New representation of a tree ①
structured model in terms of marginals

- given a Tree structured model,
  can write

$$P(X_1 \cdots X_n) = \frac{1}{Z}\left[\prod_{i \in V} \psi_i(x_i)\right]\left[\prod_{ij \in V} \varphi_{ij}(x_i, x_j)\right]$$

- we can compute marginals for this distribution easily by (say) message passing.

- Write

$\mu_i(x_i)$ for marginal of X_i

$\mu_{ij}(x_i, x_j)$ for marginal of (x_i, x_j) (which is only interesting if there is an edge $i - j$)

1a

Because they are marginals,
they <u>must</u> satisfy

$$\sum_{x_i} \mu_i(x_i) = 1$$

$$\sum_{x_i} \mu_{ij}(x_i, x_j) = \mu_j(x_j)$$

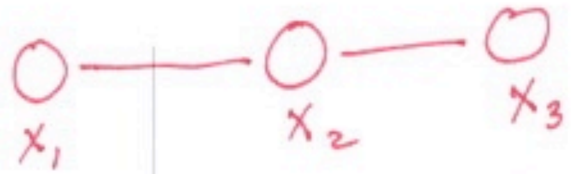
$$\sum_{x_j} \mu_{ij}(x_i, x_j) = \mu_i(x_i)$$

Claim

$$P(x_1, \dots x_N) = \prod_{i \in V} \mu_i(x_i) \cdot \prod_{i,j \in E} \frac{\mu_{ij}(x_i, x_j)}{\mu_i(x_i)\,\mu_j(x_j)}$$

is a probability distribution that has marginals μ_i, μ_{ij}

Proof: (2 examples, induction)



$$P(x_1, x_2, x_3) = \frac{\mu_{12}\,\mu_{23}}{\mu_2}$$

(← check tha
products clear

3

$$P(x_1) = \sum_{x_2} \frac{\mu_{12}}{\mu_2} \left[\sum_{x_3} \mu_{23} \right]$$

recall the message passing reasoning!

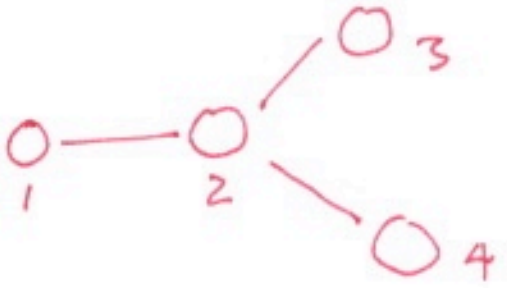
$$= \sum_{x_2} \frac{\mu_{12}}{\mu_2} \left[\mu_2 \right]$$

constraints!

$$= \mu_1$$

and $P(x_2)$, $P(x_3)$, $P(x_1, x_2)$, $P(x_2, x_3)$

work the same way.



$$p(x_1, x_2, x_3, x_4) = \frac{\mu_{12}\mu_{23}\mu_{24}}{\mu_2^2}$$

$$p(x_1) = \sum_{x_2} \frac{\mu_{12}}{\mu_2^2} \left[\sum_{x_3} \mu_{23}\right]\left[\sum_{x_4} \mu_{24}\right]$$

$$= \sum_{x_2} \frac{\mu_{12}}{\mu_2^2} \left[\mu_2\right]\left[\mu_2\right]$$

$$= \mu_1$$

etc.

<u>Induction</u> $\Rightarrow$ all this works for any tree.

Conclude :

- supply μ_i , $i \in V$
  μ_{ij} , $i,j \in E$ of some tree.
- you have supplied a P.D. on that tree.

Consequences :

- We can do ~~Karr~~ Mean field inference for a Q a tree
- Linear program on trees is easy (hardly a shock! DP works!

Q: a tree.

we must compute

$$E_Q \log P$$

and $E_Q \log Q$

1) we can compute $E_Q \log Q$

$$q(x_1 \cdots x_N) = \left[\prod_{i \in V} q_i\right]\left[\prod_{ij \in E} \frac{q_{ij}}{q_i q_j}\right]$$

$$= \frac{\prod_{ij \in E} q_{ij}}{\prod_{i \in V} q_i^{(d_i - 1)}}$$

d_i = degree of vert.

So

$E_Q \log Q$

$$= \sum_{\substack{\text{values}\\\text{of pairs}}} \left[\sum_{ij \in E} q_{ij} \log q_{ij} \right]$$

$$- \sum_{\text{values}} \left[(d_i - 1) \sum_{i \in V} q_i \log q_i \right]$$

this is tractable: eg second term is

$$\sum_{i \in V} \left[\sum_{\substack{\text{values}\\\text{at } i}} \left\{ (d_i - 1) q_i \log q_i \right\} \right]$$

etc ...

