(F, G) := \\ &\quad \text{expected}(P)(\{(x, F(x)G(x)) : x \in \text{dom}(F) \cap \text{dom}(G)\}) - \\ &\quad \text{expected}(P)(\text{filter}(\text{dom}(G), F)) \times \text{expected}(P)(\text{filter}(\text{dom}(F), G))\end{aligned}$$

The variance of given function $F \in \mathcal{X} \to \mathbf{R}$ is defined variance $\in \mathcal{P} \to ((\mathcal{X} \to \mathbf{R}) \to \mathbf{R})$, and variance$(P) \in (\text{dom}(P) \to \mathbf{R}) \to \mathbf{R}$ as

$$\text{variance}(P)(F) := \text{covariance}(P)(F, F)$$

Note that in the case of uniform *probability function*, $P = \text{dom}(F) \times \{1/|F|\}$, the variance, variance$(P)(F)$, is the population variance, not the sample variance.

The correlation of given functions $F, G \in \mathcal{X} \to \mathbf{R}$ is defined correlation $\in \mathcal{P} \to ((\mathcal{X} \to \mathbf{R}) \times (\mathcal{X} \to \mathbf{R}) \to \mathbf{R})$, and correlation$(P) \in (\text{dom}(P) \to \mathbf{R}) \times (\text{dom}(P) \to \mathbf{R}) \to \mathbf{R}$ as

$$\begin{aligned}&\text{correlation}(P)(F, G) := \\ &\quad \frac{\text{covariance}(P)(F, G)}{\sqrt{\text{var}(P)(\text{filter}(\text{dom}(G), F))} \times \sqrt{\text{var}(P)(\text{filter}(\text{dom}(F), G))}}\end{aligned}$$

where var = variance. The correlation is undefined if either variance is zero.

The moment generating function of given function $F \in \mathcal{X} \to \mathbf{R}$ and moment parameter $t \in \mathbf{R}$ is defined mgf $\in \mathcal{P} \to ((\mathcal{X} \to \mathbf{R}) \to (\mathbf{R} \to \mathbf{R}))$, and mgf$(P) \in (\text{dom}(P) \to \mathbf{R}) \to (\mathbf{R} \to \mathbf{R})$ as

$$\text{mgf}(P)(F)(t) := \text{expected}(P)(\{(x, e^{tF_x}) : x \in \text{dom}(F)\})$$

The expression $F \sim G$, where $F, G \in \mathcal{X} \to \mathbf{R}$ and $\text{dom}(F) = \text{dom}(G)$, can be formalised in terms of the covariance of the functions in a uniform *probability function*

$$F \sim G \iff \text{covariance}(\text{dom}(F) \times \{1/|F|\})(F, G) > 0$$

B.3 Function composition

A pair of functions, $F_1, F_2 \in \mathcal{X} \to \mathcal{X}$ can be composed in an outer join $F_2 \circ F_1$. Define compose $\in (\mathcal{X} \to \mathcal{X}) \times (\mathcal{X} \to \mathcal{X}) \to (\mathcal{X} \to \mathcal{X})$ as

$$\begin{aligned}&\text{compose}(F_1, F_2) := \\ &\quad \{(x_1, y_2) : (x_1, y_1) \in F_1,\ (x_2, y_2) \in F_2,\ x_2 = y_1\} \cup \\ &\quad \{(x_1, y_1) : (x_1, y_1) \in F_1,\ y_1 \notin \text{dom}(F_2)\} \cup \\ &\quad \{(x_2, y_2) : (x_2, y_2) \in F_2,\ x_2 \notin \text{dom}(F_1)\}\end{aligned}$$

Define $F_2 \circ F_1 := \text{compose}(F_1, F_2)$. The domain of a composition is the union of the domains of the arguments, $\text{dom}(F_2 \circ F_1) = \text{dom}(F_1) \cup \text{dom}(F_2)$. A sequence of compositions of functions is right associative, $F_3 \circ F_2 \circ F_1 = F_3 \circ (F_2 \circ F_1)$. A list of functions, $L \in \mathcal{L}(\mathcal{X} \to \mathcal{X})$, can be composed recursively from right to left. Define $\text{compose} \in \mathcal{L}(\mathcal{X} \to \mathcal{X}) \to (\mathcal{X} \to \mathcal{X})$ as $\text{compose}(L) := \text{compose}(\text{sequence}(\text{reverse}(L)))$ and $\text{compose} \in \mathcal{K}(\mathcal{X} \to \mathcal{X}) \to (\mathcal{X} \to \mathcal{X})$ as

$$\begin{aligned} \text{compose}((F, X)) &:= F \circ \text{compose}(X) \\ \text{compose}(\emptyset) &:= \emptyset \end{aligned}$$

A composition $L \in \mathcal{L}(\mathcal{X} \to \mathcal{X})$ is non-circular where no source subsequently appears as a target, $\forall i \in \{1 \ldots |L|\}\ (\text{dom}(L_i) \cap \text{ran}(\bigcup \text{set}(\text{drop}(i-1, L))) = \emptyset)$. A composition L is functional where unioned list is functional, $\bigcup \text{set}(L) \in \mathcal{X} \to \mathcal{X}$.

B.4 Monoidal product

The product of a monoidal set $(\prod) \in \text{P}(\mathcal{X}) \to \mathcal{X}$ requires a product operator $(*) \in \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ which is commutative, $\forall x, y \in \mathcal{X}\ (x * y = y * x)$. The monoidal product $\prod X$ is defined by choosing some enumeration $Q \in \text{enums}(X)$ and then folding over the list, $\text{flip}(Q) \in \mathcal{L}(X)$

$$\prod X = \text{fold1}((*), \text{flip}(Q))$$

The product of the empty set, $\prod \emptyset$, is undefined.

In the case of sets of sets, $(\prod) \in \text{P}(\text{P}(\mathcal{X})) \to \text{P}(\mathcal{X})$, and in the absence of an explicit monoidal operator, the operator is taken to be the union of the self. Given argument $Q \in \text{P}(\text{P}(\mathcal{X}))$, let $S = \{\text{self}(P) : P \in Q\}$. Then let the operator be $X * Y := \{x \cup y : x \in X,\ y \in Y\}$. Then $\prod Q = \prod S$. Alternatively the product can be calculated from the product of a list of sets. Choose some arbitrary enumeration of the argument Q, $X \in \text{enums}(Q)$. Then the product of a list of sets is a set of lists, $\prod \text{flip}(X) \in \text{P}(\mathcal{L}(\mathcal{X}))$. Then $\prod Q = \{\text{set}(R) : R \in \prod \text{flip}(X)\}$.

B.5 Lists, tuples and sequences

A list is defined $\mathcal{L}(\mathcal{X}) \subset \mathbf{N} \to \mathcal{X}$, where $\mathcal{X}$ is the universal set (or type variable), such that $\forall L \in \mathcal{L}(\mathcal{X})\ (L \neq \emptyset \implies \text{dom}(L) = \{1 \ldots |L|\})$. If a list is a bijection $L \in \mathbf{N} \leftrightarrow \mathcal{X} \subset \mathcal{L}(\mathcal{X})$ then this constraint can also be expressed $\text{flip}(L) \in \text{enums}(\text{ran}(L))$, which highlights the connection with enumerations.

For the subset of tuples which have elements of type $\mathcal{X}$ define list $\in \{\mathcal{X}^i : i \in \mathbf{N}\} \to \mathcal{L}(\mathcal{X})$ as list() $:= \emptyset$, list$(x) := \{(1, x)\}$, list$(x, y) := \{(1, x), (2, y)\}$, and so on. The inverse function is defined tuple $\in \mathcal{L}(\mathcal{X}) \to \{\mathcal{X}^i : i \in \mathbf{N}\}$ as tuple$(\emptyset) := ()$, tuple$(\{(1, x)\}) := (x)$, tuple$(\{(1, x), (2, y)\}) := (x, y)$, and so on.

