Bayesian networks

Chapter 14

Outline

- Syntax
- Semantics

Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

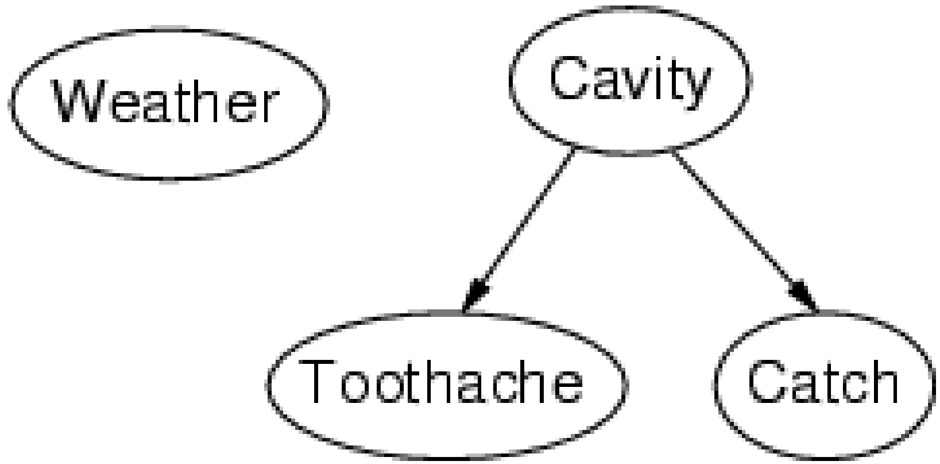
$$\mathbf{P}\,(X_i \mid \text{Parents}\,(X_i))$$

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

–

Example

- Topology of network encodes conditional independence assertions:



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

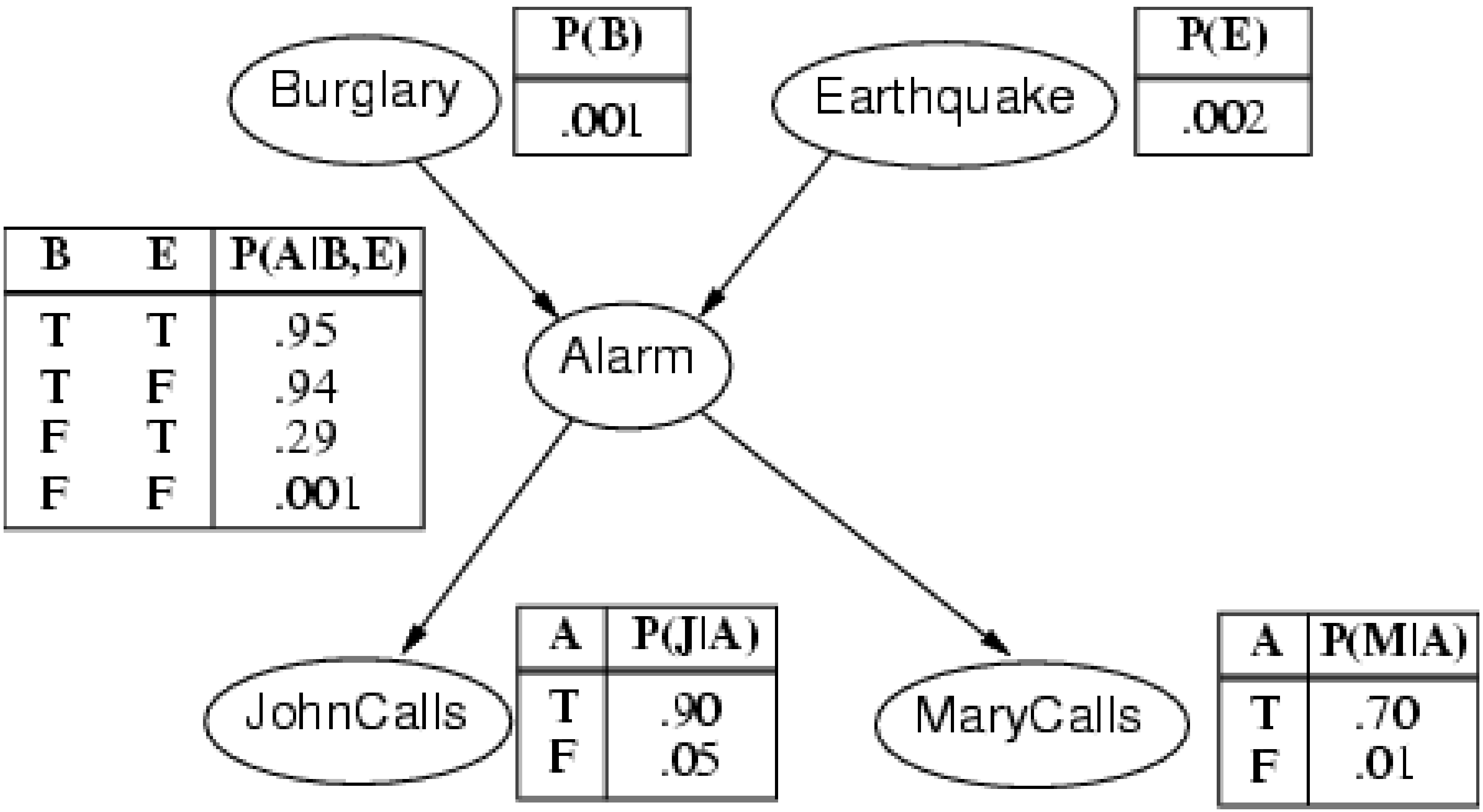
Example

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

- 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

Example contd.



