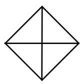


Why

Which is the optimal tree to use for tree distribution approximation?

Definition

We want to choose a tree whose corresponding approximator for the given distribution achieves minimum relative entropy with the given distribution among all tree distribution approximators. We call such a distribution an *optimal* tree approximator of the given distribution. We call a tree according to which an optimal tree approximator factors and *optimal* 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. A tree T on $\{1, \dots, n\}$ is an optimal approximator tree if and only if it is a maximal spanning tree of the mutual information graph of q.*

Proof. First, denote the optimal tree distribution approximator of q for tree T by p_T^*. Express

$$p_T^* = q_1 \prod_{i \neq 1} q_{i|\textsf{pa}\, i}$$

Second, express $d(q, p) = H(q, p) - H(q)$. Since $H(q)$ does not depend on T, p_T^* is a minimizer (w.r.t. T) of $d(q, p_T^*)$ if and only if it is a minimizer of $H(q, p_T^*)$.

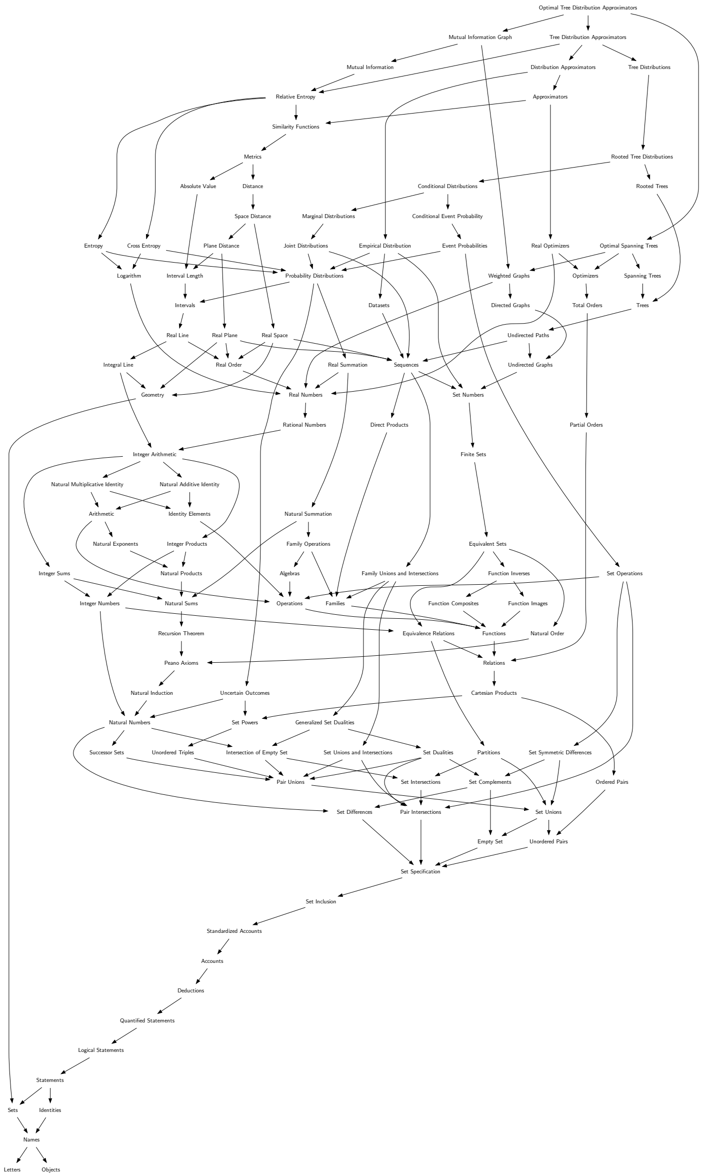
Third, express the cross entropy of p_T^* relative to q as

$$
\begin{aligned}
H(q, p_T^*) &= H(q_1) - \sum_{i \neq 1} \sum_{a \in A} q(a) \log q_{i|pai}(a_i, a_{\textsf{\textbf{pa}}\, i}) \\
&= H(q_1) - \sum_{i \neq 1} \sum_{a \in A} q(a)(\log q_{i, \textsf{\textbf{pa}}\, i}(a_i, a_{\textsf{\textbf{pa}}\, i}) - \log q_{\textsf{\textbf{pa}}\, i}(a_{\textsf{\textbf{pa}}\, i})) \\
&= H(q_1) - \sum_{i \neq 1} \sum_{a \in A} q(a)(\log q_{i, \textsf{\textbf{pa}}\, i}(a_i, a_{\textsf{\textbf{pa}}\, i}) - \log q_{\textsf{\textbf{pa}}\, i}(a_{\textsf{\textbf{pa}}\, i}) - \log q \\
&= \sum_{i=1}^{n} H(q_i) - \sum_{i \neq 1} I(q_i, q_{\textsf{\textbf{pa}}\, i}) \\
&= \sum_{i=1}^{n} H(q_i) - \sum_{\{i,j\} \in T} I(q_i, q_j)
\end{aligned}
$$

where $\textsf{\textbf{pa}}\, i$ denotes the parent of vertex i in T rooted at vertex 1 ($i = 2, \ldots, n$). For $i = 1, \ldots, n$, $H(q_i)$ does not depend on the choice of tree. Therefore selecting a tree to minimize the second term in the final expression above is equivalent to choosing a maximal spanning tree from the weighted graph with mutual information edge weights; namely, the mutual information graph of q.

$\square$

Proposition 1 says that to we should first select a maximum spanning tree of the mutual information graph of the distribution we are approximating. Then, we should pick the best approximator to q which factors according to that three.


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