The set of head/tail sequences is recursively defined in terms of null terminated pairs

$$\mathcal{K}(\mathcal{X}) := ((\mathcal{X} \setminus \{\emptyset\}) \times \mathcal{K}(\mathcal{X})) \cup \{\emptyset\}$$

Define list $\in \mathcal{K}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ as

$$\begin{aligned} \text{list}(\emptyset) &:= \emptyset \\ \text{list}((x, \emptyset)) &:= \{(1, x)\} \\ \text{list}((x, X)) &:= \{(1, x)\} \cup \{(i + 1, y) : (i, y) \in \text{list}(X)\} \end{aligned}$$

Similarly, define sequence $\in \mathcal{L}(\mathcal{X}) \to \mathcal{K}(\mathcal{X})$ as

$$\begin{aligned} \text{sequence}(\emptyset) &:= \emptyset \\ \text{sequence}(\{(i, x)\}) &:= (x, \emptyset) \end{aligned}$$

And

$$\text{sequence}(L) := (x, \text{sequence}(L \setminus \{(i, x)\}))$$

where $\{(x, i)\} \in \min(\text{flip}(L))$.

Define set $\in \mathcal{L}(\mathcal{X}) \to \text{P}(\mathcal{X})$ as

$$\text{set}(L) := \text{ran}(L)$$

We define the constructor of a list from a set as list $\in \text{P}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ without defining the method, except to say that the order is arbitrary and it is constrained such that set(list$(X)) = X$. Also define sequence $\in \text{P}(\mathcal{X}) \to \mathcal{K}(\mathcal{X})$ as sequence$(X) =$ sequence(list(X)).

Define map $\in (\mathcal{X} \to \mathcal{Y}) \times \mathcal{K}(\mathcal{X}) \to \mathcal{K}(\mathcal{Y})$ as

$$\begin{aligned} \text{map}(F, \emptyset) &:= \emptyset \\ \text{map}(F, (x, X)) &:= (F_x, \text{map}(F, X)) \end{aligned}$$

Define filter $\in (\mathcal{X} \to \mathbf{B}) \times \mathcal{K}(\mathcal{X}) \to \mathcal{K}(\mathcal{X})$ as

$$\begin{aligned} \text{filter}(F, \emptyset) &:= \emptyset \\ \text{filter}(F, (x, X)) &:= \text{if}(F(x), (x, \text{filter}(F, X)), \text{filter}(F, X)) \end{aligned}$$

Define map $\in (\mathcal{X} \to \mathcal{Y}) \times \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{Y})$ as

$$\text{map}(F, L) := \{(i, F_x) : (i, x) \in L\}$$

where $\forall (F, L) \in \text{dom}(\text{map})\ (\text{ran}(L) \in \text{dom}(F))$.

Define right associative fold $\in (\mathcal{X} \times \mathcal{Y} \to \mathcal{Y}) \times \mathcal{Y} \times \mathcal{K}(\mathcal{X}) \to \mathcal{Y}$ as

$$\begin{aligned} \text{fold}(F, y, \emptyset) &:= y \\ \text{fold}(F, y, (x, X)) &:= F(x, \text{fold}(F, y, X)) \end{aligned}$$

Define fold1 $\in (\mathcal{X} \times \mathcal{Y} \to \mathcal{Y}) \times \mathcal{K}(\mathcal{X}) \to \mathcal{Y}$ as

$$\text{fold1}(F, (x, X)) = \text{fold}(F, x, X)$$

which is defined for an non-empty sequence. Define fold $\in (\mathcal{X} \times \mathcal{Y} \to \mathcal{Y}) \times \mathcal{Y} \times \mathcal{L}(\mathcal{X}) \to \mathcal{Y}$ as $\text{fold}(F, y, L) := \text{fold}(F, y, \text{sequence}(L))$ and fold1 $\in (\mathcal{X} \times \mathcal{Y} \to \mathcal{Y}) \times \mathcal{L}(\mathcal{X}) \to \mathcal{Y}$ as $\text{fold1}(F, L) := \text{fold1}(F, \text{sequence}(L))$.

Define reverse $\in \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ as

$$\text{reverse}(L) := \{(|L| + 1 - i, x) : (i, x) \in L\}$$

Define concat $\in \mathcal{K}(\mathcal{X}) \times \mathcal{K}(\mathcal{X}) \to \mathcal{K}(\mathcal{X})$ as

$$\begin{aligned} \text{concat}(\emptyset, Y) &:= Y \\ \text{concat}((x, X), Y) &:= (x, \text{concat}(X, Y)) \end{aligned}$$

Define concat $\in \mathcal{L}(\mathcal{X}) \times \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ as

$$\text{concat}(L, M) := L \cup \{(|L| + i, x) : (i, x) \in M\}$$

Define concat $\in \mathcal{L}(\mathcal{L}(\mathcal{X})) \to \mathcal{L}(\mathcal{X})$ as

$$\text{concat}(N) := \text{fold}(\text{concat}, \emptyset, \text{sequence}(N))$$

Define head $\in \mathcal{L}(\mathcal{X}) \setminus \{\emptyset\} \to \mathcal{X}$ as

$$\text{head}(L) := x$$

where $(1, x) \in L$. head$(\emptyset)$ is undefined.

Define last $\in \mathcal{L}(\mathcal{X}) \setminus \{\emptyset\} \to \mathcal{X}$ as

$$\text{last}(L) := x$$

where $(|L|, x) \in L$. $\text{last}(\emptyset)$ is undefined.

Define $\text{tail} \in \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ as

$$\text{tail}(L) := \{(i-1, x) : (i, x) \in L,\ i > 1\}$$

Define $\text{take} \in \mathbf{N} \times \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ as

$$\text{take}(j, L) := \{(i, x) : (i, x) \in L,\ i \leq j\}$$

Define $\text{drop} \in \mathbf{N} \times \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ as

$$\text{drop}(j, L) := \{(i-j, x) : (i, x) \in L,\ i > j\}$$

A selection of a list is defined $\text{select} \in \text{P}(\mathbf{N}_{>0}) \times \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ as

$$\text{select}(Q, L) := \{(|\{j : j \in Q,\ j \leq i\}|, L_i) : i \in Q,\ i \leq |L|\}$$

Define shorthand $L_Q := \text{select}(Q, L)$ where $Q \subset \mathbf{N}_{>0}$. Then $\text{take}(j, L) = L_{\{1 \ldots j\}}$ and $\text{drop}(j, L) = L_{\{j+1 \ldots |L|\}}$.

Define $\text{sublists} \in \mathcal{L}(\mathcal{X}) \to \text{P}(\mathcal{L}(\mathcal{X}))$ as

$$\text{sublists}(L) := \{\text{take}(i, L) : i \in \{1 \ldots |L|\}\} \cup \{\emptyset\}$$

There are $|L| + 1$ sublists, $|\text{sublists}(L)| = |L| + 1$. The empty list, $\emptyset$, is a sublist of all lists. The immediate sublist is $\text{take}(|L| - 1, L) \in \text{sublists}(L)$ where $L \neq \emptyset$.

A pair of lists may be zipped together as far as the shorter list. Define $\text{zip} \in \mathcal{L}(\mathcal{X}) \times \mathcal{L}(\mathcal{Y}) \to \mathcal{L}(\mathcal{X} \times \mathcal{Y})$ as

$$\text{zip}(L, M) := \{(i, (L_i, M_i)) : i \in \{1 \ldots \text{minimum}(|L|, |M|)\}\}$$

The product of a list of sets is a set of lists, $(\prod) \in \mathcal{L}(\text{P}(\mathcal{X})) \to \text{P}(\mathcal{L}(\mathcal{X}))$. See 'Monoidal product', above. Let $\text{mul} \in \text{P}(\mathcal{X}) \times \text{P}(\mathcal{L}(\mathcal{X})) \to \text{P}(\mathcal{L}(\mathcal{X}))$ be defined as

$$\text{mul}(Q, R) := \{\text{concat}(J, \{(1, x)\}) : J \in R,\ x \in Q\}$$

Then

$$\prod L := \text{fold}(\text{mul}, \{\emptyset\}, L)$$

The product of the empty set is a set of the empty list, $\prod \emptyset = \{\emptyset\}$. The product of sequences, $(\prod) \in \mathcal{K}(\text{P}(\mathcal{X})) \to \text{P}(\mathcal{K}(\mathcal{X}))$, and the product of tuples, $(\prod) \in \text{tuples}(\text{P}(\mathcal{X})) \to \text{P}(\text{tuples}(\mathcal{X}))$, are similarly defined.

The power of a set is the product of a list of the set such that the length of the list equals the power, $X^n = \prod(\{1 \ldots n\} \times \{X\}) \subset \{L : L \in \mathcal{L}(X),\ |L| = n\}$.

