* Recap.

Compute $\Phi = \log Z$

Naive Mean Field

$$\max_{b(\cdot) \in (\Delta_{|\mathcal{X}|-1})^n} \mathbb{G}(b)$$

$\leq$

Bethe free energy on Tree

$$\max_{b(\cdot) \in LOC(T)} \mathbb{G}(b)$$

$\leq$

Gibbs free energy

$$\max_{b(\cdot) \in \Delta_{|\mathcal{X}|^n - 1}} \mathbb{G}(b)$$

$\leq$

Tree-reweighted BP

$$\sum_{T \in \mathcal{T}} \max_{b(\cdot) \in LOC(T)} \mathbb{G}_T(b)$$

$\neq$ neither lowerbound or upperbound

Region Based free energy

$$\max_{b(\cdot) \in LOC(\mathcal{G}, R)} \mathbb{G}(b)$$

$\Downarrow$

Generalized belief propagation.

Bethe free energy on non-Tree

$$\max_{b(\cdot) \in LOC(\mathcal{G})} \mathbb{G}(b)$$

$\Downarrow$

Belief Propagation

* **Bethe free energy**

- account for pairwise correlations induced by edges/factors
- exact on distribution P on trees

· Parameters: $b_i(x_i)$: approximates $P(x_i)$

$b_{ij}(x_i, x_j)$: approximates $P(x_i, x_j)$

use $b = \{b_i, b_{ij}\}$ notation

Def. Ideally we want to search over **Globally Consistent Marginals** w.r.t G.

$$MARG(G) \triangleq \left\{ b = \{\{b_i\}_{i\in V}, \{b_{ij}\}_{i,j\in E}\} \text{ s.t. } \exists P(x) \text{ with } \begin{array}{l} b_i(x_i) = \sum_{x_{-i}} P(x), \ \forall i \in V \\ b_{ij}(x_i, x_j) = \sum_{x \setminus \{i,j\}} P(x), \ \forall i,j \in E \end{array} \right\}$$

but checking if $b \in MARG$ is NP-hard.

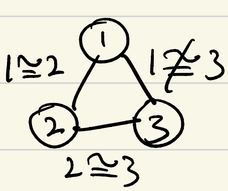
Def. Instead we propose searching over **Locally Consistent Marginals** w.r.t G

$$LOC(G) \triangleq \left\{ b \text{ s.t. } \begin{array}{ll} \sum_{x_i} b_i(x_i) = 1, & \forall i \in V \\ \sum_{x_j} b_{ij}(x_i, x_j) = b_i(x_i), & \forall i,j \in E \end{array} \right\}$$

· Local consistency does not imply Global consistency

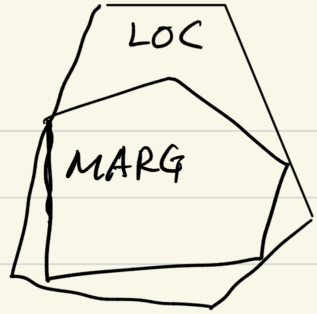
Counter example> $\mathcal{X} = \{0,1\}$, $b_1 = b_2 = b_3 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$

$$b_{12} = b_{23} = \begin{bmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{bmatrix}, \quad b_{13} = \begin{bmatrix} 0.01 & 0.49 \\ 0.49 & 0.01 \end{bmatrix}$$



we can check that $b_i(x_i) = \sum_{x_j} b_{ij}(x_i, x_j)$

· for general graph G.



· for G that is a tree.

$\forall$ Locally consistent $b = \{b_i, b_{ij}\}$, $\exists$ $P(x)$ that is globally consistent for b.

$$\tilde{P}(x) = \prod_{i \in V} b_i(x_i) \cdot \prod_{(i,j) \in E} \frac{b_{ij}(x_i, x_j)}{b_i(x_i)\, b_j(x_j)}$$

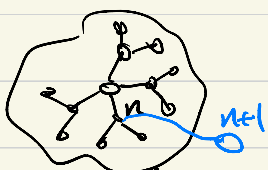
claim: $\tilde{P}(x)$ is globally consistent if G = tree.

that is $\begin{cases} b_i(x_i) = \tilde{P}(x_i), & \forall i \in V \\ b_{ij}(x_i, x_j) = \tilde{P}(x_i, x_j), & \forall (i,j) \in E \end{cases}$

proof: by induction

$n = 1$ $\quad \tilde{P}(x_1) = b(x_1)$ is globally consistent.

suppose it is true for a tree with n nodes, then we consider a tree with 1 more node that is connected to node n.



$$\tilde{P}(x_1^n, x_{n+1}) = \tilde{P}(x_1^n) \cdot \frac{b_{n,n+1}(x_n, x_{n+1})}{b_n(x_n)\, b_{n+1}(x_{n+1})} \, b_{n+1}(x_{n+1})$$

$$\tilde{P}(x_n, x_{n+1}) = \sum_{X_{V \setminus \{n, n+1\}}} \tilde{P}(x_1^n, x_{n+1}) = \tilde{P}(x_n) \cdot \frac{b_{n,n+1}}{b_n \cdot b_{n+1}} \cdot b_{n+1}$$

$$\xrightarrow{\text{induction}} = b(x_n) \cdot \frac{b_{n,n+1}}{\cancel{b_n} \cdot \cancel{b_{n+1}}} \cdot \cancel{b_{n+1}}$$

$$= b_{n,n+1}(x_n, x_{n+1}).$$

Def. Bethe free energy on a tree.

