6.825 Techniques in Artificial Intelligence

Inference in Bayesian Networks

- Exact inference
- Approximate inference

Query Types

Given a Bayesian network, what questions might we want to ask?

- Conditional probability query: P(x | e)
- Maximum a posteriori probability:
  What value of x maximizes P(x|e) ?

General question: What's the whole probability distribution over variable X given evidence e, P(X | e)?

Using the joint distribution

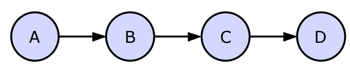
To answer any query involving a conjunction of variables, sum over the variables not involved in the query.

$$\Pr(d) = \sum_{ABC} \Pr(a,b,c,d)$$

$$= \sum_{a \in \mathrm{dom}(A)} \sum_{b \in \mathrm{dom}(B)} \sum_{c \in \mathrm{dom}(C)} \Pr(A = a \wedge B = b \wedge C = c)$$

$$\Pr(d \mid b) = \frac{\Pr(b,d)}{\Pr(b)} = \frac{\sum_{AC} \Pr(a,b,c,d)}{\sum_{ACD} \Pr(a,b,c,d)}$$

Simple Case



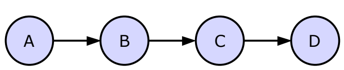
$$\Pr(d) = \sum_{ABC} \Pr(a,b,c,d)$$

$$= \sum_{ABC} \Pr(d \mid c)\Pr(c \mid b)\Pr(b \mid a)\Pr(a)$$

$$= \sum_{C} \sum_{B} \sum_{A} \Pr(d \mid c)\Pr(c \mid b)\Pr(b \mid a)\Pr(a)$$

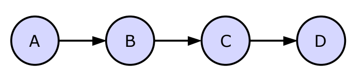
$$= \sum_{C} \Pr(d \mid c) \sum_{B} \Pr(c \mid b) \sum_{A} \Pr(b \mid a)\Pr(a)$$

Simple Case



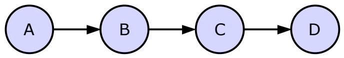
$$\Pr(d) = \sum_{C} \Pr(d \mid c) \sum_{B} \Pr(c \mid b) \underbrace{\sum_{A} \Pr(b \mid a)\Pr(a)}_{\begin{bmatrix} \Pr(b_1 \mid a_1)\Pr(a_1) & \Pr(b_1 \mid a_2)\Pr(a_2) \\ \Pr(b_2 \mid a_1)\Pr(a_1) & \Pr(b_2 \mid a_2)\Pr(a_2) \end{bmatrix}}$$

Simple Case



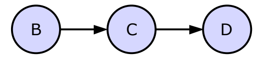
$$\Pr(d) = \sum_{C} \Pr(d \mid c) \sum_{B} \Pr(c \mid b) \underbrace{\sum_{A} \Pr(b \mid a)\Pr(a)}_{\begin{bmatrix} \sum_{A} \Pr(b_1 \mid a)\Pr(a) \\ \sum_{A} \Pr(b_2 \mid a)\Pr(a) \end{bmatrix}}$$

Simple Case



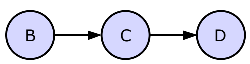
$$\Pr(d) = \sum_C \Pr(d \mid c) \sum_B \Pr(c \mid b) \underbrace{\sum_A \Pr(b \mid a)\Pr(a)}_{f_1(b)}$$

Simple Case



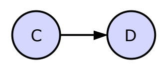
$$\Pr(d) = \sum_C \Pr(d \mid c) \underbrace{\sum_B \Pr(c \mid b) f_1(b)}_{f_2(c)}$$

Simple Case



$$\Pr(d) = \sum_C \Pr(d \mid c) \underbrace{\sum_B \Pr(c \mid b) f_1(b)}_{f_2(c)}$$

