* Recap

$$\underset{b \in LOC(T)}{\max} G(b) \leq \underset{b \in \Delta_{|X|^{|V|}}}{\max} G(b) \leq \sum_{T \in \mathcal{T}} \underset{b \in LOC(T)}{\max} G(b)$$

Bethe free energy $\leq$ Gibbs free energy $\leq$ Tree reweighted BP

Tight if the Graphical model is a Tree

$$\parallel$$

Bethe free energy

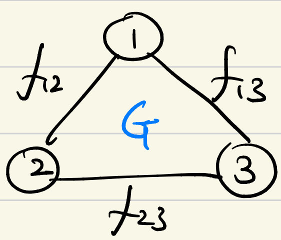
$$\underset{b \in LOC(G)}{\max} G(b)$$

$\downarrow$

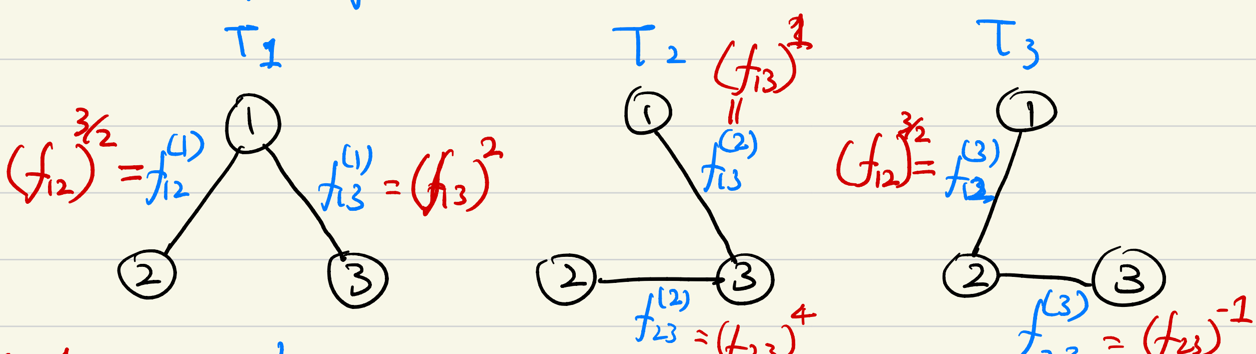
B.P.

* Tree reweighted belief propagation [more details on slides
"A new class of upper bounds on the log partition function" 2005 Wainwright, Jaakkola, Willsky.

example> Graphical model



Consider all spanning trees on G, and assign weights to each.



Rule #1: $\sum c_k = 1$ $\quad c_1 = 1/3 \quad c_2 = 1/3 \quad c_3 = 1/3$

Rule #2: for each $(i,j) \in E$, $f_{ij}(x_i, x_j) = \prod_K \left(f_{ij}^{(K)}(x_i, x_j)\right)^{c_K}$

$$\left(f_{12}^{(1)}\right)^{1/3} \times \left(f_{12}^{(3)}\right)^{1/3} = f_{12}^{3/2 \cdot 1/3} \cdot f_{12}^{3/2 \cdot 1/3} = f_{12}, \qquad \left(f_{13}^{(1)}\right)^{1/3} \cdot \left(f_{13}^{(2)}\right)^{1/3} = f_{13}$$

Def. Tree reweighted Belief Propagation.

Given $\mathcal{T} = \{T_k\}_{k=1}^{K}$ set of spanning trees.

corresponding weights $\{c_k\}$, and factors $\{\{f_{ij}^{(k)}\}_{(i,j)\in E_k}\}_k$

$$\sum c_k = 1$$

$$\log f_{ij}(x_i,x_j) = \sum_k c_k \log f_{ij}^{(k)}(x_i,x_j)$$

Then, the energy term in Gibbs free energy decomposed as

$$\mathbb{E}_b\Big[-\sum_{(i,j)\in E}\log f_{ij}(x_i,x_j)\Big] = \mathbb{E}_b\Big[-\sum_{(i,j)\in E}\sum_k c_k \log f_{ij}^{(k)}(x_i,x_j)\Big]$$

$$= \sum_k c_k \Big[\mathbb{E}_b\Big[-\sum_{(i,j)\in E_k}\log f_{ij}^{(k)}(x_i,x_j)\Big]\Big]$$

* Claim: $\log Z_G \le \min_{\substack{\{c_k\}\\ \{f_{ij}^{(k)}\}}} \sum_k c_k \underbrace{\log Z_{T_k}}$ ← easy to compute with B.P.

proof>

$$\log Z = \max_{b\in\Delta_{|\mathcal{X}|^n-1}} \mathbb{E}_b\Big[\sum_{ij}\log f_{ij}(x_i,x_j)\Big] + \text{Entropy}(b)$$

$\sum c_k = 1 \rightarrow$ $$= \max_{b\in\Delta} \sum_k c_k\, \mathbb{E}_b\Big[\sum_{(i,j)\in E_k}\log f_{ij}^{(k)}(x_i,x_j)\Big] + \sum_k \text{Entropy}(b)\cdot c_k$$

$$= \max_b \sum_k c_k \Big\{\mathbb{E}_b\Big[\sum_{E_k}\log f_{ij}^{(k)}(x_i,x_j)\Big] + \text{Entropy}(b)\Big\}$$

exchange sum/max $\rightarrow$ $$\leqq \sum_k c_k \cdot \max_{b^{(k)}\in\Delta}\Big\{\mathbb{E}_b\Big[\sum_{E_k}\log f_{ij}^{(k)}(x_i,x_j)\Big] + \text{Entropy}(b)\Big\}$$

$$= \boxed{\sum_k c_k \cdot \underbrace{\max_{b^{(k)}\in LOC(T_k)}\Big\{\mathbb{E}_b\Big[\sum_{E_k}\log f_{ij}^{(k)}(x_i,x_j)\Big] + \text{Entropy}(b)\Big\}}_{\text{B.P solves it efficiently } = \log Z_{T_k}}}$$

is well-defined U.B.

· Caveat [how do we find tightest U.B. by selecting $\{c_k\}$ $\{f_{ij}^{(k)}\}$.
spanning trees can be large.

*To solve inference problems $P(x_i)$

Variational Methods → Belief Propagation →
- Deterministic
- fast
- approximation

sampling → Gibbs Sampling →
- Randomized
- slower
- exact in the limit $N \to \infty$, but difficult to decide when to stop.

: $\hat{P}(x_i = a) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(x_i = a)$

Def. Markov Chain Monte Carlo methods.

- construct a Markov Chain $X^{(t)} \in \mathcal{X}^n \to X^{(t+1)} \in \mathcal{X}^n$
  a transition Matrix $Q \in \mathbb{R}^{|\mathcal{X}|^n \times |\mathcal{X}|^n}$
  a stationary distribution $P(x)$.
- start with $x^{(0)}$ and run M.C. until "convergence" so that $x^{(T)} \sim P(x)$.
- Repeat N times.

Q1. How do we construct such a Q? → Metropolis-Hastings Algo. → Gibbs Sampling.

Q2. How long does it take for M.C to converge? → Spectral Analysis, Path coupling.

Strategy>

Given a graphical model G, $P(x) = \frac{1}{Z} \prod_{(i,j) \in E} f_{ij}(x_i, x_j)$

State: $X^{(t)} \in \mathcal{X}^n$

initial State: $X^{(0)}$

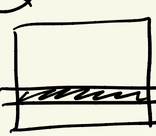
Construct transition matrix: $Q \in \mathbb{R}^{|\mathcal{X}|^n \times |\mathcal{X}|^n}$

you never actually store this matrix

[conceptually define
[sparsity.

Repeat: Sample: $X^{(t+1)} \sim (X^{(t)})^T Q$

$[\,0 \;\; 1 \;\; 0\,]$ one-hot encoding

 ← row of Q

eventually $X^{(\infty)} \sim \pi = P(x)$

stationary distribution of Q.

Def. time-Homogeneous finite-state Markov Chain

Markov chain
- State space: $\mathcal{X}^n$
- Transition Matrix $Q \in \mathbb{R}^{|\mathcal{X}|^n \times |\mathcal{X}|^n}$

$$Q_{xy} = \mathbb{P}(X^{(t+1)} = y \mid X^{(t)} = x)$$

Def. Stationary distribution of Q. is $\pi = \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}}_{|\mathcal{X}|^n}$

$$\pi^T \cdot Q = \pi^T$$

- might not be unique
- might not exist.

$X^{(t+1)} \sim X^{(t)T} \cdot Q$ then, $\lim_{t \to \infty} X^{(t)} \overset{d}{=} \pi$

Def. a Markov Chain is Reversible if $\exists \pi$, s.t. detailed balance equation is satisfied.

$$(*) \qquad \underbrace{\pi_x \cdot Q_{xy}}_{P(X^{(t)}, X^{(t+1)} = (x,y))} = \underbrace{\pi_y \cdot Q_{yx}}_{P(X^{(t)}, X^{(t+1)} = (y,x))}, \text{ for all } x, y \in \mathcal{X}^n$$

Claim. π satisfying $(*)$ is a stationary distribution of Q.

proof> $$(\pi^T Q)_y = \sum_x \pi_x Q_{xy} \underset{(*)}{=} \sum_x \pi_y \cdot Q_{yx} \underset{\substack{Q \text{ is stochastic} \\ \text{row-sum}=1}}{=} \pi_y$$

$\Rightarrow \pi^T Q = \pi^T \Rightarrow$ stationary distribution.

we will construct a M.C. Q that is reversible
- $(*)$ gives ensurance that $\pi = P(x)$
- spectral analysis to bound mixing time.

Def. Metropolis-Hastings Algorithm.

- Start with a candidate transition matrix $K \in \mathbb{R}^{|\mathcal{X}|^n \times |\mathcal{X}|^n}$

  $K = $ e.g. $\frac{1}{|\mathcal{X}|^n} \cdot \underbrace{\mathbb{1}}_{\text{all-ones vector}}\mathbb{1}^T$

- to ensure uniqueness of stationary distribution of K.

$$\left[\begin{array}{ll} K_{xx} > 0 \quad , \forall x & [\text{aperiodic}] \\ \text{for any } x, y \in \mathcal{X}^n, \exists \text{ a path } (x_0 = x, x_1, x_2, \cdots, x_m = y) & \\ \quad \text{s.t. } K_{x_i x_{i+1}} > 0. & [\text{irreducibility}] \end{array}\right.$$

what we want

(*) $Q_{xy} P(x) = Q_{yx} P(y)$

$\Updownarrow$

P is stationary dist. of Q.

What we have

$$K_{xy} P(x) \underset{\text{w.l.o.g}}{>} K_{yx} P(y)$$

the main trick is to remove "probability mass" from the larger one.

Define: $R_{xy} \triangleq \min\left\{1, \frac{P(y) K_{yx}}{P(x) K_{xy}}\right\}$

Construct $\boxed{Q_{xy} = K_{xy} \cdot R_{xy}}$

$$Q_{xx} \underset{\text{ensure } Q \text{ row sum to one.}}{=} 1 - \sum_{y \neq x} Q_{xy}.$$

interpreted as rejecting with $(1 - R_{xy})$.

Claim: $(Q, P(x))$ satisfy (*)

proof> $P(x) \cdot Q_{xy} = P_x \cdot K_{xy} \cdot R_{xy}$

suppose

$Q_{xy} = K_{xy} \cdot \frac{P_y K_{yx}}{P_x K_{xy}}$

$\Updownarrow$

$\boxed{R_{xy} < 1} \leftrightarrow \boxed{R_{yx} = 1}$

$= \cancel{P_x \cdot K_{xy}} \cdot \frac{P_y K_{yx}}{\cancel{P_x K_{xy}}}$

$= P_y \cdot K_{yx}$

$= P_y \cdot K_{yx} \cdot R_{yx} = \boxed{P_y \cdot Q_{yx}}$ $\square$