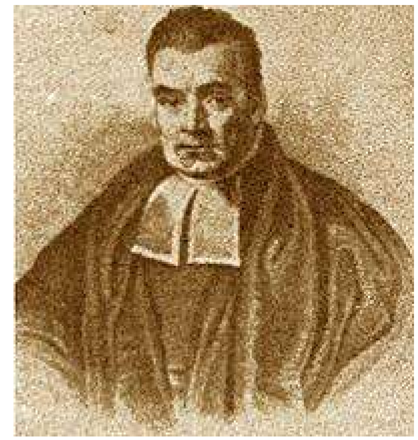
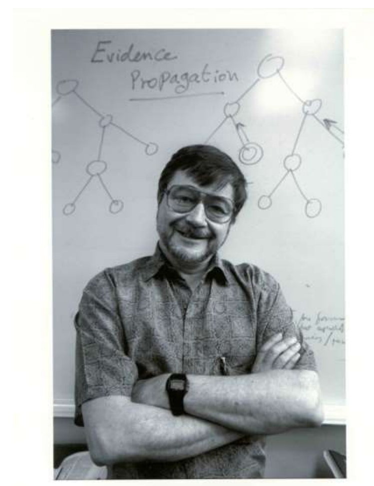
Bayesian Networks

Bayesian Network Motivation

- We want a representation and reasoning system that is based on conditional independence
  - Compact yet expressive representation
  - Efficient reasoning procedures
- Bayesian Networks are such a representation
  - Named after Thomas Bayes (ca. 1702 –1761)
  - Term coined in 1985 by Judea Pearl (1936 –)
  - Their invention changed the focus on AI from logic to probability!



Thomas Bayes



Judea Pearl

Bayesian Networks

- A Bayesian network specifies a joint distribution in a structured form

- Represent dependence/independence via a directed graph
  - Nodes = random variables
  - Edges = direct dependence

- Structure of the graph ⇔ Conditional independence relations

- Requires that graph is acyclic (no directed cycles)

- Two components to a Bayesian network
  - The graph structure (conditional independence assumptions)
  - The numerical probabilities (for each variable given its parents)

Bayesian Networks

- General form:

$$P(X_1, X_2, \ldots .X_N) = \prod_i P(X_i \mid parents(X_i))$$

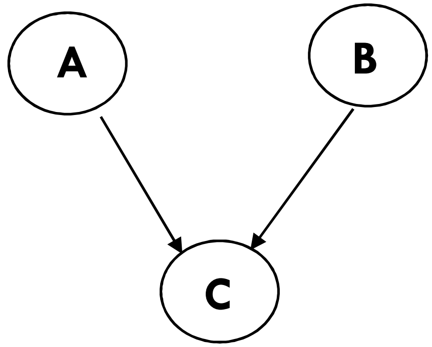
The full joint distribution

The graph-structured approximation

Example of a simple Bayesian network

$$P(X_1, X_2, \ldots . X_N) = \prod_i P(X_i \mid parents(X_i))$$

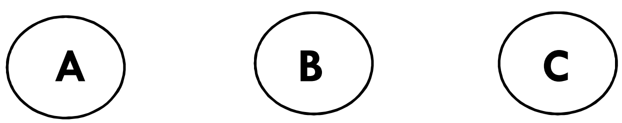
$$P(A, B, C) = P(C|A, B)P(A)P(B) \longleftrightarrow$$



- Probability model has simple factored form
- Directed edges => direct dependence
- Absence of an edge => conditional independence

- Also known as belief networks, graphical models, causal networks
- Other formulations, e.g., undirected graphical models

Examples of 3-way Bayesian Networks



Absolute Independence:
$p(A,B,C) = p(A)\, p(B)\, p(C)$

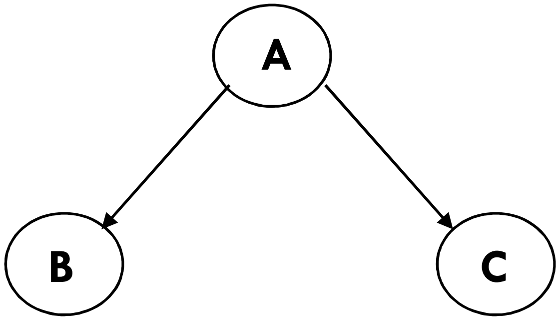
Examples of 3-way Bayesian Networks

- Conditionally independent effects:

$$p(A, B, C) = p(B|A)p(C|A)p(A)$$

- B and C are conditionally independent given A

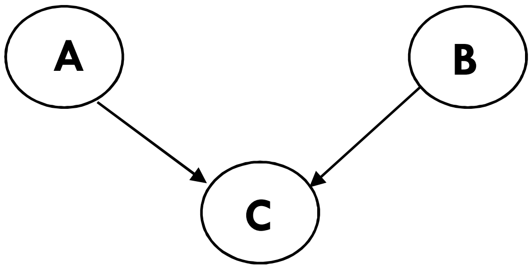
- e.g., A is a disease, and we model B and C as conditionally independent symptoms given A



Examples of 3-way Bayesian Networks

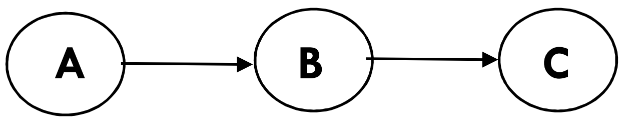
- Independent Clauses:

$$p(A, B, C) = p(C|A, B)p(A)p(B)$$



- "Explaining away" effect:
  - A and B are independent but become dependent once C is known!!
  - (we'll come back to this later)

Examples of 3-way Bayesian Networks



Markov dependence:

$p(A,B,C) = p(C|B)\ p(B|A)p(A)$

The Alarm Example

- You have a new burglar alarm installed
- It is reliable about detecting burglary, but responds to minor earthquakes
- Two neighbors (John, Mary) promise to call you at work when they hear the alarm
  - John always calls when hears alarm, but confuses alarm with phone ringing (and calls then also)
  - Mary likes loud music and sometimes misses alarm!
- Given evidence about who has and hasn't called, estimate the probability of a burglary

The Alarm Example

- Represent problem using 5 binary variables:
  - B = a burglary occurs at your house
  - E = an earthquake occurs at your house
  - A = the alarm goes off
  - J = John calls to report the alarm
  - M = Mary calls to report the alarm

- What is P(B | M, J) ?

  - We can use the full joint distribution to answer this question
    - Requires $2^5 = 32$ probabilities

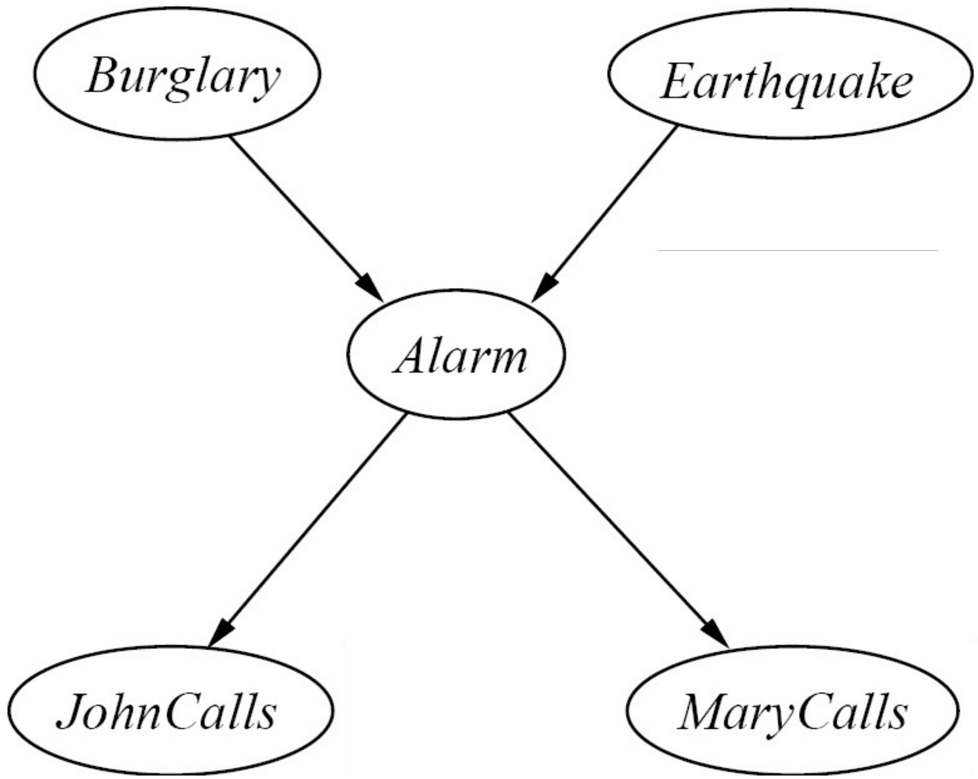
  - Can we use prior domain knowledge to come up with a Bayesian network that requires fewer probabilities?

