

Chapter 14

Probabilistic Reasoning

Motivations

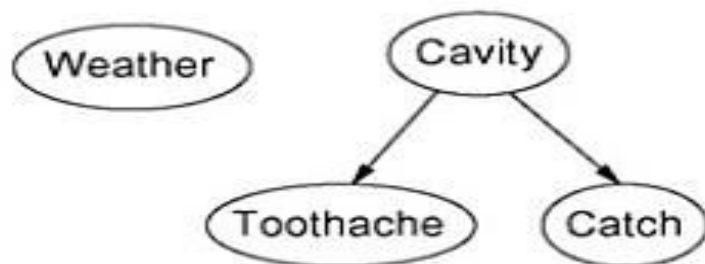
- ❑ Full joint probability distribution can answer any question but can become intractably large as number of variable increases
- ❑ Specifying probabilities for atomic events can be difficult, e.g., large set of data, statistical estimates, etc.
- ❑ Independence and conditional independence reduce the probabilities needed for full joint probability distribution.

Bayesian networks

- ❑ A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions
- ❑ A directed, acyclic graph (DAG)
- ❑ A set of nodes, one per variable (discrete or continuous)
- ❑ A set of directed links (arrows) connects pairs of nodes. X is a parent of Y if there is an arrow (direct influence) from node X to node Y .
- ❑ Each node X_i has a conditional probability distribution $P(X_i | Parents(X_i))$ that quantifies the effect of the parents on the node.
- ❑ Combinations of the topology and the conditional distributions specify (implicitly) the full joint distribution for all the variables.

✓ Causes
should
be
parents
that of
effects

Topology of network encodes conditional independence assertions:



