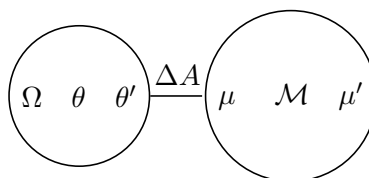


Sum-Product notes

Eliana Feasley

November 5, 2012

1 1-1 Mappings



$\mu = \Delta A \Theta$ Given any point μ there exists a θ . Given any point μ' there is a θ' .

When is the map 1-1? When the exponential family is not minimal. Variational methods are about moving from one representation to another.

When it is 1-1, $(\Delta A)^{-1}$ exists.

$$A\theta = \sup_{\mu \in \mathcal{M}} \{ \langle \theta, \mu \rangle - A^*(\mu) \}$$

This is discrete over complete sufficient statistics.

$$P_\theta(x) = \exp\{ \sum_s \theta_s x_s + \sum_{(s,t) \in E} \theta_{st} x_s x_t \}$$

$$\theta_s = \sum_{j,i} \theta_{s_{ij}} \perp_{sj} (x, s)$$

2 Pseudo-marginals

$$\mu = \begin{cases} ll\mu_s(x_s) & \text{Nodewise} \\ \mu_{st} & (x_s, x_t) \text{—Pairwise marginals of some distribution } P \end{cases}$$

μ_{st} is only defined over edges of G .

$$\mathbb{L}_G = \begin{cases} \ell T_s(x_s) & s \in V \\ T_{st} & (s, t) \in E \end{cases}$$

Norm $\sum_{x_s} T_s x_s = 1$

Marginals $\sum_{x_s} T_{st}(x_s x_t) = T_t(x_t)$

Marginals $\sum_{x_t} T_{st}(x_s x_t) = T_s(x_s)$

Prop(I) $\mathbb{L}G \supseteq M(G)$

If $\mu \in \mathcal{M}(G)$ it satisfies norm and marginal constraints $\rightarrow \mu \in \mathbb{L}G$

There are exponential constraints for μ and linear constraints for $\mathbb{L}(G)$

$$\begin{aligned} P(x_1, \dots, x_n) \mu_i(x_i) &= P(x_i) \\ \mu_{i,j}(x_i x_j) &= P(x_i, x_j) \end{aligned}$$

If graph G is a tree T , then $\mathbb{L}(T) = \mu(T)$ because nodewise and pairwise marginals are all we need.

Proof

$L(G) \supseteq \mu_G$ (Shown). We want to show that $\mathbb{L}(T) \subseteq \mu_T$. We will take a $\mu \in \mathbb{L}(T)$ such that it satisfies marginal constraints and show that it is consistent with regard to a joint distribution.

$\mu_s(x_s), s \in V$

$\mu_{st}(x_s)(x_t), (s, t) \in T$

$P_\mu(x) = \prod_{s \in V} \mu_s(x_s) \prod_{s,t \in E} \frac{\mu_{st}(x_s x_t)}{\mu_s(x_s) \mu_t(x_t)}$ such that μ is consistent with regard to P_μ .

We can compute the Bethe-entropy of the tree-structure distribution. Suppose we have a tree-structured distribution with mean parameters μ . Then we can write down the distribution in terms of μ .

$$\begin{aligned} A^*(\mu) &= \text{entropy}(P_\mu) \\ &= \mathbb{E}_{pm} |\log P_\mu(x)| \\ &= \mathbb{E}_{pm} \left| \sum_{s \in V} \log \mu_s(x_s) + \sum_{s,t \in E} \log \frac{\mu_{st}(x_s, x_t)}{\mu_s \mu_t} \right| \\ &= - \sum_s \mathbb{H}_s < \mu_s \text{right} > - \sum_{s,t \in E} I_{s,t}(\mu_{s,t}) \\ \mathbb{H}_{Bethe}(\mu) &= \sum_{s \in V} \mathbb{H}_s(\mu_s) - \sum_{s,t \in E} I_{s,t}(\mu + s, t) \\ &= -A^*(\mu) \text{ Exactly when } G \text{ is a tree} \end{aligned}$$

Computing entropy in the general case is NP-hard, but we can approximate it when G is not a tree using the variational principle.

We can approximate to $A(\theta)$ as so: $A_{Bethe}(\theta) = \sup_{\mu \in \mathbb{L}(G)} \{ \langle \theta, \mu \rangle + \mathbb{H}_{Bethe}(\mu) - I \}$

We get sum-product from the fixed point update driven solution of optimization problem I.

$$\mu_{st}(x_t) = \Psi_{st}(x_s, x_t) \prod_{v \in N(S)_t} \mu_{vs}(x_s)$$

This is a nonconvex problem. If we do coordinate descent on the dual, we can get to a *local* minimum. Regular sum-product often oscillates between very large ($O(v^2)$) parameters.

3 Reparamaterization

Suppose we solve sum-product to a local minimum τ^* (a local minimum of J). Then $P_{\tau^*}(x) = I \prod_{s \in V} T_s^*(x_s) \prod_{s,t \in E} \frac{T_{st}^*(x_s, x_t)}{T_s^*(x_s) T_t^*(x_t)}$

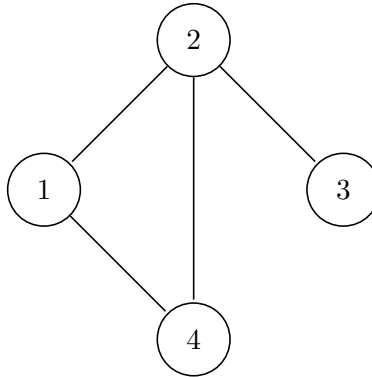
We know that if G is a tree and $T^* = \mu^*$, then $P_{T^*} = P_\theta$ and $P_{T^*}(x) = P_\theta(x)$

So it always holds that $P_\theta(x) = P_{T^*}(x)$

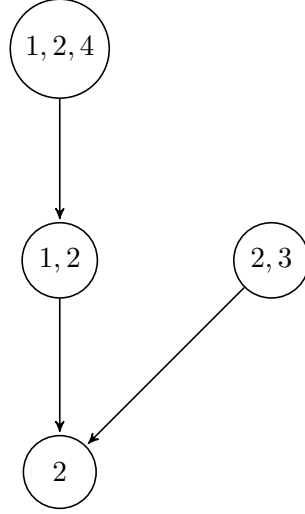
4 Hypergraphs

Define a hypergraph $G \in (V, E)$ with nodes $v \in V$ and hyperedges $e \in E$ such that $e \subseteq V$. A normal graph is a hypergraph with 2-node edges.

We can write the following graph as $E = (1, 2)(2, 3)(1, 4)(2, 4)$ or we can write $E = (1, 2, 4), (2, 3)$



Suppose we have the hypergraph with edges $E = (1, 2, 4), (2, 3), (1, 2), (2)$. Then we have a node for each superedge and we connect $h \rightarrow g$ if $g \in h$. These form a partially ordered set with regard to set inclusion.

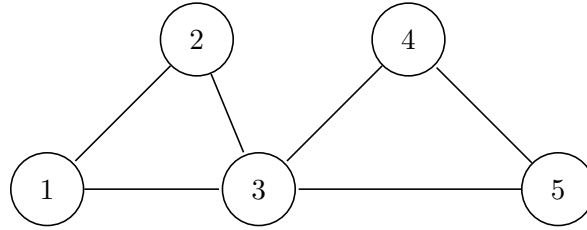


A general graphical model has subsets over cliques. The hypergraph edges correspond to cliques. We can do inference on these cliques.

Given a set of sufficient statistics $\mathbb{L}(G) = T_n(x_n)$, such that $\sum_n T_n(x_n) = 1$ we have g, h such that $f \subseteq h$:

$$\sum_{x_h} |x_g T_n(x_n)| = T_g(x_g)$$

For the following graph



$(1, 2, 3)$ and $(3, 4, 5)$ must be consistent.

$$H_{hypertree}(\mu) = \sum_{n \in E} c(h) H_h(\mu_k) \text{ where } c(h) \text{ are overcounting numbers.}$$

$$\sup_{\mu \in \mathbb{L}_G} \{ \langle \theta, \mu \rangle + H_{HT}(\mu) \}$$

Message passing here is generalized belief propagation.