Computing marginal probabilities for GRFs

$S = \{S_i \mid i \in V\}$ is a K-valued Gibbs random field w.r.t. the undirected graph (V, E)

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

Task: Compute its marginal distributions $p(s_i)$, $i \in V$ for vertices and $p(s_i, s_j)$, $\{i,j\} \in E$ for edges

We know that for exponential families

$$p_u(s) = \frac{1}{Z(u)} \exp \langle \varphi(s), u \rangle$$

implies $\mathbb{E}_u \varphi = \nabla \log Z(u)$, but this does not help.

Remark 1 The task is easy to solve if (V, E) is a tree.

A. Gibbs sampler

Let $F: K^V \to \mathbb{R}^m$ be a random vector defined on the random field $S = \{S_i \mid i \in V\}$. We can estimate its expectation by sampling:

a) generate an i.i.d. sample of realisations $s^j \sim p_u(s)$ $j = 1, \ldots, \ell$

b) approximate $\mathbb{E}_u[F] \approx \frac{1}{\ell} \sum_{j=1}^{\ell} F(s^j)$

Remark 2 If F is bounded, we may use the Hoeffding inequality to estimate the sample size ℓ required for given confidence intervals.

To sample from $p_u(s)$ $\Rightarrow$ define a homogeneous Markov chain with transition probability $T(s|s')$, $s, s' \in K^V$ s.t.

- the chain is irreducible and a-periodic.
- its stationary distribution is $p_u(s)$

Remark 3 The first condition ensures the existence and uniqueness of a limiting distribution.
A stronger and sometimes easier to prove condition is detailed balance

$$T(s|s')\,p_u(s') = T(s'|s)\,p_u(s).$$

Practically: Design a set of simpler transition prob. matrices B_m s.t. $p_u(s)$ is stationary for all of them (but not necessarily unique) and compose T by

$$T = \prod_m B_m \quad \text{or} \quad T = \sum_m \alpha_m B_m$$

and prove that T is irreducible and a-periodic

Gibbs sampler Define B_i, $i \in V$ by

$$B_i(s|\tilde{s}) = \begin{cases} p_u(s_i|s_{V\setminus i}) & \text{if } s_{V\setminus i} = \tilde{s}_{V\setminus i} \\ 0 & \text{otherwise} \end{cases}$$

Remark 4 Notice that

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

is easy to compute.

Gibbs samplers are easy to implement but slow.
Its successive realisations are correlated

$$C_F(t) = \mathrm{cov}(F_{t_0}, F_{t_0+t}) =$$

$$= \sum_{s,s'} p_u(s) F(s) T^t(s'|s) F(s') - \mathbb{E}_u^2 F$$

$$\rho_F(t) = \frac{C_F(t)}{C_F(0)} \sim e^{-t/\tau}$$

This exponential decay can be slow!

B. Mean field approximation

Main idea: approximate $p_u(s)$ by a simpler distribution, e.g. assuming that the s_i, $i \in V$ are independent

$$D_{KL}(q \| p_u) = \sum_{s \in K^V} q(s) \log \frac{q(s)}{p_u(s)} \to \min_q$$

$$\text{s.t.} \quad q(s) = \prod_{i \in V} q_i(s_i)$$

We get

$$\sum_{i \in V} \sum_{s_i \in K} q_i(s_i) \log q_i(s_i) - \sum_{ij \in E} \sum_{s_i, s_j} u_{ij}(s_i, s_j)\, q_i(s_i)\, q_j(s_j) \to \min_q$$

$$\text{s.t.} \quad \sum_{s_i \in K} q_i(s_i) = 1 \quad \forall\, i \in V$$

This task is not convex. It is however convex for a single q_i provided all other q_j, $j \neq i$ are fixed. Solve it by block-coordinate descent.

A single step reads

$$q_i(s_i) \longleftarrow \frac{1}{Z_i} \exp \sum_{j \in \mathcal{N}_i} \sum_{s_j \in K} u_{ij}(s_i, s_j)\, q_j(s_j)$$

Remark 5 The mean field approximation gives unary marginals only.

LC. Unary marginals of log-supermodular random fields

For simplicity, we consider binary valued random fields. Consider a random field $X = \{X_i \mid i \in V\}$, $X_i = 0,1$ with the joint distribution

$$p(x) = \frac{1}{Z} e^{f(x)}.$$

Example 1 If X is a random field w.r.t. a graph (V, E), then

$$f(x) = \sum_{ij \in E} f_{ij} x_i x_j + \sum_{i \in V} f_i x_i .$$

If $(V, \mathcal{C})$ is a hypergraph and X is a random field w.r.t. $(V, \mathcal{C})$, then

$$f(x) = \sum_{c \in \mathcal{C}} f_c \prod_{i \in c} x_i \qquad \square$$

We assume that f is supermodular and approximate Z as follows

$$\log Z = \log \sum_{x \in 2^V} e^{f(x)} \leq$$

$$\leq \inf_{u \in \mathbb{R}^V} \Big\{ \log \sum_{x \in 2^V} e^{\langle u, x \rangle} \;\Big|\; \langle u, x \rangle \geq f(x) \;\; \forall x \in 2^V \Big\}$$

The objective is

$$L(u) = \log \sum_{x \in 2^V}' e^{\langle u, x \rangle} = \sum_{i \in V}' \log(1 + e^{u_i})$$

The constraint is

$$\mathcal{P}_f = \{ u \in \mathbb{R}^V \mid \langle u, x \rangle \geq f(x) \ \forall x \in 2^V \}$$

and is referred to as <u>supermodular polyhedron</u> of f.
The approximation task reads

$$L(u) = \sum_{i \in V}' \log(1 + e^{u_i}) \to \inf_u$$

$$\text{s.t.} \quad u \in \mathcal{P}_f$$

and is convex. It can be solved by Frank-Wolfe algorithm, provided we can solve the linear task

$$\langle w, u \rangle \to \inf_u$$

$$\text{s.t.} \quad u \in \mathcal{P}_f$$

<u>Frank-Wolfe algoritm:</u>

- Initialise $u_0 \in \mathcal{P}_f$
- Iterate:
  - $w_k = \nabla L(u_k)$
  - solve $\langle w_k, s \rangle \to \inf_s$ s.t. $s \in \mathcal{P}_f \to s_k$
  - set $\mu = \dfrac{2}{2+k}$
  - update $u_{k+1} = u_k + \mu (s_k - u_k)$

Solving the linear task (notice that $w \in \mathbb{R}^V_+$)

$$\langle w, u \rangle \to \inf_u$$

$$\text{s.t.} \quad \langle u, x \rangle \geq f(x) \quad \forall\, x \in 2^V$$

Consider its dual

$$\sum_{x \in 2^V}' \lambda(x) f(x) \to \max_{\lambda \geq 0}$$

$$\text{s.t.} \quad \sum_{x \in 2^V}' \lambda(x) x = w$$

It can be proved that the optimal λ is non-zero only on a chain $x_1 \leq x_2 \leq \ldots \leq x_{|V|}$ at most. Moreover, this chain itself depends on w only (not on f).

The optimal solution u^* of the Frank-Wolfe algorithm gives

- The approximation for $\log Z \approx \sum_i' \log(1 + e^{u_i^*})$
- approximated unary marginals

$$P_f(x_i = 1) \approx \frac{e^{u_i^*}}{1 + e^{u_i^*}}$$