B.6 Trees

Trees, as defined here, are unordered functional relations between objects and trees

$$\text{trees}(\mathcal{X}) = \mathcal{X} \to \text{trees}(\mathcal{X})$$

The function nodes $\in \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{X} \times \text{trees}(\mathcal{X}))$ is defined

$$\text{nodes}(T) := T \cup \bigcup\{\text{nodes}(R) : (x, R) \in T\}$$

where $\text{nodes}(\emptyset) := \emptyset$.

The elements of the tree is defined elements $\in \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{X})$

$$\text{elements}(T) := \text{dom}(\text{nodes}(T))$$

The function roots $\in \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{X})$ is defined

$$\text{roots}(T) := \text{dom}(T)$$

The function leaves $\in \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{X})$ is defined

$$\text{leaves}(T) := \{x : (x, R) \in \text{nodes}(T),\ R = \emptyset\}$$

The set of pairs of elements of the tree is defined steps $\in \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{X} \times \mathcal{X})$

$$\text{steps}(T) := \{(x, y) : (x, R) \in \text{nodes}(T),\ y \in \text{roots}(R)\}$$

Define map $\in (\mathcal{X} \to \mathcal{Y}) \times \text{trees}(\mathcal{X}) \to \text{trees}(\mathcal{Y})$ as

$$\text{map}(F, T) \;:=\; \{(F(x), \text{map}(F, R)) : (x, R) \in T\}$$

where F is such that $\text{elements}(T) \subseteq \text{dom}(F)$.

Define mapNode $\in (\mathcal{X} \times \text{trees}(\mathcal{X}) \to \mathcal{Y}) \times \text{trees}(\mathcal{X}) \to \text{trees}(\mathcal{Y})$ as

$$\text{mapNode}(F, T) \;:=\; \{(F(x, R), \text{mapNode}(F, R)) : (x, R) \in T\}$$

where F is such that $\text{elements}(T) \times \text{trees}(\text{elements}(T)) \subseteq \text{dom}(F)$.

Define mapAccum $\in (\mathcal{L}(\mathcal{X}) \to \mathcal{Y}) \times \text{trees}(\mathcal{X}) \to \text{trees}(\mathcal{Y})$ as

$$\text{mapAccum}(F, T) := \text{mapAccum}(F, \emptyset, T)$$

and mapAccum $\in (\mathcal{L}(\mathcal{X}) \to \mathcal{Y}) \times \mathcal{L}(\mathcal{X}) \times \text{trees}(\mathcal{X}) \to \text{trees}(\mathcal{Y})$ as

$$\begin{aligned}&\text{mapAccum}(F, L, T) := \\ &\quad \{(F(M), \text{mapAccum}(F, M, R)) : (x, R) \in T,\ M = \text{concat}(L, \{(1, x)\})\}\end{aligned}$$

where F is such that paths$(T) \subset$ dom(F).

Define mapNodeAccum $\in (\mathcal{L}(\mathcal{X}) \times \text{trees}(\mathcal{X}) \to \mathcal{Y}) \times \text{trees}(\mathcal{X}) \to \text{trees}(\mathcal{Y})$ as

$$\text{mapNodeAccum}(F, T) := \text{mapNodeAccum}(F, \emptyset, T)$$

and mapNodeAccum $\in (\mathcal{L}(\mathcal{X}) \times \text{trees}(\mathcal{X}) \to \mathcal{Y}) \times \mathcal{L}(\mathcal{X}) \times \text{trees}(\mathcal{X}) \to \text{trees}(\mathcal{Y})$ as

$$\begin{aligned}&\text{mapNodeAccum}(F, L, T) := \\ &\quad \{(F(M, R), \text{mapNodeAccum}(F, M, R)) : (x, R) \in T,\ M = \text{concat}(L, \{(1, x)\})\}\end{aligned}$$

where F is such that elements$(T) \times$ trees(elements$(T)) \subseteq$ dom(F).

Given a pair of trees, dot $\in \text{trees}(\mathcal{X}) \times \text{trees}(\mathcal{Y}) \to \text{trees}(\mathcal{X} \times \mathcal{Y})$, returns the zipped tree of pairs,

$$\text{dot}(S, T) := \{((x, y), \text{dot}(S_x, T_y)) : (x, y) \in \text{dom}(S) \cdot \text{dom}(T)\}$$

where dot$(\emptyset, \cdot) := \{\emptyset\}$ and dot$(\cdot, \emptyset) := \{\emptyset\}$, and

$$X \cdot Y := \text{zip}(\text{flip}(\text{order}(D_{\mathcal{X}}, X)), \text{flip}(\text{order}(D_{\mathcal{Y}}, Y)))$$

where $D_{\mathcal{X}}$ and $D_{\mathcal{Y}}$ are orders on $\mathcal{X}$ and $\mathcal{Y}$.

Given a tree of pairs, distinct $\in \text{trees}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\text{trees}(\mathcal{X} \times \mathcal{Y}))$, returns the set of distinct trees such that the domains of the trees form a function, $\forall T \in \text{trees}(\mathcal{X} \times \mathcal{Y})\ \forall U \in \text{distinct}(T)\ \forall V \in \{U\} \cup \text{ran}(\text{nodes}(U))\ (\text{dom}(V) \in \mathcal{X} \to \mathcal{Y})$,

$$\begin{aligned}&\text{distinct}(T) := \\ &\quad \{U : U \subseteq \{((x, y), R) : ((x, y), S) \in T,\ R \in \text{distinct}(S)\}, \\ &\qquad\qquad \text{dom}(U) \in \text{dom}(\text{dom}(T)) :\to \text{ran}(\text{dom}(T))\}\end{aligned}$$

where distinct$(\emptyset) := \{\emptyset\}$.

Define paths $\in \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{L}(\mathcal{X}))$ as

$$\text{paths}(T) := \text{paths}(\emptyset, T)$$

where we define paths $\in \mathcal{L}(\mathcal{X}) \times \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{L}(\mathcal{X}))$

$$\text{paths}(L, T) := \bigcup \{\text{paths}(\text{concat}(L, \{(1, x)\}), R) : (x, R) \in T\}$$

where $\text{paths}(L, \emptyset) := \{L\}$.

Define tree $\in \text{P}(\mathcal{L}(\mathcal{X})) \to \text{trees}(\mathcal{X})$

$$\text{tree}(Q) := \\ \{(h, \text{tree}(\{\text{tail}(J) : J \in Q,\ (1, h) \in J\})) : h \in \{\text{head}(L) : L \in Q,\ |L| > 0\}\}$$

where $\text{tree}(\emptyset) := \emptyset$.

The function depth $\in \text{trees}(\mathcal{X}) \to \mathbf{N}$ is defined

$$\text{depth}(T) := \text{maxr}(\{(L, |L|) : L \in \text{paths}(T)\})$$

The depth of the empty tree is defined as zero, $\text{depth}(\emptyset) := 0$.

Define places $\in \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{L}(\mathcal{X}) \times \text{trees}(\mathcal{X}))$ as

$$\text{places}(T) := \text{places}(\emptyset, T)$$

where we define places $\in \mathcal{L}(\mathcal{X}) \times \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{L}(\mathcal{X}) \times \text{trees}(\mathcal{X}))$

$$\text{places}(L, T) := \\ \bigcup \{\{(M, R)\} \cup \text{places}(M, R) : (x, R) \in T,\ M = \text{concat}(L, \{(1, x)\})\}$$

Define $\text{places}(\cdot, \emptyset) := \emptyset$. The places function is related to nodes, $\{(L_{|L|}, R) : (L, R) \in \text{places}(T)\} = \text{nodes}(T)$. The places function is related to paths, $\{L : (L, R) \in \text{places}(T),\ R = \emptyset\} = \text{paths}(T)$. The places function can be defined in terms of mapNodeAccum, $\text{places}(T) = \text{elements}(\text{mapNodeAccum}(\text{id}, T))$, where $\forall x \in \mathcal{X}\ (\text{id}(x) = x)$.