We can also compute

$$-E_Q \log P.$$

$$P(H|X) = \frac{1}{Z} \exp\left[-\sum_{ij} \theta_{ij}(H_i, H_j) - \sum_i \theta_i(H_i)\right]$$

$$-\log P = \log Z + \sum_{ij} \theta_{ij}(H_i, H_j) + \sum_j \theta_i(H_i).$$

constant; don't care cause we are minimizing in H.

$$-E_Q \log P = \log Z$$

$$+ \left[\sum_{ij\in E} \left\{\sum_{values} q_{ij}\left[\theta_{ij} + \theta_i + \theta_j\right]\right\}\right]$$

$$+ \left[\sum_{i\in V} \left[\sum_{values} (d_i - 1)\, q_i\, \theta_i\right]\right]$$

Notice this looks like entropy

rewrite

$$\tau_{ij} = \theta_{ij} + \theta_i + \theta_j$$

$$\tau_i = \theta_i$$

We get

$$\min \left[\sum_{ij \in \mathcal{E}} \left[\sum_{\text{values}} q_{ij} \left(\log q_{ij} + \tau_{ij} \right) \right] \right.$$

$$\left. + \sum_{i \in V} \left[\sum_{\text{values}} q_i \left(\log q_i + \tau_i \right) \right] \right]$$

where q_{ij}, q_i are marginals

$\mathcal{E}$ are edges of tree

- τ_{ij}, τ_i are <u>known</u>
- want q_{ij}, q_i

Marginals satisfy

$$\sum_i q_{ij} = q_j$$

$$\sum_j q_{ij} = q_i$$

$$\sum_i q_i = 1$$

- We have a constrained, continuous optimization problem
- Lagrangian, Lagrange Mults, etc.

L.M.'s associated w/

λ_{Σ_i} — $\sum_j q_{ij} = q_i$

λ_{Σ_j} — " $\sum_i q_{ij} = q_j$

λ_{v_i} — $\sum_i q_i = 1$

Notice:

λ_{ε_i} is a vector

λ_{V_i} is a scalar

at soln,

$$\left[\frac{\partial L}{\partial q_{ij}}\right]_{uv} = 0 = \left[\log q_{V_{ij}}\right]_{uv} + 1 + \tau_{ij} + \sum \left[\lambda_{\varepsilon_i}\right]_u + \sum \left[\lambda_{\varepsilon_j}\right]_v$$

$q_{V_{ij}}$ is a table; this is the u,v'th component

Sum over all edges attached to V_i

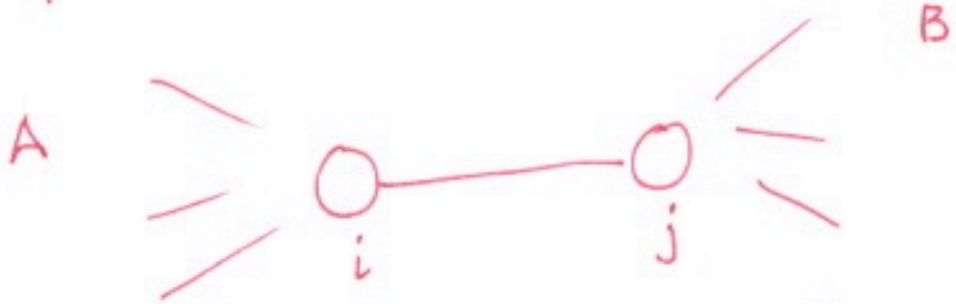
" j

so

$$\left[q_{ij}\right]_{uv} \propto \left[e^{-\tilde{c}_{ij}}\right]_{uv} \cdot \left[e^{\sum \lambda_{\varepsilon_i}}\right]_u \cdot \left[e^{\sum \lambda_{\varepsilon_j}}\right]_v$$

(Q: what is const. of proportionality?)

Compare with B.P.



$$\left[q_{ij}\right]_{uv} \propto \left[\varphi_{ij}\, \psi_i\, \psi_j\right]_{uv} \cdot \left[\prod_{\substack{\text{all inc} \\ \text{to } i}} m_{ia}\right]_u \cdot \left[\prod_{\substack{\text{all inc} \\ \text{to } j}} m_{jb}\right]_v$$

Conclusion:

- log messages $\equiv$ ~~log~~ L.M.s

- we can fit a var. model of a single tree.
- extract MAP from Q (easy!)
- loopy BP "like" fitting a tree without attending to what tree