Why

We approximate a distribution with a distribution that factors according to a given tree. Such a distribution requires tabulating fewer numbers to express the probability of outcome.

Definition

We will use the relative entropy as a criterion of approximation. Given a distribution over a product of finite sets and a tree, we want to find the optimal approximator among distributions which factor according to the tree. We call such a distribution a *tree distribution approximator* or *tree approximator* of the given distribution for the tree. We call the tree the *approximator tree*.

Result

Prop. 1. *Let $A_1, \ldots, A_n$ 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 $\{1, \ldots, n\}$. 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 ac-

cording 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 $(i = 2, \ldots, n)$.

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

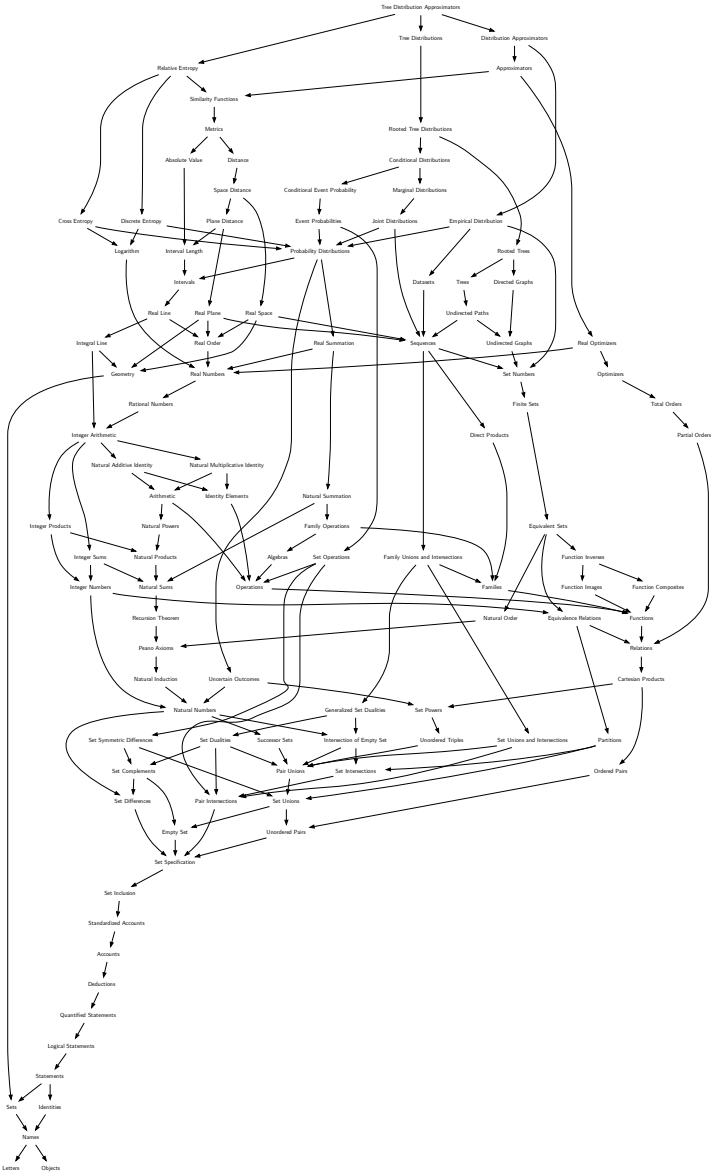
$$\begin{aligned} H(q,p) &= -\sum_{a \in A} q(a) \log p(a) \\ &= -\sum_{a \in A} q(a) \left(\log p_1(a_1) + \sum_{i \neq 1} \log p_{i|\mathrm{pa}\, i}(a_i, a_{\mathrm{pa}\, i}) \right) \\ &= H(q_1, p_1) + \sum_{i \neq 1} \sum_{a_{\mathrm{pa}\, i} \in A_{\mathrm{pa}\, i}} q_{\mathrm{pa}\, i}(a_{\mathrm{pa}\,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

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

□

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


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