Why

We want to approximate a given distribution with one which factors according to a tree.

Definition

Given $q : A \to [0, 1]$, we want to find a distribution p on A and tree T on $\{1, \dots, n\}$ to

$$\begin{aligned} \text{minimize} \quad & d_{kl}(q, p) \\ \text{subject to} \quad & p \text{ factors according to } T. \end{aligned}$$

where d_{kl} is the relative entropy as a criterion of approximation. We call such a distribution a *tree distribution approximator* (or *tree approximator*) and we call the tree the *approximator tree.*

Result

Prop. 1. *Let $A_1, \dots, A_n$ be finite non-empty sets. Define $A = \prod_{i=1}^n A_i$. Let $q : A \to [0, 1]$ a distribution and T a tree on $\{1, \dots, n\}$. The distribution $p^*_T : A \to [0, 1]$ defined by*

$$p^*_T = q_1 \prod_{i \neq 1} q_{i|\mathrm{pa}\, i}$$

minimizes the relative entropy with q among all distributions on A which factor according to T.

Proof. Let $p : A \to [0, 1]$ be a distribution which factors according to T. First, express

$$p = p_1 \prod_{i \neq i} p_{i|\mathrm{pa}\, i}$$

where $\mathrm{pa}\, i$ is the parent of vertex i in T rooted at vertex 1 ($i = 2, \dots, n$).

Second, recall that the relative entropy of q with p is $H(q,p) - H(q)$. Since $H(q)$ does not depend on p, p is a minimizer of the relative of q with p if and only if p is a minimizer of $H(q,p)$.

Third, express

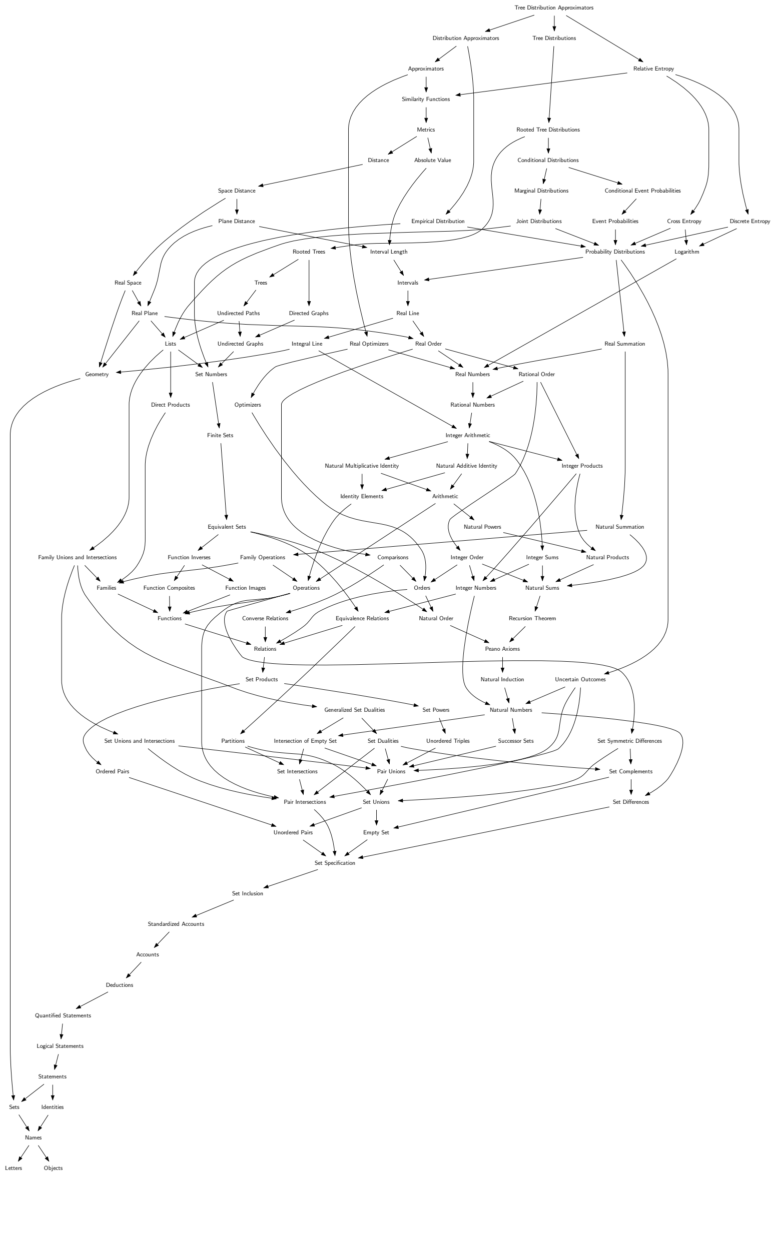
$$
\begin{aligned}
H(q,p) &= -\sum_{a \in A} q(a) \log p(a) \\
&= -\sum_{a \in A} q(a) \left(\log p_1(a_1) + \sum_{i \neq 1} \log p_{i|\mathrm{pa}\, i}(a_i, a_{\mathrm{pa}\, i}) \right) \\
&= H(q_1, p_1) + \sum_{i \neq 1} \sum_{a_{\mathrm{pa}\, i} \in A_{\mathrm{pa}\, i}} q_{\mathrm{pa}\, i}(a_{\mathrm{pa}\, i}) H\big(q_{i|\mathrm{pa}\, i}(\cdot, a_{\mathrm{pa}\, i}), p_{i|\mathrm{pa}\, i}(\cdot, a_{\mathrm{pa}\, i})\big)
\end{aligned}
$$