Recall, Gibbs free energy for general $b(x)$ is defined as

$$G_{f_{total}}(b) = -\underbrace{\mathbb{E}_b[-\log f_{total}(X)]}_{\text{Energy}} + \underbrace{\mathbb{E}_b[-\log b(x)]}_{\text{Entropy.}}$$

we evaluate it on $b(x) = \prod_{i\in V} b_i(x_i) \prod_{(i,j)\in E} \frac{b_{ij}(x_i,x_j)}{b_i(x_i)\, b_j(x_j)}$

$$\text{Energy} = -\sum_{(i,j)\in E} \sum_{x_i,x_j} b_{ij}(x_i,x_j) \cdot \log f_{ij}(x_i,x_j)$$

$$\text{Entropy} = \mathbb{E}_b\left[-\log\left(\prod_i b_i(x_i) \prod_{(i,j)} \frac{b_{ij}(x_i,x_j)}{b_i(x_i)\, b_j(x_j)}\right)\right]$$

$$= \sum_i \sum_{x_i} \left(-b_i(x_i) \log b_i(x_i)\right) - \underbrace{\sum_{(i,j)\in E} \sum_{x_i,x_j} \left(\log b_{ij} - \log b_i - \log b_j\right) b_{ij}}_{\sum_{(i,j)\in E} \sum_{x_i x_j} b_{ij} \log b_{ij} - \sum_{i\in V} \sum_{x_i} \deg(i)\, b_i \log b_i}$$

Bethe free energy

$$\mathbb{F}(b) \triangleq G_{f_{tree}}(b) = -\text{Energy} + \text{Entropy}$$

(f_{tree}: Tree)

$$= \sum_{(i,j)\in E} \sum_{x_i,x_j} b_{ij}(x_i,x_j)\left(\log f_{ij}(x_i,x_j) - \log b_{ij}(x_i,x_j)\right) + \sum_{i\in V} (\deg(i)-1) \sum_{x_i} b_i(x_i) \log b_i(x_i)$$

claim. if G is a tree then,

$$\sup_{b\in LOC(G)} \mathbb{F}(b) = \sup_{b\in MARG(G)} G(b) = \Phi \quad : \log \text{ partition function.}$$

this is called Bethe variational problem.

if you apply this formula $\mathbb{F}(b)$ and optimize for a non-tree G, then it is called Bethe approximation.

* From Bethe free energy to belief propagation.

claim. fixed points of BP are one-to-one correspondence with stationary points of Bethe variational problem.

Further, BP messages $\{ m_{i\to j}(x_i) \}$ are simple exponentials of Lagrangian variables $\{ \lambda^*_{i,j}(x_i) \}$.

proof.

Define. Lagrangian multipliers λ_i for $\sum_{x_i} b_i(x_i) = 1$

$\lambda_{i\to j}(x_i)$ for $\sum_{x_j} b_{ij}(x_i, x_j) = b(x_i)$

$$\text{Lagrangian}(b, \lambda) = \mathbb{F}(b) - \sum_i \lambda_i \left\{ \sum_{x_i} b(x_i) - 1 \right\} - \sum_{(i,j)} \sum_{x_i} \lambda_{i\to j}(x_i) \left\{ \sum_{x_j} b_{ij}(x_i, x_j) - b_i(x_i) \right\}$$

taking the derivative,

$$\nabla_{b_{ij}(x_i,x_j)} L(b,\lambda) = -1 - \log b_{ij}(x_i, x_j) + \log f_{ij}(x_i, x_j) - \lambda_{i\to j}(x_i) - \lambda_{j\to i}(x_j)$$

$$\nabla_{b_i(x_i)} L(b,\lambda) = -(1 - \deg(i)) \log [b_i(x_i) \cdot e] - \lambda_i + \sum_{j\in\partial i} \lambda_{i\to j}(x_i)$$

Setting them to zero,

$$b^*_{ij}(x_i, x_j) = f_{ij}(x_i, x_j) \exp\{ -1 - \lambda_{i\to j}(x_i) - \lambda_{j\to i}(x_j) \}$$

$$b^*_i(x_i) \propto \exp\left\{ -\frac{1}{d_i - 1} \sum_{j\in\partial i} \lambda_{i\to j}(x_i) \right\}$$

$$\sum_{x_j} b^*_{ij}(x_i, x_j) = b^*_i(x_i)$$

We change variables as: $m_{i\to j}(x_i) \propto e^{-\lambda_{i\to j}(x_i)}$

$$b^*_{ij}(x_i, x_j) \propto m_{i\to j}(x_i) f_{ij}(x_i, x_j) m_{j\to i}(x_j)$$

$$b^*_i(x_i) \propto \prod_{j\in\partial i} \left\{ (m_{i\to j}(x_i))^{\frac{1}{\deg(i)-1}} \right\}$$

$$\sum_{x_j} b^*_{ij}(x_i, x_j) = b^*_i(x_i)$$

To show equivalence btw BP vs. Bethe free
we start with.

