CSCI 561 - Foundation for Artificial Intelligence

Discussion Section (Week 13) Bayesian Networks

PROF WEI-MIN SHEN WMSHEN@USC.EDU

Two Major Components in Probability

Probability Distribution Model

- **Variables, Value Assignments (possible worlds)**
- **Represented as a table or a graph**

Inferences that can be made from the model

1. **Sum rule:** P(a) + P(~a) = 1
2. **Product rule:** P(ab) = P(a|b)P(b) = P(b|a)P(a) // Bayes
3. **Conditional**
4. **Marginalization**
5. **Normalization**

Independence

Based on the Product Rule: $P(AB)=P(A)P(B|A)=P(B)P(A|B)$

Absolute Independence

A and B are <u>independent</u> iff

$P(AB) = P(A)\,P(B)$

$P(A|B) = P(A)$

$P(B|A) = P(B)$

Conditional Independence

A and B are <u>conditional independent on C</u> iff

$P(AB \mid C) = P(A \mid C)\,P(B \mid C)$

$P(A \mid BC) = P(A \mid C)$

$P(B \mid AC) = P(B \mid C)$

Bayes' Rule

Product rule: $P(a \wedge b) = P(a \mid b)\, P(b)$

$$P(a \wedge b) = P(b \mid a)\, P(a)$$

$$P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$$

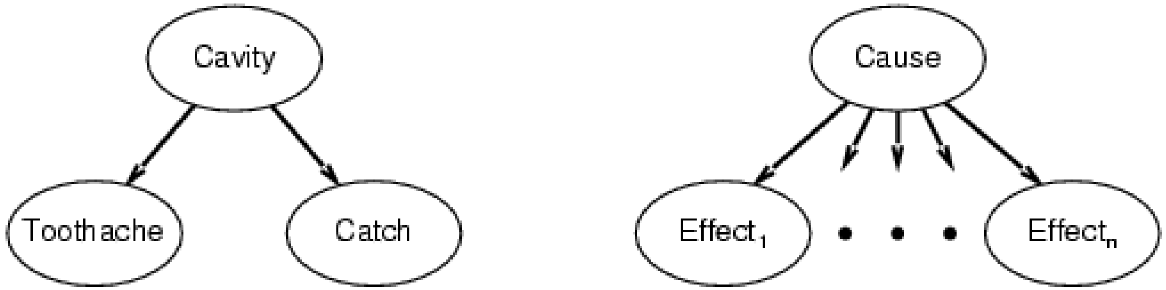
$\Rightarrow$ **Bayes' rule:** $P(a \mid b) = P(b \mid a)\, P(a) \,/\, P(b)$

Rev. Thomas Bayes
c. 1701 - 1761

Combining Evidence (for Diagnosis

P(*Cavity* | *toothache,catch*)
= αP(*toothache* ∧ *catch* | *Cavity*) P(*Cavity*) **[Bayes' Rule]**
= αP(*toothache* | *Cavity*) P(*catch* | *Cavity*) P(*Cavity*) **[Cond. Ind.]**

This is an example of a *naïve Bayes* model:
$\mathbf{P}(\text{Cause}, \text{Effect}_1, \ldots, \text{Effect}_n) = \mathbf{P}(\text{Cause})\, \Pi_i \mathbf{P}(\text{Effect}_i|\text{Cause})$



- **Cost of diagnostic reasoning now grows linearly rather than exponentially in number of conditionally independent effects**

Called naïve, because often used when the effects are not completely conditionally independent given the cause

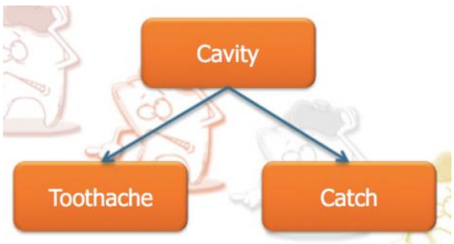
Why Need Bayesian Networks?

