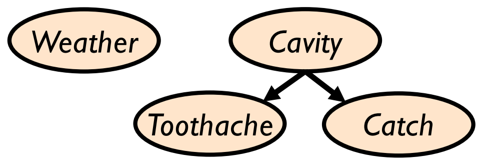
Bayes Nets
AI Class 10 (Ch. 14.1–14.4.2; skim 14.3)



Based on slides by Dr. Marie desJardin. Some material also adapted from slides by Matt E. Taylor @ WSU, Lise Getoor @ UCSC, Dr. P. Matuszek @ Villanova University, and Weng-Keen Wong at OSU. Based in part on www.csc.calpoly.edu/~fkurfess/Courses/CSC-481/W02/Slides/Uncertainty.ppt .

1

Probability, redux

- Worlds, random variables, events, sample space
- **Joint probabilities** of multiple connected variables
- **Conditional probabilities** of a variable, given another variable(s)
- **Marginalizing out** unwanted variables
- **Inference** from the joint probability

The big idea: figuring out the probability of variable(s) taking certain value(s)

2

Review: Bayesian Diagnostic Reasoning

- Bayes' rule says that
  - $P(H_i \mid E_1, \ldots, E_m) = P(E_1, \ldots, E_m \mid H_i)\, P(H_i) / P(E_1, \ldots, E_m)$
- Assume each piece of evidence E_i is conditionally independent of the others, **given** a hypothesis H_i, then:
  - $P(E_1, \ldots, E_m \mid H_i) = \prod^{l}_{j=1} P(E_j \mid H_i)$
- If we only care about relative probabilities for the H_i, then we have:
  - $P(H_i \mid E_1, \ldots, E_m) = \alpha\, P(H_i) \prod^{l}_{j=1} P(E_j \mid H_i)$

3

Next Up

- Bayesian networks
  - Network structure and independence
- Inference in Bayesian networks
  - Exact inference
  - Approximate inference

4

Review: Independence

What does it mean for A and B to be **independent**?

- $P(A) \perp\!\!\!\perp P(B)$
- A and B do not affect each other's probability
- $P(A \wedge B) = P(A)\, P(B)$

5

Review: Conditioning

What does it mean for A and B to be **conditionally independent given C?**

- A and B don't affect each other **if C is known**
- $P(\mathrm{A} \wedge \mathrm{B} \mid \mathrm{C}) = P(\mathrm{A} \mid \mathrm{C})\, P(\mathrm{B} \mid \mathrm{C})$

6

Review: Bayes' Rule

What is **Bayes' Rule**?

$$P(H_i \mid E_j) = \frac{P(E_j \mid H_i)P(H_i)}{P(E_j)}$$

What's it useful for?

- Diagnosis
- Effect is perceived, want to know (probability of) cause

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

R&N, 495–496

7

Review: Bayes' Rule

What is **Bayes' Rule**?

$$P(H_i \mid E_j) = \frac{P(E_j \mid H_i)P(H_i)}{P(E_j)}$$

What's it useful for?

- Diagnosis
- Effect is perceived, want to know (probability of) cause

$$P(hidden \mid observed) = \frac{P(observed \mid hidden)P(hidden)}{P(observed)}$$

R&N, 495–496

8

Review: Joint Probability

- What is the **joint probability** of A and B?
  - $P(\text{A,B})$
- The probability of **any pair** of legal assignments.
  - Generalizing to > 2, of course
- Booleans: expressed as a matrix/table

	alarm	¬ alarm
burglary	0.09	0.01
¬ burglary	0.1	0.8

≡

A	B	
T	T	0.09
T	F	0.1
F	T	0.01
F	F	0.8

- Continuous domains: probability functions

9

Review: Bayes' Nets: Big Picture

- Problems with full joint distribution tables as our probabilistic models:
  - Joint gets **way** too big to represent explicitly
    - Unless there are only a few variables
  - Hard to learn (estimate) anything empirically about more than a few variables at a time

	A		¬A	
	E	**¬E**	**E**	**¬E**
B	0.01	0.08	0.001	0.009
¬B	0.01	0.09	0.01	0.79

Slides derived from Matt E. Taylor, U Alberta

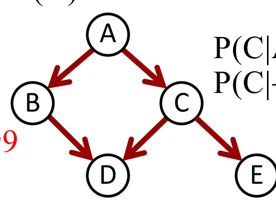
10

Review: Bayes' Nets

- Bayesian Network BN: **BN = (DAG, CPD)**
  - **DAG**: directed acyclic graph (BN's structure)
  - **CPD**: conditional probability distribution (BN's parameters)

P(A) = 0.001

P(B|A) = 0.3
P(B|¬A) = 0.001
P(¬B|A) = 0.7
P(¬B|¬A) = 0.999



P(C|A) = 0.2
P(C|¬A) = 0.005

P(E|C) = 0.4
P(E|¬C) = 0.002

P(D|B,C) = 0.1
P(D|B,¬C) = 0.01
P(D|¬B,C) = 0.01
P(D|¬B,¬C) = 0.00001

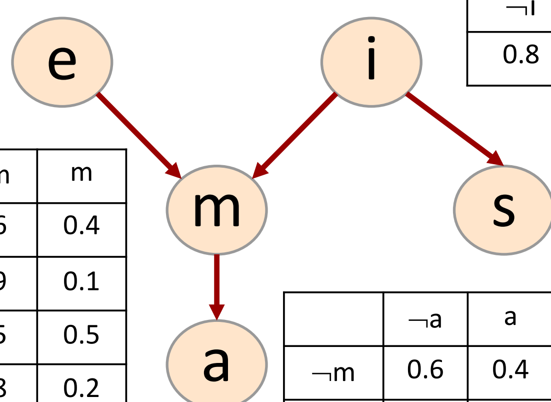
11

Review: Bayes' Nets

- P(a, m, i, e, s) = P(a | m) * P(m | i, e) * P(i) * P(e) * P(s | i)

¬e	e
0.7	0.3

¬i	i
0.8	0.2



	¬m	m
¬i, ¬e	0.6	0.4
¬i, e	0.9	0.1
i, ¬e	0.5	0.5
i, e	0.8	0.2

	¬s	s
¬i	0.75	0.25
i	0.4	0.6

	¬a	a
¬m	0.6	0.4
m	0.9	0.1

www.upgrad.com/blog/bayesian-network-example/

12

The Chain Rule

- $P(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n) =$ $P(\alpha_1) \times$
  $P(\alpha_2 \mid \alpha_1) \times$
  $P(\alpha_3 \mid \alpha_1 \wedge \alpha_2) \times \ldots \times$
  $P(\alpha_n \mid \alpha_1 \wedge \cdots \wedge \alpha_{n-1})$

  $= \prod_{i=1..n} P(\alpha_i \mid \alpha_1 \wedge \cdots \wedge \alpha_{i-1})$

  $= P(x_1,\ldots,x_n) = \Pi_{i=1}^{n} P(x_i \mid \pi_i)$

artint.info/html/ArtInt_143.html

13

The Chain Rule

$$P(x_1,\ldots,x_n) = \Pi_{i=1}^{n} P(x_i \mid \pi_i)$$

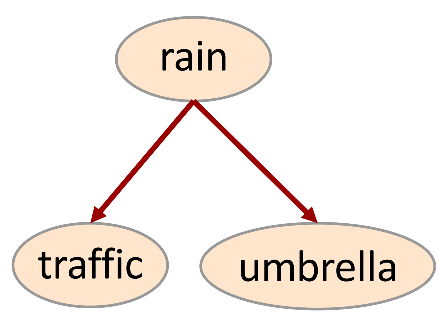
- Decomposition: $P(x_1,\ldots,x_n) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x2)\ldots$

  $P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$
  $P(\text{Rain})\ P(\text{Traffic} \mid \text{Rain})\ P(\text{Umbrella} \mid \text{Rain}, \text{Traffic})$

- With assumption of conditional independence:

  $P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$
  $P(\text{Rain})\ P(\text{Traffic} \mid \text{Rain})\ P(\text{Umbrella} \mid \text{Rain})$

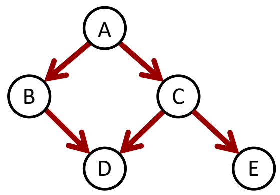
- Bayes' nets express conditional independences
  - (Assumptions)



Slides derived from Matt E. Taylor, U Alberta

14

Chaining: Example



Computing the joint probability for all variables is easy:

$$
\begin{aligned}
P(a, b, c, d, e) &= P(e \mid a, b, \boldsymbol{c}, d)\, P(a, b, c, d) && \leftarrow \text{By product rule} \\
&= P(e \mid c)\, P(a, b, c, d) && \leftarrow \text{By conditional independence assumption} \\
&= P(e \mid c)\, P(d \mid a, \boldsymbol{b}, \boldsymbol{c})\, P(a, b, c) \\
&= P(e \mid c)\, P(d \mid b, c)\, P(c \mid \boldsymbol{a}, b)\, P(a, b) \\
&= P(e \mid c)\, P(d \mid b, c)\, P(c \mid a)\, P(b \mid a)\, P(a)
\end{aligned}
$$

We're reducing distributions–P(x,y)–to single values.

15

Topological Semantics

- A node is **conditionally independent** of its non-descendants given its parents
- A node is **conditionally independent** of all other nodes in the network given its parents, children, and children's parents (also known as its **Markov blanket**)
  - (For much later: a method called d-separation can be applied to decide whether a set of nodes X is independent of a set Y, given a third set Z)

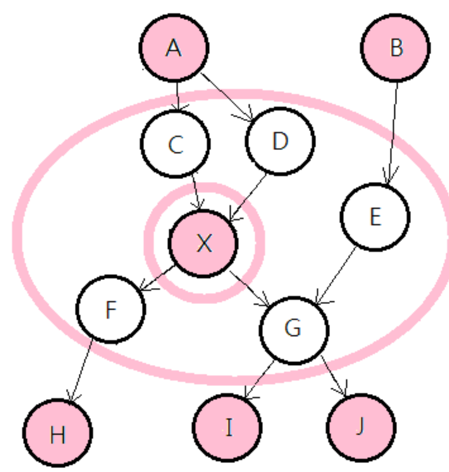
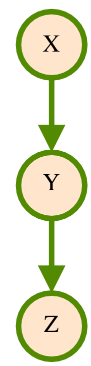


Image: mitsai1974.github.io/DevBlog/2018/07/11/bayesian-ml-net-profound

16

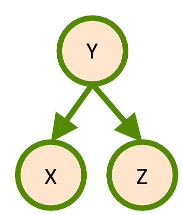
Independence and Causal Chains

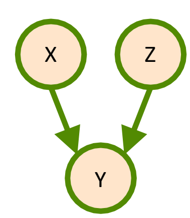
- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, we can it prove using algebra (tedious)
  - If no, can prove it with a counter-example
- Question: are X and Z necessarily independent?
  - No.
  - Ex: Clouds (X) cause rain (Y), which causes traffic (Z)
  - X can influence Z, Z can influence X (via Y)
- This configuration is a "causal chain"



Two More Main Patterns

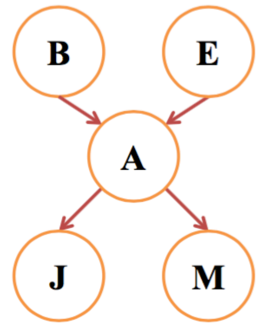
- Common Cause:
  - Y causes X and Y causes Z
  - Are X and Z independent? **No**
  - Are X and Z independent given Y? **Yes**
- Common Effect:
  - Two causes of one effect
  - Are X and Z independent? **Yes**
  - Are X and Z independent given Y?
  - No!
  - Observing an effect "activates" influence between possible causes.





Conditionality Example

- Hidden: A, B, E. You don't know:
  - If there's a burglar.
  - If there was an earthquake.
  - If the alarm is going off.
- Observed: J and M.
  - John and/or Mary have some chance of calling if the alarm rings.
  - You know who called you.

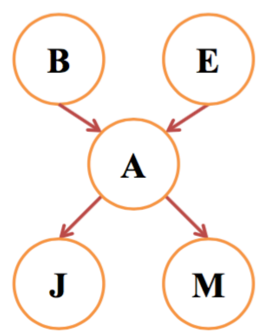


Slides derived from Matt E. Taylor, WSU

19

Conditionality Example 2

- At first:
  - Is the probability of John calling affected by whether there's an earthquake?
  - Is the probability of Mary calling affected by John calling?
- Your alarm is going off!
  - Is the probability of Mary calling affected by John calling?

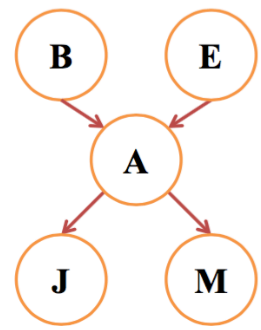


Slides derived from Matt E. Taylor, WSU

20

Conditionality Example 3

- At first:
  - Is whether there's an earthquake affected by whether there's a burglary in progress (and vice versa)?
- Your alarm is going off!
  - Does the probability a burglary is happening depend on whether there's an earthquake?



Slides derived from Matt E. Taylor, WSU

21

Representational Extensions

- Conditional probability tables (CPTs) for large networks can require a large number of parameters
  - $O(2^k)$ where k is the branching factor of the network
- There are ways of compactly representing CPTs
  - Deterministic relationships
  - Noisy-OR
  - Noisy-MAX
- What about continuous variables?
  - Discretization
  - Use density functions (usually mixtures of Gaussians) to build hybrid Bayesian networks (with discrete and continuous variables)

23

Bayes' Net Inference




Some material borrowed from Lise Getoor

24

24

Inference Tasks

- **Simple queries:** Compute posterior marginal $P(X_i \mid E=value)$
  - E.g., P(NoGas | Gauge=*empty*, Lights=*on*, Starts=*false*)
- **Conjunctive queries:**
  - $P(X_i, X_j \mid E=value) = P(X_i \mid E=value)\, P(X_j \mid X_i, E=value)$
- **Optimal decisions:**
  - *Decision networks* include utility information
  - Probabilistic inference gives P(outcome | action, evidence)
- **Value of information:** Which evidence should we seek next?
- **Sensitivity analysis:** Which probability values are most critical?
- **Explanation:** Why do I need a new starter motor?

25

Direct Inference with BNs

- Instead of computing the joint, suppose we just want the probability for one variable.
- Exact methods of computation:
  - **Enumeration**
  - **Variable elimination**
  - **Join trees:** get the probabilities associated with every query variable

27

Inference by Enumeration

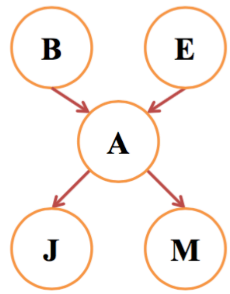
Reminder: P(E) is known (observed), so 1/P(E) is a constant that makes everything sum to 1: the *normalizing constant*

- Add all of the terms (atomic event probabilities) from the full joint distribution
- If **E** are the evidence (observed) variables and **Y** are the other (unobserved) variables, then:
  - $P(X \mid \mathbf{E}) = \alpha\, P(X, \mathbf{E}) = \alpha \sum P(X, \mathbf{E}, \mathbf{Y})$
- Each P(X, **E, Y**) term can be computed using the chain rule
- Computationally expensive!

28

Example 1: Enumeration

- Recipe:
  - State the marginal probabilities you need
  - Figure out ALL the atomic probabilities you need
  - Calculate and combine them
- Example:
  - $P(+b \mid +j, +m) = \dfrac{P(+b, +j, +m)}{P(+j, +m)}$

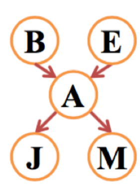


29

Example 1 cont'd

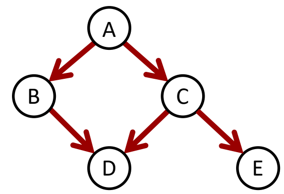
$$P(+b,+j,+m) =$$
$$P(+b)P(+e)P(+a|+b,+e)P(+j|+a)P(+m|+a)+$$
$$P(+b)P(+e)P(-a|+b,+e)P(+j|-a)P(+m|-a)+$$
$$P(+b)P(-e)P(+a|+b,-e)P(+j|+a)P(+m|+a)+$$
$$P(+b)P(-e)P(-a|+b,-e)P(+j|-a)P(+m|-a)$$

P(+m | +b, +e)?



30

Example 2: Enumeration



- $P(x_i) = \Sigma_{\pi_i} P(x_i \mid \pi_i) \, P(\pi_i)$
- Say we want to know $P(D{=}t)$
- Only E is *given* as true
- $P(d \mid e) = \alpha \, \Sigma_{ABC} P(a, b, c, d, e)$ *(reminder: $\alpha = 1/P(e)$)*
  $= \alpha \, \Sigma_{ABC} P(a) \, P(b \mid a) \, P(c \mid a) \, P(d \mid b, c) \, P(e \mid c)$
- With simple iteration, that's a lot of repetition!
  - $P(e \mid c)$ has to be recomputed every time we iterate over C=true

31

Variable Elimination

- Basically just enumeration with caching of local calculations
- Linear for polytrees (singly connected BNs)
- Potentially exponential for multiply connected BNs
  - **Exact inference in Bayesian networks is NP-hard!**
- Join tree algorithms are an extension of variable elimination methods that compute posterior probabilities for all nodes in a BN simultaneously

32

Variable Elimination Approach

- General idea:
- Write query in the form

$$P(X_n, e) = \sum_{x_k} \cdots \sum_{x_3} \sum_{x_2} \prod_i P(x_i \mid pa_i)$$

  - Note that there is no α term here
  - It's a conjunctive probability, not a conditional probability…
- Iteratively
  - Move all irrelevant terms outside of innermost sum
  - Perform innermost sum, getting a new term
  - Insert the new term into the product

33

Variable Elimination: Example

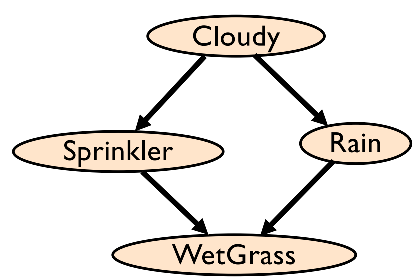
$$\mathrm{P(w)} = \sum_{\mathrm{r,s,c}} \mathrm{P(w \mid r,s)P(r \mid c)P(s \mid c)P(c)}$$

$$= \sum_{\mathrm{r,s}} \mathrm{P(w \mid r,s)} \sum_{\mathrm{c}} \mathrm{P(r \mid c)P(s \mid c)P(c)}$$

$$= \sum_{\mathrm{r,s}} \mathrm{P(w \mid r,s)f_1(r,s)}$$

$\mathrm{f_1(r,s)}$

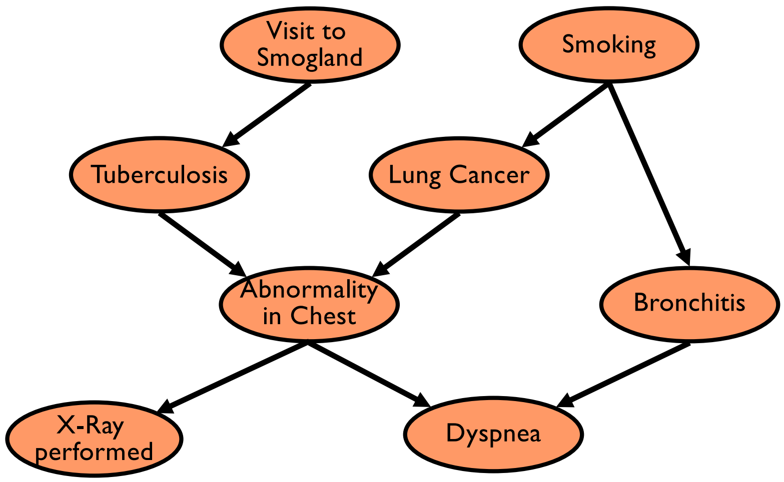
"factors"



34

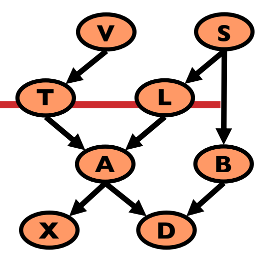
A More Complex Example

- "Lungs" network:



35

Lungs 1



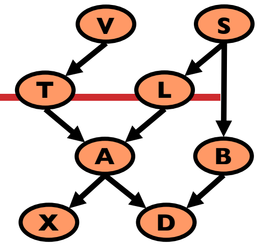
- We want to compute *P(d)*
- Need to eliminate: *v,s,x,t,l,a,b*

Initial factors:

$$P(v)P(s)P(t \mid v)P(l \mid s)P(b \mid s)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$

36

Lungs 2



- We want to compute *P(d)*
- Need to eliminate: *v,s,x,t,l,a,b*

Initial factors:

$$\underline{P(v)}P(s)\underline{P(t \mid v)}P(l \mid s)P(b \mid s)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$

Eliminate: *v*

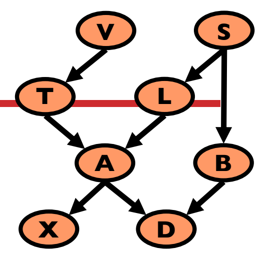
Compute: $f_v(t) = \sum_v P(v)P(t \mid v)$

$$\Rightarrow \underline{f_v(t)}P(s)P(l \mid s)P(b \mid s)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$

- Note: $f_v(t) = P(t)$
- Result of elimination is not **necessarily** a probability term

37

Lungs 3



- We want to compute *P(d)*
- Need to eliminate: *s,x,t,l,a,b*

Initial factors:

$$P(v)P(s)P(t \mid v)P(l \mid s)P(b \mid s)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$

$$\Rightarrow f_v(t)\underline{P(s)}\underline{P(l \mid s)}\underline{P(b \mid s)}P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$

Eliminate: *s*

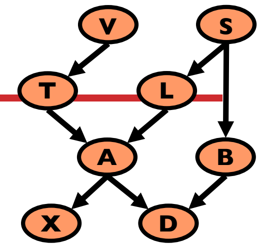
Compute: $f_s(b,l) = \sum_s P(s)P(b \mid s)P(l \mid s)$

$$\Rightarrow f_v(t)\underline{f_s(b,l)}P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$

- Summing on *s* results in a factor with two arguments $f_s(b,l)$
- In general, result of elimination may be a function of several variables

38

Lungs 4



- We want to compute *P(d)*
- Need to eliminate: *x,t,l,a,b*

Initial factors

$$P(v)P(s)P(t \mid v)P(l \mid s)P(b \mid s)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$
$$\Rightarrow f_v(t)P(s)P(l \mid s)P(b \mid s)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$

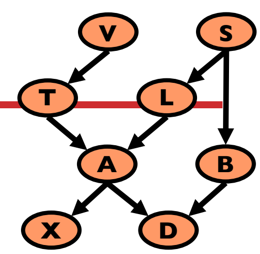
Eliminate: x $\quad \Rightarrow f_v(t)f_s(b,l)P(a \mid t,l)P(x \mid a)P(d \mid \underline{a,b})$

Compute: $f_x(a) = \sum_x P(x \mid a)$

$$\Rightarrow f_v(t)f_s(b,l)\underline{f_x(a)}P(a \mid t,l)P(d \mid a,b)$$

39

Lungs 5



- We want to compute *P(d)*
- Need to eliminate: *t,l,a,b*

Initial factors $\quad P(v)P(s)P(t \mid v)P(l \mid s)P(b \mid s)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$

$$\Rightarrow f_v(t)P(s)P(l \mid s)P(b \mid s)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$
$$\Rightarrow \underline{f_v(t)}f_s(b,l)f_x(a)\underline{P(a \mid t,l)}P(d \mid a,b)$$

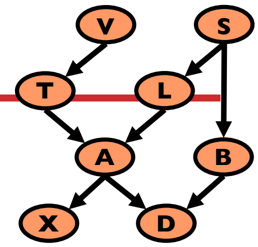
Eliminate: t

Compute: $f_t(a,l) = \sum_t f_v(t)P(a \mid t,l)$

$$\Rightarrow f_s(b,l)f_x(a)\underline{f_t(a,l)}P(d \mid a,b)$$

40

Lungs 6



- We want to compute *P(d)*
- Need to eliminate: *l,a,b*

Initial factors $P(v)P(s)P(t\,|\,v)P(l\,|\,s)P(b\,|\,s)P(a\,|\,t,l)P(x\,|\,a)P(d\,|\,a,b)$

$$\Rightarrow f_v(t)P(s)P(l\,|\,s)P(b\,|\,s)P(a\,|\,t,l)P(x\,|\,a)P(d\,|\,a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)P(a\,|\,t,l)P(x\,|\,a)P(d\,|\,a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)f_x(a)P(a\,|\,t,l)P(d\,|\,a,b)$$
$$\Rightarrow f_s(b,l)f_x(a)f_t(a,l)P(d\,|\,a,b)$$

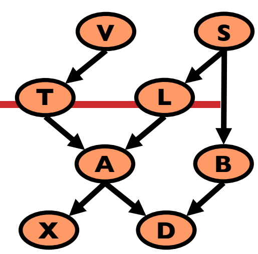
Eliminate: l

Compute: $f_l(a,b) = \sum_l f_s(b,l)f_t(a,l)$

$$\Rightarrow f_l(a,b)f_x(a)P(d\,|\,a,b)$$

41

Lungs Finale



- We want to compute *P(d)*
- Need to eliminate: *b*

Initial factors $P(v)P(s)P(t\,|\,v)P(l\,|\,s)P(b\,|\,s)P(a\,|\,t,l)P(x\,|\,a)P(d\,|\,a,b)$

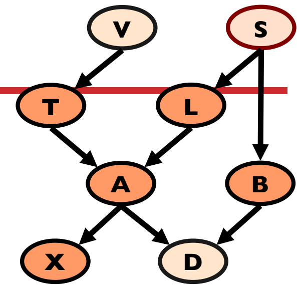
$$\Rightarrow f_v(t)P(s)P(l\,|\,s)P(b\,|\,s)P(a\,|\,t,l)P(x\,|\,a)P(d\,|\,a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)P(a\,|\,t,l)P(x\,|\,a)P(d\,|\,a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)f_x(a)P(a\,|\,t,l)P(d\,|\,a,b)$$
$$\Rightarrow f_s(b,l)f_x(a)f_t(a,l)P(d\,|\,a,b)$$
$$\Rightarrow f_l(a,b)f_x(a)P(d\,|\,a,b) \Rightarrow f_a(b,d) \Rightarrow f_b(d)$$

Eliminate: a,b

Compute: $f_a(b,d) = \sum_a f_l(a,b)f_x(a)p(d\,|\,a,b) \quad f_b(d) = \sum_b f_a(b,d)$

42

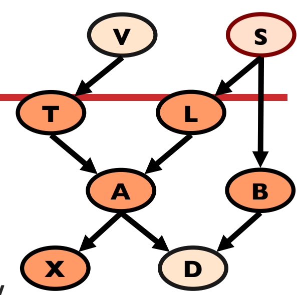
Dealing with Evidence



- How do we deal with evidence?
  - And what is "evidence?"
  - Variables whose value has been observed
- Suppose we are given evidence: $V = t, S = f, D = t$
- We want to compute $P(L, V = t, S = f, D = t)$

44

Dealing with Evidence



- We start by writing the factors:

$$P(v)P(s)P(t \mid v)P(l \mid s)P(b \mid s)P(a \mid t,l)P(x \mid a)P(d \mid a,b)$$

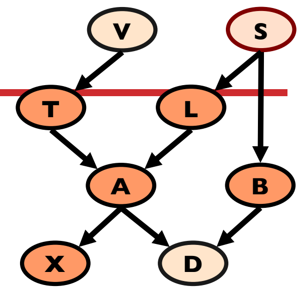
- Since we know that V = t, we don't need to eliminate V
- Instead, we can replace the factors P(V) and P(T |V) with

$$f_{P(V)} = P(V = t) \quad f_{p(T|V)}(T) = P(T \mid V = t)$$

- These "select" appropriate parts of original factors given evidence
- Note that $f_{P(V)}$ is a constant, so does not appear in elimination of other variables

45

Dealing with Evidence

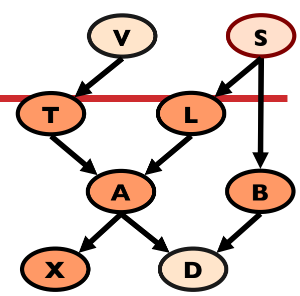


- So now…
  - Given evidence $V = t,\ S = f,\ D = t$
  - Compute $P(L,\ V = t,\ S = f,\ D = t\ )$
  - Initial factors, after setting evidence:

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a \mid t,l) P(x \mid a) f_{P(d|a,b)}(a,b)$$

46

Dealing with Evidence



- Given evidence $V = t,\ S = f,\ D = t$, we want to compute $P(L,\ V = t,\ S = f,\ D = t\ )$
- Initial factors, after setting evidence:

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a \mid t,l) P(x \mid a) f_{P(d|a,b)}(a,b)$$

- Eliminating x, we get

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a \mid t,l) f_x(a) f_{P(d|a,b)}(a,b)$$

- Eliminating t, we get

$$f_{P(v)} f_{P(s)} f_{P(l|s)}(l) f_{P(b|s)}(b) f_t(a,l) f_x(a) f_{P(d|a,b)}(a,b)$$

- Eliminating a, we get

$$f_{P(v)} f_{P(s)} f_{P(l|s)}(l) f_{P(b|s)}(b) f_a(b,l)$$

- Eliminating b, we get

$$f_{P(v)} f_{P(s)} f_{P(l|s)}(l) f_b(l)$$

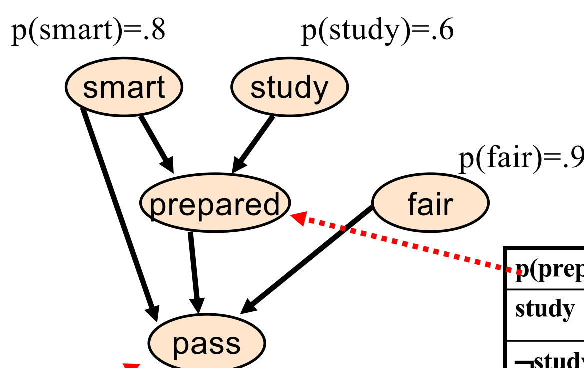
47

Variable Elimination Algorithm

- Let $X_1, \ldots, X_m$ be an ordering on the non-query variables
- For i = m, …, 1 $\sum_{X_1}\sum_{X_2}\ldots\sum_{X_m}\prod_j P(X_j \mid Parents(X_j))$
  - In the summation for X_i, leave only factors mentioning X_i
  - Multiply the factors, getting a factor that contains a number for each value of the variables mentioned, including X_i
  - Sum out X_i, getting a factor f that contains a number for each value of the variables mentioned, not including X_i
  - Replace the multiplied factor in the summation

48

Exercise: Variable Elimination



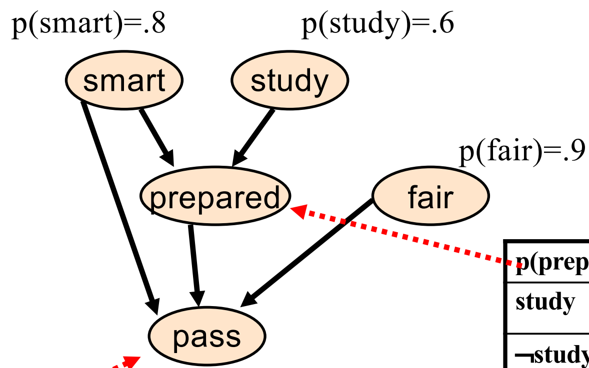
p(prep|…)	smart	¬smart
study	.9	.7
¬study	.5	.1

p(pass|…)	smart		¬smart	
	prep	¬prep	prep	¬prep
fair	.9	.7	.7	.2
¬fair	.1	.1	.1	.1

Query: What is the probability that a student **studied**, given that they pass the exam?

49

Exercise: Variable Elimination



p(prep|…)	**smart**	**¬smart**
study	.9	.7
¬study	.5	.1

p(pass|…)	**smart**		**¬smart**	
	prep	**¬prep**	**prep**	**¬prep**
fair	.9	.7	.7	.2
¬fair	.1	.1	.1	.1

Query: What is the probability that a student **is smart**, given that they pass the exam?

50

Summary

- Bayes nets
  - Structure
  - Parameters
  - Conditional independence
  - Chaining
- BN inference
  - Enumeration
  - Variable elimination
  - Sampling methods

51