$$\prod_{k \in \partial i \setminus j} \left\{ \sum_{x_k} b_{ik}^*(x_i, x_k) \right\} \overset{\text{Local Consistency}}{=} \prod_{k \in \partial i \setminus j} b_i^*(x_i) = b_i^*(x_i)^{\deg(i)-1}$$

Stationarity $\Rightarrow$ $S||$ $\qquad$ $|| \leftarrow$ Stationary $F(b)$

$$\prod_{k \in \partial i \setminus j} \left\{ m_{i \to k}\cancel{(x_i)} \cdot \sum_{x_k} m_{k \to i}(x_k) \cdot f_{ik}(x_i, x_k) \right\} \propto \prod_{k \in \partial i} m_{i \to k}\cancel{(x_i)}$$

$\Downarrow$ $\qquad$ $\Downarrow$ cancelling

$$m_{i \to j}(x_i)$$

$$\prod_{k \in \partial i \setminus j} \sum_{x_k} m_{k \to i}(x_k) f_{ik}(x_i, x_k) \propto$$

$\Downarrow$

We arrive at BP update

Recap:

$$\underset{b \in LOC(T)}{\max} G(b) \underset{\text{Bethe free energy}}{} \overset{\le}{=} \underset{\text{Gibbs free energy}}{} \max_b G(b)$$

$\uparrow$ Tight if $P(x)$ defined on a tree.

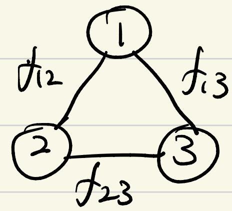
SS

Bethe free energy

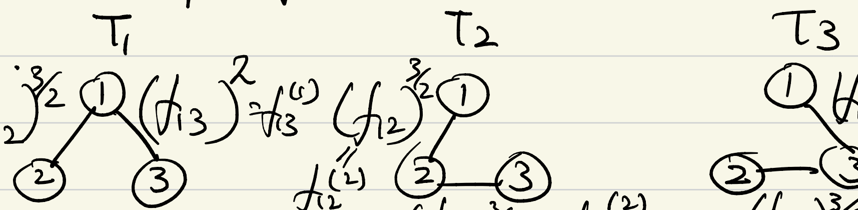
$$\max_{b \in LOC(G)} G(b) \longleftrightarrow$$ finding stationary point of this optimization is equivalent to finding a fixed point of BP update.

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

example> original graphical model



consider all spanning trees. (and weights on those trees).



$f_{12}^{(1)} = (f_{12})^{3/2}$, $(f_{13})^{2} = f_{13}^{(1)}$, $(f_{12})^{3/2} = f_{12}^{(2)}$, $(f_{23})^{3/2} = f_{23}^{(2)}$, $(f_{13})^{1} = f_{13}^{(3)}$, $(f_{23})^{3/2} = f_{23}^{(3)}$

$c_1 = \frac{1}{3}$ $\quad c_2 = \frac{1}{3}$ $\quad c_3 = \frac{1}{3}$

$$f_{12} = (f_{12}^{(1)})^{c_1} \cdot (f_{12}^{(2)})^{c_2} = (f_{12})^{\frac{3}{2}\cdot\frac{1}{3}} \cdot (f_{12})^{\frac{3}{2}\cdot\frac{1}{3}}$$

$$f_{13} = (f_{13}^{(1)})^{c_1} \cdot (f_{13}^{(3)})^{c_3} = f_{13}^{2/3} \cdot f_{13}^{1/3}$$

in general, consider all spanning trees $T_k \in \mathcal{T}(G)$.
assign weights c_k such that. $\sum_{k=1}^{|\mathcal{T}(G)|} c_k = 1$ &

$$\log f_{ij}(x_i, x_j) = \sum_{k=1}^{|\mathcal{T}(G)|} c_k \log f_{ij}^{(k)}(x_i, x_j)$$

then, the energy term in Gibbs free energy decomposes as

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

$$= \sum_k c_k \underbrace{\mathbb{E}_b\Big[-\sum_{(i,j)\in E} \log f_{ij}^{(k)}(x_i, x_j)\Big]}_{\text{Energy on one Tree model.}}$$

* claim: $\log Z \leq \sum_K C_K \cdot \log Z_K$

proof)

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

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

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

$$\underset{\substack{\uparrow \\ \text{Graphical} \\ \text{model on } T_K}}{=} \sum_K C_K \cdot \underbrace{\max_{b \in \text{LOC}(T_K)} \Big\{ \mathbb{E}_b\Big[\sum_{E_K} \log f_{ij}^{(K)}(x_i, x_j)\Big] + \text{Entropy}(b) \Big\}}_{\text{Can be solved exactly with BP.}}$$

- Caveat
  - C_K, $f_{ij}^{(K)}$ need to be chosen to minimize the RHS
  - # of spanning trees can explode