A better representation for the Fully Joint Probability Distribution Model (take advantage of "variable independence")

Fully Joint Distribution Table

Cavity	Catch	Toothache	Logic Truth	Probability
0	0	0	{0,1}	0.576
0	0	1	{0,1}	0.064
0	1	0	{0,1}	0.144
0	1	1	{0,1}	0.016
1	0	0	{0,1}	0.008
1	0	1	{0,1}	0.012
1	1	0	{0,1}	0.072
1	1	1	{0,1}	0.108

Bayesian Network



$\mathbf{P}(X_1, X_2, \ldots, X_n)$

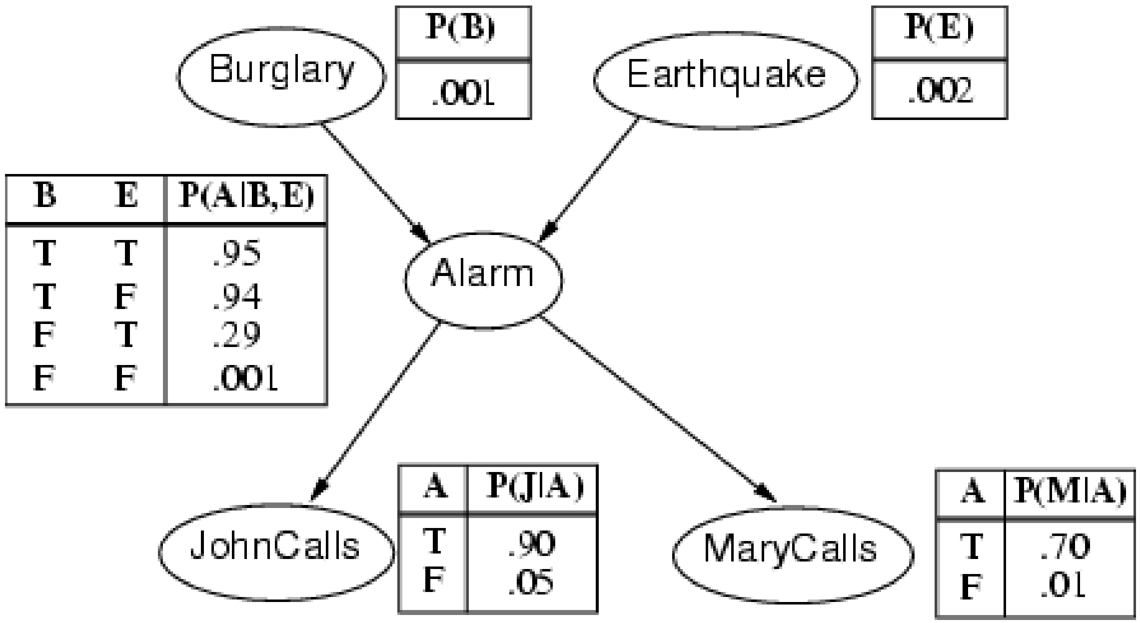
Size = $O(d^n)$
d: variable domain

$\Pi_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$

Size = $O(nd^k)$
k: # of parents

$P(X_1,X_2,\ldots,X_n) = P(X_1|X_2,\ldots X_n)P(X_2,\ldots,X_n)$
$= P(X_1|X_2,\ldots X_n)P(X_2|X_3,\ldots,X_n)P(X_3,\ldots,X_n)$
$= \ldots$
$= P(X_1|X_2,\ldots X_n)P(X_2|X_3,\ldots,X_n) \ldots P(X_n)$

Alarm Example



Only one value needed for X_i in each row because, for boolean variables, P(*false*)=1-P(*true*)