Simple Case



$$\Pr(d) = \sum_C \Pr(d \mid c) f_2(c)$$

Variable Elimination Algorithm

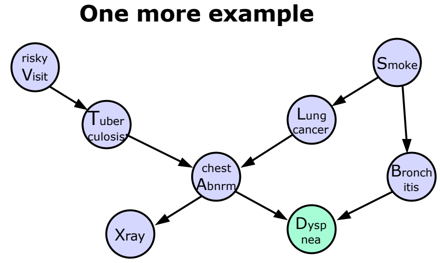
Given a Bayesian network, and an *elimination order* for the non-query variables, compute

$$\sum_{X_1} \sum_{X_2} \ldots \sum_{X_m} \prod_j \Pr(x_j \mid Pa(x_j))$$

For i = m downto 1

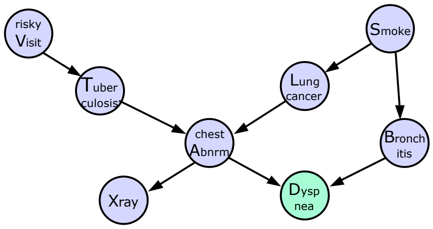
- remove all the factors that mention X_i
- multiply those factors, getting a value for each combination of mentioned variables
- sum over X_i
- put this new factor into the factor set

One more example



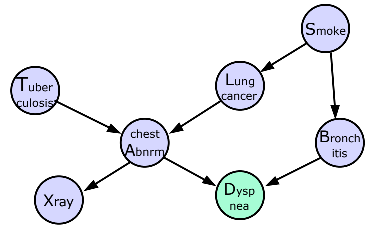
$$\Pr(d) = \sum_{A,B,L,T,S,X,V} \Pr(d \mid a,b)\Pr(a \mid t,l)\Pr(b \mid s)\Pr(l \mid s)\Pr(s)\Pr(x \mid a)\Pr(t \mid v)\Pr(v)$$

One more example



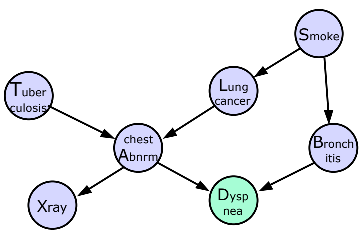
$$\Pr(d) = \sum_{A,B,L,T,S,X} \Pr(d \mid a,b)\Pr(a \mid t,l)\Pr(b \mid s)\Pr(l \mid s)\Pr(s) \Pr(x \mid a)\underbrace{\sum_{V} \Pr(t \mid v)\Pr(v)}_{f_1(t)}$$

One more example



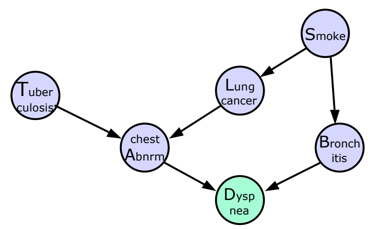
$$\Pr(d) = \sum_{A,B,L,T,S,X} \Pr(d \mid a,b)\Pr(a \mid t,l)\Pr(b \mid s)\Pr(l \mid s)\Pr(s) \Pr(x \mid a) f_1(t)$$

One more example



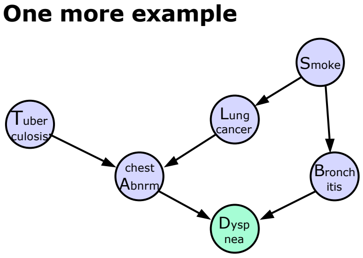
$$\Pr(d) = \sum_{A,B,L,T,S} \Pr(d \mid a,b)\Pr(a \mid t,l)\Pr(b \mid s)\Pr(l \mid s)\Pr(s) f_1(t) \underbrace{\sum_{X} \Pr(x \mid a)}_{1}$$

One more example



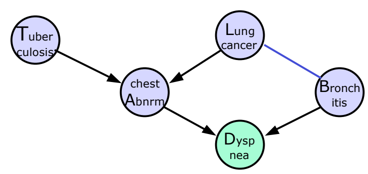
$$\Pr(d) = \sum_{A,B,L,T,S} \Pr(d \mid a,b)\Pr(a \mid t,l)\Pr(b \mid s)\Pr(l \mid s)\Pr(s) f_1(t)$$

One more example



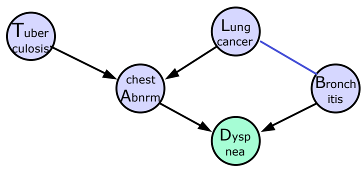
$$\Pr(d) = \sum_{A,B,L,T} \Pr(d \mid a,b)\Pr(a \mid t,l) f_1(t) \underbrace{\sum_{S} \Pr(b \mid s)\Pr(l \mid s)\Pr(s)}_{f_2(b,l)}$$

One more example



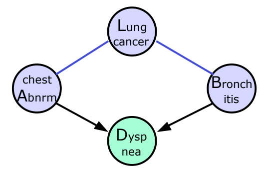
$$\Pr(d) = \sum_{A,B,L,T} \Pr(d \mid a,b)\Pr(a \mid t,l) f_1(t) f_2(b,l)$$

One more example



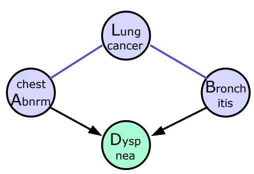
$$\Pr(d) = \sum_{A,B,L} \Pr(d \mid a,b) f_2(b,l) \underbrace{\sum_{T} \Pr(a \mid t,l) f_1(t)}_{f_3(a,l)}$$

One more example



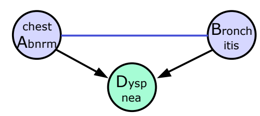
$$\Pr(d) = \sum_{A,B,L} \Pr(d \mid a,b) f_2(b,l) f_3(a,l)$$

One more example



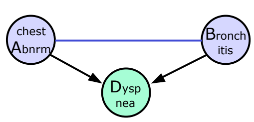
$$\Pr(d) = \sum_{A,B} \Pr(d \mid a,b) \underbrace{\sum_{L} f_2(b,l) f_3(a,l)}_{f_4(a,b)}$$

One more example



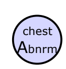
$$\Pr(d) = \sum_{A,B} \Pr(d \mid a,b) f_4(a,b)$$

One more example



$$\Pr(d) = \sum_{A} \underbrace{\sum_{B} \Pr(d \mid a,b) f_4(a,b)}_{f_5(a)}$$

One more example



$$\Pr(d) = \sum_{A} f_5(a)$$

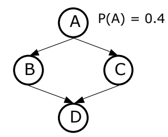
Properties of Variable Elimination

- Time is exponential in size of largest factor
- Bad elimination order can generate huge factors
- NP Hard to find the best elimination order
- Even the best elimination order may generate large factors
- There are reasonable heuristics for picking an elimination order (such as choosing the variable that results in the smallest next factor)

- Inference in polytrees (nets with no cycles) is linear in size of the network (the largest CPT)
- Many problems with very large nets have only small factors, and thus efficient inference

Sampling

To get approximate answer we can do stochastic simulation (sampling).



A	B	C	D
T	F	T	T
...			

Estimate:

P*(D|A) = #D,A / #A

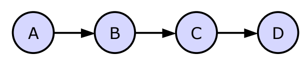
- Flip a coin where P(T)=0.4, assume we get T, use that value for A
- Given A=T, lookup P(B|A=T) and flip a coin with that prob., assume we get F
- Similarly for C and D
- We get one sample from joint distribution of these four vars

Estimation

- Some probabilities are easier than others to estimate
- In generating the table, the rare events will not be well represented
- P(Disease| spots-on-your-tongue, sore toe)
- If spots-on-your-tongue and sore toe are not root nodes, you would generate a huge table but the cases of interest would be very sparse in the table
- Importance sampling lets you focus on the set of cases that are important to answering your question

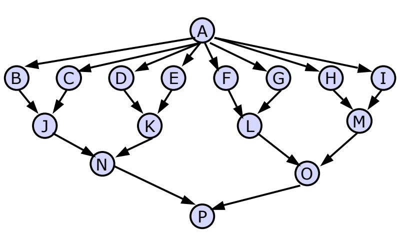
Recitation Problem

- Do the variable elimination algorithm on the net below using the elimination order A,B,C (that is, eliminate node C first). In computing P(D=d), what factors do you get?
- What if you wanted to compute the whole marginal distribution P(D)?



Another Recitation Problem

Find an elimination order that keeps the factors small for the net below, or show that there is no such order.



The Last Recitation Problem (in this lecture)

Bayesian networks (or related models) are often used in computer vision, but they almost always require sampling. What happens when you try to do variable elimination on a model like the grid below?