CS 188: Artificial Intelligence
Spring 2006

Lecture 17: Bayes' Nets III
3/16/2006

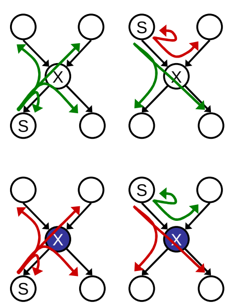
Dan Klein – UC Berkeley

Today

- Last time:
  - Bayes nets
  - Conditional independence
- Today:
  - More conditional independence
  - Inference to answer queries

Reachability (the Bayes' Ball)

- Correct algorithm:
  - Start at source node
  - Try to reach target by search
  - States: node, along with previous arc
  - Successor function:
    - Unobserved nodes:
      - To any child
      - To any parent if coming from a child (or start)
    - Observed nodes:
      - From parent to parent
  - If you can't reach a node, it's conditionally independent of the start node



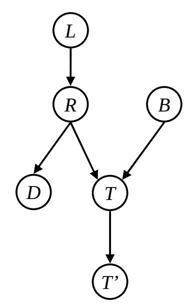
Example

$L \perp\!\!\!\perp T' | T$ *Yes*

$L \perp\!\!\!\perp B$ *Yes*

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$



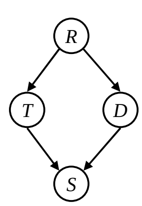
Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I'm sad
- Questions:

$T \perp\!\!\!\perp D$

$T \perp\!\!\!\perp D | R$ *Yes*

$T \perp\!\!\!\perp D | R, S$



Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- The Bayes' ball algorithm (aka d separation)
- A Bayes net may have other independencies that are not detectable until you inspect its specific distribution

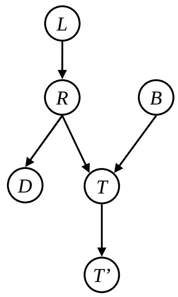
Inference

- Inference: calculating some statistic from a joint probability distribution
- Examples:
  - Posterior marginal probability:

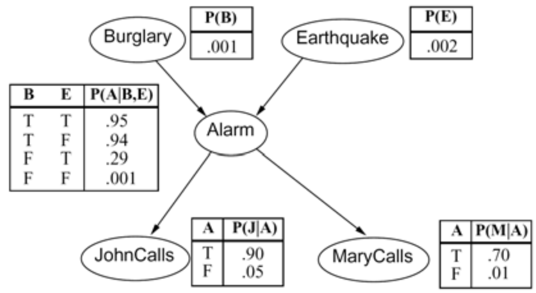
$$P(Q|E_1 = e_1, \ldots E_k = e_k)$$

  - Most likely explanation:

$$\text{argmax}_q \; P(Q = q|E_1 = e_1 \ldots)$$



Reminder: Alarm Network



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

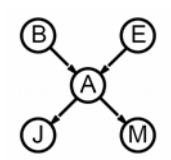
A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Atomic Inference

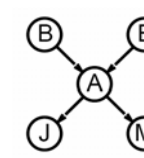
- Given unlimited time, inference in BNs is easy
- Recipe:
  - State the marginal probabilities you want
  - Figure out ALL the atomic probabilities you need
  - Calculate and combine them
- Example:

$$P(b|j,m) = \frac{P(b,j,m)}{P(j,m)}$$



Example

$$P(b|j,m) = \frac{P(b,j,m)}{P(j,m)}$$



$$\begin{aligned} P(b,j,m) = \; & P(b,e,a,j,m) + \\ & P(b,\bar{e},a,j,m) + \\ & P(b,e,\bar{a},j,m) + \\ & P(b,\bar{e},\bar{a},j,m) \\ = & \sum_{e,a} P(b,e,a,j,m) \end{aligned}$$

Where did we use the BN structure?

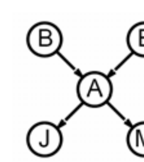
We didn't!

Example

$$\begin{aligned} P(b,j,m) = & \\ & P(b)P(e)P(a|b,e)P(j|a)P(m|a) + \\ & P(b)P(e)P(\bar{a}|b,e)P(j|\bar{a})P(m|\bar{a}) + \\ & P(b)P(\bar{e})P(a|b,\bar{e})P(j|a)P(m|a) + \\ & P(b)P(\bar{e})P(\bar{a}|b,\bar{e})P(j|\bar{a})P(m|\bar{a}) \end{aligned}$$

Example

$$P(b|j,m) = \frac{P(b,j,m)}{P(j,m)}$$



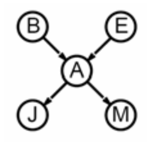
$$P(b,j,m) = \sum_{e,a} P(b,e,a,j,m)$$

$$P(\bar{b},j,m) = \sum_{e,a} P(\bar{b},e,a,j,m)$$

$$\begin{pmatrix} P(b,j,m) \\ P(\bar{b},j,m) \end{pmatrix} \xrightarrow{\text{Normalize}} \begin{pmatrix} P(b|j,m) \\ P(\bar{b}|j,m) \end{pmatrix}$$

