CSCI 3202: Intro to Artificial Intelligence
Lecture 17: Bayesian Inference and Sampling

Rachel Cox
Department of Computer Science

HW4 - now posted
- Due Friday March 20th

Problem1 - hand computations

Problem2 - very similar to the Inclass - BayesNet.ipynb

Problem3 - Bayesian network with modification.

Bayesian Networks

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

chain rule

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

Node ordering: write in such a way that

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

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

entire chain rule

This statement:

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

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

because of conditional independence

Bayesian Networks

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

entire joint probability distribution $\boxed{P(B, E, A, J, M) = P(B) \cdot P(E) \cdot P(A|B,E) P(J|A) P(M|A)}$

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

$p(\neg B) = 1 - .001$

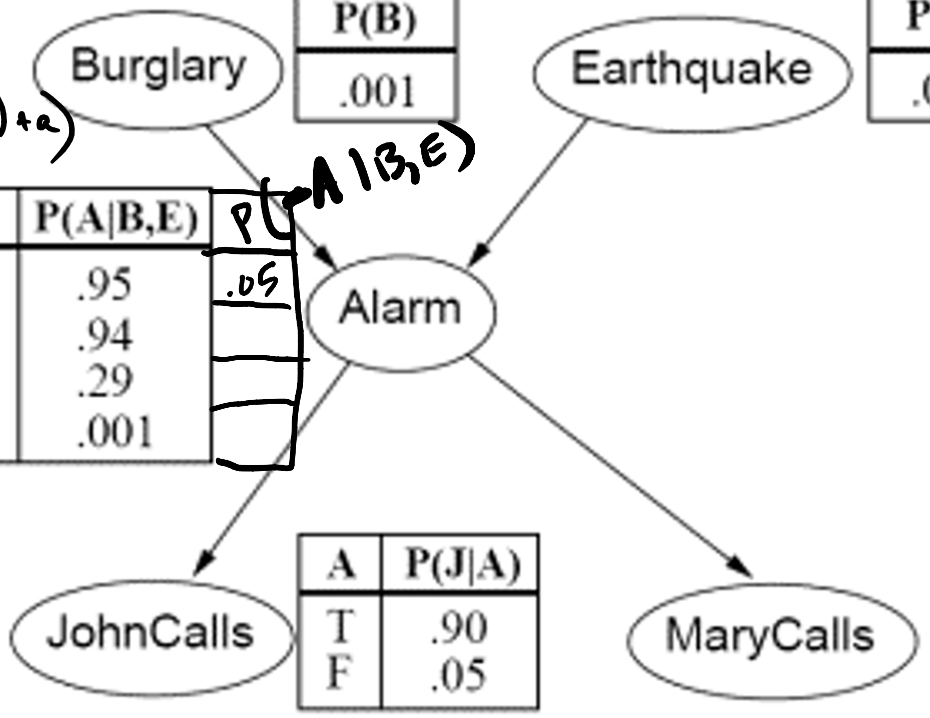
Example: Burglary or earthquake?

P(¬j, ¬m, +a, +b, +e) = ?

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

$= (0.001)(0.002) \cdot (.95)(1-0.90)(1-0.70)$

$\boxed{= 0.00000057}$



P(B)
.001

P(E)
.002

B	E	P(A|B,E)	P(¬A|B,E)
T	T	.95	.05
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: 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:

