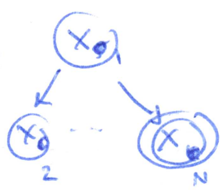
4/16/15
Lec 21

Reviewing variable elimination:
List ← cond & marg. distributions, and pick I an order of elimination
for X_i: $m_{X_i} \leftarrow \sum_{X_i} \prod_j list_j^i [X_i]$, ~~having added all distributions~~ having removed from List all distr containing X_i and adding them to list^i.

insert m_{X_i} into list,

Runtime "poly in n"? but not linear.
but actually depends on # of values that each var can take.

example:



And consider $I = X_{N-1} X_{N-2} \cdots X_1$ (A)

vs. $I' = X_1 \cdots X_{N-1}$ (B)

claim: B is better: b/c:

(A) $P(X_1, \ldots, X_N) = P(X_1) \prod_{j=2}^{N} P(X_j | X_1)$

so what would m_{X_1} be? $\sum_{X_1} P(X_1) \prod_{j=2}^{N} P(X_j | X_1)$

it's a fn of all the vars $X_2 \cdots X_N$.

so will get:

$P(X_N) = \underbrace{\sum_{X_{N-1}} \sum_{X_{N-2}} \cdots \sum_{X_2}}_{\text{which is how many possibilities?}} m_{X_1}(X_2 \cdots X_N)$

2^{N-1}

(B) ~~$P(X_1, \ldots, X_N) = P($~~
$P(X_N) = \sum_{X_1} \sum_{X_2} \cdots \sum_{X_{N-1}} P(X_1) \prod_{j=2}^{N} P(X_j | X_1)$

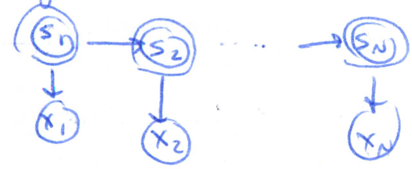
but X_{N-1} only shows up in one of the ~~terms~~ summations.
~~so~~ in which case $\sum_{X_{N-1}} P(X_{N-1} | X_1) = 1$.

this results in $P(X_N) = \sum_{X_1} P(X_1) P(X_N | X_1)$

so this is "linear" time.

but picking a good ordering is NP complete.

another example: HMMs.



here, it might be less obvious what is the right order.

let's try variable elimination, and assume we're given $x_1, \dots, x_n$, and ~~what's~~ my q is, what is $P(s_t | x_1, \dots, x_n)$?

Can we get this down to marginal distr? yes: $P(s_t, x_1, \dots, x_n) / P(x_1, \dots, x_n)$

(usually you don't actually need to compute $P(x_1, \dots, x_n)$, but just normalize afterwards).

q: which vars should we eliminate? a: all the remaining hidden states.

q: what order?

a: t is distinguished, so we should consider $(1, \dots, t-1)$ and $(N, N-1, \dots, t+1)$

$$P(s_t, x_1 \dots x_T) = \sum_{s_1 \dots s_{t-1}, s_N \dots s_{t+1}} P(s_1, \dots, s_N, x_1, \dots, x_T)$$

$$= \sum_{s_{t-1}} \sum \cdots \sum_{s_1} \sum_{s_{t+1}} \cdots \sum_{s_N} \underbrace{P(s_1, \dots, s_N, x_1, \dots, x_T)}$$

which gives us: $P(s_1) P(x_1 | s_1) P(s_2 | s_1) P(x_2 | s_2) \dots$

if you consider the $\sum_{s_N}$ summation, the only thing that contains it is $P(s_N | s_{N-1}) P(x_N | s_N)$.

let's now introduce forward and backward probs α & β.

$$\beta_N(\underbrace{s_{N-1}, x_N}_{\text{fnct...}}) = \sum_{s_N} P(s_N | s_{N-1}) P(x_N | s_N)$$

Similarly, $\beta_{N-1}(s_{N-2}, x_{N-1}, x_N)$ ~~which stuck~~ it

since it depends on β_N.

eventually, this gives us:

$$P(s_t, x_1 \dots x_N) = \beta_{t+1}(s_{t+1}, x_{t+1}, \dots, x_N)$$ ~~which will be $P(s_t, x_t, \dots,$~~

should this be t? yes.

$$\times \underbrace{\sum_{s_{t-1}} \cdots \sum_{s_1} P(s_1) P(x_1 | s_1) P(x_2 | s_2) \cdots P(s_t | s_{t-1})}$$

and you can do the same process, but left to right instead.

~~β_N = sum over s_N 0~~

$= \cancel{\sum_{s_1} P(s_2|s_1) P(x_1, s_1)}$

that is, have $\alpha_1(x_1) = \sum_{s_1}^{s_2} P(s_1) P(x_1|s_1) P(s_2|s_1)$

i.e., sum over what involves s_1.

similarly, $\alpha_2(x_2, s_3, x_1) = \sum_{s_2} \alpha_1(s_2|x_1) \underline{P(s_3|s_2) P(x_2|s_2)}$ → I guess this, too

will then get: $P(s_t, x_1, \ldots, x_N) = \beta_{t+1}(s_t, \ldots, x_N \ldots x_{t+1}) \alpha_{t-1}(s_t, x_1 \ldots x_{t-1}) P(x_t|s_t)$

so... this was meant to be an eg of variable elimination ("Lillian will do sthg similar, but not quite variable elim.")
→ note that you can think of this as if neighbors are passing msgs to each other.

Now, generally: what if we need more than one variable's marginals?
each node is doing add'n & multiplicat'n.

so: belief propagat'n:

spse we have observat'ns $X_j = x_j$, let the evidence vector (fn): $E_{X_j}(x) = \mathbb{1}\{x_j = x\}$

for unobserved vars, set $E_{X_j}(x) = 1$. for all x.

At round 0, all msgs btwn nodes are 1.

<he wants us to do this next for HMMs>

· during runs, we have incoming & outgoing msgs from parents & children.

ex: incoming msg to node from children:

$$\lambda(x) = E_{X_i}(x) \prod_{\substack{j \in \text{children} \\ (X_i)}} \lambda_{X_j}(x)$$

β for HMMs.   → β for HMMs: ie s_{t+1}, x_t.

[slides: for HMM, $\Pi \to \alpha$.]

can show that you get convergence of this for trees.

<for next lecture: ~~inference for sampling~~
inference from sampling>