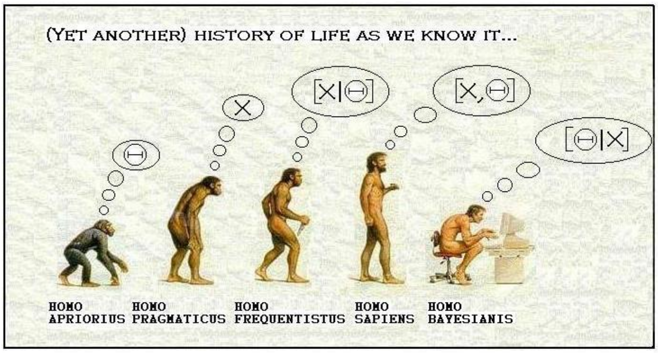
CSE 473

Lecture 21
(Chapter 14)

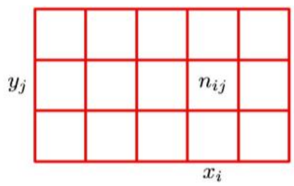
Bayesian Networks



(Courtesy Mike West)

Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

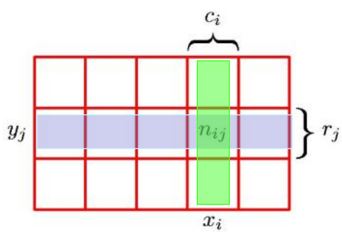


Total number of events = *N*

Marginal Probability

$$P(X = x_i) = \sum_j P(x_i, y_j) = \frac{c_i}{N}$$

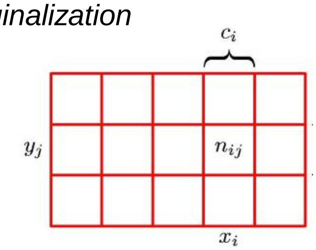
$$P(Y = y_j) = \sum_i P(x_i, y_i) = \frac{r_j}{N}$$



Summing out a variable is called *marginalization*

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$



Recall: Bayes' Rule

$$P(x,y) = P(x \mid y)P(y) = P(y \mid x)P(x)$$

i.e.

$$P(x \mid y) = \frac{P(y \mid x)\ P(x)}{P(y)}$$

3

Normalization in Bayes' Rule

$$P(x \mid y) = \frac{P(y \mid x)\ P(x)}{P(y)} = \alpha\, P(y \mid x)\, P(x)$$

$$\alpha = \frac{1}{P(y)} = \frac{1}{\sum_x P(y,x)} = \frac{1}{\sum_x P(y \mid x)P(x)}$$

α is called the normalization constant

(can be calculated by summing over numerator values)

4

Why is Bayes rule useful?

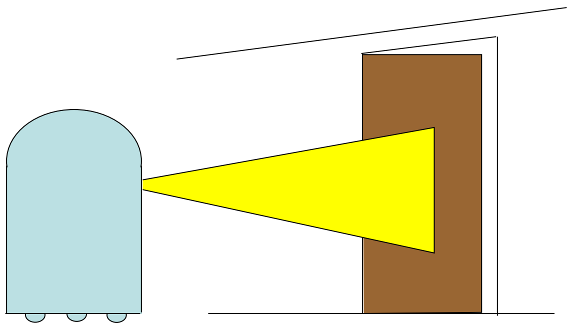
Allows diagnostic reasoning from causal information:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

5

Example 1: State Estimation

- Suppose a robot obtains measurement z
- What is *P(doorOpen|z)?*



Causal vs. Diagnostic Reasoning

- *P(open|z)* is diagnostic.
- *P(z|open)* is causal.
- Often causal knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge to diagnose a situation:

count frequencies!

$$P(open \mid z) = \frac{P(z \mid open) P(open)}{P(z)}$$

State Estimation Example

- *Suppose:* $P(z|open) = 0.6$ $\quad P(z|\neg open) = 0.3$
- $P(open) = P(\neg open) = 0.5$

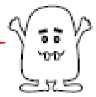
$$P(open \mid z) = \frac{P(z \mid open) P(open)}{P(z \mid open) p(open) + P(z \mid \neg open) p(\neg open)}$$

$$P(open \mid z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{0.30}{0.45} = 0.67$$

Measurement z raises the probability that the door is open from 0.5 to 0.67

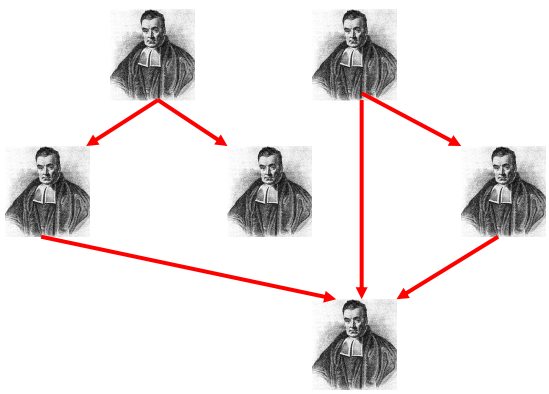
Is there a general representation scheme for efficient probabilistic inference?

