5. Supervised learning of GRFs

A. Maximum likelihood estimators

Given: A family of p.d.s $p_u(x)$, $x \in X$, $u \in \mathcal{U}$ and a sample $\mathcal{X}_n = (x_1, \dots, x_n)$, where x_i are generated i.i.d. from $p_{u_0}(x)$ with unknown $u_0 \in \mathcal{U}$

Task: Estimate the unknown parameter u_0

Maximum likelihood estimator

$$u_* \in \operatorname*{argmax}_{u \in \mathcal{U}} \frac{1}{n} \sum_{i=1}^{n} \log p_u(x_i) = \operatorname*{argmax}_{u \in \mathcal{U}} L(u; \mathcal{X}_n)$$

Consistency of MLEs

$$\begin{array}{ccc} L(u; \mathcal{X}_n) & \xrightarrow[n \to \infty]{P} & L(u) = \sum\limits_{x \in X} p_{u_0}(x) \log p_u(x) \\ \Big\downarrow \operatorname*{argmax}_{u \in \mathcal{U}} & & \Big\downarrow \operatorname*{argmax}_{u \in \mathcal{U}} \\ u_*(\mathcal{X}_n) & \xrightarrow[n \to \infty]{P} & u_0 \end{array}$$

a) $L(u; \mathcal{X}_n) \xrightarrow[n \to \infty]{P} L(u)$? Yes, WLLN

$$P(|L(u; \mathcal{X}_n) - L(u)| > \varepsilon) \xrightarrow{n \to \infty} 0 \quad \forall u \in \mathcal{U}, \forall \varepsilon > 0$$

b) $u_0 \in \operatorname*{argmax}_{u \in \mathcal{U}} L(u)$? Yes, Shannon:

For any two p.d.s p, q, $D_{KL}(p \| q) \geq 0$
with equality iff $p \equiv q$

c) $u_*(\mathcal{X}_n) \xrightarrow[n\to\infty]{P} u_0$? Yes, if one of the following holds

1) $\mathcal{U}$ is finite
2) $\mathcal{U}$ is compact and $L(u;\mathcal{X}_n)$ converges <u>uniformly</u>, i.e.

$$P\Big(\sup_{u\in\mathcal{U}} \big|L(u;\mathcal{X}_n) - L(u)\big| > \varepsilon\Big) \xrightarrow{n\to\infty} 0$$

3) $\mathcal{U}$ is a convex set, $u_0 \in \mathrm{int}(\mathcal{U})$ and $L(u;\mathcal{X}_n)$ is concave in u for all $\mathcal{X}_n$

B. <u>Maximum likelihood learning for Gibbs random fields</u>

$S = \{S_i \mid i \in V\}$ is a Gibbs random field w.r.t. to the graph structure (V, E). I.e. its distribution is

$$p_u(s) = \frac{1}{Z(u)} \exp \sum_{\{ij\}\in E} u_{ij}(s_i, s_j)$$

but the parameters u_{ij} are unknown.
We are given a sample of realisations

$$\mathcal{T}_\ell = \{s^j \in \mathcal{S} \mid j = 1, \dots, \ell\}$$

i.i.d. generated from the model and want to estimate its parameters.

A K-valued GRF on the graph (V, E) is an exponential family $\Rightarrow$ we may write for the MLE

$$L(u; \mathcal{T}_\ell) = \frac{1}{\ell} \sum_{s\in\mathcal{T}_\ell} \log p_u(s) =$$

$$L(u; \mathcal{T}_\ell) = \frac{1}{\ell} \sum_{s \in \mathcal{T}_\ell} \langle \varphi(s), u \rangle - \log Z(u)$$

- $L(u; \mathcal{T}_\ell)$ is concave in u. If in addition $\exists\ p(s) > 0$ s.t. $\mathbb{E}_p(\varphi) = \frac{1}{\ell} \sum_{s \in \mathcal{T}_\ell} \varphi(s)$ then $\Rightarrow$ MLE is consistent.

- How to solve the task

$$L(u; \mathcal{T}_\ell) = \frac{1}{\ell} \sum_{s \in \mathcal{T}_\ell} \langle \varphi(s), u \rangle - \log Z(u) \to \max_u$$

for a GRF? E.g. by gradient ascend, where

$$\nabla L(u; \mathcal{T}_\ell) = \frac{1}{\ell} \sum_{s \in \mathcal{T}_\ell} \varphi(s) - \mathbb{E}_u(\varphi)$$

Find a GRF with u such that its pairwise marginal distr. coincide with the pairwise marginal statistics of the sample

Use e.g. sampling or Bethe approximation to compute $\mathbb{E}_u(\varphi)$ i.e. the pairwise marginals in each iteration of the gradient ascend.

LC. Pseudo-likelihood estimators for GRFs

Can we do better & simpler? Besag, 1975 $\Rightarrow$

Remember that a GRF on a graph (V, E) is defined by fixing the family of cond. distr.

$$p_u(s_i \mid s_{\mathcal{N}_i}) = \frac{1}{Z_i(u, s_{\mathcal{N}_i})} \exp \sum_{j \in \mathcal{N}_i} u_{ij}(s_i, s_j)$$

see sec. 4. Gibbs sampler.

Use pseudo-likelihood instead of likelihood

$$\tilde{L}(u; \mathcal{T}_\ell) = \frac{1}{\ell} \sum_{s \in \mathcal{T}_\ell} \sum_{i \in V} \log p_u(s_i \mid s_{\mathcal{N}_i}) \to \max_u$$

$$\tilde{L}(u; \mathcal{T}_\ell) =$$

$$= \frac{1}{\ell} \sum_{s \in \mathcal{T}_\ell} \sum_{i \in V} \sum_{j \in \mathcal{N}_i} u_{ij}(s_i, s_j) - \frac{1}{\ell} \sum_{s \in \mathcal{T}_\ell} \sum_{i \in V} \log Z_i(u, s_{\mathcal{N}_i})$$

$$= 2 \sum_{\{i,j\} \in E} \frac{1}{\ell} \sum_{s \in \mathcal{T}_\ell} u_{ij}(s_i, s_j) - \sum_{i \in V} \frac{1}{\ell} \sum_{s \in \mathcal{T}_\ell} \log Z_i(u, s_{\mathcal{N}_i})$$

- Computing $\tilde{L}(u; \mathcal{T}_\ell)$ and $\nabla L(u; \mathcal{T}_\ell)$ has complexity $\mathcal{O}(\ell |E| |K|^2)$
- $\tilde{L}(u; \mathcal{T}_\ell)$ is concave $\Rightarrow$ it can be proved that pseudo-likelihood estimators are consistent.
- However its variance is higher as compared with (exact) MLE