Constructing a Bayesian Network: Step 1

- Order the variables in terms of causality (may be a partial order)
  - e.g., {E, B} -> {A} -> {J, M}

- Use these assumptions to create the graph structure of the Bayesian network

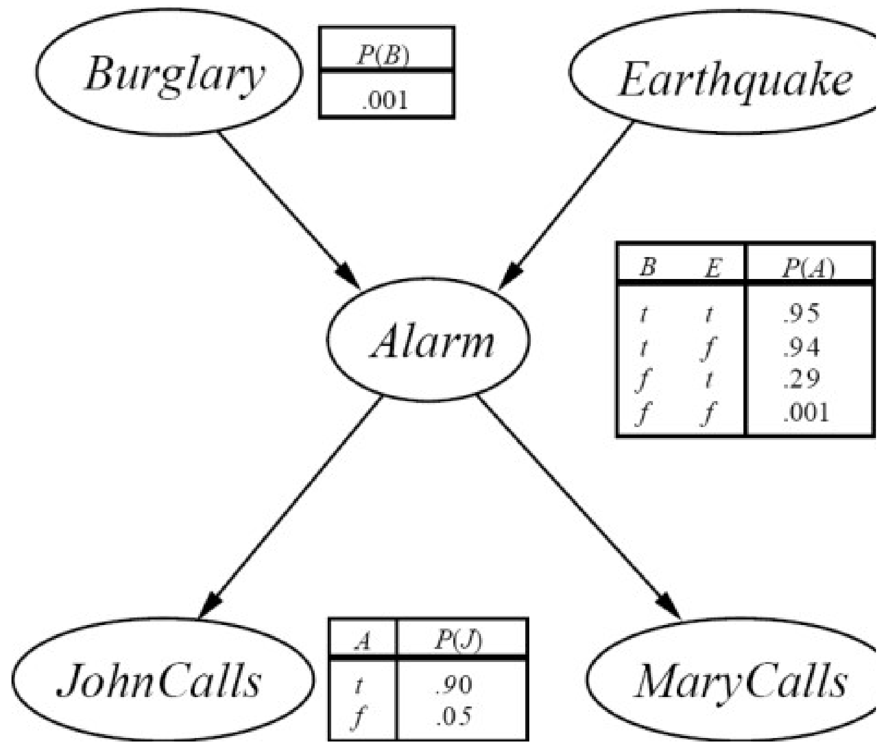
The Resulting Bayesian Network



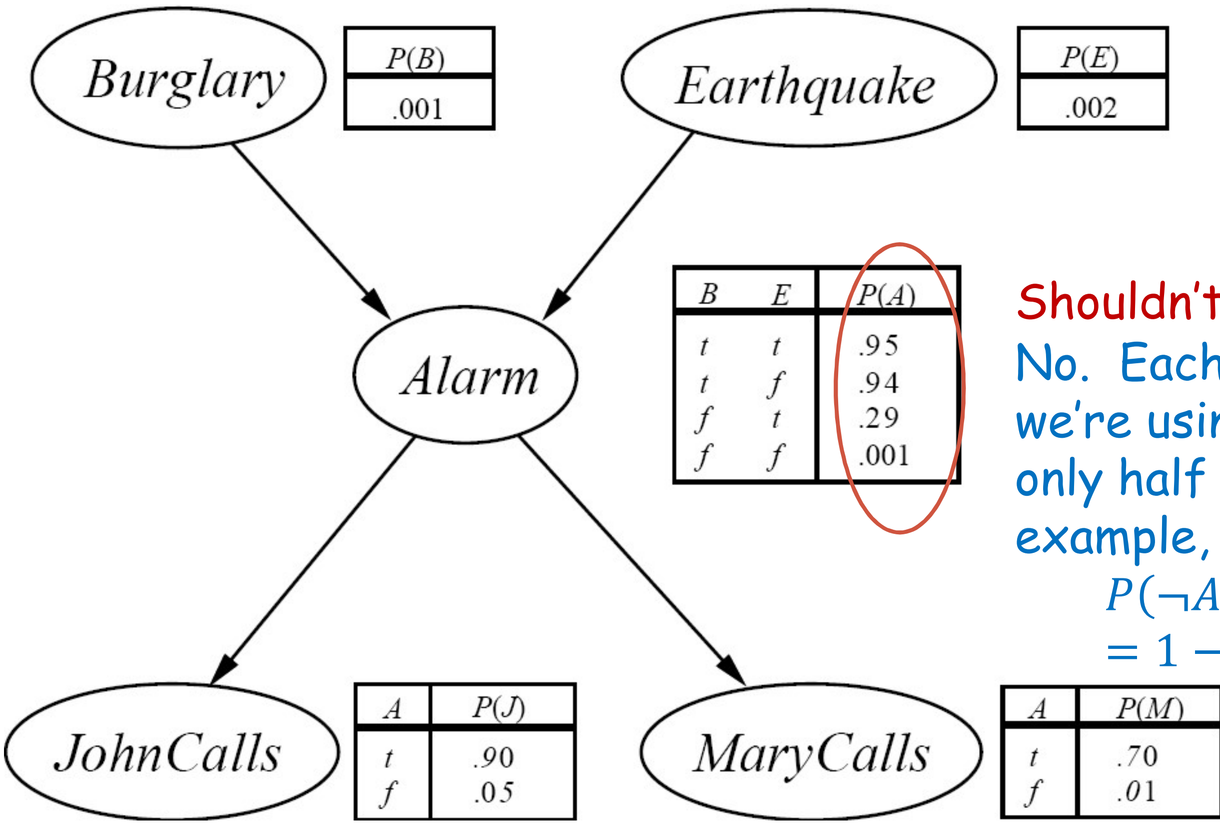
network topology reflects *causal* knowledge

