Tree Density Approximators

1 Why

Can we approximate a density by a tree density similar to how we approximated a distribution with by a tree distribution.

2 Definition

We will use the differential relative entropy as a criterion of approximation. Given a density of R^n and a tree, we want to find the optimal approximator among densities which factor according to a tree. We call such a density an *approximator* of the given density for the tree We call such a density an *approximator of* the given density *for* the given tree.

3 Result

Proposition 1. *Let $g : \mathsf{R}^n \to \mathsf{R}$ be a density and T be a tree on $\{1, \dots, n\}$. The density $f_T^* : \mathsf{R}^d \to \mathsf{R}$ defined by*

$$f_T^* = g_1 \prod_{i \neq 1} g_{i|pa_i}$$

minimizes the differential relative entropy with q among all densities on R^n which factor according to T (pa_i is the parent of i in T, $i = 2, \dots, n$).

Proof. Let $f : \mathsf{R}^d \to \mathsf{R}$ be a density factoring according to T. First, express

$$f = f_1 \prod_{i=1} f_{i|\mathrm{pa}_i}$$

Second, recall that $d(g, f) = h(g, f) - h(g)$. Since $h(g)$ does not depend on f, f is a minimizer of $d(g, f)$ if and only if f is a minimizer of $h(g, f)$.

Third, express

$$
\begin{aligned}
h(g, f) &= -\int_{\mathsf{R}^d} g \log f \\
&= -\int_{\mathsf{R}^d} g(x)\left(\log f_i(x_i) + \sum_{i \neq 1} \log fi \mid \mathrm{pa}_i(x_i, x_{\mathrm{pa}_i})\right) dx \\
&= h(g_1, f_1) + \sum_{i \neq 1}\left(\int_{\mathsf{R}} g_{\mathrm{pa}_i}(\xi) h\big(g_{i|\mathrm{pa}_i}(\cdot, \xi), f_{i|\mathrm{pa}_i}(\cdot, \xi)\big) d\xi\right)
\end{aligned}
$$

which separates across f_1 an $f_{i|\mathrm{pa}_i}(\cdot, \xi)$ for $i = 1, \ldots, n$ and $\xi \in \mathsf{R}$. In particular, since $g_{pai} \geq 0$, we can minimize the integrand pointwise.

Fourth, recall $h(\cdot, \cdot) \geq 0$ and is zero on repeated pairs. So $f_1 = g_1$ and $f_{i|\mathrm{pa}_i} = g_{i|\mathrm{pa}_i}$ are solutions.

$\square$