Yes!



9

Enter…Bayesian networks



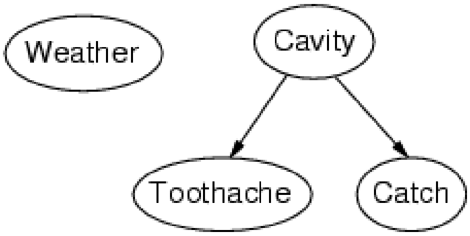
10

What are Bayesian networks?

- Simple, graphical notation for conditional independence assertions
  - Allows compact specification of full joint distributions

Example: Back at the Dentist's

- Topology of network encodes conditional independence assertions:



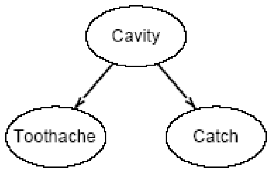
- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent of each other given *Cavity*

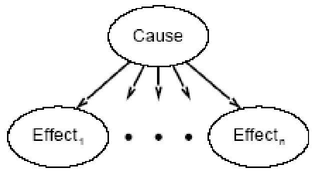
Conditional Independence and the “Naïve Bayes Model”

$$
\begin{aligned}
&\mathbf{P}(Cavity|toothache \wedge catch)\\
&= \alpha\,\mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity)\\
&= \alpha\,\mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}
$$

This is an example of a *naive Bayes* model:

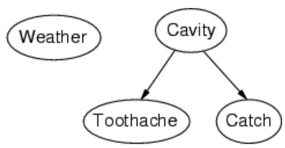
$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i\mathbf{P}(Effect_i|Cause)$$





Total number of parameters is *linear* in n

Bayesian networks



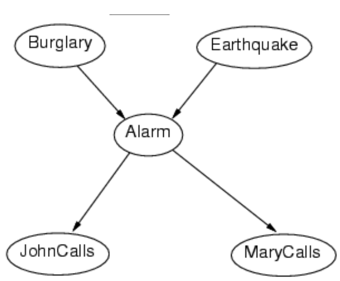
- Syntax:
  - set of nodes, one per random variable
  - directed, acyclic graph (link ≈ "directly influences")
  - conditional distribution for each node given its parents:
    $P(X_i \mid Parents(X_i))$
- For discrete variables, conditional distribution = conditional probability table (CPT) = distribution over X_i for each combination of parent values

Example 2: Burglars and Earthquakes

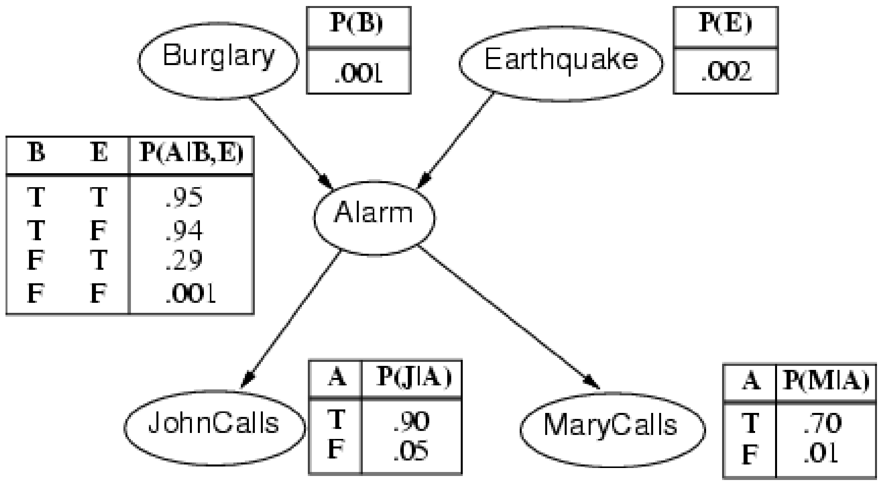
- You are at a “Done with the AI class” party.
- Neighbor John calls to say your home alarm has gone off (but neighbor Mary doesn't).
- Sometimes your alarm is set off by minor earthquakes.
- Question: Is your home being burglarized?