Inference by Enumeration

- Atomic inference is extremely slow!
- Slightly clever way to save work:
  - Move the sums as far right as possible
  - Example:



$$P(b,j,m) = \sum_{e,a} P(b,e,a,j,m)$$

$$= \sum_{e,a} P(b)P(e)P(a|b,e)P(j|a)P(m|a)$$

$$= P(b)\sum_{e} P(e)\sum_{a} P(a|b,e)P(j|a)P(m|a)$$

Example

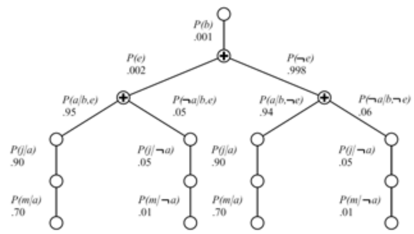
$$\begin{aligned} P(b,j,m) = \; & P(b)P(e)P(a|b,e)P(j|a)P(m|a) + \\ & P(b)P(e)P(\bar{a}|b,e)P(j|\bar{a})P(m|\bar{a}) + \\ & P(b)P(\bar{e})P(a|b,\bar{e})P(j|a)P(m|a) + \\ & P(b)P(\bar{e})P(\bar{a}|b,\bar{e})P(j|\bar{a})P(m|\bar{a}) \end{aligned}$$

$$\oplus \qquad \oplus$$

$$P(b)\begin{cases} P(e)\begin{cases} P(a|b,e) & P(j|a)P(m|a) \\ P(\bar{a}|b,e) & P(j|\bar{a})P(m|\bar{a}) \end{cases} \\ P(\bar{e})\begin{cases} P(a|b,\bar{e}) & P(j|a)P(m|a) \\ P(\bar{a}|b,\bar{e}) & P(j|\bar{a})P(m|\bar{a}) \end{cases} \end{cases}$$

Evaluation Tree

- View the nested sums as a computation tree:

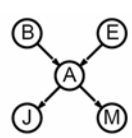


- Still repeated work: calculate P(m | a) P(j | a) twice, etc.

Variable Elimination: Idea

- Lots of redundant work in the computation tree!
- We can save time if we cache all partial results
- This is the basic idea behind variable elimination

Basic Objects



- Track objects called factors
- Initial factors are local CPTs

$$\underbrace{P(B)}_{f_B(B)} \qquad \underbrace{P(J|A)}_{f_J(A,J)} \qquad \underbrace{P(A|B,E)}_{f_A(A,B,E)}$$

- During elimination, create new factors
- Anatomy of a factor:

4 numbers, one for each value of D and E

$$f_{A\bar{B}\bar{C}D}(D,E)$$

Variables introduced

Variables summed out

Argument variables, always non-evidence variables

Basic Operations

- First basic operation: join factors
- Combining two factors:
  - Just like a database join
  - Build a factor over the union of the domains
- Example:

$$f_1(A,B) \times f_2(B,C) \Longrightarrow f_3(A,B,C)$$

$$f_3(a,b,c) = f_1(a,b)\cdot f_2(b,c)$$

$$\text{“}P(a,b|c) = P(a|b)\cdot P(b|c)\text{”}$$

Basic Operations

- Second basic operation: marginalization
- Take a factor and sum out a variable
  - Shrinks a factor to a smaller one
  - A projection operation
- Example:

$$f_{\bar{A}B}(b) = \sum_a f_{AB}(a,b)$$

$$\text{“}P(b) = \sum_a P(a,b)\text{”}$$

Example

$P(b, j, m)$

$$= \underbrace{P(b)}_{B} \textstyle\sum_e \underbrace{P(e)}_{E} \textstyle\sum_a \underbrace{P(a|b,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M}$$

$$= f_B(b) \sum_e f_E(e) \sum_a f_A(a,b,e) f_J(a) f_M(a)$$

$$= f_B(b) \sum_e f_E(e) \sum_a f_{AJM}(a,b,e)$$

$$= f_B(b) \sum_e f_E(e) f_{\bar{A}JM}(b,e)$$

Example

$P(b, j, m)$

$$= f_B(b) \sum_e f_E(e) f_{\bar{A}JM}(b,e)$$

$$= f_B(b) \sum_e f_{\bar{A}EJM}(b,e)$$

$$= f_B(b) f_{\bar{A}\bar{E}JM}(b)$$

$$= f_{\bar{A}B\bar{E}JM}(b)$$

General Variable Elimination

- Query: $P(Q|E_1 = e_1, \dots E_k = e_k)$
- Start with initial factors:
  - Local CPTs (but instantiated by evidence)
- While there are still hidden variables (not Q or evidence):
  - Pick a hidden variable H
  - Join all factors mentioning H
  - Project out H
- Join all remaining factors and normalize

Example

$P(B|j,m) \propto P(B,j,m)$

$$\underbrace{P(B)}_{f_B(B)} \quad \underbrace{P(E)}_{f_E(E)} \quad \underbrace{P(A|B,E)}_{f_A(A,B,E)} \quad \underbrace{P(j|A)}_{f_J(A)} \quad \underbrace{P(m|A)}_{f_M(A)}$$

Choose A

$$f_A(A,B,E),\ f_J(A),\ f_M(A) \xrightarrow{\times} f_{AJM}(A,B,E) \xrightarrow{\Sigma} f_{\bar{A}JM}(B,E)$$

$$\boxed{f_B(B) \quad f_E(E) \quad f_{\bar{A}JM}(B,E)}$$

Example

$$\boxed{f_B(B) \quad f_E(E) \quad f_{\bar{A}JM}(B,E)}$$

Choose E

$$f_E(E),\ f_{\bar{A}JM}(B,E) \xrightarrow{\times} f_{\bar{A}EJM}(B,E) \xrightarrow{\Sigma} f_{\bar{A}\bar{E}JM}(B)$$

$$\boxed{f_B(B) \quad f_{\bar{A}\bar{E}JM}(B)}$$

Finish

$$f_B(B),\ f_{\bar{A}\bar{E}JM}(B) \xrightarrow{\times} f_{\bar{A}B\bar{E}JM}(B) \xrightarrow{\text{Normalize}} P(B|j,m)$$