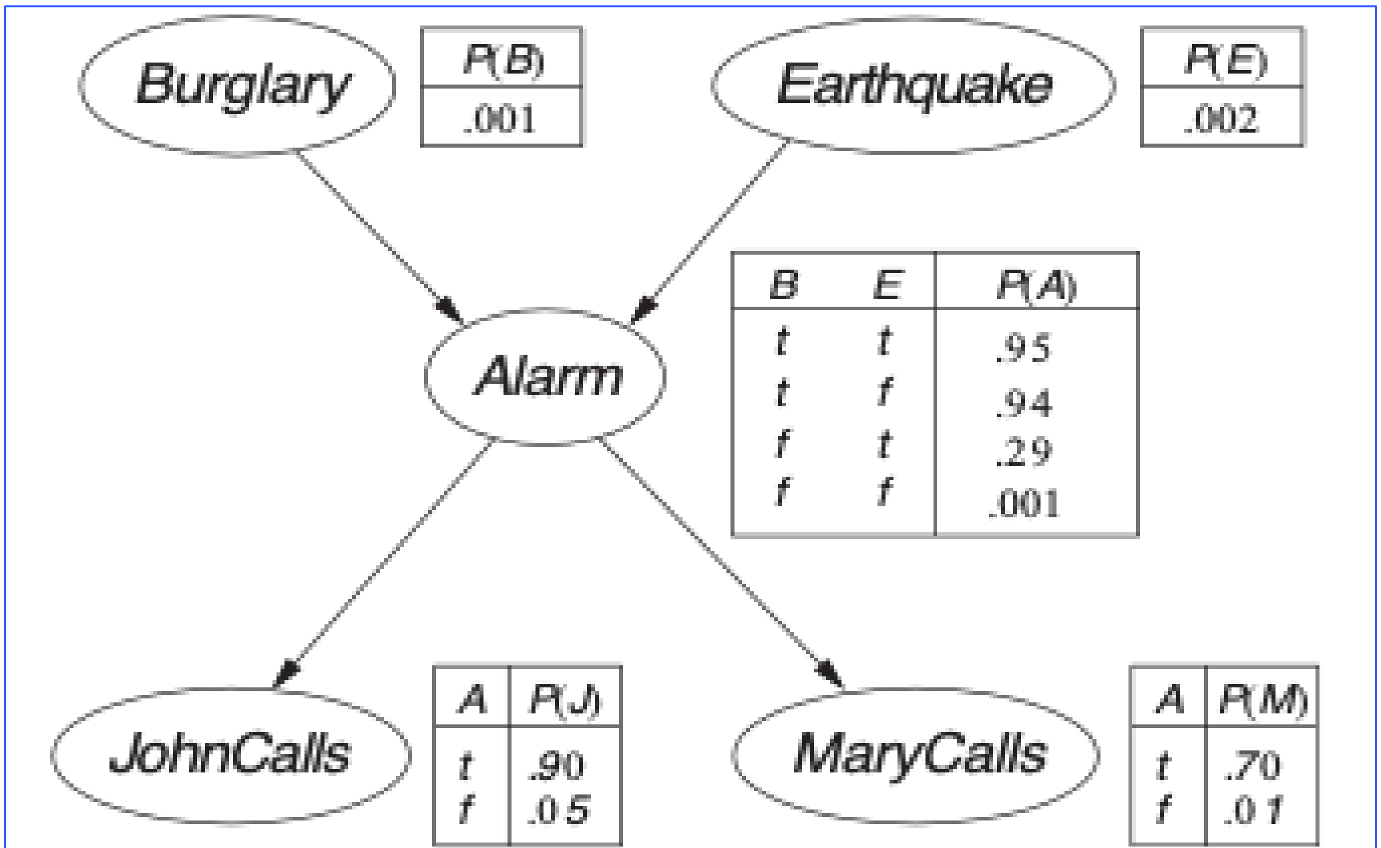
All factors (possibly infinite) not explicitly mentioned are implicitly incorporated into probabilities

- Bird could fly through window pane, power could fail, …

Semantics

If correct, the network represents the full joint distribution:

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



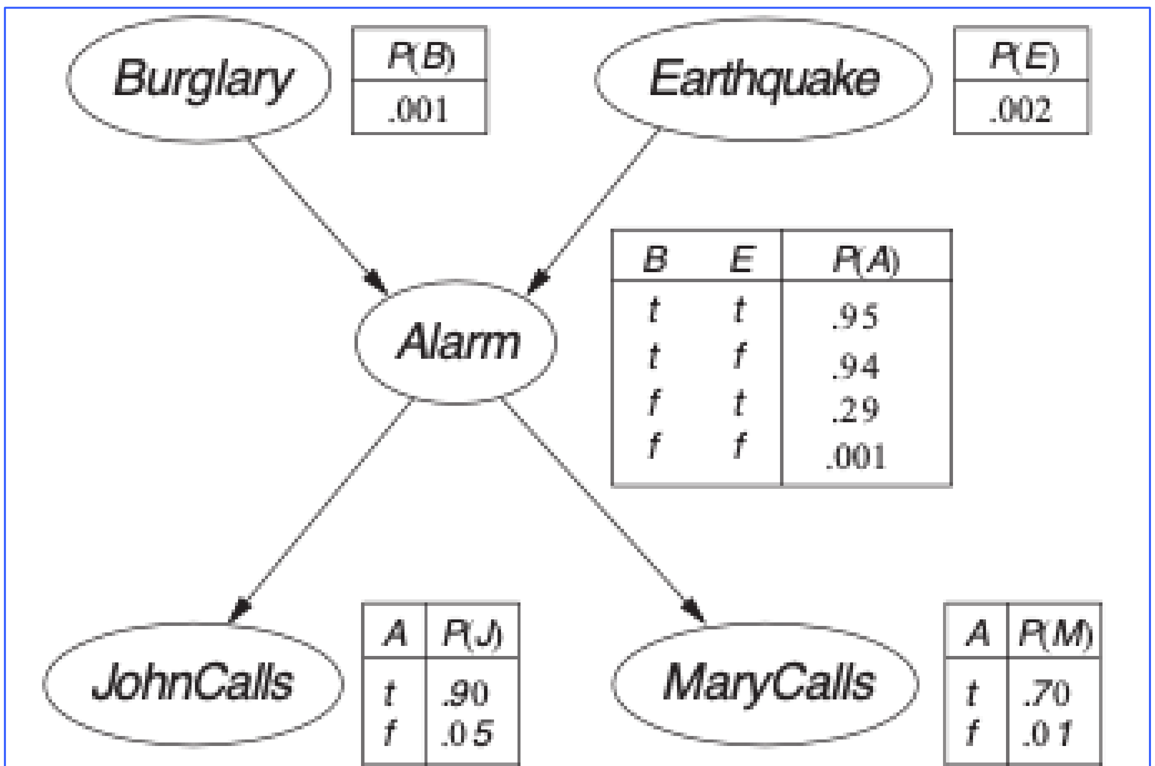
$$\mathbf{P}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \mathbf{P}(x_i \mid parents(X_i))$$

E.g., the probability of a complete false alarm (no burglary or earthquake) with two calls is:

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

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

$$= .9 \times .7 \times .001 \times .999 \times .998 \approx .000063$$



General Enumeration Algorithm

Given any query *P(H* | *E*), you can solve it by the following:

Brute force calculation of $P(H \mid E)$ is done by:

1. Apply the conditional probability rule.

$$P(H \mid E) = P(H \wedge E) \,/\, P(E)$$

2. Apply the marginal distribution rule to the unknown vertices $\mathbf{U}$.

$$P(H \wedge E) = \sum_{\mathbf{U}=\mathbf{u}} P(H \wedge E \wedge \mathbf{U} = \mathbf{u})$$

Only those dependent variables based on the net

3. Apply joint distribution rule for Bayesian networks.

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

Enumeration in Bayesian Networks

Compute probabilities from Bayesian network as if from FJPT, but without explicitly constructing the table

- Otherwise would lose benefit of decomposing full table into network

Consider simple query on burglary network

$\mathbf{P}(b \mid j,m)$

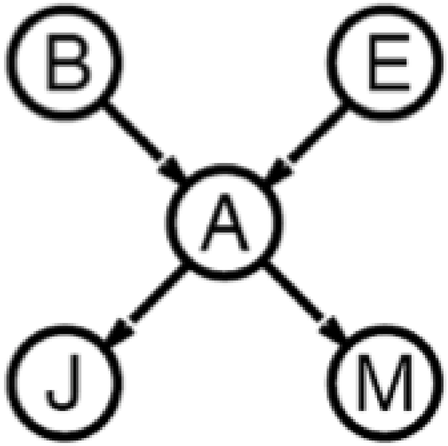
$= \mathbf{P}(b,j,m)/P(j,m)$

$= \alpha\mathbf{P}(b,j,m)$

$= \alpha\Sigma_e\Sigma_a\mathbf{P}(b,e,a,j,m)$

$= \alpha\Sigma_e\Sigma_a\mathbf{P}(b)\, P(e)\, \mathbf{P}(a \mid b,e)\, P(j \mid a)\, P(m \mid a)$