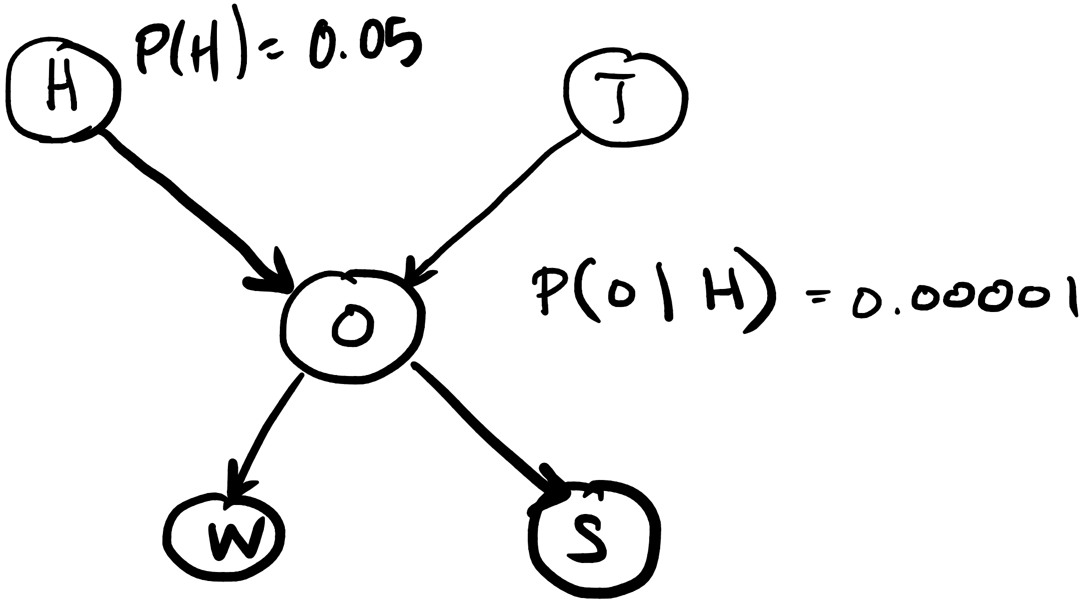
- Choose a minimal set of parents $\text{parents}(X_i) \subseteq \{X_{i-1}, X_{i-2}, \ldots, X_2, X_1\}$
  such that $P(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_1) = P(x_i \mid \text{parents}(X_i))$
- For each parent, insert arcs (links) from parent to X_i
- Write down CPT $P(X_i \mid \text{parents}(X_i))$

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.

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



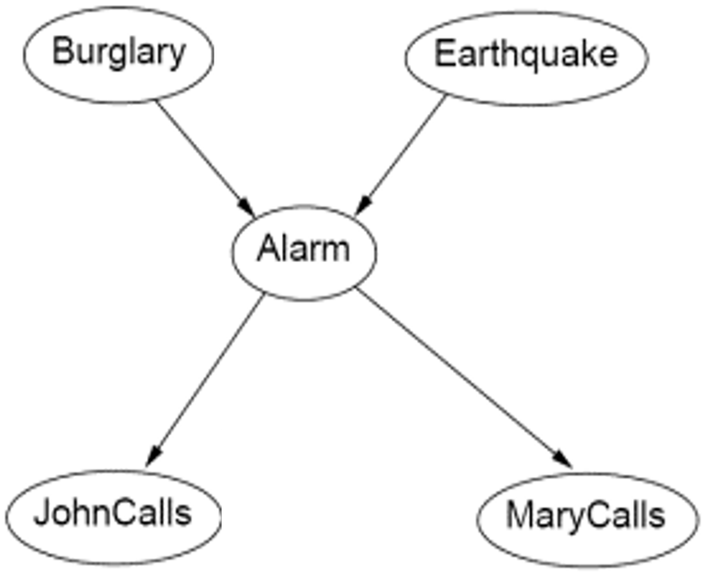
Bayesian Networks: Construction

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}



$$P(B, E, A, J, M) = P(J, M, A, E, B)$$

$$P(B)P(E)P(A|B,E) \cdot P(J|A)P(M|A)$$

$$P(J)P(M)(P(A|J,M,E) \ldots$$

Bayesian Networks: Construction with Python Example

What do we need to describe a Bayesian Network? Nodes, Arcs, Conditional Probability Tables (CPTs)

particular_bayes_net = BayesNet([list of nodes: (‘Name’, ‘Parents, [T, F], {dict cpt})

class BayesNet:
- generic “tree class” that will interact with a more specific Node class
- Read in the “Node Specifications”, like what’s listed above.
- Add Nodes using class BayesNode

class BayesNode:
- node = BayesNode(name=name, parents = parents, values = values, cpt = cpt)

We’ll have a **probability function** - calculates the probability of seeing a particular BayesNode variable when the parents’ values are given as “evidence”

We’ll have a **class PDF_discrete**

Lastly, we’ll need a way to compute probabilities and ”ask” things of our Bayesian Network: **enumeration_ask**, **enumeration_all** pseudocode in textbook