Weather is independent of the other variables

Toothache and *Catch* are conditionally independent given *Cavity*

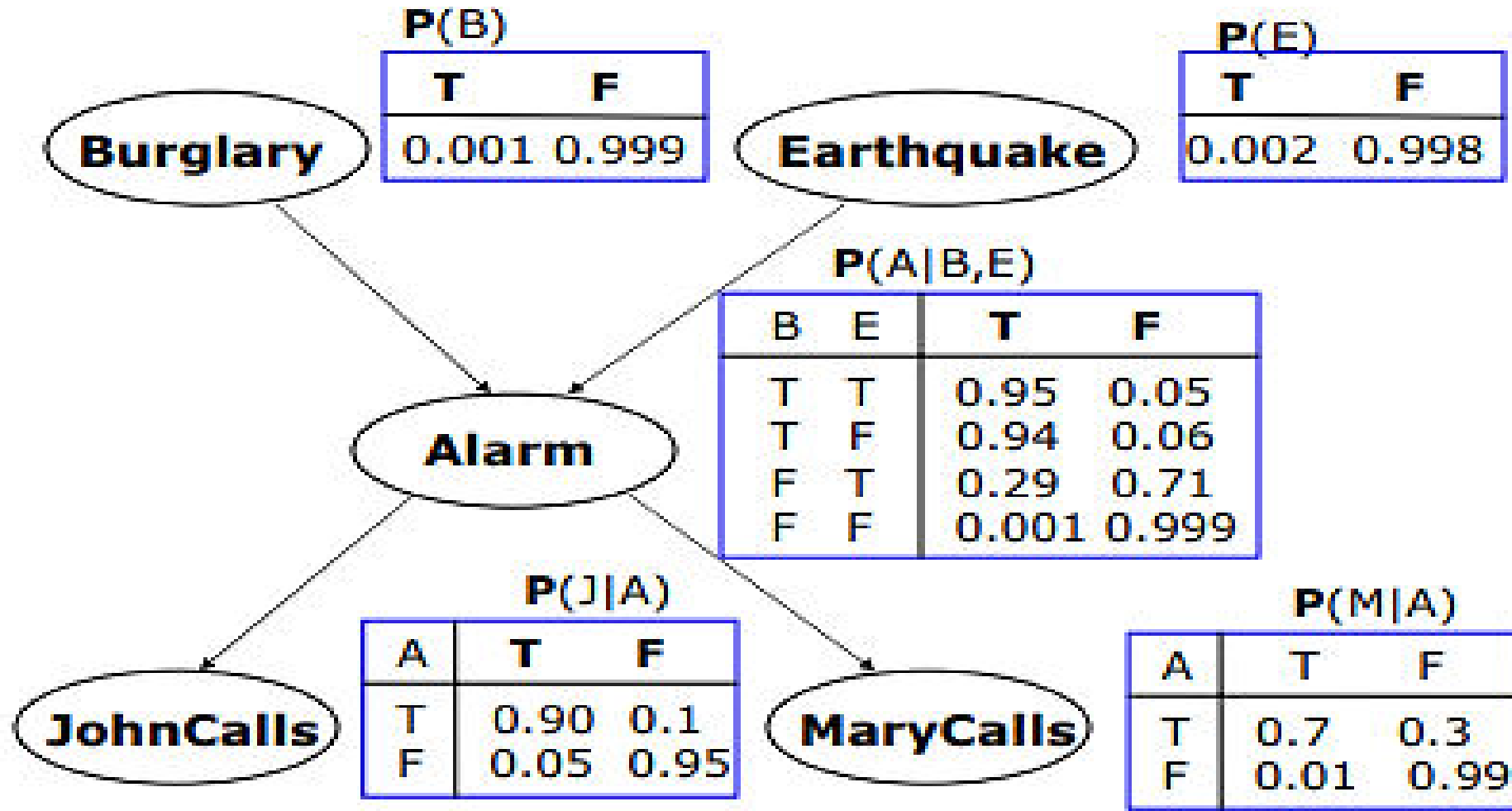
Example: Burglar alarm system

- I have a burglar alarm installed at home
 - It is fairly reliable at detecting a burglary, but also responds on occasion to minor earth quakes.
- I also have two neighbors, John and Mary
 - They have promised to call me at work when they hear the alarm
 - John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too.
 - Mary likes rather loud music and sometimes misses the alarm altogether.
- Bayesian networks variables:
 - *Burglar, Earthquake, Alarm, JohnCalls, MaryCalls*



Bayesian belief network.

Variable Order= [John, Mary, Alarm, Burglar, Earthquake]



Compactness of Bayesian networks

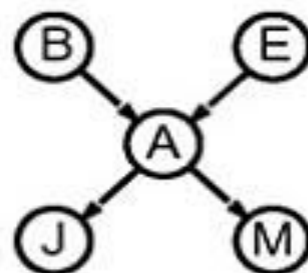
A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values

Each row requires one number p for $X_i = \text{true}$ (the number for $X_i = \text{false}$ is just $1 - p$)

If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers

I.e., grows linearly with n , vs. $O(2^n)$ for the full joint distribution

For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5 - 1 = 31$)



Global Semantics

“Global” semantics defines the full joint distribution as the product of the local conditional distributions:

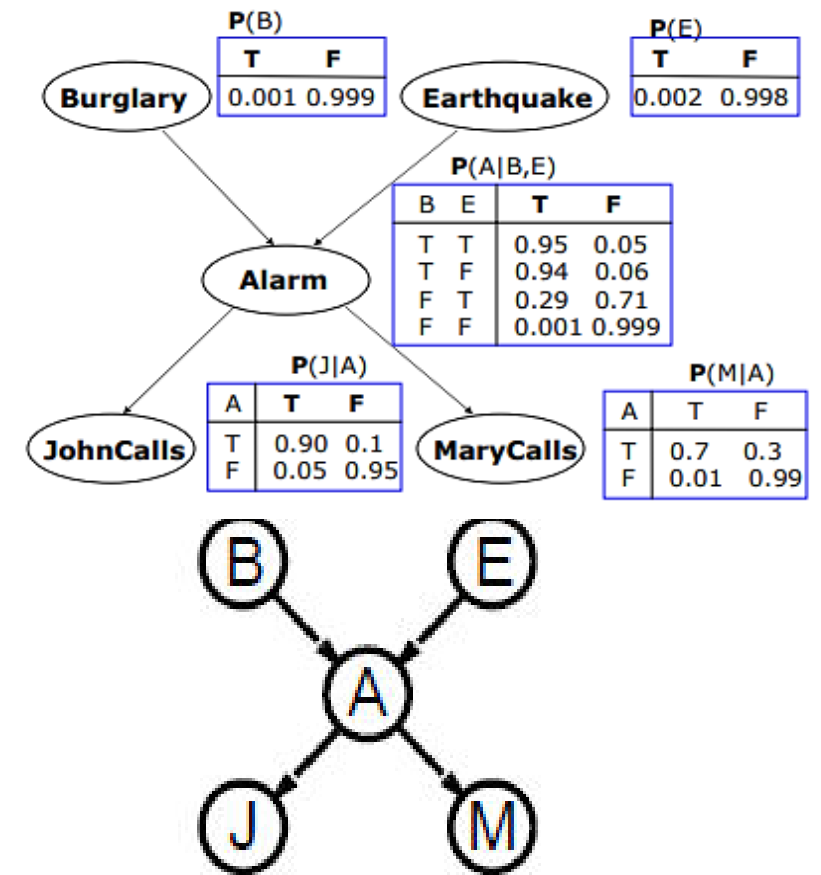
$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$$

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

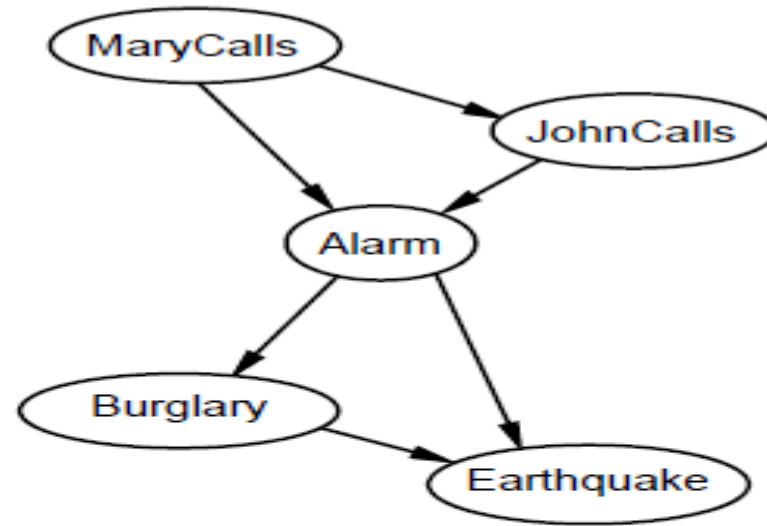
$$= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e)$$

$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$$

$$\approx 0.00063$$



Example



Deciding conditional independence is hard in noncausal directions

(Causal models and conditional independence seem hardwired for humans!)

Assessing conditional probabilities is hard in noncausal directions

Network is less compact: $1 + 2 + 4 + 2 + 4 = 13$ numbers needed

Example

$$P(A) = \sum_{b \in \{B, !B\}} \sum_{e \in \{E, !E\}} P(A, b, e)$$

Recall from conditional probability: $P(A, B) = P(A|B) \cdot P(B)$

So $P(A, B, E) = P(A|B, E) \cdot P(B, E)$,
for example.

Which means:

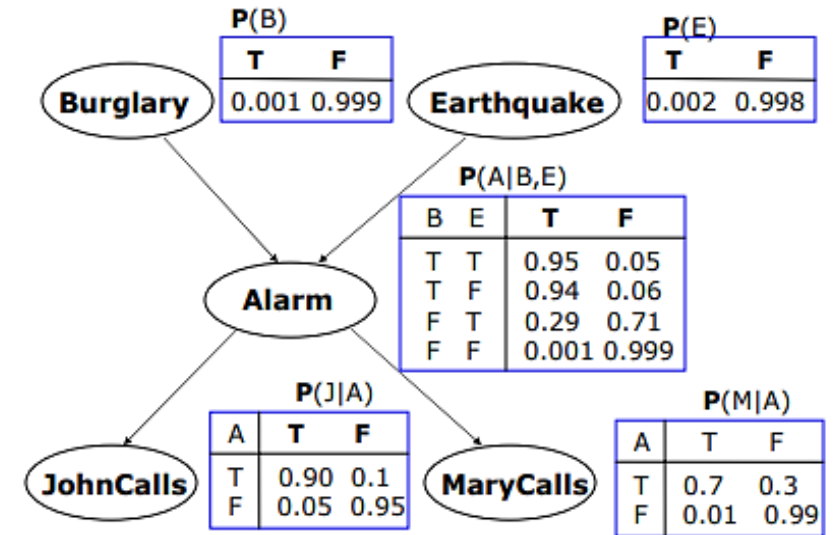
$$P(A) = P(A|B, E)P(B, E) + P(A|B, !E)P(B, !E) + P(A|!B, E)P(!B, E) + P(A|!B, !E)P(!B, !E)$$

This is because there are 4 ways in which A can happen.

A can happen if events B and E happen, if B happens and not E, if E happens and not B, or if neither of them happen.

Working that out (and assuming B and E are independent), we have:

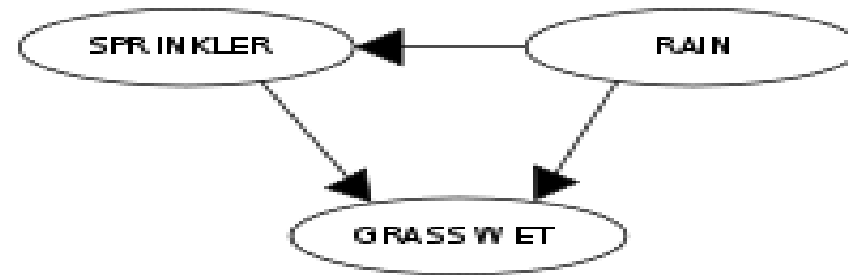
$$P(A) = (0.95 * 0.001 * 0.002) + (0.94 * 0.001 * 0.998) + (0.29 * 0.999 * 0.002) + (0.001 * 0.999 * 0.998) = 0.002516442$$



Example:2

Two events can cause grass to be wet: an active sprinkler or rain. Rain has a direct effect on the use of the sprinkler (namely that when it rains, the sprinkler usually is not active). This situation can be modeled with a Bayesian network. Each variable has two possible values, T (for true) and F (for false).

RAIN	SPRINKLER	
	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	T	F
	0.2	0.8

SPRINKLER	RAIN	GRASS WET	
		T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

What is the probability that it is raining, given the grass is wet?

What is the probability that it is raining, given the grass is wet?

- We can then calculate, for example:

$$P(\text{it is raining} \mid \text{grass is wet}) = \frac{P(\text{it is raining AND grass is wet})}{P(\text{grass is wet})}$$
$$= \frac{\sum_{\text{sprinkler} \in \{T, F\}} P(\text{grass is wet} = T \text{ AND sprinkler AND raining} = T)}{\sum_{\text{sprinkler} \in \{T, F\}, \text{ raining} \in \{T, F\}} P(\text{grass is wet} = T \text{ AND sprinkler AND raining})}$$

The joint probability Distribution formula is

$$\Pr(G, S, R) = \Pr(G \mid S, R) \Pr(S \mid R) \Pr(R)$$

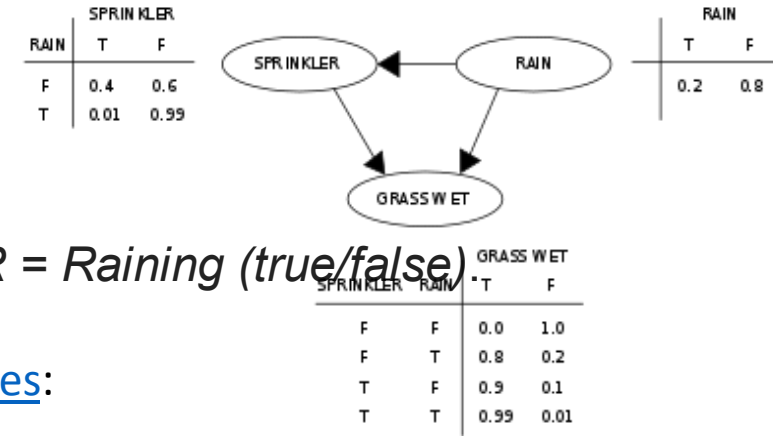
where G = Grass wet (true/false), S = Sprinkler turned on (true/false), and R = Raining (true/false).

The joint probability Distribution formula is

$$\Pr(G, S, R) = \Pr(G|S, R) \Pr(S|R) \Pr(R)$$

where $G = \text{Grass wet (true/false)}$, $S = \text{Sprinkler turned on (true/false)}$, and $R = \text{Raining (true/false)}$

by using the [conditional probability](#) formula and summing over all [random variables](#):



$$\Pr(R = T|G = T) = \frac{\Pr(G = T, R = T)}{\Pr(G = T)} = \frac{\sum_{S \in \{T, F\}} \Pr(G = T, S, R = T)}{\sum_{S, R \in \{T, F\}} \Pr(G = T, S, R)}$$

Using the expansion for the joint probability function $\Pr(G, S, R)$ and the conditional probabilities from CPT

$$\begin{aligned} \Pr(G = T, S = T, R = T) &= \Pr(G = T|S = T, R = T) \Pr(S = T|R = T) \Pr(R = T) \\ &= 0.99 \times 0.01 \times 0.2 \\ &= 0.00198. \end{aligned}$$

$$\Pr(R = T|G = T) = \frac{0.00198_{TTT} + 0.1584_{TFT}}{0.00198_{TTT} + 0.288_{TTF} + 0.1584_{TFT} + 0.0_{TFF}} = \frac{891}{2491} \approx 35.77\%.$$

Conditional Probability

Q1: You, your Father and Mother lineup randomly in a queue to take a memorable picture at your Convocation. Find the $P(A/B)$ such that
A= Daughter on one end, B= Father in Middle

Q2: Consider the following contingency table

	RIGHT-HANDED	LEFT-HANDED	TOTAL
MALE	0.41	0.08	0.49
FEMALE	0.45	0.06	0.51
TOTAL	0.86	0.14	1

Find the Probability that a randomly selected person is

A: A Male given that he is right handed.

B: Right handed given that he is a male.

C: A Female given that she is left handed.

D: Are the events being a female and being left handed Independent?
Justify

Naïve Bays

Bayesian Classifiers

- Approach:
 - compute the posterior probability $P(C \mid A_1, A_2, \dots, A_n)$ for all values of C using the Bayes theorem

$$P(C \mid A_1 A_2 \dots A_n) = \frac{P(A_1 A_2 \dots A_n \mid C) P(C)}{P(A_1 A_2 \dots A_n)}$$

- Choose value of C that maximizes
$$P(C \mid A_1, A_2, \dots, A_n)$$
 - Equivalent to choosing value of C that maximizes
$$P(A_1, A_2, \dots, A_n \mid C) P(C)$$
- How to estimate $P(A_1, A_2, \dots, A_n \mid C)$?

Naïve Bayes Classifier

- Assume independence among attributes A_i when class is given:
 - $P(A_1, A_2, \dots, A_n | C) = P(A_1 | C_j) P(A_2 | C_j) \dots P(A_n | C_j)$
 - Can estimate $P(A_i | C_j)$ for all A_i and C_j .
 - New point is classified to C_j if $P(C_j) \prod P(A_i | C_j)$ is maximum.

How to Estimate Probabilities from Data?

<i>Tid</i>	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Class: $P(C) = N_c/N$
 - e.g., $P(\text{No}) = 7/10$,
 $P(\text{Yes}) = 3/10$
- For discrete attributes: $P(A_i | C_k) = |A_{ik}|/N_c$
 - where $|A_{ik}|$ is number of instances having attribute A_i and belongs to class C_k
 - Examples:
 $P(\text{Status}=\text{Married} | \text{No}) = 4/7$
 $P(\text{Refund}=\text{Yes} | \text{Yes})=0$