Define subpaths $\in \text{trees}(\mathcal{X}) \to \text{P}(\mathcal{L}(\mathcal{X}))$ as

$$\text{subpaths}(T) := \text{dom}(\text{places}(T))$$

The set of subtrees can be constructed from the monoidal product of the sublists of the paths. Define subtrees $\in \text{trees}(\mathcal{X}) \to \text{P}(\text{trees}(\mathcal{X}))$ as

$$\text{subtrees}(T) := \{\text{tree}(Q) : Q \in \prod \{\text{sublists}(L) : L \in \text{paths}(T)\}\} \cup \{\emptyset\}$$

The empty tree, $\emptyset$, is a subtree of all trees. The set of immediate subtrees is the subset where exactly one path is an immediate sublist, $\{\text{tree}(\text{paths}(T) \setminus \{L\} \cup \{\text{take}(|L|-1, L)\}) : L \in \text{paths}(T)\} \subseteq \text{subtrees}(T)$ where $T \neq \emptyset$. The cardinality of the nodes of immediate subtrees is reduced by one, $\{S : S \in \text{subtrees}(T),\ |\text{nodes}(S)| = |\text{nodes}(T)| - 1\}$ where $T \neq \emptyset$.

An n-ary tree T is such that $\forall X \in \{T\} \cup \text{ran}(\text{nodes}(T))\ (X \neq \emptyset \implies |X| = n)$.

Define list trees $\text{listTrees}(\mathcal{X}) = \mathcal{L}(\mathcal{X} \times \text{listTrees}(\mathcal{X}))$. List trees are useful for path-dependent traversal. List trees are constructed from trees given an order $D \in \mathcal{X} \leftrightarrow \mathbf{N}$. The function $\text{order} \in (\mathcal{X} \leftrightarrow \mathbf{N}) \times \mathrm{P}(\mathcal{X}) \to (\mathcal{X} \leftrightarrow \mathbf{N})$, defined below, facilitates the construction of lists, $\text{inverse}(\text{order}(D, Y)) \in \mathcal{L}(Y)$. Define $\text{listTree} \in (\mathcal{X} \leftrightarrow \mathbf{N}) \times \text{trees}(\mathcal{X}) \to \text{listTrees}(\mathcal{X})$ as

$$\text{listTree}(D, T) := \{(i, (x, \text{listTree}(D, T(x)))) : (x, i) \in \text{order}(D, \text{dom}(T))\}$$

The converse function $\text{tree} \in \text{listTrees}(\mathcal{X}) \to \text{trees}(\mathcal{X})$ is defined

$$\text{tree}(L) := \{(x, \text{tree}(M)) : (i, (x, M)) \in L,\ \text{set}(L) \in \mathcal{X} \to \text{listTrees}(\mathcal{X})\}$$

A list tree can be concatenated into a list in a depth-first traversal. Define $\text{concat} \in \text{listTrees}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ as

$$\text{concat}(L) := \text{fold}(\text{accum}, \emptyset, L)$$

where $\text{accum} \in (\mathcal{X} \times \text{listTrees}(\mathcal{X})) \times \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ is defined

$$\text{accum}((x, M), Q) := \text{concat}(\text{concat}(Q, \{(1, x)\}), \text{concat}(M))$$

B.7 Binary maps

Binary maps are defined

$$\mathcal{B}(\mathcal{X}) \subset (\mathbf{N} \times \mathcal{X} \times \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{X})) \cup \{\emptyset\}$$

The function $\text{function} \in \mathcal{B}(\mathcal{X}) \to (\mathbf{N} \to \mathcal{X})$ is defined

$$\begin{aligned} \text{function}((m, x, L, R)) &:= \{(m, x)\} \cup \text{function}(L) \cup \text{function}(R) \\ \text{function}(\emptyset) &:= \emptyset \end{aligned}$$

The function $\text{domain} \in \mathcal{B}(\mathcal{X}) \to \mathrm{P}(\mathbf{N})$ is defined as

$$\text{domain}(B) := \text{dom}(\text{function}(B))$$

The function range $\in \mathcal{B}(\mathcal{X}) \to \mathrm{P}(\mathcal{X})$ is defined as

$$\mathrm{range}(B) := \mathrm{ran}(\mathrm{function}(B))$$

The function depth $\in \mathcal{B}(\mathcal{X}) \to \mathbf{N}$ is defined

$$\begin{aligned} \mathrm{depth}((\cdot, \cdot, L, R)) &:= 1 + \mathrm{maximum}(\mathrm{depth}(L), \mathrm{depth}(R)) \\ \mathrm{depth}(\emptyset) &:= 0 \end{aligned}$$

where $\mathrm{maximum}(a, b) := \mathrm{if}(a < b, b, a)$.

Binary maps are constrained such that

$$\forall (m, \cdot, L, \cdot) \in \mathcal{B}(\mathcal{X})\ \forall (l, \cdot) \in \mathrm{function}(L)\ (l < m)$$

and

$$\forall (m, \cdot, \cdot, R) \in \mathcal{B}(\mathcal{X})\ \forall (r, \cdot) \in \mathrm{function}(R)\ (m < r)$$

The function mapBinary $\in (\mathbf{N} \to \mathcal{X}) \to \mathcal{B}(\mathcal{X})$ is constrained such that $\forall Q \in \mathbf{N} \to \mathcal{X}$ $(\mathrm{function}(\mathrm{mapBinary}(Q)) = Q)$ and such that $\forall M \in \mathcal{B}(\mathcal{X}) \setminus \{\emptyset\}$ $(\mathrm{depth}(M) \leq \log_2|\mathrm{function}(M)| + 1)$.

The function find $\in \mathcal{B}(\mathcal{X}) \times \mathbf{N} \to \mathcal{X}$ is defined

$$\begin{aligned} \mathrm{find}((m, x, L, R), i) &:= \mathrm{if}(i = m, x, \mathrm{if}(i < m, \mathrm{find}(L, i), \mathrm{find}(R, i))) \\ \mathrm{find}(\emptyset, i) &:= \emptyset \end{aligned}$$

Note that if the binary map contains the empty set, $\emptyset \in \mathrm{range}(M)$, then the find is ambiguous.

A binary map can be represented in a list, $L \in \mathcal{L}(\mathbf{N} \times \mathcal{X} \times \mathbf{N} \times \mathbf{N})$. L is constrained such that $\forall (i, (m, x, p, q)) \in L$ $(p \neq 0 \implies i < p \leq |L|)$, $\forall (i, (m, x, p, q)) \in L$ $(q \neq 0 \implies i < q \leq |L|)$ and $\mathrm{bin}(L) \in \mathcal{B}(\mathcal{X})$. The constructor is defined $\mathrm{bin}(\emptyset) := \emptyset$ and $\mathrm{bin}(L) := (m, x, \mathrm{if}(p \neq 0, \mathrm{bin}(\mathrm{drop}(p - 1, L)), \emptyset), \mathrm{if}(q \neq 0, \mathrm{bin}(\mathrm{drop}(q - 1, L)), \emptyset))$ where $(m, x, p, q) = L_1$.

A binary map can represent a set where there exists an enumeration on the domain. For example, consider order D on some set X, $D \in \mathrm{enums}(X)$, and binary map $B \in \mathcal{B}(X)$ such that $\mathrm{flip}(\mathrm{function}(B)) \subseteq D$, then $\mathrm{function}(B) \in \mathbf{N} \leftrightarrow X$ and $|\mathrm{function}(B)| = |\mathrm{range}(B)|$.

The binary map type, $\mathcal{B}(\mathcal{X})$, as defined above is an algebraic data type representing a function that is constrained such that the domain is a subset of the natural numbers, $\text{domain}(B) \subset \mathbf{N}$ where $B \in \mathcal{B}(\mathcal{X})$. The comparison operator, $(\leq) \in \mathbf{N} \times \mathbf{N} \to \mathbf{B}$, is that of the natural numbers. A generalisation would be to supply a partially ordered set $\mathcal{Y}$ as the superset of the domain and implement the find comparison with the poset relation $\mathcal{Y} \times \mathcal{Y}$ so that $(\leq) \in \mathcal{Y} \times \mathcal{Y} \to \mathbf{B}$. The set of poset binary maps, $\text{mapBinaryPosets}(\mathcal{Y}, \mathcal{X})$, is a type class having a constructor of a poset binary map which supplies the poset relation and the function

$$\text{mapBinaryPoset} \in \text{P}(\mathcal{Y} \times \mathcal{Y}) \times (\mathcal{Y} \to \mathcal{X}) \to \text{mapBinaryPosets}(\mathcal{Y}, \mathcal{X})$$

Given poset relation R, which is such that $\forall (a, b) \in R \ (a \leq b)$ and $\text{dom}(R) \cup \text{ran}(R) = \mathcal{Y}$, and function $F \in \mathcal{Y} \to \mathcal{X}$, let $B = \text{mapBinaryPoset}(R, F) \in \text{mapBinaryPosets}(\mathcal{Y}, \mathcal{X})$. Then $\text{function}(B) = F \in \text{domain}(B) \to \text{range}(B)$. The binary map type, $\mathcal{B}(\mathcal{X})$, is therefore the special case of the poset of natural numbers, $B = \text{mapBinaryPoset}(\{(i, j) : i, j \in \mathbf{N},\ i \leq j\}, F) \in \text{mapBinaryPosets}(\mathbf{N}, \mathcal{X})$ where $F \in \mathbf{N} \to \mathcal{X}$. The logarithmic constraint on the depth holds for poset binary maps just as it does for natural number binary maps. Implementing a binary map with a poset may be representationally convenient if the natural number encoding of the domain is too large, for example in cases of factorial complexity.

B.8 Definition of powerset

The powerset function $\text{P} = \text{powerset} \in \text{P}(\mathcal{X}) \to \text{P}(\text{P}(\mathcal{X}))$ is the set of all subsets of the argument

$$\text{P}(A) := \{X : X \subseteq A\}$$

B.9 Definition of function predicate

The isfunc returns true if the given relation is functional. Let $\mathcal{X}$ be the universal set, then $\text{isfunc} \in \text{P}(\mathcal{X} \times \mathcal{X}) \to \mathbf{B}$

$$\text{isfunc}(A) := \forall (a, b), (p, q) \in A \ (a = p \implies b = q)$$

or equivalently

$$\text{isfunc}(A) := |\{a : (a, b) \in A\}| = |A|$$

Empty relations are defined as functional.

B.10 Definition of cross operators

Define $(\times) \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{Y}) \to \mathrm{P}(\mathcal{X} \times \mathcal{Y})$ as the cartesian cross of sets to create a relation,

$$A \times B := \{(x, y) : x \in A,\ y \in B\}$$

B.11 Definition of mapping operators

Define $(\to) \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{Y}) \to \mathrm{P}(\mathcal{X} \to \mathcal{Y})$ as a powerset of functional relations

$$A \to B := \{X : X \in \mathrm{P}(A \times B),\ \mathrm{isfunc}(X)\}$$

So $A \to B \subseteq \mathrm{P}(A \times B) \in \mathrm{P}(\mathrm{P}(A \times B))$. Thus we can type the function operator as $(\to) \subset (\mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{Y})) \times \mathrm{P}(\mathrm{P}(\mathcal{X} \times \mathcal{Y}))$.

$A \to B$ is sometimes denoted B^A.

And $(\leftrightarrow) \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{Y}) \to \mathrm{P}(\mathcal{X} \leftrightarrow \mathcal{Y})$ is a powerset of bi-directional functional relations

$$A \leftrightarrow B := \{X : X \in \mathrm{P}(A \times B),\ \mathrm{isfunc}(X),\ \mathrm{isfunc}(\mathrm{flip}(X))\}$$

B.12 Total specifiers

Given any relation function $(\otimes) \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{Y}) \to \mathrm{P}(\mathcal{X} \times \mathcal{Y})$, define the left total subsets $(: \otimes) \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{Y}) \to \mathrm{P}(\mathcal{X} \times \mathcal{Y})$ as

$$A : \otimes\ B := \{X : X \in A \otimes B,\ \mathrm{dom}(X) = A\}$$

Similarly for right total $(\otimes :) \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{Y}) \times \mathrm{P}(\mathcal{X} \to \mathcal{Y})$ as

$$A \otimes : B := \{X : X \in A \otimes B,\ \mathrm{ran}(X) = B\}$$

And for both $(: \otimes :) \in \mathrm{P}(\mathcal{X}) \times \mathrm{P}(\mathcal{Y}) \to \mathrm{P}(\mathcal{X} \times \mathcal{Y})$ as

$$A : \otimes : B := \{X : X \in A \otimes B,\ \mathrm{dom}(X) = A,\ \mathrm{ran}(X) = B\}$$

For example

$$A :\to B := \{X : X \in \mathrm{P}(A \times B),\ \mathrm{isfunc}(X),\ \mathrm{dom}(X) = A\}$$

or

$$A :\leftrightarrow: B := \\ \{X : X \in \mathrm{P}(A \times B),\ \mathrm{isfunc}(X),\ \mathrm{isfunc}(\mathrm{flip}(X)),\ \mathrm{dom}(X) = A,\ \mathrm{ran}(X) = B\}$$

B.13 Partitions

The partition function B is the set of all partitions of the argument. A partition is a set of non-empty disjoint subsets, called components, which union to equal the argument, $\mathrm{B} \in \mathrm{P}(\mathcal{X}) \to \mathrm{P}(\mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}))$

$$\mathrm{B}(A) := \\ \{X : X \subseteq (\mathrm{P}(A) \setminus \{\emptyset\}),\ \bigcup X = A,\ (\forall C, D \in X\ (C \neq D \implies C \cap D = \emptyset))\}$$

Define $\mathrm{B}(\emptyset) := \emptyset$.

The weak partition function B' includes component sets that contain the empty set

$$\mathrm{B}'(A) := \mathrm{B}(V) \cup \{Y \cup \{\emptyset\} : Y \in \mathrm{B}(V)\}$$

where $\mathrm{B}'(\emptyset) := \{\{\emptyset\}\}$.

The Bell number function $\mathrm{bell} \in \mathbf{N}_{>0} \to \mathbf{N}_{>0}$ is defined

$$\mathrm{bell}(n) := |\mathrm{B}(\{1 \ldots n\})|$$

The fixed cardinality partition function S is the special case of the partition function in which all partitions have the given cardinality. Define $\mathrm{S} \in \mathrm{P}(\mathcal{X}) \times \mathbf{N}_{>0} \to \mathrm{P}(\mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}))$ as

$$\mathrm{S}(A, k) := \{P : P \in \mathrm{B}(A),\ |P| = k\}$$

The Stirling number of the second kind $\mathrm{stir} \in \mathbf{N}_{>0} \times \mathbf{N} \to \mathbf{N}_{>0}$ is is the cardinality of the fixed cardinality partition function S,

$$\mathrm{stir}(n, k) := |\mathrm{S}(\{1 \ldots n\}, k)|$$

So $\bigcup_{k \in \{1 \ldots n\}} \{\mathrm{S}(A, k) = \mathrm{B}(A)$ and $\sum_{k \in \{1 \ldots n\}} \mathrm{stir}(n, k) = \mathrm{bell}(n)$.

The partition function cardinality function $\mathrm{bellcd} \in \mathbf{N}_{>0} \to (\mathcal{L}(\mathbf{N}) \to \mathbf{N})$ computes the histogram of the histograms of the component cardinalities. The partition function cardinality function is such that $\mathrm{bellcd}(n) \in (\{1 \ldots n\} :\to \{0 \ldots n\}) \to \{0 \ldots n\}$. It is defined

$$\mathrm{bellcd}(n) := \\ \{(L, \frac{n!}{\prod_{(k,m) \in L} (k!)^m m!}) : L \in \prod \{1 \ldots n\} \times \{\{0 \ldots n\}\},\ \sum_{(k,m) \in L} mk = n\}$$

The partition function cardinality function recovers the Bell number, $\sum(c : (\cdot, c) \in \text{bellcd}(n)) = \text{bell}(n)$.

The partition function cardinality function may be constrained such that the partitions have fixed cardinality. The function stircd $\in \mathbf{N}_{>0} \times \mathbf{N}_{>0} \to (\mathcal{L}(\mathbf{N}) \to \mathbf{N})$ computes the special case of the histogram of the histograms of the component cardinalities. The fixed cardinality partition function cardinality function is such that $\text{stircd}(n, i) \in (\{1 \ldots n\} :\to \{0 \ldots n\}) \to \{0 \ldots n\}$. It is defined

$$\begin{aligned}
&\text{stircd}(n, i) := \\
&\quad \{(L, \frac{n!}{\prod_{(k,m) \in L} (k!)^m m!}) : L \in \prod \{1 \ldots n\} \times \{\{0 \ldots n\}\}, \\
&\qquad\qquad \sum_{(k,m) \in L} mk = n,\ \sum_{(\cdot,m) \in L} m = i\}
\end{aligned}$$

or, equivalently in terms of the weak composition function, $\text{C}' \in \text{P}(\mathcal{X}) \times \mathbf{N} \to \text{P}(\mathcal{X} \to \mathbf{N})$,

$$\begin{aligned}
&\text{stircd}(n, i) := \\
&\quad \{(L, \frac{n!}{\prod_{(k,m) \in L} (k!)^m m!}) : L \in \text{C}'(\{1 \ldots n\}, i),\ \sum_{(k,m) \in L} mk = n\}
\end{aligned}$$

The fixed cardinality partition function cardinality function recovers the Stirling number of the second kind, $\sum(c : (\cdot, c) \in \text{stircd}(n, i)) = \text{stir}(n, i)$. The union of the fixed cardinality partition function cardinality function equals the partition function cardinality function, $\bigcup_{i \in \{1 \ldots n\}} \text{stircd}(n, i) = \text{bellcd}(n)$.

The self-partition or uni-partition $S \in \text{B}(A)$ of non-empty $A \in \text{P}(\mathcal{X}) \setminus \{\emptyset\}$ is the special case $S = \{\{a\} : a \in A\}$. The cardinality of the self-partition is $|S| = |A|$. Define self $\in \text{P}(\mathcal{X}) \to \text{P}(\text{P}(\mathcal{X}))$ as $\text{self}(X) := \{\{x\} : x \in X\}$. Define the shorthand $X^{\{\}} = \text{self}(X)$.

The unary-partition $N \in \text{B}(A)$ is the special case $N = \{A\}$. The cardinality of the unary-partition is $|N| = 1$. The unary-partition is the only partition of a singleton set, $\text{B}(\{a\}) = \{\{\{a\}\}\}$. Define unary $\in \text{P}(\mathcal{X}) \to \text{P}(\text{P}(\mathcal{X}))$ as $\text{unary}(X) := \{X\}$.

Binary-partitions $Q \in \text{B}(A)$ are such that $Q \in \{\{X,\ A \setminus X\} : X \in \text{P}(A),\ X \neq \emptyset,\ X \neq A\}$. The cardinality of a binary-partition is $|Q| = 2$.

A partition P is a parent of another partition Q if each component in Q intersects with exactly one component in P. Define parent $\in \mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}) \times \mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}) \to \mathbf{B}$

$$\begin{aligned}&\mathrm{parent}(P,Q) := \\ &\quad (\bigcup P = \bigcup Q) \wedge (\{(D,C) : D \in Q,\ C \in P,\ D \cap C \neq \emptyset\} \in Q \to P)\end{aligned}$$

An equivalent definition is

$$\mathrm{parent}(P,Q) := (\bigcup P = \bigcup Q) \wedge (Q = \bigcup\{R : C \in P,\ R \in \mathrm{B}(C),\ R \subseteq Q\})$$

The set of parents can be constructed explicitly. Define parents $\in \mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}) \to \mathrm{P}(\mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}))$ as

$$\mathrm{parents}(Q) := \{\{\bigcup C : C \in X\} : X \in \mathrm{B}(Q)\}$$

so that $\forall X \subset \mathcal{X}\ \forall P, Q \in \mathrm{B}(X)\ (\mathrm{parent}(P,Q) \iff P \in \mathrm{parents}(Q))$. The cardinality of the set of parents is $|\mathrm{parents}(Q)| = \mathrm{bell}(|Q|) \leq \mathrm{bell}(|\bigcup Q|)$.

A partition of disjoint sets can be exploded into a partition. Define explode $\in \mathrm{P}(\mathrm{P}(\mathrm{P}(\mathcal{X}))) \to \mathrm{P}(\mathrm{P}(\mathcal{X}))$ as $\mathrm{explode}(X) := \{\bigcup C : C \in X\}$. If X is a partition of disjoint sets Q, $X \in \mathrm{B}(Q)$ where $Q \in \mathrm{B}(\bigcup Q)$, then the exploded result is a partition, $\mathrm{explode}(X) \in \mathrm{B}(\bigcup Q)$, and the exploded cardinality is unchanged $|\mathrm{explode}(X)| = |X|$. The set of parents can be defined in terms of explode, $\mathrm{parents}(Q) = \{\mathrm{explode}(X) : X \in \mathrm{B}(Q)\}$.

Define a partition sequence as a list $L \in \mathcal{L}(\mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}))$ of partitions such that each element in the list is a parent of the next

$$|L| \geq 2 \implies (\forall i \in \{1 \ldots |L|-1\}\ (\mathrm{parent}(L_i, L_{i+1})))$$

Similarly define a reverse partition sequence as a list $L \in \mathcal{L}(\mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}))$ of partitions such that each element in the list is a child of the next

$$|L| \geq 2 \implies (\forall i \in \{1 \ldots |L|-1\}\ (\mathrm{parent}(L_{i+1}, L_i)))$$

Define a partition tree as a tree $T \in \mathrm{trees}(\mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}))$ of partitions such that each node in the tree is a parent-child

$$\forall (P,Q) \in \mathrm{nodes}(T)\ (Q \neq \emptyset \implies \mathrm{parent}(P,Q))$$

and define a reverse partition tree as a tree $T \in \mathrm{trees}(\mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}))$ of partitions such that each node in the tree is a child-parent

$$\forall (P,Q) \in \mathrm{nodes}(T)\ (Q \neq \emptyset \implies \mathrm{parent}(Q,P))$$

A partition P can be viewed as a disjoint set of equivalent classes. A weighting can be assigned to each element of the partition, $x \in \bigcup P$, equal to its fraction of the cardinality of the component containing it, $1/|C|$, where $x \in C \in P$. Define weights $\in \mathrm{P}(\mathrm{P}(\mathcal{X}) \setminus \{\emptyset\}) \to (\mathcal{X} \to \mathbf{Q}_{>0})$ as

$$\mathrm{weights}(P) := \{(x, 1/|C|) : C \in P,\ x \in C\}$$

which is defined if P is a partition, $P \in \mathrm{B}(\bigcup P)$. The weights are such that $\mathrm{sum}(\mathrm{weights}(P)) = |P| \leq |\bigcup P|$.

B.14 Compositions

A composition is a functional relation between some set of objects and the natural numbers $\mathcal{X} \to \mathbf{N}$. Given a set of objects $X \subset \mathcal{X}$ and a total $n \in \mathbf{N}$, composition functions return all compositions such that each has domain X and sums to n. Define the composition function, which excludes 0, as

$$\mathrm{C} \in \mathrm{P}(\mathcal{X}) \times \mathbf{N}_{>0} \to \mathrm{P}(\mathcal{X} \to \mathbf{N}_{>0})$$

and the weak composition function, which includes 0, as

$$\mathrm{C}' \in \mathrm{P}(\mathcal{X}) \times \mathbf{N} \to \mathrm{P}(\mathcal{X} \to \mathbf{N})$$

such that

$$\forall C \in \mathrm{C}(X, n) \cup \mathrm{C}'(X, n)\ ((\mathrm{dom}(C) = X) \wedge (\sum_{x \in X} C_x = n))$$

By implication, composition function $\mathrm{C}(X, n)$ is constrained

$$\forall C \in \mathrm{C}(X, n)\ (n \geq |X|)$$

The cardinality of the composition function $\mathrm{C}(\{1 \ldots k\}, n)$ is

$$|\mathrm{C}(\{1 \ldots k\}, n)| = \frac{(n-1)!}{(k-1)!\ (n-k)!}$$

The cardinality of the weak composition function $\mathrm{C}'(\{1 \ldots k\}, n)$ is

$$|\mathrm{C}'(\{1 \ldots k\}, n)| = \frac{(n+k-1)!}{(k-1)!\ n!}$$

Note that composition, as defined here, is not to be confused with composition of functions.

B.15 Relation functions

Let $\mathcal{X}$ and $\mathcal{Y}$ be universal sets, then the domain is defined $\text{dom} \in \text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{X})$

$$\text{dom}(A) := \{x : (x, y) \in A\}$$

and the range, $\text{ran} \in \text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{Y})$

$$\text{ran}(A) := \{y : (x, y) \in A\}$$

The inverse, $\text{flip} \in \text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{Y} \times \mathcal{X})$

$$\text{flip}(A) := \{(y, x) : (x, y) \in A\}$$

The function $\text{filter} \in \text{P}(\mathcal{X}) \times \text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{X} \times \mathcal{Y})$ is defined

$$\text{filter}(X, A) := \{(x, y) : (x, y) \in A,\ x \in X\}$$

The inverse of a function, $\text{inverse} \in (\mathcal{X} \to \mathcal{Y}) \to (\mathcal{Y} \to \text{P}(\mathcal{X}))$ is defined

$$\begin{aligned} \text{inverse}(F) &:= \{(y, \text{ran}(\text{filter}(\{y\}, \text{flip}(F)))) : y \in \text{ran}(F)\} \\ &= \{(y, \{x : x \in \text{dom}(F),\ F(x) = y\}) : y \in \text{ran}(F)\} \end{aligned}$$

The inverse is sometimes denoted F^{-1}. Note that this is a different definition from the convention $\{(Y, \{x : x \in \text{dom}(F),\ F(x) \in Y\}) : Y \subseteq \text{ran}(F)\} \in \text{P}(\mathcal{Y}) \to \text{P}(\mathcal{X})$.

B.16 Dot operator

The set of all bidirectional mappings between two sets of the same cardinality $(\cdot) \in \text{P}(\mathcal{X}) \times \text{P}(\mathcal{Y}) \to \text{P}(\mathcal{X} \leftrightarrow \mathcal{Y})$

$$X \cdot Y := \{Z : Z \in X \leftrightarrow Y,\ |X| = |Y|,\ |Z| = |Y|\}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are the universal set.

The outer dot product is defined where the cardinality of the left argument is greater than or equal to that of the right, $(\cdot =) \in \text{P}(\mathcal{X}) \times \text{P}(\mathcal{Y}) \to \text{P}(\mathcal{X} \leftrightarrow \mathcal{Y})$

$$X \cdot = Y := \{Z : Z \in X \leftrightarrow Y,\ |X| \geq |Y|,\ |Z| = |Y|\}$$

B.17 Selections

The set of subsets of given cardinality of a given set, $\text{selections} \in \text{N} \times \text{P}(\mathcal{X}) \to \text{P}(\text{P}(\mathcal{X}))$ is defined

$$\text{selections}(k, X) := \{Y : Y \subseteq X,\ |Y| = k\}$$

There are $n!/(k!(n-k)!)$ of these where $n = |X|$.

B.18 Enumerations

The set of all enumerations enums $\in \mathrm{P}(\mathcal{X}) \to \mathrm{P}(\mathcal{X} \leftrightarrow \mathbf{N})$

$$\mathrm{enums}(X) := X \cdot \{1 \ldots |X|\}$$

The cardinality of this function is the number of permutations of its argument $|\mathrm{enums}(X)| = |X|!$.

An order D on some set X is a choice of the enumerations, $D \in \mathrm{enums}(X)$. Given the order, any subset can be *enumerated*, define order $\in (\mathcal{X} \leftrightarrow \mathbf{N}) \times \mathrm{P}(\mathcal{X}) \to (\mathcal{X} \leftrightarrow \mathbf{N})$

$$\mathrm{order}(D, Y) := \{(y, |\{(z, i) : (z, i) \in Q,\ i \leq j\}|) : (y, j) \in Q\}$$

where $Q = \{(y, D_y) : y \in Y\}$. So $\mathrm{order}(D, Y) \in \mathrm{enums}(Y)$.

B.19 Normalisation

Normalising a real valued function $F \in \mathcal{X} \to \mathbf{R}$ is defined normalise $\in (\mathcal{X} \to \mathbf{R}) \to (\mathcal{X} \to \mathbf{R})$ as

$$\mathrm{normalise}(F) := \{(x, r/\mathrm{sum}(F)) : (x, r) \in F\} \in \mathcal{X} \to \mathbf{R}$$

Normalising is undefined if $\mathrm{sum}(F) = 0$. Define notation $\hat{F} := \mathrm{normalise}(F)$. If the function is rational valued, $F \in \mathcal{X} \to \mathbf{Q}$, then the normalised function is also rational valued, $\hat{F} \in \mathcal{X} \to \mathbf{Q}$, and hence a *probability function*, $\hat{F} \in \mathcal{P}$.

B.20 Aggregation and inclusion functions

The function singleton $\in \mathcal{X} \to \mathrm{P}(\mathcal{X})$, sometimes called tip, creates a singleton set, $\mathrm{singleton}(x) := \{x\}$. The converse function only $\in \mathrm{P}(\mathcal{X}) \to \mathcal{X}$ is defined $\mathrm{only}(X) := x$ where $|X| = 1$ and $X = \{x\}$. The function only is undefined if $|X| \neq 1$.

Given a relation, the count function creates a functional relation between the domain of the argument and the count of the elements corresponding. Let $\mathcal{X}$ and $\mathcal{Y}$ be the universal set, count $\in \mathrm{P}(\mathcal{X} \times \mathcal{Y}) \to (\mathcal{X} \to \mathbf{N})$

$$\mathrm{count}(A) := \{(a, |\{q : (p, q) \in A, p = a\}|) : a \in \mathrm{dom}(A)\}$$

Given a relation $\mathcal{X} \times \mathcal{Y}$ such that an order operator, $\mathrm{enums}(\mathcal{Y})$, or partially ordered set, $(\mathcal{Y}, \leq)$, is defined on the range, $\mathcal{Y}$, the min and max functions returns the minimum/maximum subset, max $\in \mathrm{P}(\mathcal{X} \times \mathcal{Y}) \to (\mathcal{X} \to \mathcal{Y})$

$$\max(A) := \{(x, y) : (x, y) \in A, (\forall (r, s) \in A\ (s \leq y))\}$$

We also define the convenience functions

$$\text{maxd}(A) := \text{dom}(\max(A))$$

and

$$\text{maxr}(A) := m$$

where $\{m\} = \text{ran}(\max(A))$. Note that maxr is undefined for empty sets.

For minimum, $\min \in \text{P}((\mathcal{X} \times \mathcal{Y})) \to (\mathcal{X} \to \mathcal{Y})$

$$\min(A) := \{(x,y) : (x,y) \in A, (\forall (r,s) \in A\ (s \geq y))\}$$

We also define $\text{mind}(A)$ and $\text{minr}(A)$ similarly.

Similar to the min and max functions are the $\text{bottom}(n)$ and $\text{top}(n)$ functions. Define $\text{top} \in \mathbf{N}_{>0} \to (\text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{X} \times \mathcal{Y}))$

$$\begin{aligned} \text{top}(n)(A) &:= \text{top}(\text{maximum}(n - |\max(A)|, 0))(A \setminus \max(A)) \cup \max(A) \\ \text{top}(0)(A) &:= \emptyset \end{aligned}$$

Define $\text{bottom} \in \mathbf{N}_{>0} \to (\text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{X} \times \mathcal{Y}))$

$$\begin{aligned} \text{bottom}(n)(A) &:= \text{bottom}(\text{maximum}(n - |\min(A)|, 0))(A \setminus \min(A)) \cup \min(A) \\ \text{bottom}(0)(A) &:= \emptyset \end{aligned}$$

Thus $\text{top}(1) = \max$, $\text{bottom}(1) = \min$, and $\text{top}(|A|)(A) = \text{bottom}(|A|)(A) = A$. Define $\text{topd}(n)(A) := \text{dom}(\text{top}(n)(A))$ and similarly $\text{bottomd}(n)(A) := \text{dom}(\text{bottom}(n)(A))$.

Given a zero element in the range, $0 \in \mathcal{Y}$, define inclusion function $\text{zero} \in \text{P}(\mathcal{X} \times \mathcal{Y}) \to (\mathcal{X} \to \{0\})$

$$\text{zero}(A) := \{(x,y) : (x,y) \in A,\ y = 0\}$$

Define inclusion function $\text{nonzero} \in \text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{X} \times (\mathcal{Y} \setminus \{0\}))$

$$\text{nonzero}(A) := \{(x,y) : (x,y) \in A,\ y \neq 0\}$$

Define inclusion function $\text{positive} \in \text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{X} \times \mathcal{Y}_{\geq 0})$

$$\text{positive}(A) := \{(x,y) : (x,y) \in A,\ y \geq 0\}$$

Define inclusion function $\text{negative} \in \text{P}(\mathcal{X} \times \mathcal{Y}) \to \text{P}(\mathcal{X} \times \mathcal{Y}_{<0})$

$$\text{negative}(A) := \{(x,y) : (x,y) \in A,\ y < 0\}$$

Define aggregation function sum $\in \mathrm{P}(\mathcal{X} \times \mathbf{R}) \to \mathbf{R}$

$$\mathrm{sum}(A) := \sum q : (x, q) \in A$$

Define arithmetic average as average $\in \mathrm{P}(\mathcal{X} \times \mathbf{R}) \to \mathbf{R}$

$$\mathrm{average}(A) := \mathrm{sum}(A)/|A|$$

which is defined where $|A| \geq 0$.

Define product $\in \mathrm{P}(\mathcal{X} \times \mathbf{R}) \to \mathbf{R}$

$$\mathrm{product}(A) := \prod q : (x, q) \in A$$

B.21 Convenience functions

Ceiling of positive reals ceil $\in \mathbf{R} \to \mathbf{N}$

$$\mathrm{ceil}(r) := \mathrm{minr}(\{(n, n) : n \in \mathbf{N}, n \geq r\})$$

Floor of positive reals flr $\in \mathbf{R} \to \mathbf{N}$

$$\mathrm{flr}(r) := \mathrm{maxr}(\{(n, n) : n \in \mathbf{N}, n \leq r\})$$

Greater of a pair maximum $\in \mathbf{R} \times \mathbf{R} \to \mathbf{R}$

$$\mathrm{maximum}(a, b) := \mathrm{if}(a < b, b, a)$$

Lesser of a pair minimum $\in \mathbf{R} \times \mathbf{R} \to \mathbf{R}$

$$\mathrm{minimum}(a, b) := \mathrm{if}(a < b, a, b)$$

B.22 Big O definition

The Big O function imposes an upper bound. It requires both a functional map to a real and a real multiplier, $\mathrm{O} \in (\mathcal{X} \to \mathbf{R}) \times \mathbf{R}_{>0} \to \mathrm{P}(\mathcal{X} \to \mathbf{R}_{\geq 0})$

$$\begin{aligned}&\mathrm{O}(A, m) := \\ &\quad \{X : X \in \mathrm{P}(\{(x, r) : x \in \mathrm{dom}(A),\ r \in \mathbf{R},\ 0 \leq r \leq mA_x\}),\ \mathrm{isfunc}(X)\}\end{aligned}$$

Define $\mathrm{O} \in (\mathcal{X} \to \mathbf{R}) \to \mathrm{P}(\mathcal{X} \to \mathbf{R})$

$$\mathrm{O}(A) := \mathrm{O}(A, 1)$$

Similarly Big Omega function imposes an lower bound. It requires a functional map to a real and a real multiplier, $\Omega \in (\mathcal{X} \to \mathbf{R}) \times \mathbf{R}_{>0} \to \mathrm{P}(\mathcal{X} \to \mathbf{R})$

$$\Omega(A, m) := \\ \{X : X \in \mathrm{P}(\{(x, r) : x \in \mathrm{dom}(A),\ r \in \mathbf{R},\ r \geq mA_x\}),\ \mathrm{isfunc}(X)\}$$

Define $\Omega \in (\mathcal{X} \to \mathbf{R}) \to \mathrm{P}(\mathcal{X} \to \mathbf{R})$

$$\Omega(A) := \Omega(A, 1)$$

B.23 Let quantifier

A lozenge $\Diamond$ is used to signify the binding of a variable amongst quantifiers. For example $\forall a \in A\ \Diamond b = \mathrm{f}(a)\ \exists c \in C(b)\ (\ldots)$.

B.24 Set builder notation

This expression

$$Z = \{\mathrm{z}(x) : x \in X,\ \mathrm{p}(x)\}$$

where $\mathrm{p}(x)$ is a predicate, is shorthand for

$$\forall x \in X\ (\mathrm{p}(x) \iff \mathrm{z}(x) \in Z)$$

And

$$Z = \{\mathrm{z}(x, y) : x \in X,\ y \in Y,\ \mathrm{p}(x),\ \mathrm{q}(y)\}$$

is shorthand for

$$\forall x \in X\ \forall y \in Y\ (\mathrm{p}(x) \wedge \mathrm{q}(y) \iff \mathrm{z}(x, y) \in Z)$$

and so on.

B.25 if function

Let $\mathbf{B}$ be the set of booleans. Let $\mathcal{X}$ be the universal set. The logical switch function, $\mathrm{if} \in \mathbf{B} \times \mathcal{X} \times \mathcal{X} \to \mathcal{X}$

$$\forall b \in \mathbf{B}\ \forall x, y \in \mathcal{X}\ ((b \implies \mathrm{if}(b, x, y) = x) \wedge (\neg b \implies \mathrm{if}(b, x, y) = y))$$

B.26 Function definition

Let func be a functional relation with a type definition $\text{func} \in \mathcal{A} \to \mathcal{Z}$. Consider the expression

$$\text{func}(a) := \text{z}(a)$$

where $\text{z}(a)$ is an expression with free variable a that evaluates to an element in $\mathcal{Z}$. In the case where func is a total function, that is, where there are no other constraints imposed on the domain of the function $\text{dom}(\text{func}) = \mathcal{A}$, then the expression is shorthand for

$$\forall a \in \mathcal{A}\ \exists \text{z}(a) \in \mathcal{Z}\ ((a, \text{z}(a)) \in \text{func})$$

In the case that func is a partial function, that is, where there are other constraints $\text{r}(\text{func}) \in \mathbf{B}$ on the function, then the domain is taken (ambiguously) to be one of the largest possible subsets

$$\text{func} \in \text{maxd}(\{(F, |\text{dom}(F)|) : F \in \mathcal{A} \to \mathcal{Z},\ \text{r}(F),\ (\forall (a, b) \in F\ (b = z(a)))\})$$

Similarly, with the same caveat for constraints on the domain, if the type definition is $\text{func} \in \mathcal{A} \times \mathcal{B} \to \mathcal{Z}$ then the expression

$$\text{func}(a, b) := \text{z}(a, b)$$

is shorthand for

$$\forall a \in \mathcal{A}\ \forall b \in \mathcal{B}\ \exists \text{z}(a, b) \in \mathcal{Z}\ (((a, b), \text{z}(a, b)) \in \text{func})$$

and so on.

The identity function $\text{id} \in \mathcal{X} \to \mathcal{X}$ is defined

$$\text{id}(x) := x$$

B.27 Natural numbers

The set of natural numbers $\mathbf{N}$ is taken to include 0. The set $\mathbf{N}_{>0}$ excludes 0.

Define $\text{encode} \in \mathbf{N} \leftrightarrow \mathcal{L}(\text{bits})$ which encodes a natural number in the shortest list such that

$$\forall (i, L) \in \text{encode}\ ((L = \emptyset \vee \text{last}(L) = 1) \wedge (i = \sum (2^{j-1} b : (j, b) \in L)))$$

Define $\text{decode} \in \mathcal{L}(\text{bits}) \leftrightarrow \mathbf{N}$ as $\text{decode} = \text{flip}(\text{encode})$.

Define $\text{space} \in \mathbf{N}_{>0} \to \ln \mathbf{N}_{>0}$ as $\text{space}(n) := \ln n$. The length of the encoded natural number is an approximation to the *space*

$$(|\text{encode}(n)| - 1) \ln 2 \leq \text{space}(n) < (|\text{encode}(n)|) \ln 2$$

B.28 Booleans and bits

The set of boolean values $\mathbf{B}$ is not specified except to constrain the cardinality $|\mathbf{B}| = 2$.

The set of *bits* is defined $\text{bits} = \{0, 1\} \subset \mathbf{N}$. Define $\text{encode} \in \mathbf{B} \to \text{bits}$ as $\text{encode}(b) := \text{if}(b, 1, 0)$, and define $\text{decode} \in \text{bits} \to \mathbf{B}$ as $\text{decode}(i) := i = 1$.

201803070530.PAP1