which separates across p_1 an $p_{i|\mathrm{pa}\, i}(\cdot, a_{\mathrm{pa}\, i})$ for $i = 2, \dots, n$ and $a_{pa_i} \in A_{\mathrm{pa}\, i}$.

Fourth, recall $H(\cdot, \cdot) \geq 0$ and is zero on repeated pairs. By this, we mean, for example, $H(p_1, p_1) = 0$. So $p_1 = q_1$ and $p_{i|\mathrm{pa}\, i} = q_{i|\mathrm{pa}\, i}$ are solutions.

$\square$

Proposition 1 states the form of an optimal approximator given a tree. A natural next question is to select the tree.


Tree Distribution Approximators
Distribution Approximators
Tree Distributions
Approximators
Relative Entropy
Similarity Functions
Metrics
Rooted Tree Distributions
Distance
Absolute Value
Conditional Distributions
Space Distance
Marginal Distributions
Conditional Event Probabilities
Plane Distance
Empirical Distribution
Joint Distributions
Event Probabilities
Cross Entropy
Discrete Entropy
Rooted Trees
Interval Length
Probability Distributions
Logarithm
Real Space
Trees
Intervals
Real Plane
Undirected Paths
Directed Graphs
Real Line
Lists
Undirected Graphs
Integral Line
Real Optimizers
Real Order
Real Summation
Geometry
Set Numbers
Real Numbers
Rational Order
Direct Products
Optimizers
Rational Numbers
Finite Sets
Integer Arithmetic
Natural Multiplicative Identity
Natural Additive Identity
Integer Products
Identity Elements
Arithmetic
Equivalent Sets
Natural Powers
Natural Summation
Family Unions and Intersections
Function Inverses
Family Operations
Comparisons
Integer Order
Integer Sums
Natural Products
Families
Function Composites
Function Images
Operations
Orders
Integer Numbers
Natural Sums
Functions
Converse Relations
Equivalence Relations
Natural Order
Recursion Theorem
Relations
Peano Axioms
Set Products
Natural Induction
Uncertain Outcomes
Generalized Set Dualities
Set Powers
Natural Numbers
Set Unions and Intersections
Partitions
Intersection of Empty Set
Set Dualities
Unordered Triples
Successor Sets
Set Symmetric Differences
Ordered Pairs
Set Intersections
Pair Unions
Set Complements
Pair Intersections
Set Unions
Set Differences
Unordered Pairs
Empty Set
Set Specification
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Sets
Identities
Names
Letters
Objects