Constructing a Bayesian Network: Step 2

- Fill in conditional probability tables (CPTs)
  - One for each node
  - 2^p entries, where p is the number of parents
- Where do these probabilities come from?
  - Expert knowledge
  - From data (relative frequency estimates)
  - Or a combination of both



The Bayesian network



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

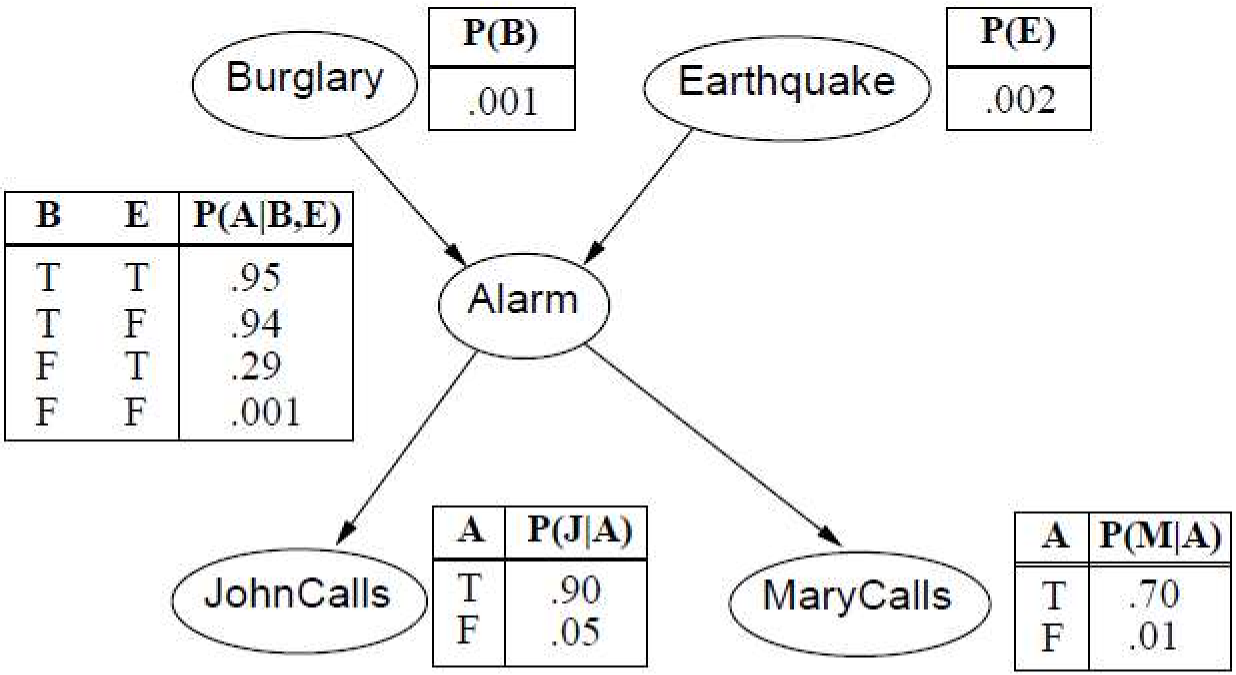
A	P(M)
t	.70
f	.01

Shouldn't these add up to 1?

No. Each row adds up to 1, and we're using this to let us show only half of the table. For example,

$$P(\neg A|B,E) = 1 - P(A|B,E)$$
$$= 1 - 0.95 = 0.05$$

The Bayesian network



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

Find the probability:

1. that there is an alarm and burglary (no earthquake), and neither John called nor Mary called.
2. that John calls