Bayesian Networks: Construction with Python Example

Example: Find $P(Burglary = T \mid JohnCalls = T, MaryCalls = T)$

Entire Joint Probability Distribution:

$$P(B,E,A,J,M) = P(B)P(E)\,P(A \mid B,E)\,P(J \mid A)\,P(M \mid A) \quad ★$$

$$P(+B \mid +J,+M) = \frac{P(+B,+J,+M)}{P(+J,+M)}$$

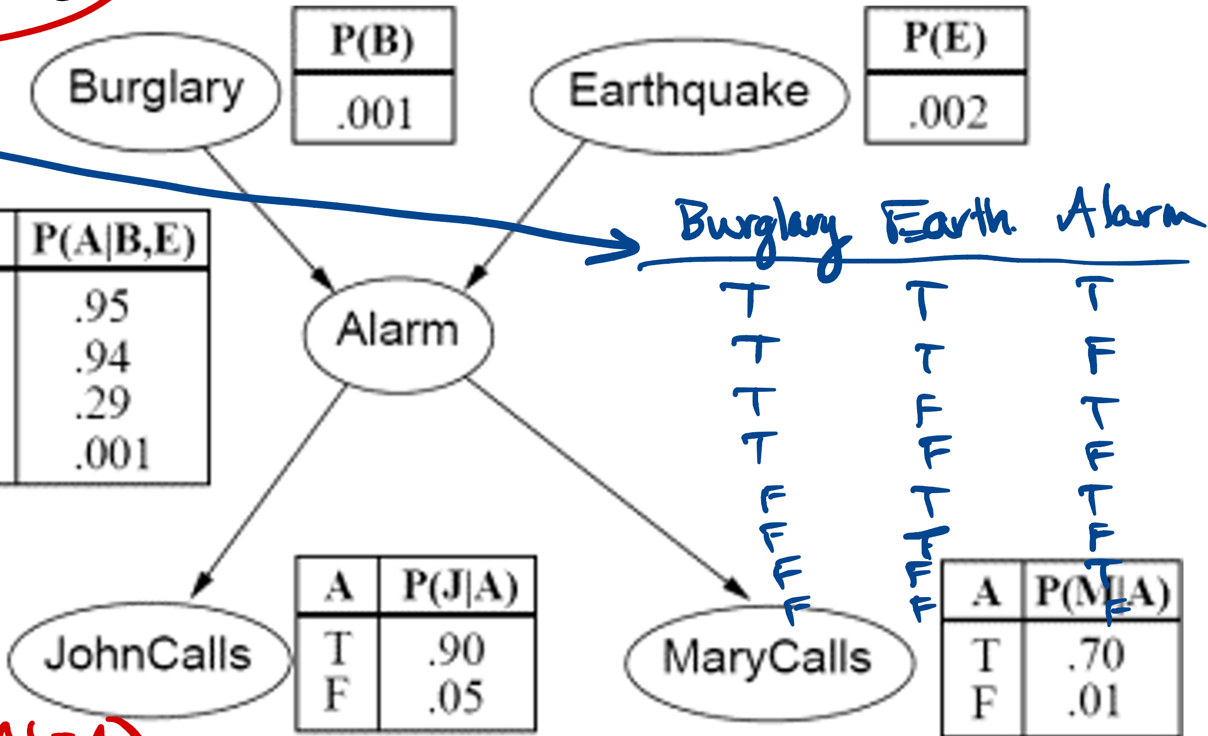
By def. of conditional probability.

$$= \frac{P(+B,\pm E,\pm A,+J,+M)}{P(\pm B,\pm E,\pm A,+J+M)}$$

$$= \frac{P(+B,+E,+A,+J+M) + P(+B,+E,-A,+J+M) + P(+B,-E,+A,+J+M) + P(+B,-E,-A,+J,+M)}{P(\pm B,\pm E,\pm A,+J+M)}$$

$$P(+B,+E,-A,+J,+M) = P(+B)P(+E)\cdot P(-A \mid +B,+E)\cdot P(+J \mid -A)\,P(+M \mid -A)$$

Earthquake	Alarm
T	T
T	F
F	T
F	F



P(B)
.001

P(E)
.002

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

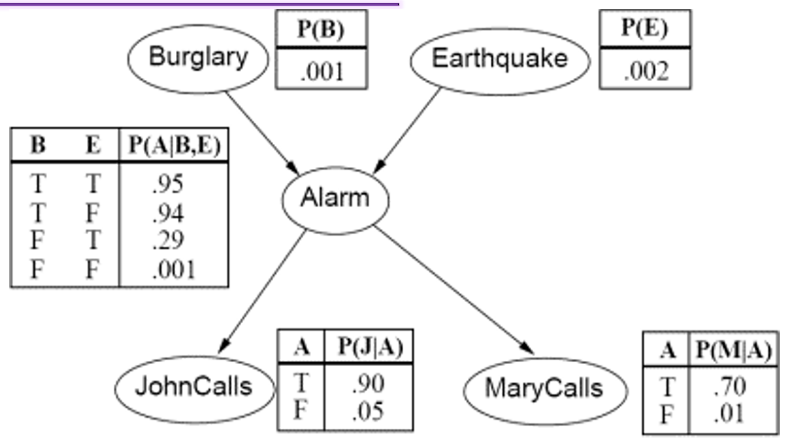
A	P(J|A)
T	.90
F	.05

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

Burglary	Earth	Alarm
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

Bayesian Networks: Construction with Python Example

Example: Find $P(Burglary = T \mid JohnCalls = T, MaryCalls = T)$



Entire Joint Probability Distribution:

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

$$P(Burglary = T \mid JohnCalls = T, MaryCalls = T) = \frac{P(Burglary = T, JohnCalls = T, MaryCalls = T)}{P(JohnCalls = T, MaryCalls = T)}$$

$$= \frac{P(Burglary = T, \boldsymbol{Earthquake = T\ or\ F, Alarm = T\ or\ F}, JohnCalls = T, MaryCalls = T)}{P(\boldsymbol{Burglary = T\ or\ F, Earthquake = T\ or\ F, Alarm = T\ or\ F}, JohnCalls = T, MaryCalls = T)}$$

$$= \frac{P(+B,+E,+A,+J,+M) + P(+B,+E,-A,+J,+M) + P(+B,-E,+A,+J,+M) + P(+B,-E,-A,+J,+M)}{P(+B,+E,+A,+J,+M) + P(+B,+E,-A,+J,+M) + P(+B,-E,+A,+J,+M) + P(+B,-E,-A,+J,+M) + P(-B,+E,+A,+J,+M) + P(-B,+E,-A,+J,+M) + P(-B,-E,+A,+J,+M) + P(-B,-E,-A,+J,+M)}$$

$$= \frac{p(+B)p(+E)p(+A|+B,+E)p(+J|+A)p(+M|+A) + p(+B)p(+E)p(-A|+B,+E)p(+J|-A)p(+M|-A) + p(+B)p(-E)p(+A|+B,-E)p(+J|+A)p(+M|+A) + p(+B)p(-E)p(-A|+B,-E)p(+J|-A)p(+M|-A)}{\begin{array}{l} p(+B)p(+E)p(+A|+B,+E)p(+J|+A)p(+M|+A) + p(+B)p(+E)p(-A|+B,+E)p(+J|-A)p(+M|-A) + p(+B)p(-E)p(+A|+B,-E)p(+J|+A)p(+M|+A) + p(+B)p(-E)p(-A|+B,-E)p(+J|-A)p(+M|-A) \\ + p(-B)p(+E)p(+A|-B,+E)p(+J|+A)p(+M|+A) + p(-B)p(+E)p(-A|-B,+E)p(+J|-A)p(+M|-A) + p(-B)p(-E)p(+A|-B,-E)p(+J|+A)p(+M|+A) + p(-B)p(-E)p(-A|-B,-E)p(+J|-A)p(+M|-A) \end{array}}$$

$$= \frac{(0.001)(0.002)(0.95)(0.90)(0.70) + (0.001)(0.002)(1-0.95)(0.05)(0.01) + (0.001)(1-0.002)(0.94)(0.90)(0.70) + (0.001)(1-0.002)(1-0.94)(0.05)(0.01)}{\begin{array}{l} (0.001)(0.002)(0.95)(0.90)(0.70) + (0.001)(0.002)(1-0.95)(0.05)(0.01) + (0.001)(1-0.002)(0.94)(0.90)(0.70) + (0.001)(1-0.002)(1-0.94)(0.05)(0.01) \\ + (1-0.001)(0.002)(0.29)(0.90)(0.70) + (1-0.001)(0.002)(1-0.29)(0.05)(0.01) + (1-0.001)(1-0.002)(0.001)(0.90)(0.70) + (1-0.001)(1-0.002)(1-0.001)(0.05)(0.01) \end{array}}$$

$$= \frac{0.0000012 + 0.000000001 + 0.0005910156 + 0.00000002994}{0.0000012 + 0.000000001 + 0.0005910156 + 0.00000002994 + 0.0003650346 + 0.00000070929 + 0.00062811126 + 0.000498002499} = \frac{0.00059224654}{0.002084104189} = \mathbf{0.284173}$$

Bayesian Networks: Canonical Cases

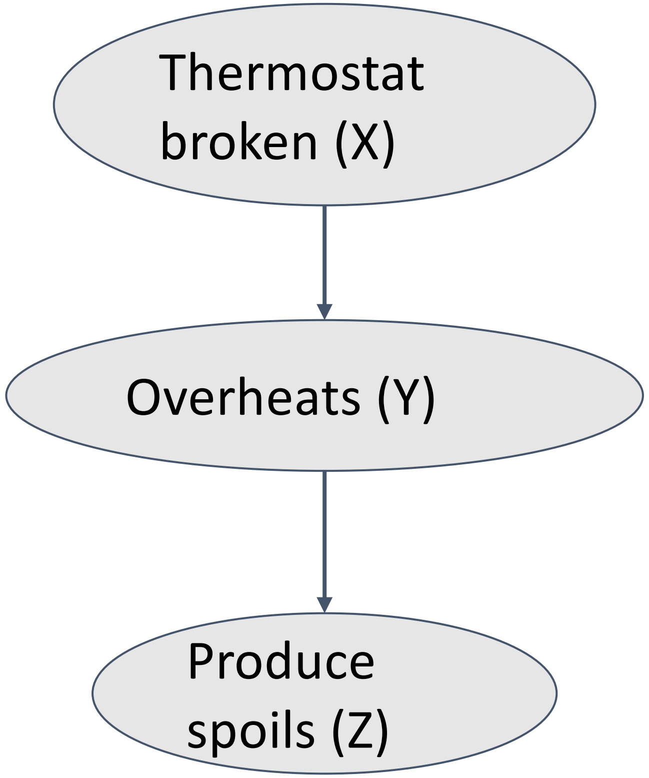
Important Bayes net question: Are two nodes independent *given* certain evidence?

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

Example: Are X and Z necessarily independent?

$\boxed{\text{no}}$

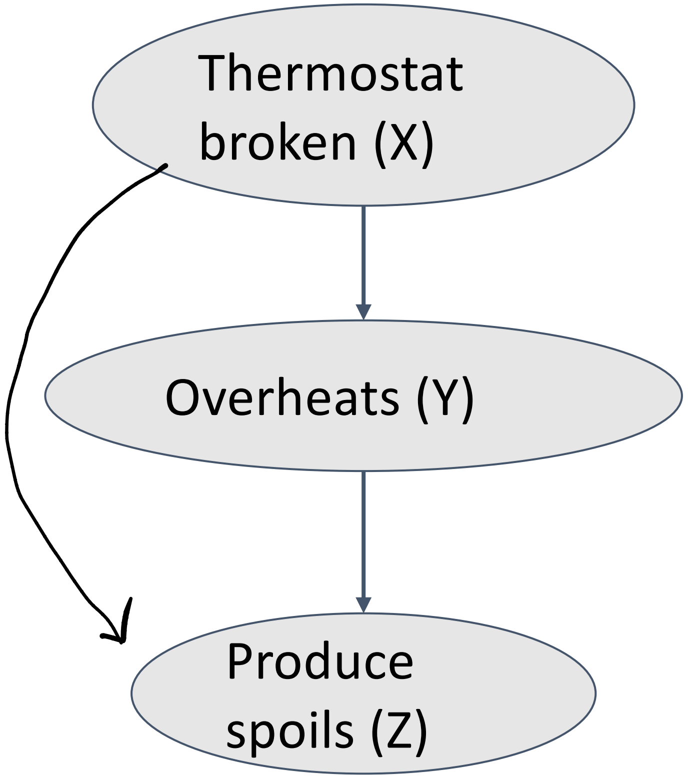
Causal chain



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

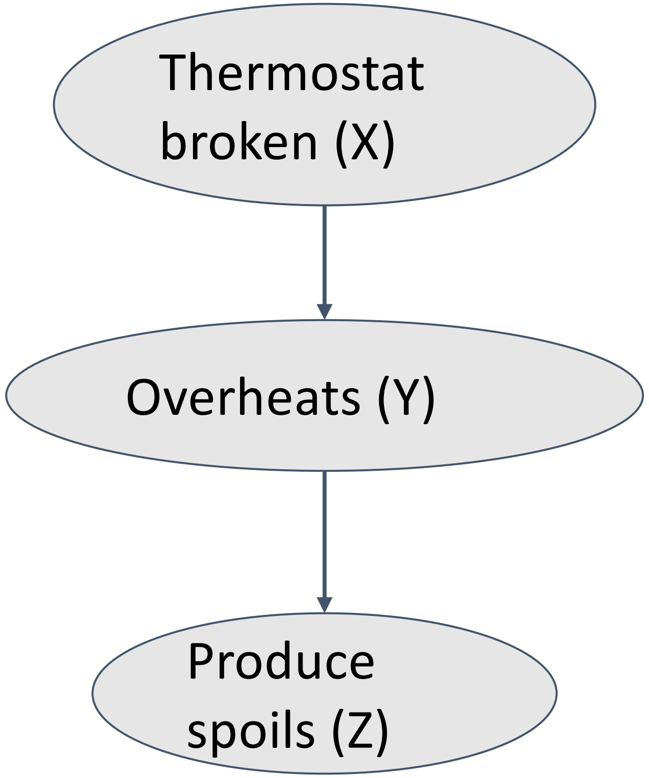
$\prod P(x_i \mid \text{parents}(x_i))$

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

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

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

$$P(X,Y,Z) = P(X)\,P(Y \mid X)\,P(Z \mid Y)$$



$$P(Z \mid X,Y) = \frac{P(Z,X,Y)}{P(X,Y)}$$

$$= \frac{P(\cancel{X})\,P(Y\cancel{\mid X})\,P(Z \mid Y)}{P(\cancel{X})\,P(Y\cancel{\mid X})}$$

$$= P(Z \mid Y)$$

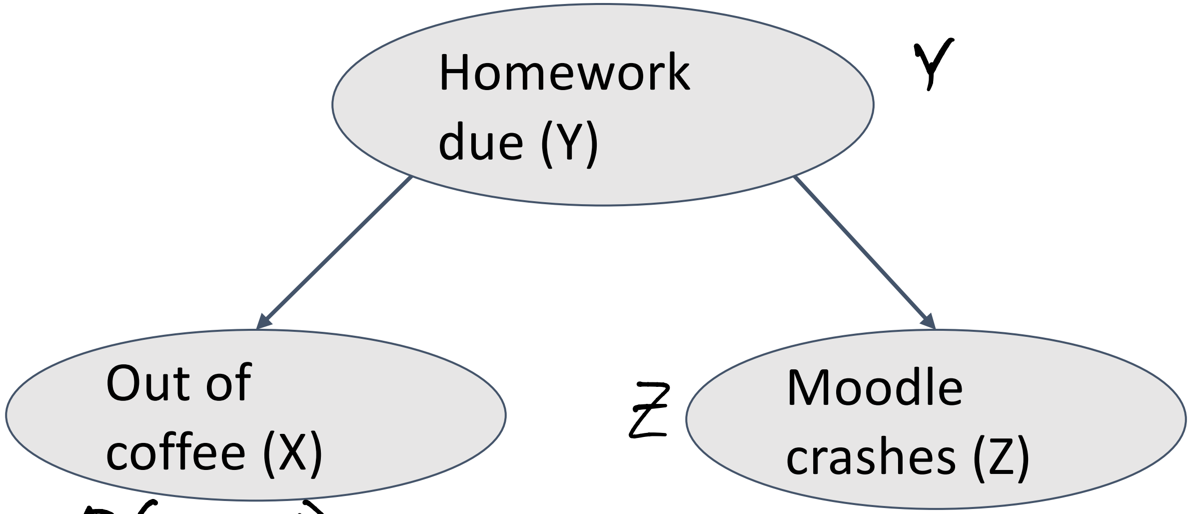
$\Rightarrow$ X and Z are conditionally independent given Y.

Bayesian Networks: Canonical Cases

"common cause"

Common cause is another canonical case.

