Bayes IV
(Exact Inference in Bayesian Nets)

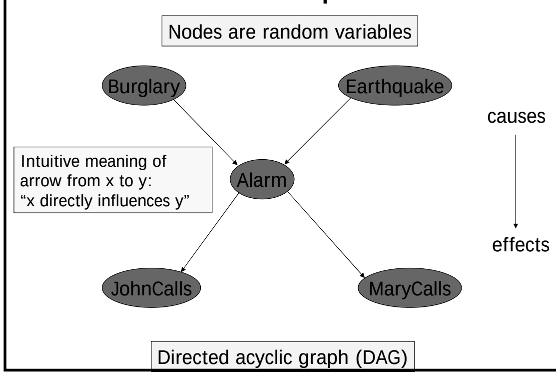
Lecture #19

11/06/08

Announcements

- Programming assignment #2 is now past due
- Late assignments accepted until the 10th
  - 20% penalty

Review: A Simple Network



Review: Conditional independence

- Two variables *X* and *Y* are **conditionally independent** given *Z* if
  - $P(X = x|Y = y,Z=z) = P(X = x|Z=z)$ for all values x,y,z
  - That is, learning the values of *Y* does not change prediction of *X* once we know the value of *Z*
  - Notation: $I(X ; Y | Z)$

Where are we?

- We know the basics of probability, e.g.
  - The probability of an event P(e)
  - Conditional independence: P(e;f | x)
- We know how to represent networks of causal events
- But… *how do we infer probabilities given a causal network and a set of observations?*

If Mary calls and John calls, do I believe there was a burglary?

Two Steps:

1. First, we need to be able to calculate P(X) for an arbitrary node X in a Bayes Net *without* any evidence
2. Then we will compute P(X|E) = P(X&E)/P(E)

Inference in BN

- Set E of *evidence variables* that are *observed*, e.g., {JohnCalls,MaryCalls}
- *Query variable* X, e.g., *Burglary*, for which we would like to know the posterior probability distribution P(X|E)

J	M	P(B|...)
T	T	?

Distribution conditional to the observations made

Inference in BN

Simplest case:

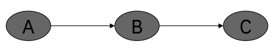
$$P(B) = \sum_A P(A,B) = \sum_A P(B|A)P(A)$$

For Boolean variables:

P(b) = P(b|a)P(a) + P(b|¬a) P(¬a)

P(¬b) = P(¬b|a)P(a) + P(¬b|¬a) P(¬a)

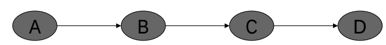
Inference on a chain



$$P(C) = \sum_{A,B} P(C \mid B) P(B \mid A) P(A)$$

$$P(C) = \sum_{B} P(C \mid B) \sum_{A} (B \mid A) P(A)$$

Inference on a chain



- For a chain with n nodes where each node has k possible values the complexity is *O(n k^2).*

Inference on a chain

Summary:

$$P(D) = \sum_{A,B,C} P(A,B,C,D) = \sum_{A,B,C} P(D|C)P(C|B)P(B|A)P(A) = \sum_C P(D|C) \sum_B P(C|B) \sum_A P(B|A)P(A)$$

Inference on a chain

We can perform variable elimination using a different order:

$$P(D) = \sum_{A,B,C} P(A)P(B \mid A)P(C \mid B)P(D \mid C)$$

$$P(D) = \sum_A P(A) \sum_B P(B \mid A) \sum_C P(C \mid B)P(D \mid C)$$

$$P(D) = \sum_A P(A) \sum_B P(B \mid A) f_c(B,D)$$

$$P(D) = \sum_A P(A) f_B(A,D)$$

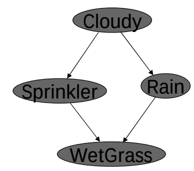
Variable Elimination

- Suppose we're interested in $P(X_k)$
- Write query in the form

$$P(X_k) = \sum_{X_n,\ldots,X_{k+1},X_{k-1},\ldots,X_1} \prod_i P(X_i | Pa(X_i))$$

- Iteratively
  - Move all irrelevant terms outside of innermost sum
  - Perform innermost sum, getting a new term
  - Insert the new term into the product

Inference Ex. 2

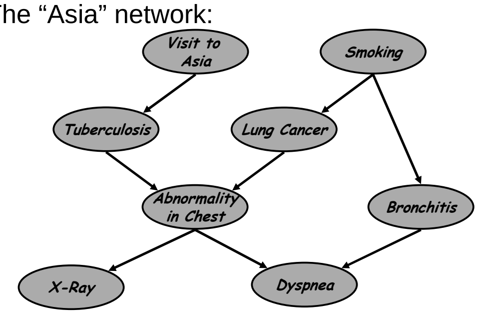


$$P(W) = \sum_{R,S,C} P(w | r,s)P(r | c)P(s | c)P(c)$$
$$= \sum_{R,S} P(w | r,s) \sum_{C} P(r | c)P(s | c)P(c)$$
$$= \sum_{R,S} P(w | r,s) f_C(r,s)$$

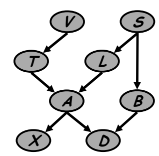
$f_C(R,S)$

A More Complex Example

The "Asia" network:



Example (cont)

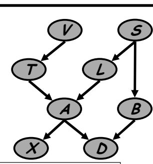


Suppose we want to compute P(d)

The form of the joint distribution:

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

Example (cont)



Need to eliminate: v,s,x,t,l,a,b

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

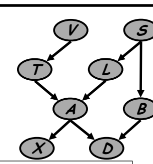
Eliminate: v

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

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

Note: $f_v(t) = P(t)$
In general, result of elimination is not necessarily a probability term

Example (cont)



Need to eliminate: s,x,t,l,a,b

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

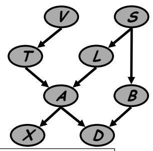
Eliminate: s

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

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

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

Example (cont)



Need to eliminate: *x,t,l,a,b*

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

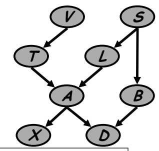
Eliminate: *x*

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

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

Note: $f_x(a) = 1$ for all values of *a*. In fact, every variable that is not an ancestor of the query or evidence variables can be ignored!

Example (cont)



Need to eliminate: *t,l,a,b*

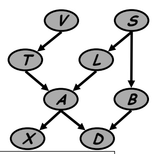
$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$
$$\Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)P(a|t,l)P(x|a)P(d|a,b)$$
$$\Rightarrow \underline{f_v(t)}f_s(b,l)f_x(a)\underline{P(a|t,l)}P(d|a,b)$$

Eliminate: *t*

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

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

Example (cont)



Need to eliminate: *l,a,b*

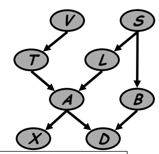
$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$
$$\Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)P(a|t,l)P(x|a)P(d|a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)f_x(a)P(a|t,l)P(d|a,b)$$
$$\Rightarrow \underline{f_s(b,l)}f_x(a)\underline{f_t(a,l)}P(d|a,b)$$

Eliminate: *l*

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

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

Example (cont)



Need to eliminate: *a,b*

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

Eliminate: *a,b*

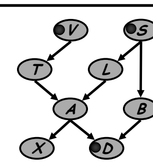
Compute:

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

Variable Elimination

- We now understand variable elimination as a sequence of **rewriting** operations
- Computation depends on order of elimination

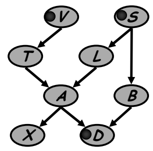
Dealing with Evidence



- How do we deal with evidence?
- Suppose get evidence $V = t, S = f, D = t$ and want to compute $P(L | V = t, S = f, D = t)$

$$P(L|V=t,S=f,D=t)=\frac{P(L,V=t,S=f,D=t)}{P(V=t,S=f,D=t)}$$

Dealing with Evidence



- We start by writing the factors:

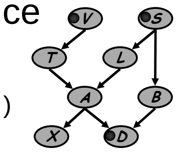
$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$

- 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

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

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

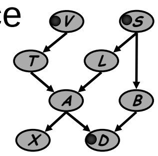
Dealing with Evidence



- Compute P(L, V = t, S = f, D = t)
- Initial factors, after setting evidence:

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

Dealing with Evidence



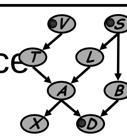
- Compute P(L, V = t, S = f, D = t)
- Initial factors, after setting evidence:

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

- Eliminating X:

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

Dealing with Evidence



- Initial factors, after setting evidence:

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

- Eliminating x, we get

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

- 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

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

Variable Elimination

- Compute the probability of X_k given values to evidence variables E

$$P(X_k, E) = \sum_{\text{non query, non-evidence variables}} \prod_i P(X_i|Pa(X_i))$$

- Algorithm is same as before, with no need to perform summation with respect to evidence variables.

Complexity of inference

Thm:

Computing P(X = x) in a Bayesian network is NP-hard

Approaches to inference

- Exact inference
  - Variable elimination
  - Join tree algorithm
- Approximate inference
  - Simplify the structure of the network to make exact inference efficient (variational methods, loopy belief propagation)
- Probabilistic methods
  - Stochastic simulation / sampling methods
  - Markov chain Monte Carlo methods