

# Tree Density Approximators

## 1 Why

We can approximate a density with a tree density similar to how we can approximate a distribution with a tree distribution.

### 2 Definition

We use the differential relative entropy as a criterion of approximation. An *optimal tree approximator* of density for a tree is a density which factors according to a tree and minimizes its differential relative entropy with the given density.

#### 3 Notation

Let  $g: \mathbb{R}^n \to \mathbb{R}$  be a density and T be a tree on  $\{1, \ldots, n\}$ . An optimal tree approximator of g for T is a density f that factors according to T and minimizes d(g, f). In other words, given g and T we want to find f to

minimize d(g, f)subject to f factors according to T.

#### 4 Result

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

$$f_T^* = g_1 \prod_{i \neq 1} g_{i|\mathbf{pa}_i|}$$

minimizes the differential relative entropy with g among all densities on  $\mathbb{R}^n$  which factor according to T ( $\mathbf{pa}_i$  is the parent of i in T rooted at vertex 1, i = 2, ..., n).

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

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

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

$$h(g, f) = -\int_{\mathbf{R}^d} g \log f$$

$$= -\int_{\mathbf{R}^d} g(x) \left( \log f_i(x_i) + \sum_{i \neq 1} \log f_{i|\mathbf{pa}_i}(x_i, x_{\mathbf{pa}_i}) \right) dx$$

$$= h(g_1, f_1) + \sum_{i \neq 1} \left( \int_{\mathbf{R}} g_{\mathbf{pa}_i}(\xi) h(g_{i|\mathbf{pa}_i}(\cdot, \xi), f_{i|\mathbf{pa}_i}(\cdot, \xi)) d\xi \right)$$

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

Fourth, recall  $h(\phi, \psi) \geq 0$  for densities  $\phi, \psi$  of any dimension, and zero if  $\phi = \psi$ . So  $f_1 = g_1$  and  $f_{i|\mathbf{pa}_i} = g_{i|\mathbf{pa}_i}$  are solutions.