* Recap. Inference Task of computing $\mathbb{P}(X_1 = x_1 \mid Y_1^m)$

① Sampling

$$X^{(i)} \sim \mathbb{P}(X \mid Y_1^m)$$
$$\mathcal{X}^n$$

$$\hat{\mathbb{P}}(X_1 = x_1 \mid Y_1^m) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(x_1^{(i)} = x_1)$$

N ↑ # Samples drawn

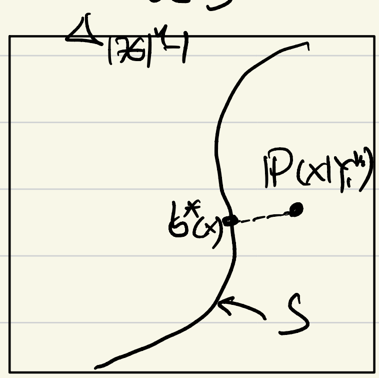
$$\pm\, O\left(\tfrac{1}{\sqrt{N}}\right)$$

② Variational Bayes

$\mathbb{P}(X \mid Y_1^m)$ ← test distribution $b(x) \in S$ ← simple family

functionals.

$$\min_{b \in S} D_{KL}(b \,\|\, \mathbb{P}(x \mid Y_1^m))$$
$$= \max_{b \in S} \log Z - D_{KL}(b \,\|\, \mathbb{P}(x \mid Y_1^m))$$



⟺ find Stationary Point of this optimization

Calculus of Variations

* Punchline Variational Bayes gives B.P.

$$\text{Compute } \Phi = \log Z$$

Naive Mean field
$$\max_{b \in \{\Delta_{|\mathcal{X}|-1}\}^n} G(b)$$
$b = b_1 \cdot b_2 \cdots b_n$
↓
Mean Field Equations.

$\leqq$

Bethe free energy on Tree
$$\max_{b \in LOC(T)} G(b)$$

$\leqq$

Gibbs free energy ($= \log Z$)
$$\max_{b \in \Delta_{|\mathcal{X}|^n - 1}} G(b)$$

$\leqq$

Tree-reweighted BP
$$\sum_{T \in \mathcal{T}} \max_{b \in LOC(T)} G_T(b)$$

Regionalized BP ⇓ Generalized BP.

neither LB or UB

Bethe free energy on non-Tree G
$$\max_{b \in LOC(G)} G(b)$$
No longer a valid dist.

⇓ exactly BP update

* **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.

$$\mathrm{MARG}(G) \triangleq \left\{ b = \{\{b_i\}_{i\in V}, \{b_{ij}\}_{i,j\in E}\} \text{ s.t. } \exists\, P'(x) \text{ with } \begin{array}{l} b_i(X_i) = \sum_{X_{-i}} P'(X), \ \forall i \in V \\ b_{ij}(X_i, X_j) = \sum_{X \setminus \{i,j\}} P'(x), \ \forall i,j \in E \end{array} \right\}$$

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

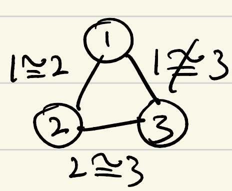
Def. Instead we propose searching over Locally Consistent Marginals w.r.t G

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

· Local consistency does not imply Global consistency

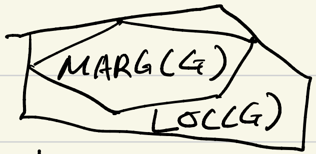
Counter example> $\mathcal{X} = \{0,1\}$, $b_1 = b_2 = b_3 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$

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



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

- for general graph G.



- for a tree T

Local Consistency implies Global Consistency if the graph is a tree.

for G that is a tree, suppose we know the singleton & pairwise marginals then we can recover the joint $P(x)$ uniquely.

$\{\tilde{P}(x_i)\}_{i\in V}$'s, $\{\tilde{P}_{ij}(x_i, x_j)\}_{(i,j)\in E_T}$, then

$$\tilde{P}(x_1 \cdots x_n) = \prod_{i\in V} \tilde{P}_i(x_i) \cdot \prod_{(i,j)\in E} \frac{\tilde{P}_{ij}(x_i, x_j)}{\tilde{P}_i(x_i)\,\tilde{P}_j(x_j)} \qquad (*)$$

claim. $b(x) \triangleq \prod_{i\in V} b_i(x_i) \cdot \prod_{(i,j)\in E} \frac{b_{ij}(x_i, x_j)}{b_i(x_i)\, b_j(x_j)}$ is globally consistent

if $G=(V,E)$ is a tree, i.e., $b_i(x_i) = \sum_{x_{\sim i}} b(x)$

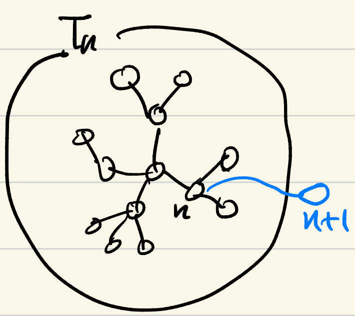
$b_{ij}(x_i, x_j) = \sum_{x_{V\setminus\{i,j\}}} b(x)$

proof) by induction,

$n=1$, $b(x_1) = b_1(x_1)$ is globally consistent.

