Bayesian networks

Bayes' Theorem

Simple:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

General:

$$P(Y|X,E) = \frac{(X|Y,E)P(Y|E)}{P(X|E)}$$

Bayesian networks

- A simple, graphical notation (directed acyclic graph) for:
  - Showing conditional independence
  - Calculating full joint distributions.

- It is a causal structure.

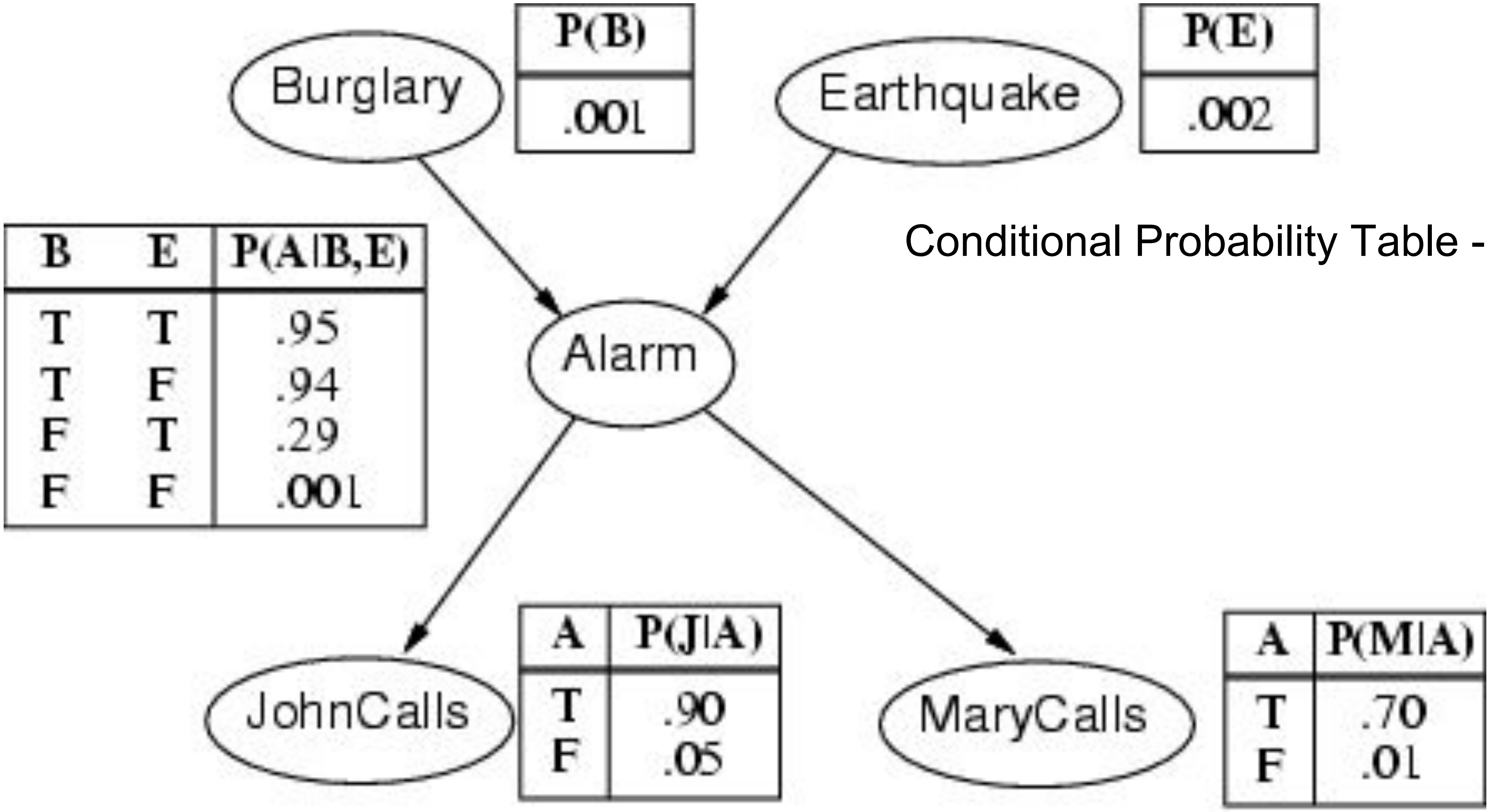
Example

- 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.



Conditional Probability Table - CPT)

P(B)
.001

P(E)
.002

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

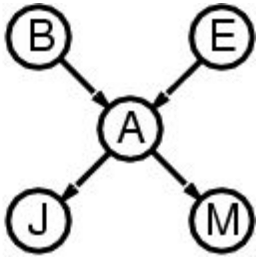
A	P(J|A)
T	.90
F	.05

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

Semantics

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

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

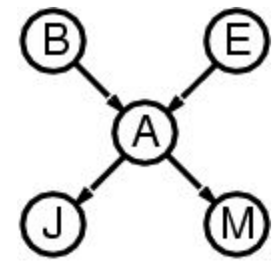


e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

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

Calculating joint Prob

- **P(A|B) = P(A∩B) ÷ P(B)**



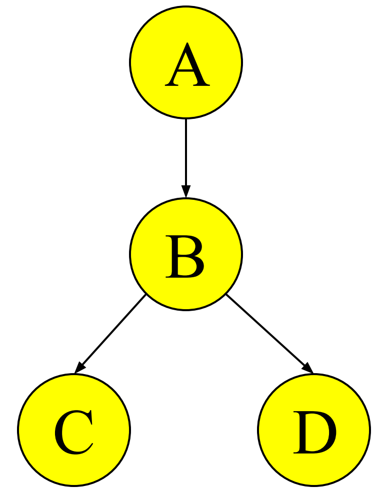
P(J|M) = P(J ∩ M) ÷ P(M)

= P(JMABE) + ….+ P(JM~A~B~E)

$$\frac{P(JMABE) + \cdots + P(JM\bar{A}\bar{B}\overline{E})}{P(MJABE) + 16\ combis}$$

P(JMABE) = P(J|A*M|A*P(A|BE)*P(B)*P(E)   and so on…..

Inference Example



Supposed we know that A=true.
What is more probable C=true or D=true?
For this we need to compute
P(C=t | A =t) and P(D=t | A =t).
Let us compute the first one.

$$P(C=t \mid A=t) = \frac{P(A=t, C=t)}{P(A=t)} = \frac{\sum_{b,d} P(A=t, B=b, C=t, D=d)}{P(A=t)}$$

A	**P(A)**
false	0.6
true	0.4

A	**B**	**P(B|A)**
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

B	**D**	**P(D|B)**
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

B	**C**	**P(C|B)**
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

The pros and cons

- Exact inference is feasible in small to medium-sized networks
- Exact inference in large networks takes a very long time
- We resort to approximate inference techniques which are much faster and give pretty good results

How to get BN and CPT

- Get an expert to design it
- Learn it from data