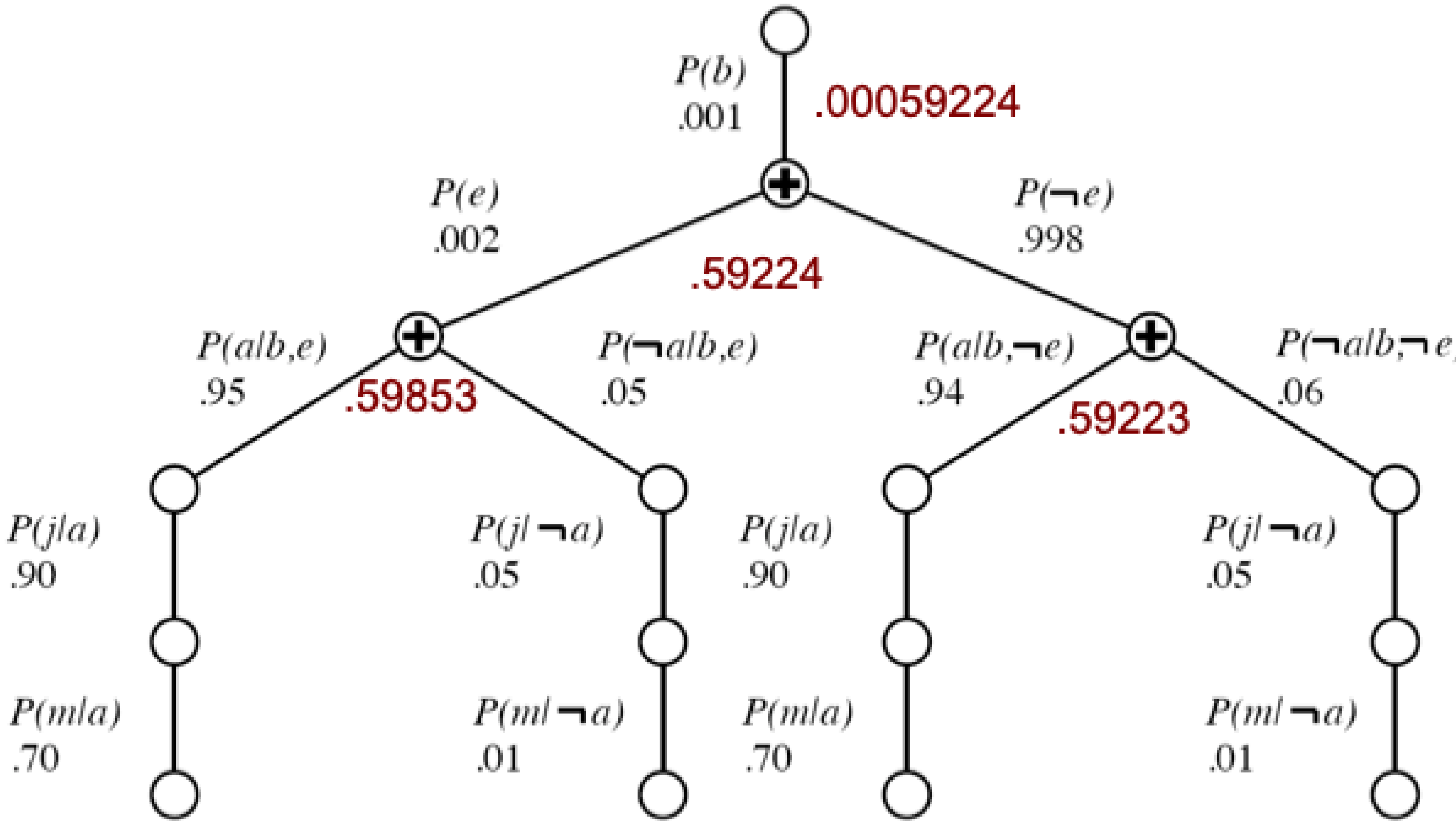
$= \alpha\mathbf{P}(b)\Sigma_e\, P(e)\, \Sigma_a\, \mathbf{P}(a \mid b,e)\, P(j \mid a)\, P(m \mid a)$



