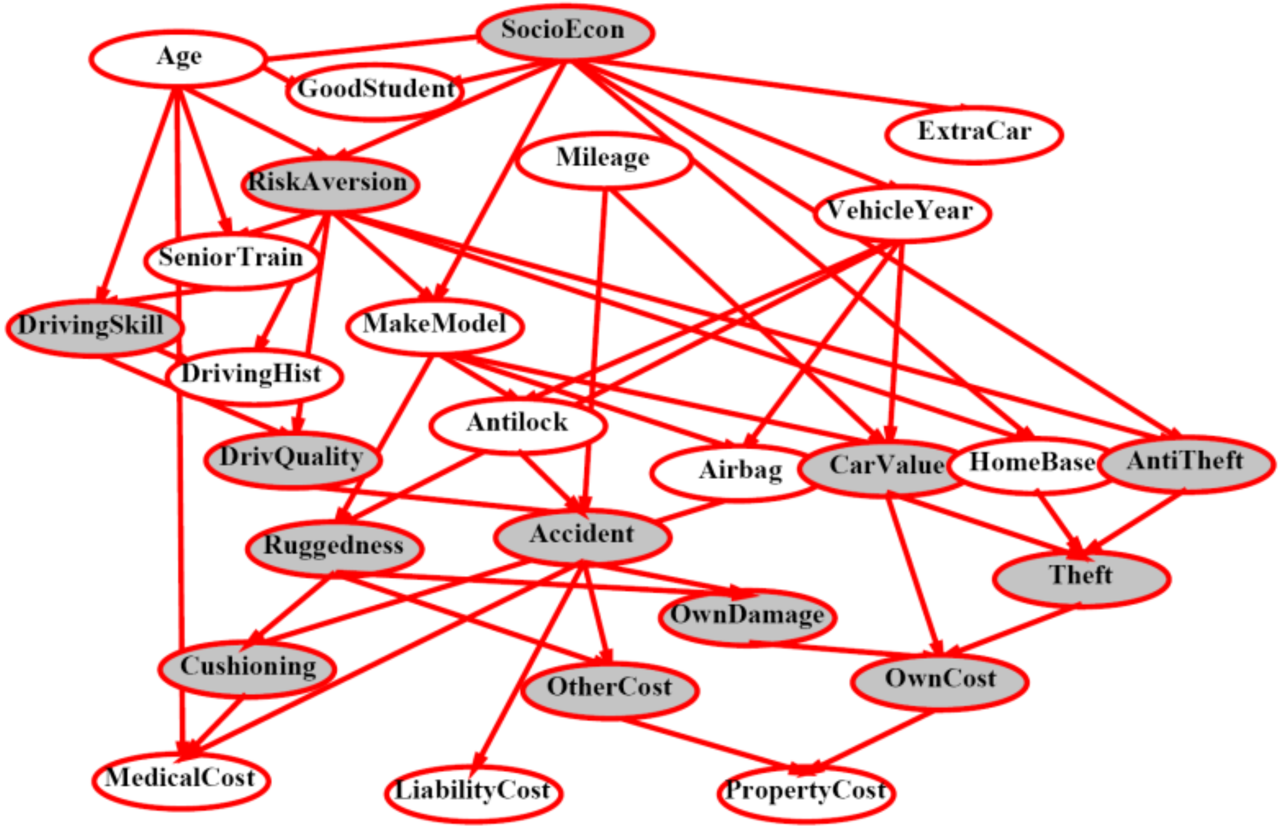
CSCI 3202: Intro to Artificial Intelligence

Lecture 15 & 16: Introduction to Bayesian Networks

Rachel Cox

Department of Computer Science



Uncertainty

Probabilistic reasoning gives us a framework for managing uncertain beliefs and knowledge.

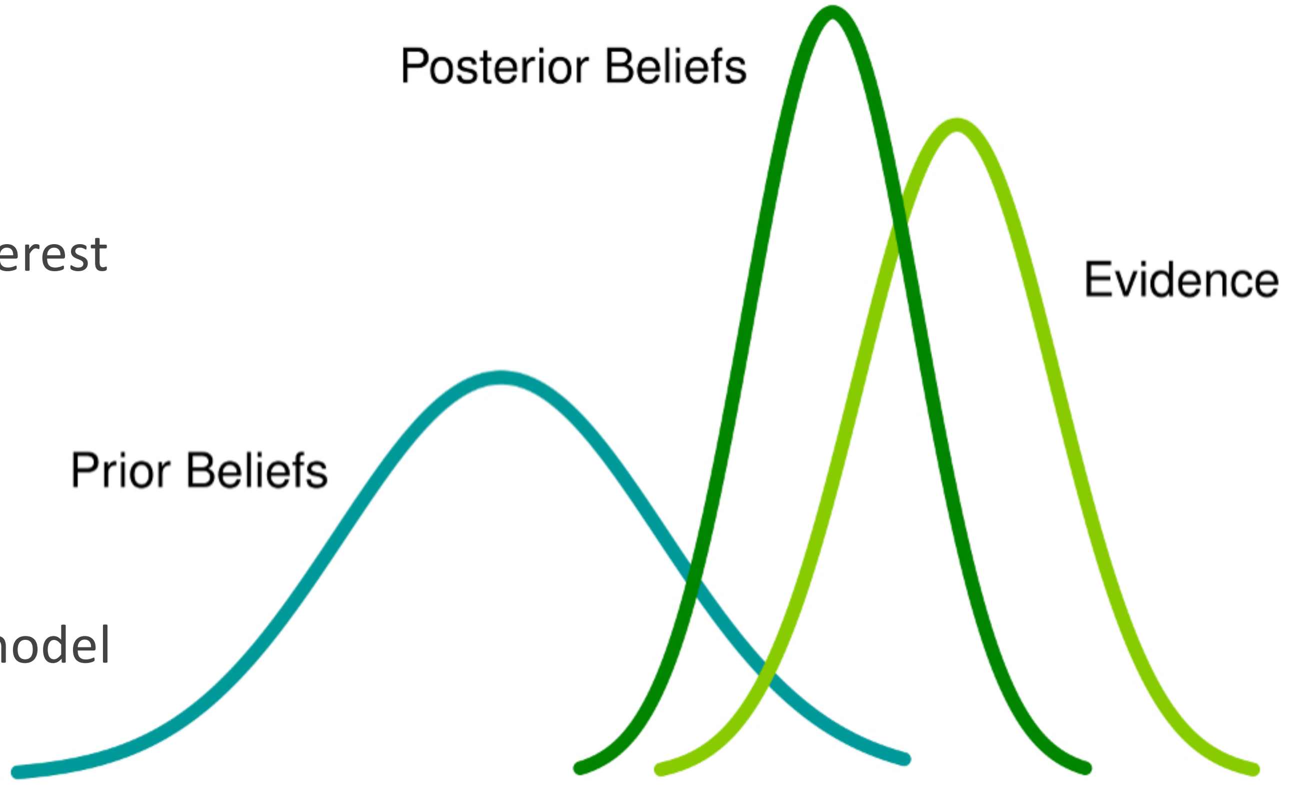
In general:

- **Observed variables (evidence)**: Agent knows certain things about the state of the world (e.g. sensor readings or symptoms)
- **Unobserved variables**: Agent needs to reason about other aspects (e.g. where an object is or what disease is present)
- **Model**: Agent knows something about how the known variables relate to the unknown variables

Bayesian Reasoning

The whole Bayesian statistical framework can be summarized as follows:

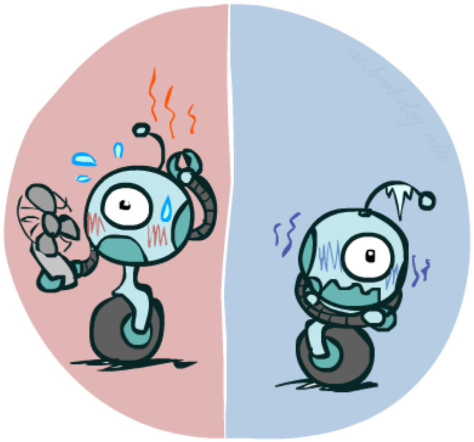
- What is your process of interest?
- Get **data** (evidence) for your process of interest
- Build a **model** $\eta(\theta)$ of this process
  Depends on uncertain parameters θ
- Formalize your *a priori* knowledge of the model parameters θ in **prior distributions**, $P(\theta)$
- **Update** your prior knowledge using the match between your model and the data



Probability Distributions

- Associate a probability with each value

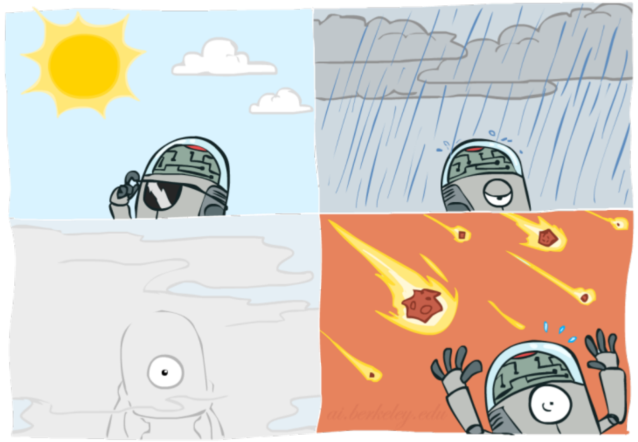
- Temperature:



$P(T)$

T	P
hot	0.5
cold	0.5

- Weather:



$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Joint Distributions

- A *joint distribution* over a set of random variables: $X_1, X_2, \ldots, X_n$ specifies a real number for each assignment (or *outcome*):

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n) \text{ or equivalently: } P(x_1, x_2, \ldots, x_n)$$

  - Must obey: $P(x_1, x_2, \ldots, x_n) \geq 0$ and

$$\sum_{(x_1, x_2, \ldots, x_n)} P(x_1, x_2, \ldots, x_n) = 1$$

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

- Size of distribution if n variables with domain sizes d?
  - For all but the smallest distributions, impractical to write out!

Probability Models

- A probability model is a joint distribution over a set of random variables

- Probability models:
  - (Random) variables with domains
  - Assignments are called *outcomes*
  - Joint distributions: say whether assignments (outcomes) are likely
  - *Normalized:* sum to 1.0
  - Ideally: only certain variables directly interact

Distribution over T,W

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(t) = \sum_w P(t, w)$$

$P(T)$

T	P
hot	
cold	

$$P(w) = \sum_t P(t, w)$$

$P(W)$

W	P
sun	
rain	

e.g. $$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$

Probability Models

Example:

- P(+x, +y) ?
- P(+x) ?
- P(-y OR +x) ?

$$P(X, Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Marginal Distributions

$P(X,Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

$$P(x) = \sum_y P(x,y)$$

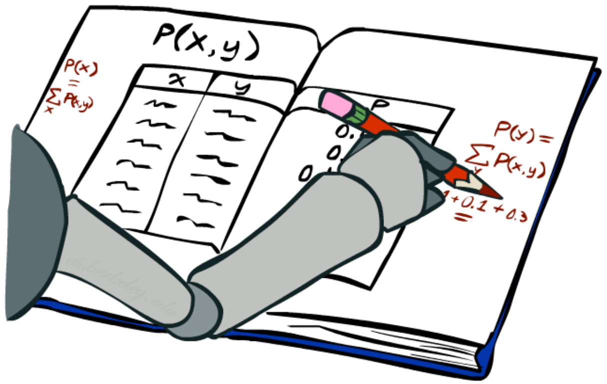
$$P(y) = \sum_x P(x,y)$$

$P(X)$

X	P
+x	
-x	

$P(Y)$

Y	P
+y	
-y	



Conditional Probabilities

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

- P(+x | +y) ?

- P(-x | +y) ?

- P(-y | +x) ?

Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others

Conditional Distributions

$P(W|T)$

$P(W|T = hot)$

W	P
sun	0.8
rain	0.2

$P(W|T = cold)$

W	P
sun	0.4
rain	0.6

Joint Distribution

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Independence

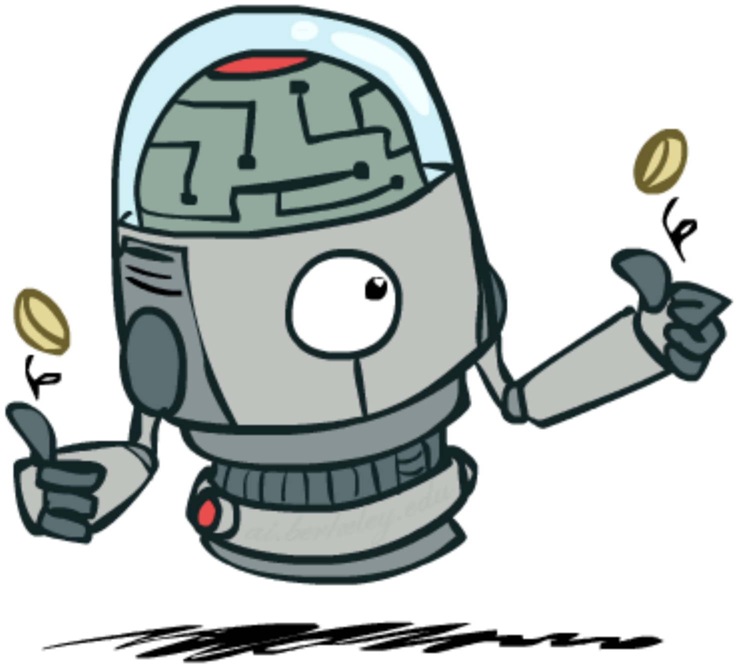
- Two variables are *independent* if:

$$\forall x, y : P(x, y) = P(x)P(y)$$

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form:

$$\forall x, y : P(x|y) = P(x)$$

- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?



Independence

$$P(T)$$

T	P
hot	0.5
cold	0.5

$$P_1(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W)$$

W	P
sun	0.6
rain	0.4

$$P_2(T, W)$$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

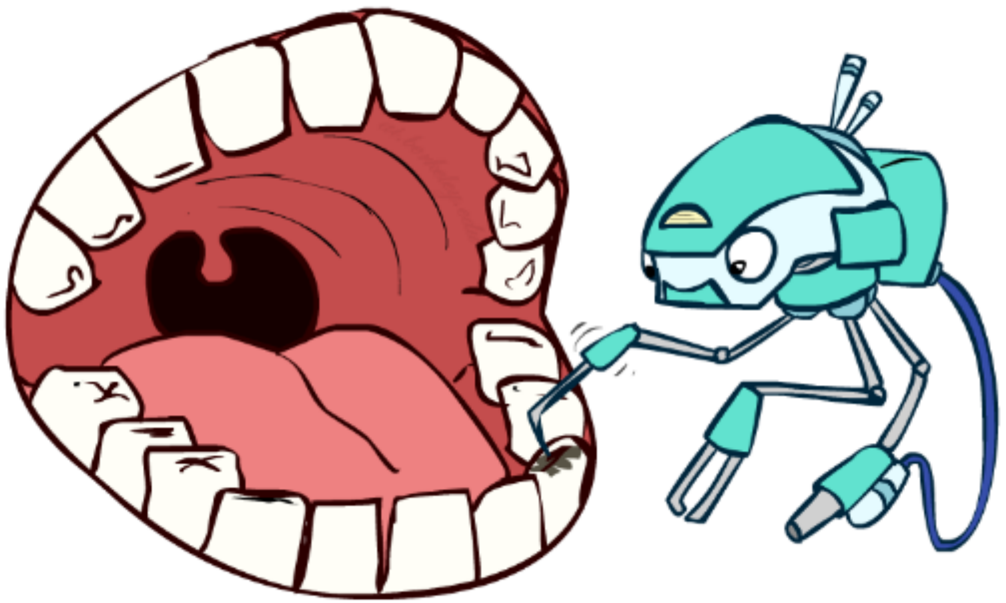
Conditional Independence

- P(Toothache, Cavity, Catch)

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)

- The same independence holds if I don't have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)

- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)



- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily

Conditional Independence

- Unconditional (absolute) independence very rare (why?)
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- X is conditionally independent of Y given Z

if and only if: $\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$

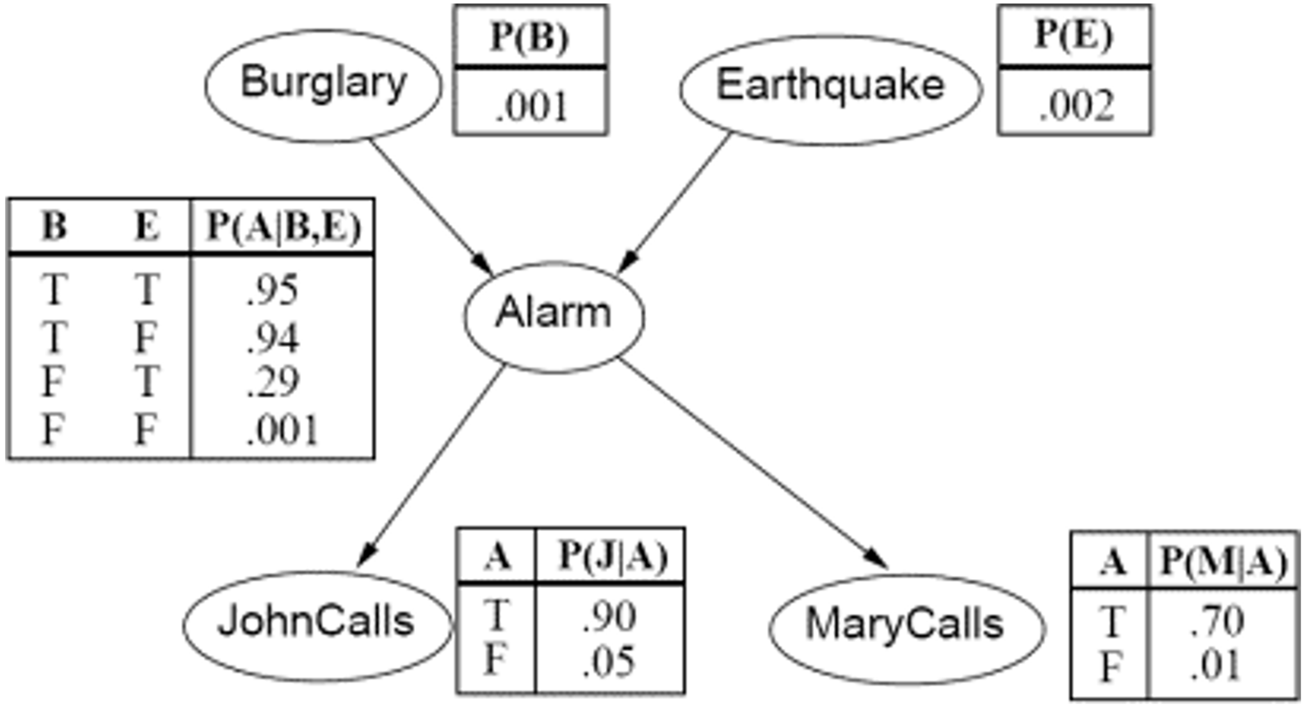
or, equivalently, if and only if $\forall x, y, z : P(x|z, y) = P(x|z)$

Bayesian Networks

The point of Bayes nets is to **represent full joint probability distributions**, *and*

to encode an interrelated set of **conditional independence/probability statements**

- Consists of **nodes** (events), and
- **conditional probability tables (CPTs)**, relating those events
- Describe how variables interact **locally**
- Chain together local interactions to estimate **global**, **indirect** interactions



Bayesian Networks

The point of Bayes nets is to **represent full joint probability distributions**, *and*

to encode an interrelated set of **conditional independence/probability statements**

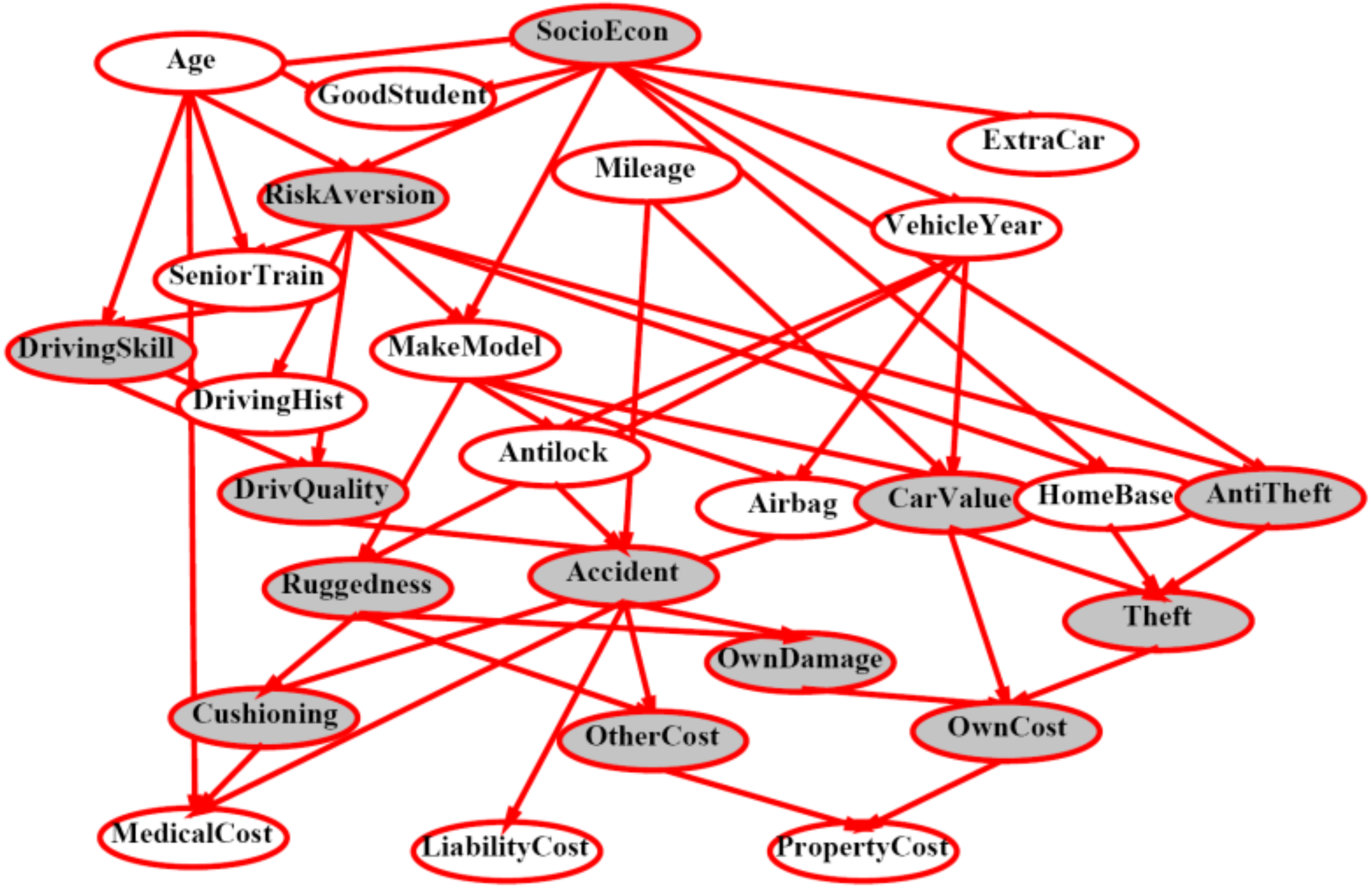
Example: Represent the full joint distribution for P(Traffic, Rain, Umbrella)

Trivial decomposition:

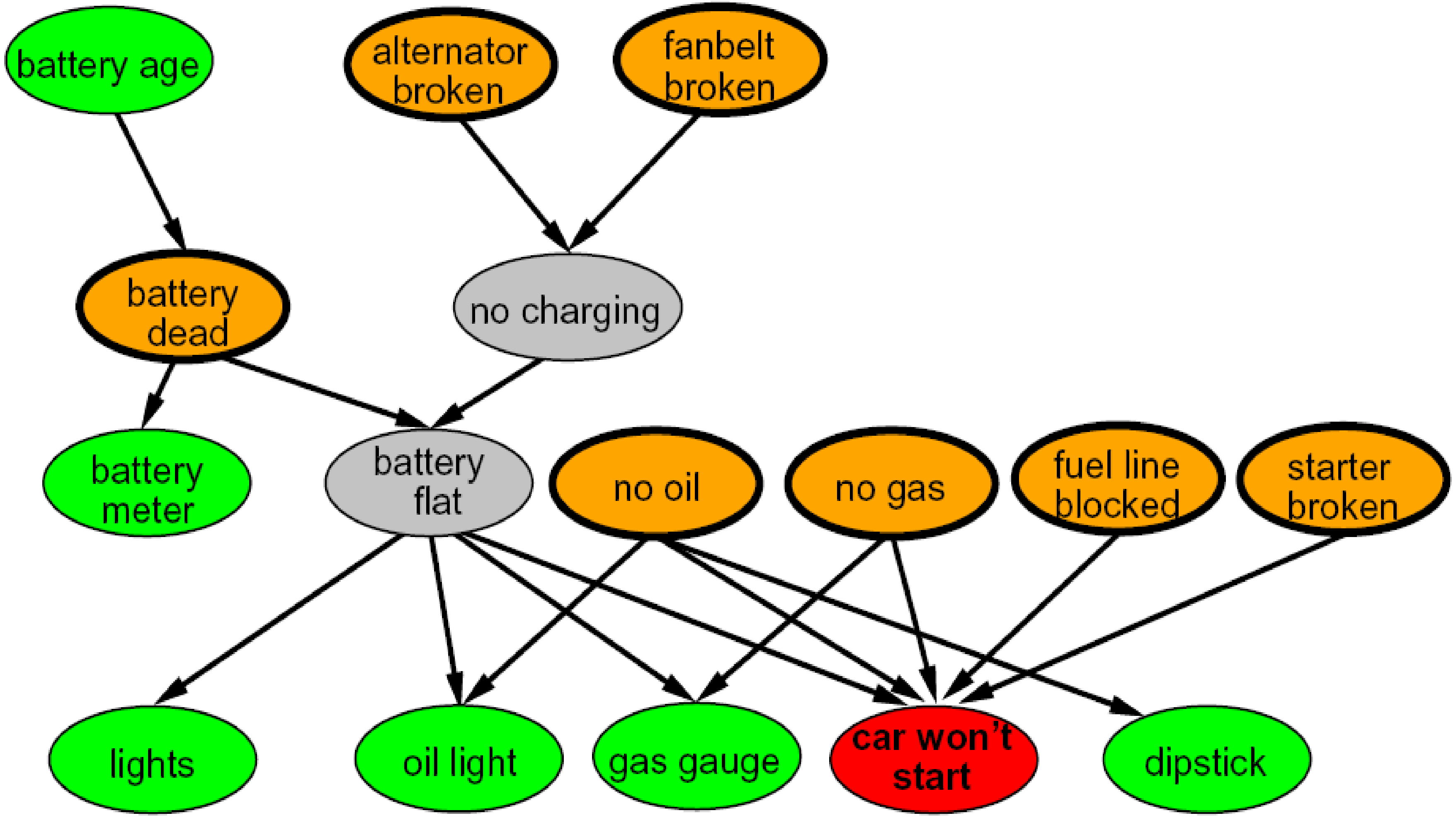
Conditional independence:

❖ **Bayes nets (graphical models) help us compactly express conditional independence assumptions/relationships.**

Graphical Model Notation



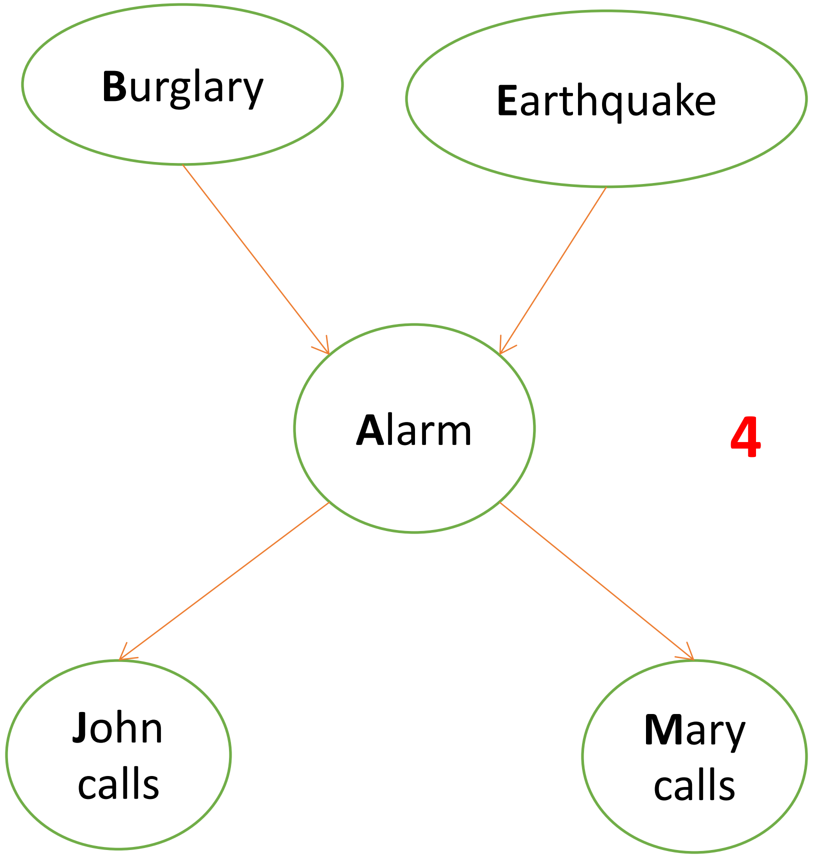
Graphical Model Notation



Bayesian Networks

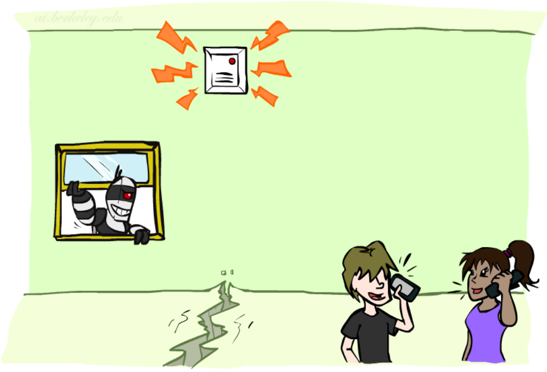
P(B)	
true	false
0.001	0.999

1



1

P(E)	
true	false
0.002	0.998



4

B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

A	P(J|A)	
	true	false
true	0.9	0.1
false	0.05	0.95

2

2

A	P(M|A)	
	true	false
true	0.7	0.3
false	0.01	0.99

Number of free parameters in each CPT:

Parent domain sizes $d_1,\ldots,d_k$

Child domain size d

Each table row must sum to 1

$(d-1)\ \Pi_i\ d_i$

Bayesian Networks

Example: Coin flips

Bayesian Networks

Example: Coin flips

Reminder:

X and *Y* are **independent** if $\forall x\ \forall y\ P(x, y) = P(x)\ P(y)$ $$X \perp\!\!\!\perp Y$$

X and *Y* are **conditionally independent given *Z*** if $\forall x\ \forall y\ \forall z\ P(x, y \mid z) = P(x \mid z)\ P(y \mid z)$

$$X \perp\!\!\!\perp Y \mid Z$$

Bayesian Networks

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

$P(x_1, x_2, \ldots, x_n) = ?$

$$
\begin{aligned}
P(x_1, x_2, \ldots, x_n) &= P(x_n \mid x_{n-1}, x_{n-2}, \ldots, x_2, x_1) P(x_{n-1}, x_{n-2}, \ldots, x_2, x_1) \\
&= P(x_n \mid x_{n-1}, x_{n-2}, \ldots, x_2, x_1) P(x_{n-1} \mid x_{n-2}, \ldots, x_2, x_1) P(x_{n-2}, \ldots, x_2, x_1) \\
&= \ldots \\
&= P(x_n \mid x_{n-1}, x_{n-2}, \ldots, x_2, x_1) P(x_{n-1} \mid x_{n-2}, \ldots, x_2, x_1) \ldots P(x_3 \mid x_2, x_1) P(x_2 \mid x_1) P(x_1) \\
&= \prod_{i=1}^{n} P(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_2, x_1)
\end{aligned}
$$

Bayesian Networks

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_2, x_1) = \ldots?$$

Node ordering: write in such a way that

$$\text{parents}(X_i) \subseteq \{X_{i-1}, X_{i-2}, \ldots, X_2, X_1\}$$

→

$$\prod_{i=1}^{n} P(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_2, x_1) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

This statement:

$$P(X_i \mid X_{i-1}, X_{i-2}, \ldots, X_2, X_1) = P(X_i \mid \text{parents}(X_i))$$

is key: **each node is conditionally independent of its other predecessors, given its parents**

Bayesian Networks

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

Example: Traffic

$P(+r, \neg t) = ?$

$P(R)$

+r	1/4
¬r	3/4

$P(T|R)$

+r →

+t	3/4
¬t	1/4

¬r →

+t	1/2
¬t	1/2

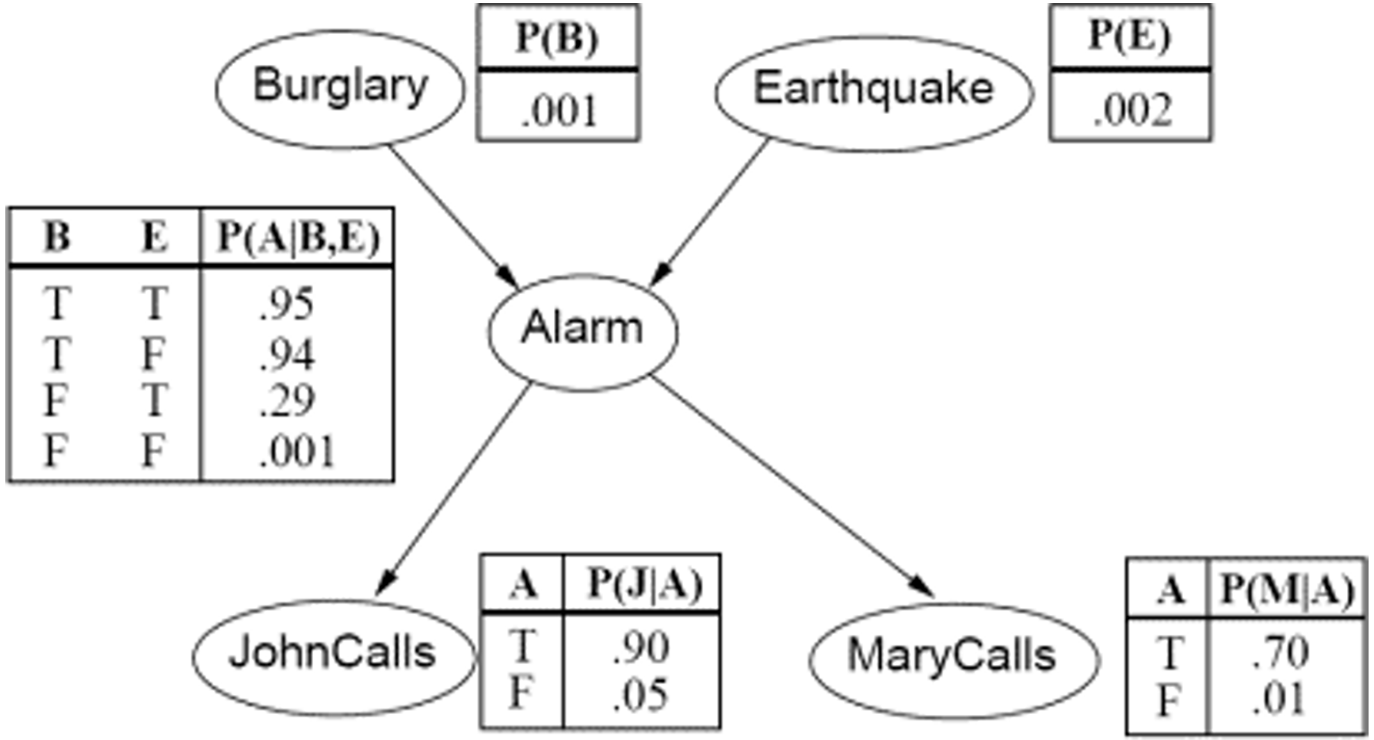
Bayesian Networks

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

Example: Burglary or earthquake?

P(¬*j*, ¬*m*, +*a*, +*b*, +*e*) = ?



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

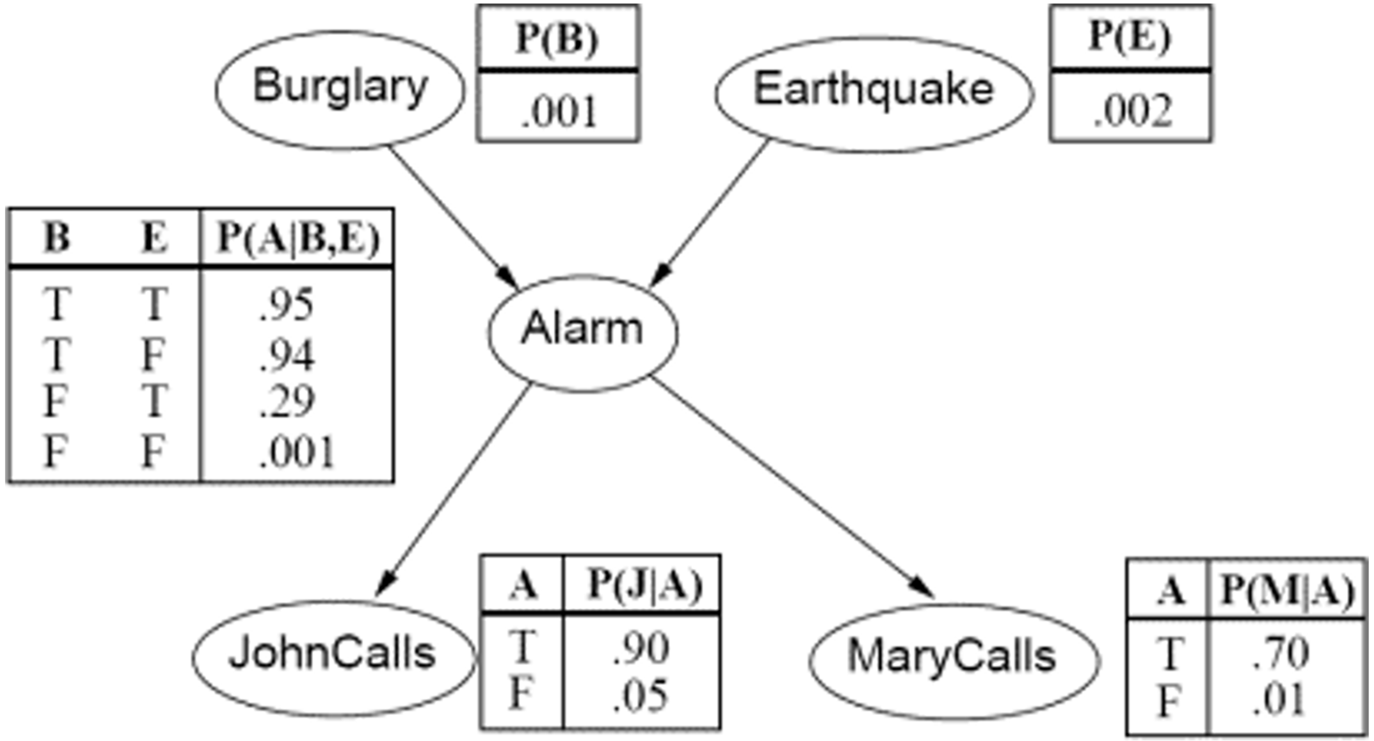
Bayesian Networks: Compactness

Full joint distribution for n Boolean nodes would require 2^n probabilities

- 5 nodes here → $2^5 = 32$

Bayes net representation:

- Permit $k = 2$ parents per node
- $n = 5$ nodes
- $5 \cdot 2^2 = 20$



Bayesian Networks: Compactness

Full joint distribution for n Boolean nodes would require 2^n probabilities

- 5 nodes here → $2^5 = 32$

Bayes net representation:

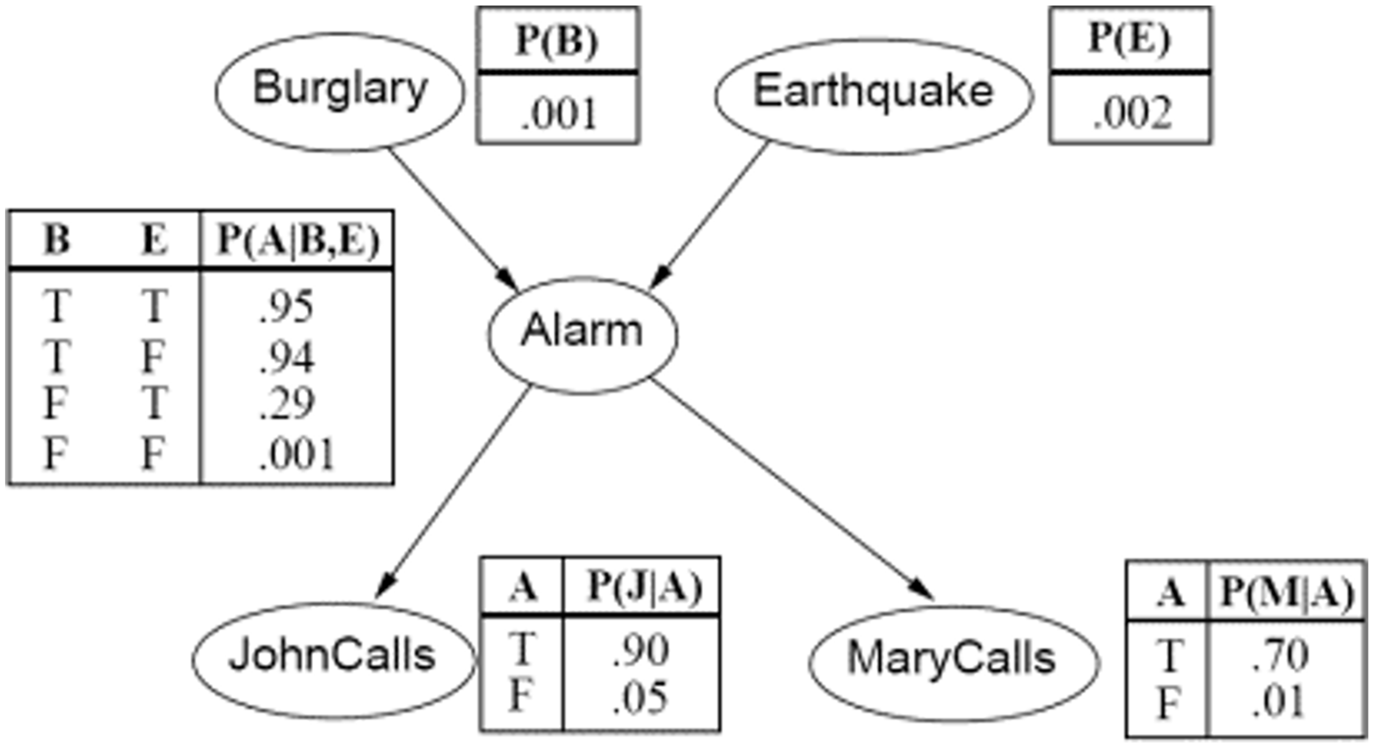
- Permit $k = 2$ parents per node
- $n = 5$ nodes
- $5 \cdot 2^2 = 20$

More extreme example:

With 5 parents per node and 30 nodes, that's

Bayes net: $n \cdot 2^k = 30 \cdot 2^5 = 960$

vs Naive: $2^{30} \approx 1{,}000{,}000{,}000$



Bayesian Networks: Construction

Show a "flow" from cause to effect: **Pearl's Network Construction Algorithm**

Nodes: What is the set of variables we need to model?

Order them: $\{X_1, X_2, X_3, \ldots, X_n\}$

Best if ordered such that **causes precede effects**

Links: For each node X_i , do:

- Choose a minimal set of parents $\text{parents}(X_i) \subseteq \{X_{i-1}, X_{i-2}, \ldots, X_2, X_1\}$ such that $P(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_1) = P(x_i \mid \text{parents}(X_i))$
- For each parent, insert arcs (links) from parent to X_i
- Write down CPT $P(X_i \mid \text{parents}(X_i))$

Bayesian Networks: Construction

Example: Suppose we have an old motorcycle that might either blow a head gasket (H) or have a broken thermometer (T). Either one would cause the bike to overheat (O). If the bike overheats, then it might blow smoke (S) and/or run weak (W).

Construct a Bayesian network for this situation.

Bayesian Networks: Construction

Example: Suppose we have an old motorcycle that might either blow a head gasket (H) or have a broken thermometer (T). Either one would cause the bike to overheat (O). If the bike overheats, then it might blow smoke (S) and/or run weak (W).

Construct a Bayesian network for this situation.

1. Node ordering: {H, T, O, W, S}
2. Insert arcs

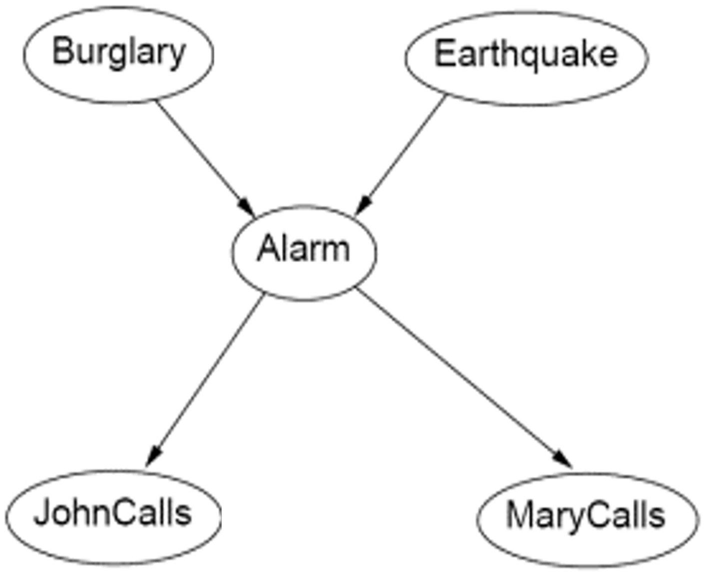
Bayesian Networks: Construction

Here, we chose to put the causes before effects:

{Burglary, Earthquake, Alarm, JohnCalls, MaryCalls}

What if instead we did the following?

{MaryCalls, JohnCalls, Alarm,
Burglary, Earthquake}

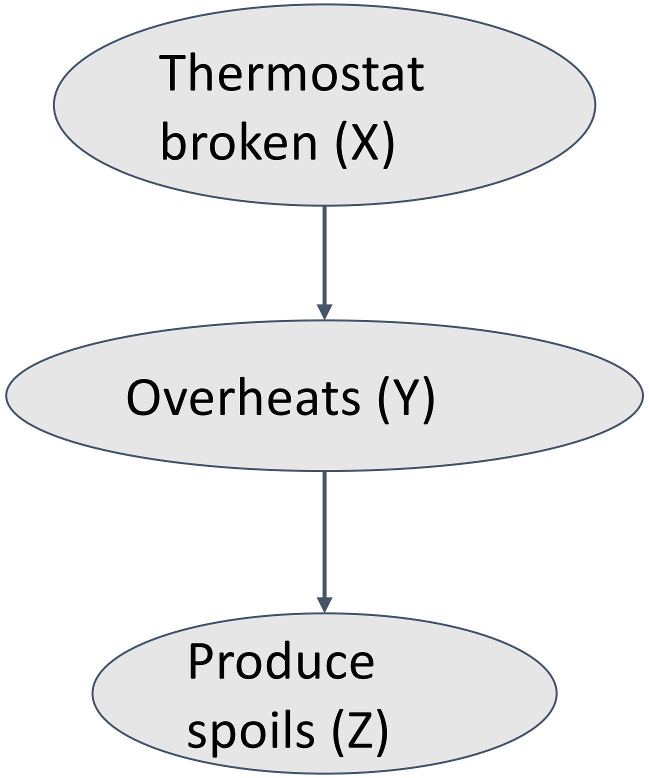


Bayesian Networks: Canonical Cases

Important Bayes net question: Are two nodes independent *given* certain evidence?

- If yes -- can prove using algebra
- If no -- can prove using a counterexample

Example: Are X and Z necessarily independent?



Bayesian Networks: Canonical Cases

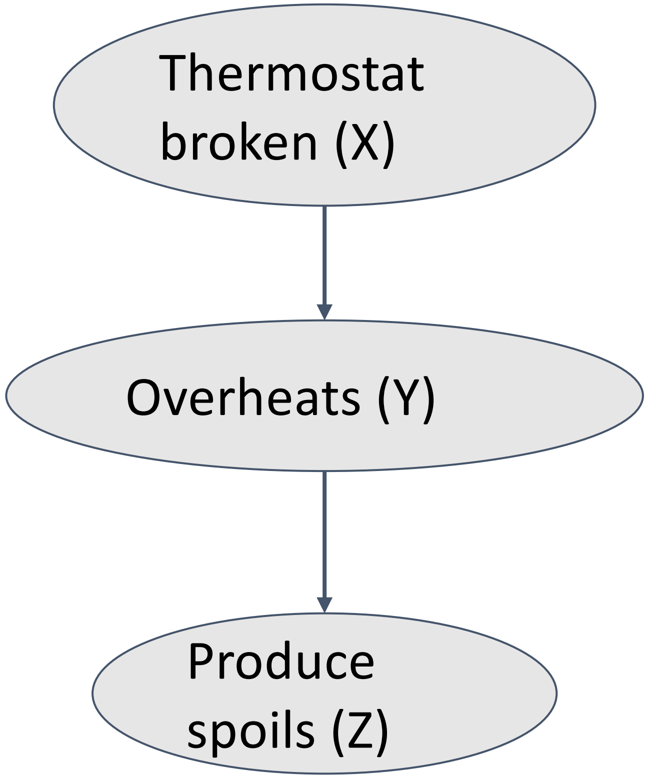
Important Bayes net question: Are two nodes independent *given* certain evidence?

- If yes -- can prove using algebra
- If no -- can prove using a counterexample

Example: Are X and Z necessarily independent?

No!

- X certainly influences Y, which influences Z
- Also, knowledge of Z influences beliefs about X (through Y)



Bayesian Networks: Canonical Cases

Important Bayes net question: Are two nodes independent *given* certain evidence?

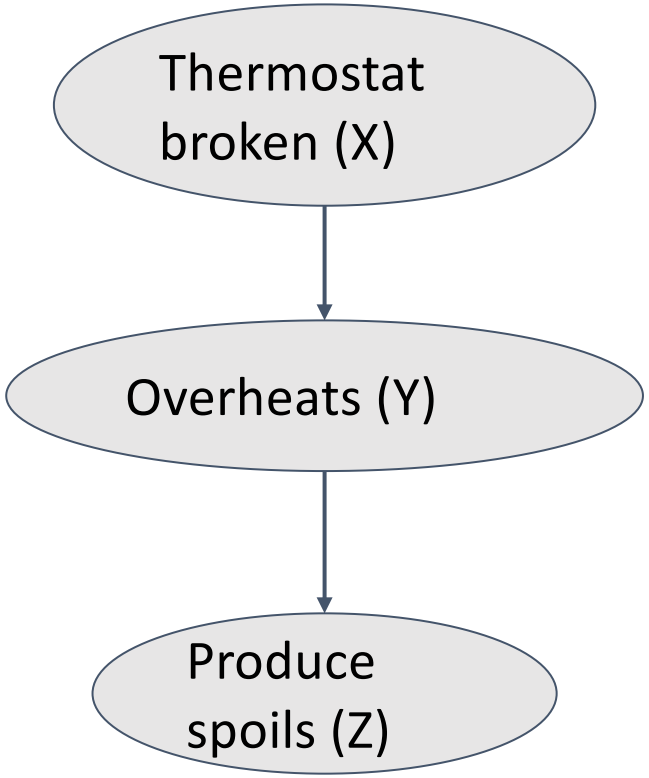
- If yes -- can prove using algebra
- If no -- can prove using a counterexample

Example: Are X and Z necessarily independent?

No!

- X certainly influences Y, which influences Z
- Also, knowledge of Z influences beliefs about X (through Y)

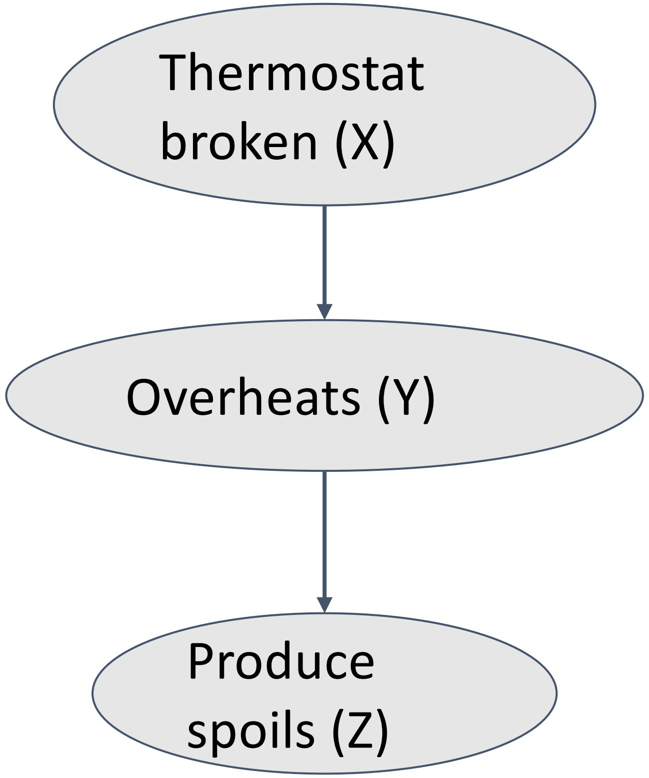
Example, rebooted: What about X and Z, ***given*** Y?



Bayesian Networks: Canonical Cases

Example, rebooted: What about X and Z, ***given*** Y?

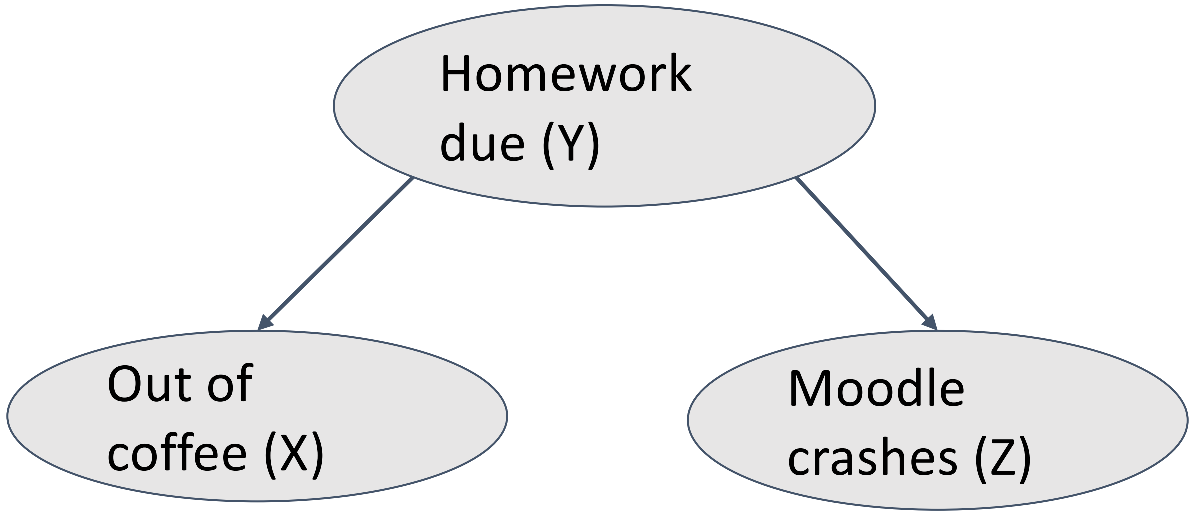
This is a canonical case is called a **causal chain**



Bayesian Networks: Canonical Cases

Common cause is another canonical case.

→ Two effects, from the same cause



Example: Are X and Z independent?

Are X and Z independent ***given*** Y?

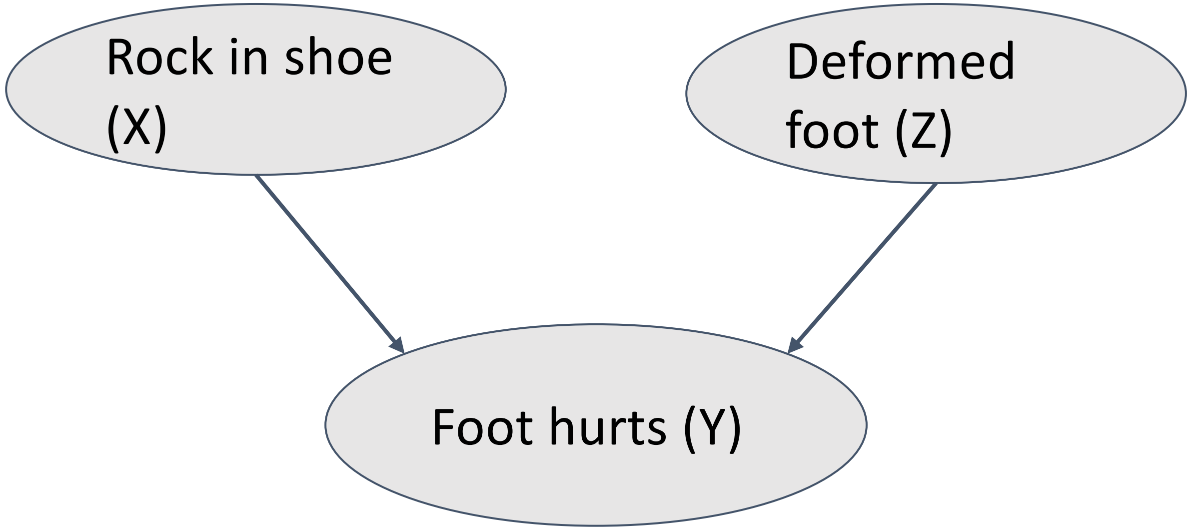
Bayesian Networks: Canonical Cases

Common effect is the third canonical case.

→ One effect, two possible causes

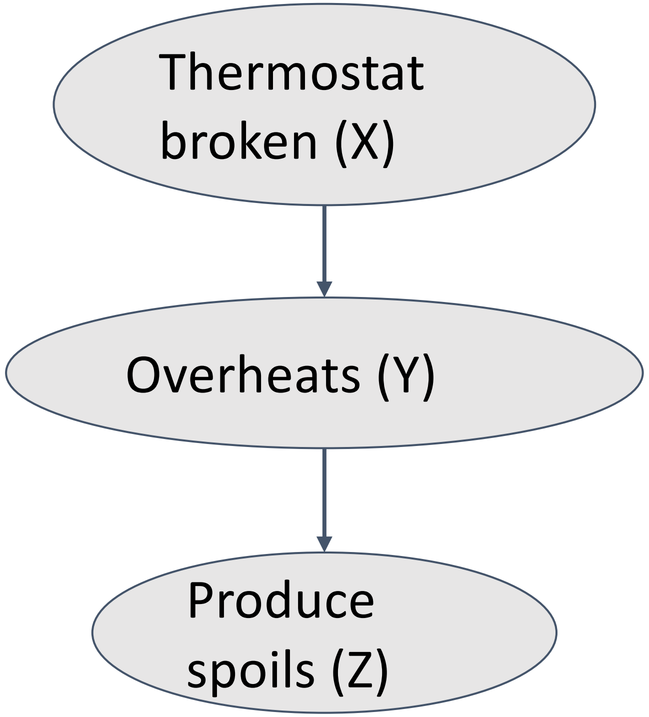
Example: Are X and Z independent?

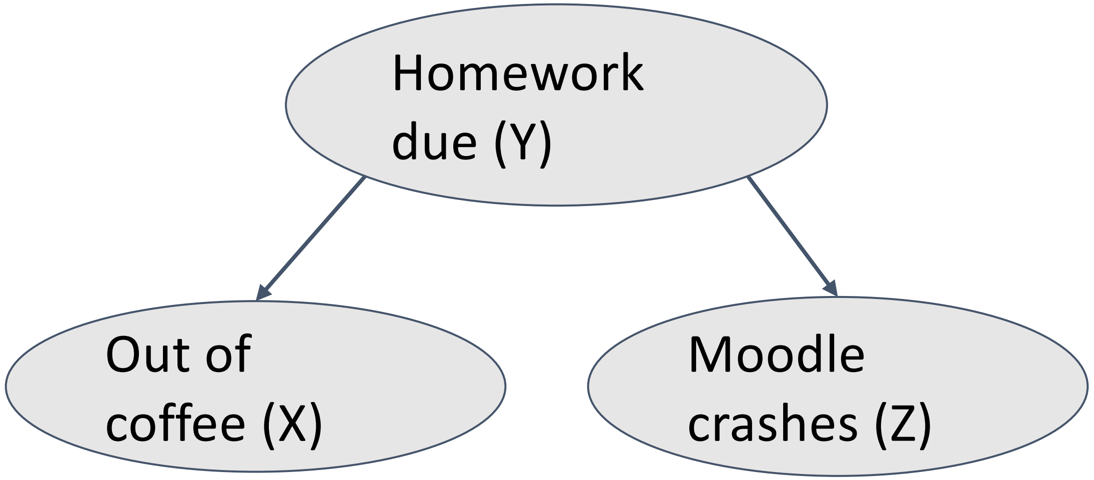
Are X and Z independent ***given*** Y?

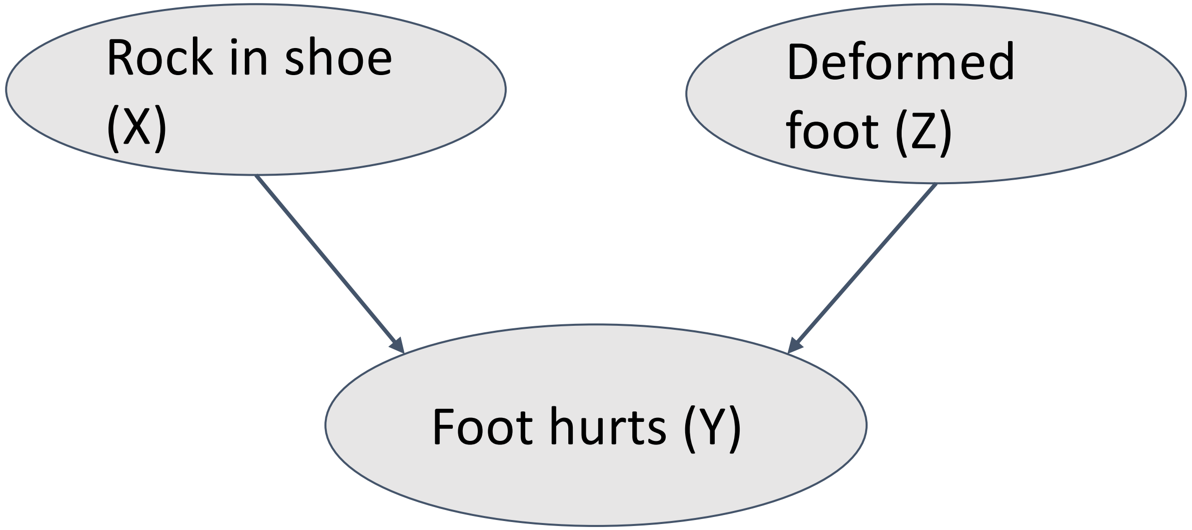


Causal vs Diagnostic Modeling

Causal:







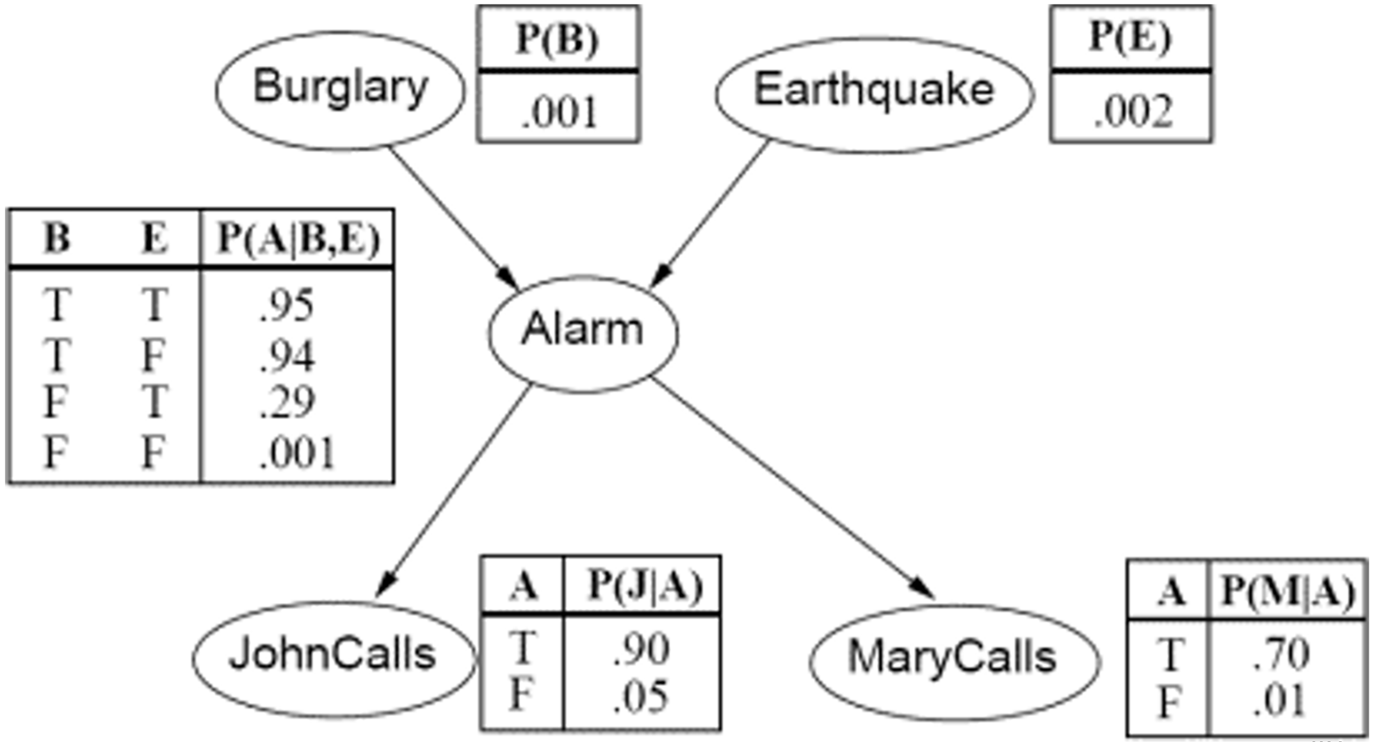
Diagnostic: observing an effect leads to competition between possible causes

→ *diagnose* which is most likely

Bayesian Networks: “Explaining Away”

Suppose we know that the alarm has gone off.

Suppose we find out later that we have been robbed



Next Time

- *Notebook Day on Monday*
- *Sampling using Bayesian networks on Wednesday!*