Example 2: Burglars and Earthquakes

- Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*
- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call



Bayesian Network for Burglars and Earthquakes



P(B)
.001

P(E)
.002

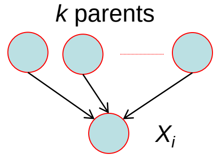
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

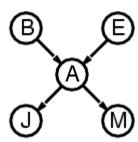
A	P(M|A)
T	.70
F	.01

Compact Representation of Probabilities in Bayesian Networks

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values



- Each row requires one number p for X_i = *true* (the other number for X_i = *false* is just *1-p*)
- If each variable has no more than k parents, an n-variable network requires $O(n \cdot 2^k)$ numbers
  - This grows linearly with n vs. $O(2^n)$ for full joint distribution

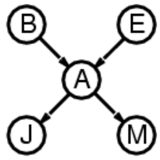


- For burglar network, 1+1+4+2+2 = 10 numbers (vs. 2^5-1 = 31 numbers) for full joint distribution

Bayesian Network Semantics

- Full joint distribution is defined as product of local conditional distributions:

$$P(X_1, \ldots, X_n) = \pi_{i=1}^{n} P(X_i \mid Parents(X_i))$$



- e.g., Joint probability of all variables being true = ?

  $P(j \wedge m \wedge a \wedge b \wedge e)$

  $= P(j \mid a)\, P(m \mid a)\, P(a \mid b, e)\, P(b)\, P(e)$

- Similarly, $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

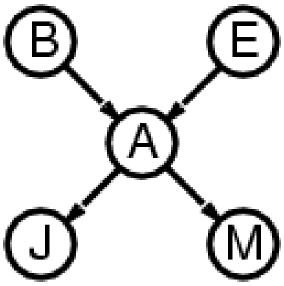
  $= P(j \mid a)\, P(m \mid a)\, P(a \mid \neg b, \neg e)\, P(\neg b)\, P(\neg e)$

Probabilistic Inference in BNs

- The graphical independence representation yields efficient inference schemes
- We generally want to compute
  - $P(X|E)$ where E is evidence from sensory measurements etc. (known values for variables)
  - Sometimes, may want to compute just $P(X)$
- One simple inference algorithm:
  - *variable elimination (VE)*

What is the probability of burglary given that John and Mary called?

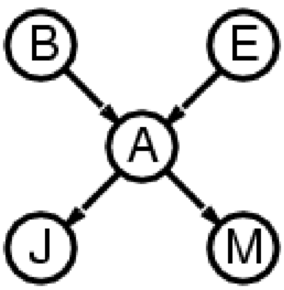
Compute P(B=true | J=true, M=true)



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

21

Computing P(B=true | J=true, M=true)



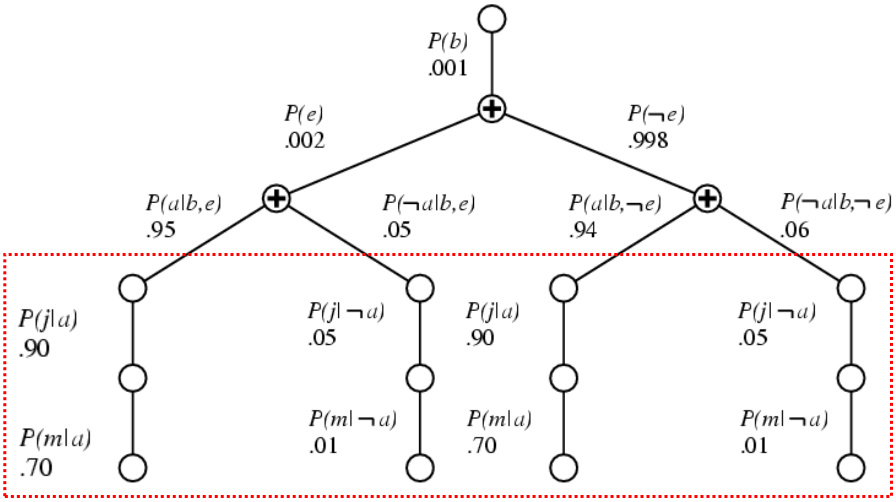
$P(b|j,m) = \alpha\, \Sigma_{e,a}\, P(b,j,m,e,a)$

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

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

22

Structure of Computation



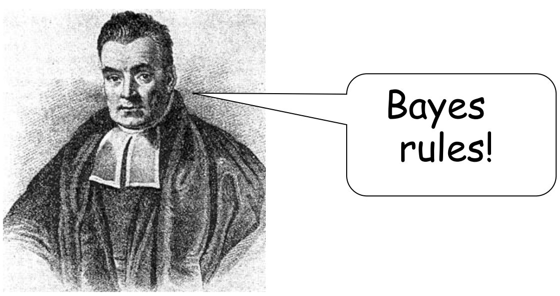
$P(b|j,m) = \alpha\, P(b)\, \Sigma_e P(e)\, \Sigma_a P(a|b,e) P(j|a) P(m|a)$

Repeated computations ⇒ use dynamic programming

23

Next Time

- Inference Algorithms
  - Variable Elimination (VE)
- Hidden Markov Models
- To Do:
  - Project 3 due Sunday before midnight



24