Compute by proceeding through terms in a depth-first fashion, multiplying and adding CPT entries as we go

Evaluation Tree for P(*b* | *j,m*)

$$P(b)\Sigma_e P(e)\, \Sigma_a P(a \mid b,e)\; P(j \mid a)\; P(m \mid a) = .00059224$$



Exercise 14.1

We have a bag of three biased coins a, b, and c with probabilities of coming up heads of 20%, 60%, and 80%, respectively. One coin is drawn randomly from the bag (with equal likelihood of drawing each of the three coins), and then the coin is flipped three times to generate the outcomes X_1, X_2, and X_3.

a. Draw the Bayesian network corresponding to this setup and define the necessary CPTs.

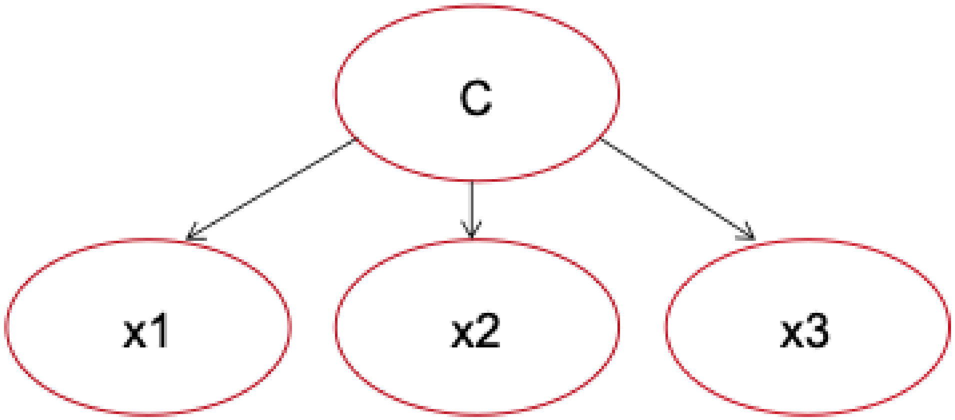
b. Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

Exercise 14.1

With the random variable C denoting which coin {a, b, c} we drew, the network has C at the root and X1, X2, and X3 as children.

The CPT for C is:

C	$P(C)$
a	$1/3$
b	$1/3$
c	$1/3$



The CPT for X_i given C are the same, and equal to:

C	X_1	$P(C)$
a	$heads$	0.2
b	$heads$	0.6
c	$heads$	0.8

Exercise 14.1

The coin most likely to have been drawn from the bag given this sequence is the value of C with greatest posterior probability P(C|2 heads, 1 tails).

P(C|2 heads, 1 tails) = P(2 heads, 1 tails|C)P(C)/P(2 heads, 1 tails)

- 1/P (2 heads, 1 tails) is independent of C

  $\propto$ P(2heads,1tails|C)P(C)

- P(C) is independent of the value of C, by hypothesis, equal to 1/3.

  $\propto$ P(2 heads, 1 tails|C)

Exercise 14.1

X_1, X_2, and X_3 are conditionally independent given C, so for example

$P(X_1 = \text{tails}, X_2 = \text{heads}, X_3 = \text{heads}|C = a)$

$= P(X_1 = \text{tails}|C = a)P(X_2 = \text{heads}|C = a)P(X_3 = \text{heads}|C = a)$

$= 0.8 \times 0.2 \times 0.2 = 0.032$

$P(\text{2heads, 1tails}|C = a) = 3 \times 0.032 = 0.096$.

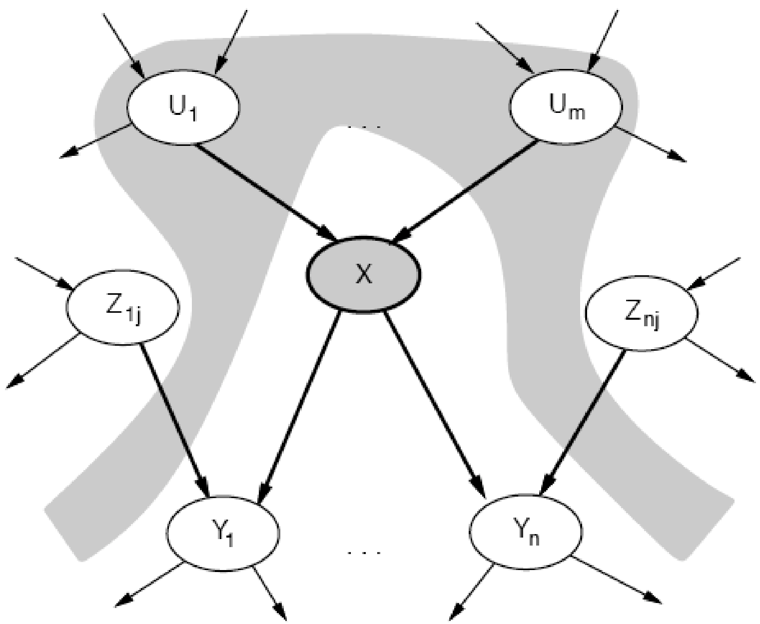
Note that we would get the same probability above for any ordering of 2 heads and 1 tails.

$P(\text{2heads, 1tails}|C = b) = 0.432$
$P(\text{2heads, 1tails}|C = c) = 0.384$

showing that coin b is most likely to have been drawn.
Alternatively, one could directly compute the value of P(C|2 heads, 1 tails).

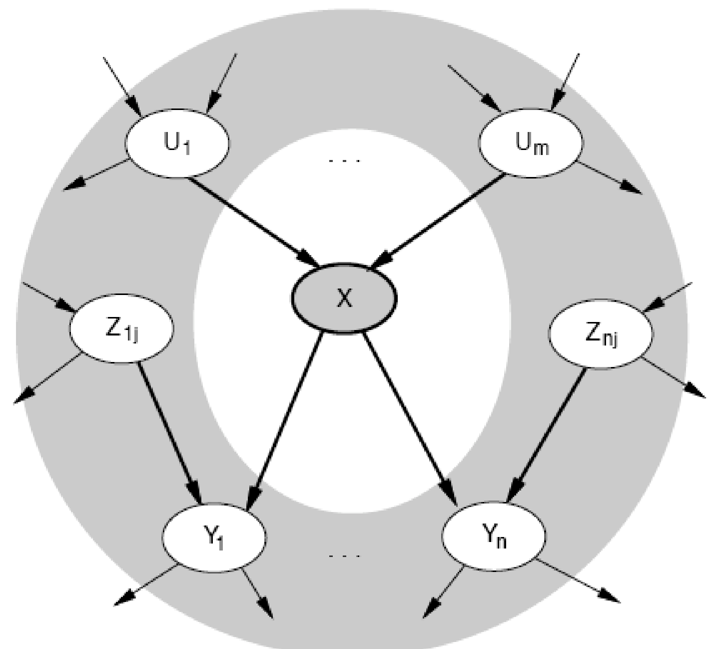
Conditional Independence of Nodes

A node is conditionally independent of its *nondescendents* given its *parents*



A node is conditionally independent of *all others* given its *Markov blanket*

- Parents, children and children's parents



Bayesian Networks

Given this network calculate the following probabilities. Give both the formula and calculations with values. These questions are designed so that they can be answered with a minimum of computation. If you find yourself doing a copious amount of computation for each part, step back and consider whether there is simpler way to deduce the answer.

1. P(a,¬b,c,¬d)

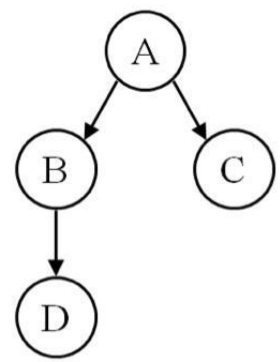
P (a)P (¬b|a)P (c|a)P (¬d|¬b)
= 0.1 × 0.5 × 0.4 × 0.8 = 0.016

$P(A)$

+a	0.1
¬a	0.9

$P(B|A)$

+a	+b	0.5
	¬b	0.5
¬a	+b	0.8
	¬b	0.2



$P(C|A)$

+a	+c	0.4
	¬c	0.6
¬a	+c	0.7
	¬c	0.3

$P(D|B)$

+b	+d	0.9
	¬d	0.1
¬b	+d	0.2
	¬d	0.8

Bayesian Networks

2. P(b)

$P(b) = \sum_{A=\{a,\neg a\}} P(A)P(b|$

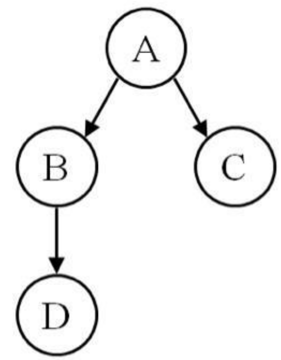
$= 0.1 \times 0.5 + 0.9 \times 0.8 =$

$P(A)$

+a	0.1
¬a	0.9

$P(B|A)$

+a	+b	0.5
	¬b	0.5
¬a	+b	0.8
	¬b	0.2



$P(C|A)$

+a	+c	0.4
	¬c	0.6
¬a	+c	0.7
	¬c	0.3

$P(D|B)$

+b	+d	0.9
	¬d	0.1
¬b	+d	0.2
	¬d	0.8

3. P(a|b)

$P(a|b) = P(a,b) / P(b) = P(a)P(b|a) / P(b)$

$= 0.1 \times 0.5 / .77 = 0.064935$

Bayesian Networks

4. P(d|a)

$P(d|a) = \sum_{B=\{b,\neg b\}} P(d|B)p(B|a)$

// think: why no P(a) here?

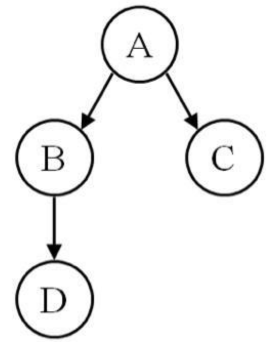
$= 0.9 \times 0.5 + 0.2 \times 0.5 = 0.55$

$P(A)$

+a	0.1
¬a	0.9

$P(B|A)$

+a	+b	0.5
	¬b	0.5
¬a	+b	0.8
	¬b	0.2



$P(C|A)$

+a	+c	0.4
	¬c	0.6
¬a	+c	0.7
	¬c	0.3

$P(D|B)$

+b	+d	0.9
	¬d	0.1
¬b	+d	0.2
	¬d	0.8

5. P(d|a,c)

From the conditional independence properties of the graph, $D \perp C|\{A\}$. Hence, $P(d|a,c) = p(d|a)$

$= 0.55$

What you should know

Probability formulas:
Product rule: $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$
$\Rightarrow$ **Bayes' rule:** $P(a \mid b) = P(b \mid a)\, P(a) / P(b)$
Conditional probability: $P(a \mid b) = P(a \wedge b) / P(b)$

- **What is independence? What is conditional independence? Why are they needed for reasoning about uncertainty?**
- **What is Bayes rule? How is this addressing combining evidence for diagnosis?**
  - **Bayesian networks provide a natural representation for (causally induced) conditional independence**
  - **Topology + CPTs = compact representation of joint distribution**
  - **Why do we need approximate inference? What are some approximate inference techniques?**
  - **Are there limits to probabilistic reasoning?**
  - **How can we reasoning probabilistic over time?**

Want more?

Try exercise 13.4,7,8,13,15, 14.2,8 in AIMA