suppose it is true for a tree with n nodes.

then we consider a tree with node $n+1$ attached to node n



$$\underline{b(x_1^n, x_{n+1})} = \underline{b(x_1^n)} \cdot \frac{b_{n,n+1}(x_n, x_{n+1})}{b_n(x_n)}.$$

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

induction

$$= \cancel{b_n(x_n)} \cdot \frac{b_{n,n+1}(x_n, x_{n+1})}{\cancel{b_n(x_n)}}$$

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

Bethe free energy

Recall, Gibbs free energy for $b(x)$

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

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

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

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

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

Def. **Bethe Free Energy**

$$\mathbb{F} : LOC(G) \to \mathbb{R}, \quad b \mapsto \mathbb{F}(b) = G\left(\prod_{i\in V} b_i(x_i) \prod_{(i,j)\in E} \frac{b_{ij}(x_i,x_j)}{b_i(x_i)\, b_j(x_j)}\right)$$

from Graphical Model f_{total} that is not necessarily a tree

$$= -\text{Energy} + \text{Entropy}$$

$$\mathbb{F}(b) = -\text{Energy} + \text{Entropy}$$

$$= \sum_{ij\in E} \sum_{x_i x_j} b_{ij}(x_i,x_j)\left(\log f_{ij}(x_i,x_j) - \log b_{ij}(x_i,x_j)\right)$$

$$+ \sum_{i\in V} \sum_{x_i} (\deg(i) - 1) \cdot \sum_{x_i} b_i(x_i) \log b_i(x_i)$$

claim: If G is a tree, then

$$\max_{b\in LOC(G)} \mathbb{F}(b) = \max_{b\in \Delta_{|X|^n-1}} G_{f_{total}}(b) = \Phi = \log Z$$

This is called Bethe Variational Problem

If you apply this to G that is not a tree, then it is called Bethe approximation.

For a graph not necessarily a tree

Claim: fixed points of BP updated are one-to-one correspondence with stationary points of Bethe variational problem. Further, BP messages $\{m_{i\to j}(x_i)\}$ are simple exponential of $\{\lambda^*_{i,j}(x_i)\}$ Lagrangian multipliers.

proof>

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

$\lambda_{i\to j}(x_i)$ for $\sum_{x_j} b_{ij}(x_i,x_j) = b_i(x_i)$ $\quad (i,j)\in \overleftrightarrow{E}$

$$\text{Lagrangian}(b,\lambda) = \underset{\substack{\parallel \\ \sum_{ij}\sum_{x_i x_j} b_{ij}(\log f_{ij} - \log b_{ij}) + \sum_i (\deg(i)-1)\sum_{x_i} b_i \log b_i}}{\mathbb{F}(b)} - \sum_i \lambda_i \left(\sum_{x_i} b_i(x_i) - 1\right) - \sum_{(i,j)\in E}\sum_{x_i} \lambda_{i\to j}(x_i)\cdot\left\{\sum_{x_j} b_{ij}(x_i,x_j) - b_i(x_i)\right\}$$

take derivative

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

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

Setting them to zero,

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

$$b_i^*(x_i) \underset{\substack{\uparrow \\ \text{dropping all constant scaling}}}{\propto} \exp\left\{-\frac{1}{\deg(i)-1}\cdot\sum_{j\in\partial i}\lambda^*_{i\to j}(x_i)\right\}$$

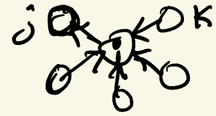
$$\sum_{x_j} b^*_{ij}(x_i,x_j) = b^*_i(x_i)$$

we change variables, $m_{i\to j}(x_i) \propto e^{-\lambda^*_{i\to j}(x_i)}$

① $b^*_{ij}(x_i,x_j) \propto f_{ij}(x_i,x_j)\cdot m_{i\to j}(x_i)\cdot m_{j\to i}(x_i)$

② $b^*_i(x_i) \propto \prod_{j\in\partial i} \left(m_{i\to j}(x_i)\right)^{\frac{1}{\deg(i)-1}}$

③ $\sum_{x_j} b^*_{ij}(x_i,x_j) = b_i^*(x_i)$



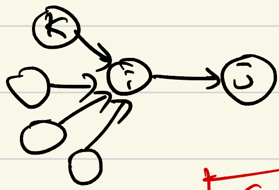
$$\prod_{k\in\partial i\setminus j}\left\{\sum_{x_k} b^*_{ik}(x_i,x_k)\right\} \overset{③}{=} \prod_{k\in\partial i\setminus j} b_i^*(x_i) = \left(b_i^*(x_i)\right)^{\deg(i)-1}$$

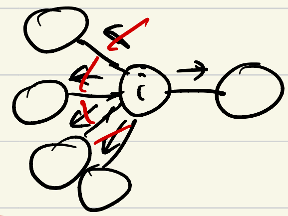
$$\| ①$$

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

$$\| ②$$

$$\prod_{k\in\partial i} m_{i\to k}(x_i)$$





$\boxed{\text{Cancel } m_{i\to k} \text{ for } k\neq j}$

$$\|$$

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

This is the BP update