P(B)
.001

P(E)
.002

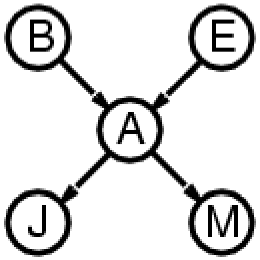
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

Compactness

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

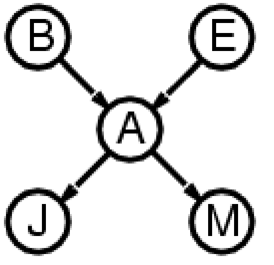


- Each row requires one number p for $X_i = true$ (the number for $X_i = false$ is just $1-p$)
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers
- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution
- For burglary net, 1 + 1 + 4 + 2 + 2 = 10 numbers (vs. 2^5-1 = 31)

Semantics

The full joint distribution is defined as the product of the local conditional distributions:

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



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

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

•

Constructing Bayesian networks

- 1. Choose an ordering of variables $X_1, \ldots, X_n$
- 2. For i = 1 to n
  - add X_i to the network
  - select parents from $X_1, \ldots, X_{i-1}$ such that

$$\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots X_{i-1})$$

This choice of parents guarantees:

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

$$= \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$

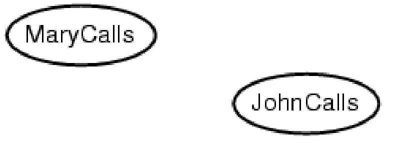
(by construction)

(chain rule)

–

Example

- Suppose we choose the ordering *M, J, A, B, E*

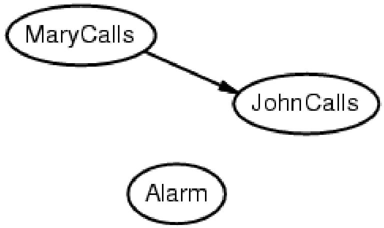


$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

-

Example

- Suppose we choose the ordering *M, J, A, B, E*



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

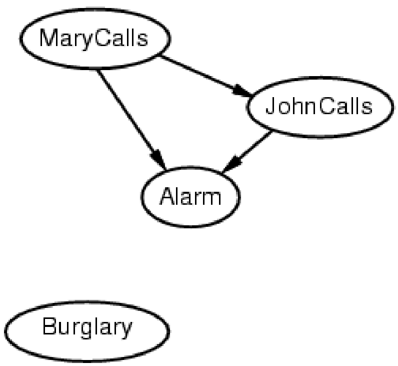
$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$?

No

-

Example

- Suppose we choose the ordering *M, J, A, B, E*



$\mathbf{P}(J \mid M) = \mathbf{P}(J)?$

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)?$ $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)?$ **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)?$

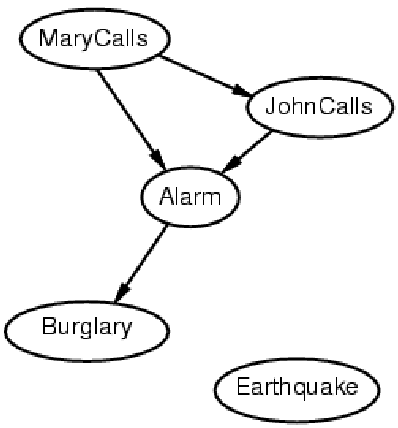
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)?$

No

-

Example

- Suppose we choose the ordering M, J, A, B, E



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$? **Yes**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$? **No**

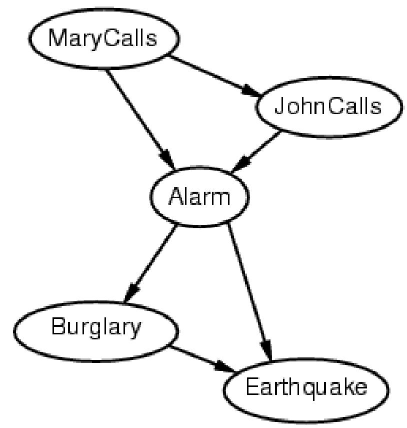
$\mathbf{P}(E \mid B, A ,J, M) = \mathbf{P}(E \mid A)$?

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A, B)$?

No

Example

- Suppose we choose the ordering M, J, A, B, E



P$(J \mid M)$ = ***P***(J)?

P$(A \mid J, M)$ = ***P***$(A \mid J)$? ***P***$(A \mid J, M)$ = ***P***(A)? **No**

P$(B \mid A, J, M)$ = ***P***$(B \mid A)$? **Yes**

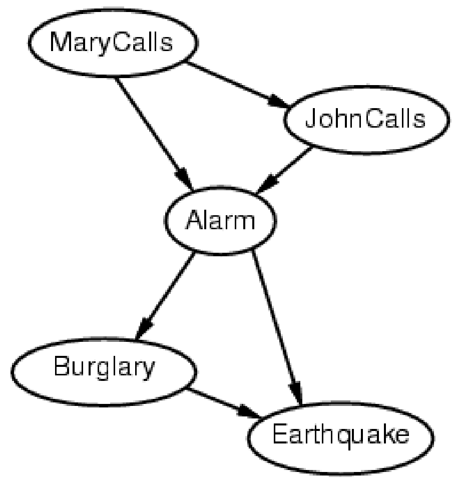
P$(B \mid A, J, M)$ = ***P***(B)? **No**

P$(E \mid B, A ,J, M)$ = ***P***$(E \mid A)$? **No**

P$(E \mid B, A, J, M)$ = ***P***$(E \mid A, B)$? **Yes**

No

Example contd.



- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed
-
-

Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct