Why

What is the optimal tree approximator of a multivariate normal density?

Result

Proposition 1. *Let $g : \mathbf{R}^n \to \mathbf{R}$ be a normal density with mean $\mu \in \mathbf{R}^d$ and covariance $\Sigma \in \mathbf{S}^d_{++}$. The normal density $f^*_T : \mathbf{R}^d \to \mathbf{R}$ with mean μ and precision matrix P defined by*

- $P_{11} = \Sigma_{11}^{-1} + \sum_{\mathrm{pa}\, j=1} \Sigma_{j1}^2 \Sigma_{11}^{-2} \Sigma_{j|1}^{-1}$
- *for* $i = 2, \ldots, d$, $P_{ii} = \Sigma_{i|\mathrm{pa}\, i}^{-1} + \sum_{\mathrm{pa}\, j=i} \Sigma_{ji}^2 \Sigma_{ii}^{-2} \Sigma_{j|i}^{-1}$
- $i, j = 1, \ldots d$ *and* $i = \mathrm{pa}\, j$, $P_{ij} = P_{ji} = -\Sigma_{ji} \Sigma_{jj}^{-1} \Sigma_{j|i}^{-1}$

where $\mathrm{pa}\, i$ *is the parent of* i *in an optimal approximator tree* T *(*$i = 2, \ldots, n$*) is an optimal tree approximator of* g*.*

Proof. Using Proposition 1 of Best Tree Density Approximators, express an optimal tree approximator of g by

$$(1/c) \exp\left(-\frac{1}{2}\left(\Sigma_{11}^{-1} \bar{x}_1^2 + \sum_{i \neq 1} \left(\bar{x}_i - \Sigma_{i,\mathrm{pa}\, i} \Sigma_{\mathrm{pa}\, i,\mathrm{pa}\, i}^{-1} \bar{x}_{\mathrm{pa}\, i}\right)^2 \Sigma_{i|\mathrm{pa}\, i}^{-1}\right)\right)$$

where $\bar{x}_i = x_i - \mu_i$ and $c = \sqrt{(2\pi)^d \Sigma_{11} \prod_{i \neq 1} \Sigma_{i|\mathrm{pa}\, i}}$.

Second, express the quadratic in the exponential as

$$\Sigma_{11}^{-1} \bar{x}_1^2 + \sum_{i \neq 1} \left[\Sigma_{i|\mathrm{pa}\, i}^{-1} \bar{x}_i^2 - 2\Sigma_{i,\mathrm{pa}\, i} \Sigma_{\mathrm{pa}\, i,\mathrm{pa}\, i}^{-1} \Sigma_{i|\mathrm{pa}\, i}^{-1} \bar{x}_i \bar{x}_{\mathrm{pa}\, i} + \Sigma_{i,\mathrm{pa}\, i}^2 \Sigma_{\mathrm{pa}\, i,\mathrm{pa}\, i}^{-2} \Sigma_{i|\mathrm{pa}\, i}^{-1} \bar{x}_{\mathrm{pa}\, i}^2\right]$$

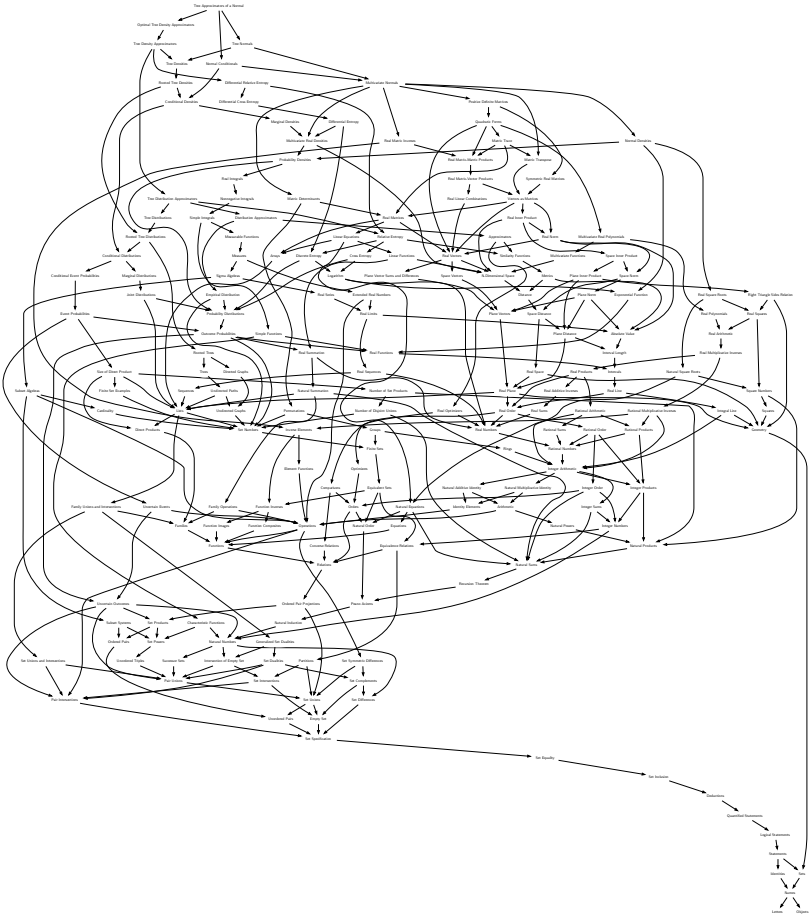
With P defined as earlier, we can express the above as $\bar{x}^\top P \bar{x}$.

Third, note that c is $\sqrt{(2\pi)^d \det P^{-1}}$ since f^*_T is a density and so integrates to one. □

Notice that f^*_T is a tree normal density.

Empirical normal

In particular, notice that we can approximate the empirical normal density of a dataset with a density that factors according to a tree.


Tree Approximants of a Normal
Optimal Tree Density Approximants
Tree Density Approximants
Tree Normals
Normal Coordinate
Tree Kernels
Normal Tree Density
Conditional Entropy
Differential Relative Entropy
Differential Cross Entropy
Multivariate Normals
Positive Definite Matrices
Marginal Densities
Marginal Entropy
Multivariate Beta Matrices
Real Matrix Inverse
Wishart Density
Real Matrix Determinant
Matrix Cofactor
Real Matrix Inverse Function
Symmetric Real Matrices
Real Cross Correlations
Real Matrix Determinant
Real Integrals
Stochastic Approximation
Tree Distribution Approximants
Stochastic Approximation
Tree Distribution
Random Tree Distributions
Conditional Entropy
Mutual Information
Conditional Distribution
Marginal Distributions
Conditional Error Probabilities
Error Probabilities
Marginal Distribution
Probability Measures
Measure Probabilities
Measure Functions
Measure Sets
Measure Unit
Measure Unit
Measure Unit