→ Two effects, from the same cause



Y

X

Z

$P(X \mid Y)$

$P(Z \mid Y)$

Example: Are X and Z independent? $\boxed{\text{no}}$

Are X and Z independent ***given*** Y? $\boxed{\text{yes}}$

$$P(X,Y,Z) = P(Y)\,P(X \mid Y)\,P(Z \mid Y)$$

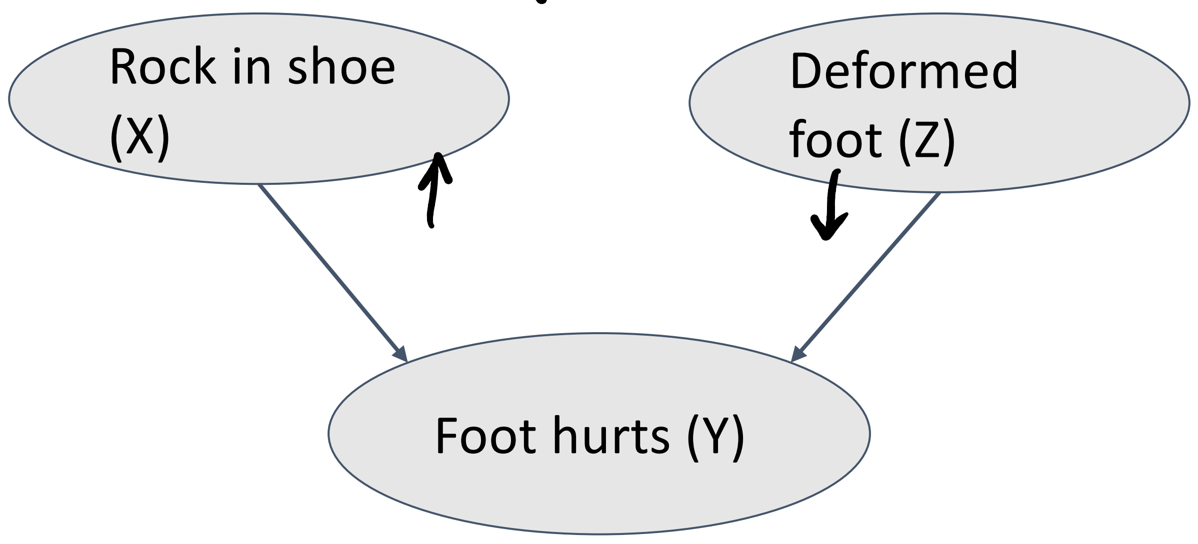
$$P(Z \mid X,Y) = \frac{P(Z,X,Y)}{P(X,Y)} = \frac{\cancel{P(Y)}\,\cancel{P(X \mid Y)}\,P(Z \mid Y)}{\cancel{P(Y)}\,\cancel{P(X \mid Y)}} = P(Z \mid Y)$$

Bayesian Networks: Canonical Cases

common effect.

Common effect is the third canonical case.

→ One effect, two possible causes



yes

Example: Are X and Z independent?

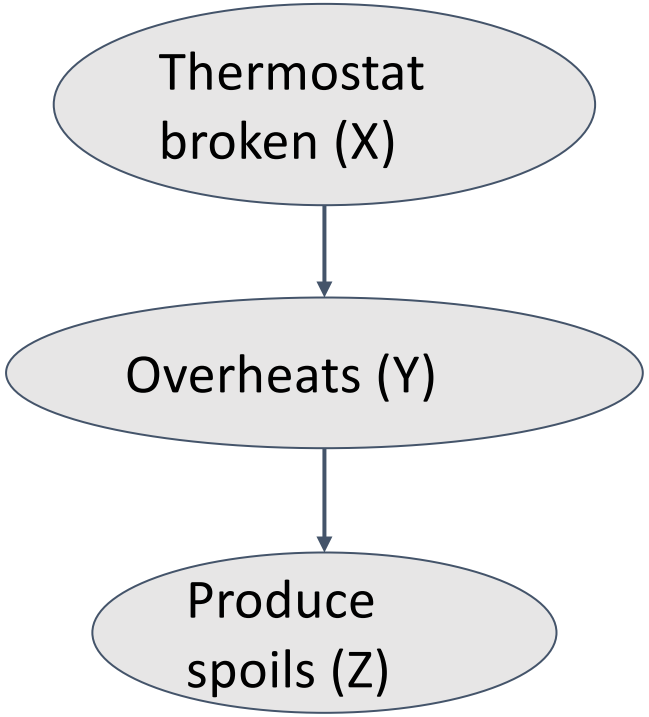
$\boxed{\text{Yes}}$

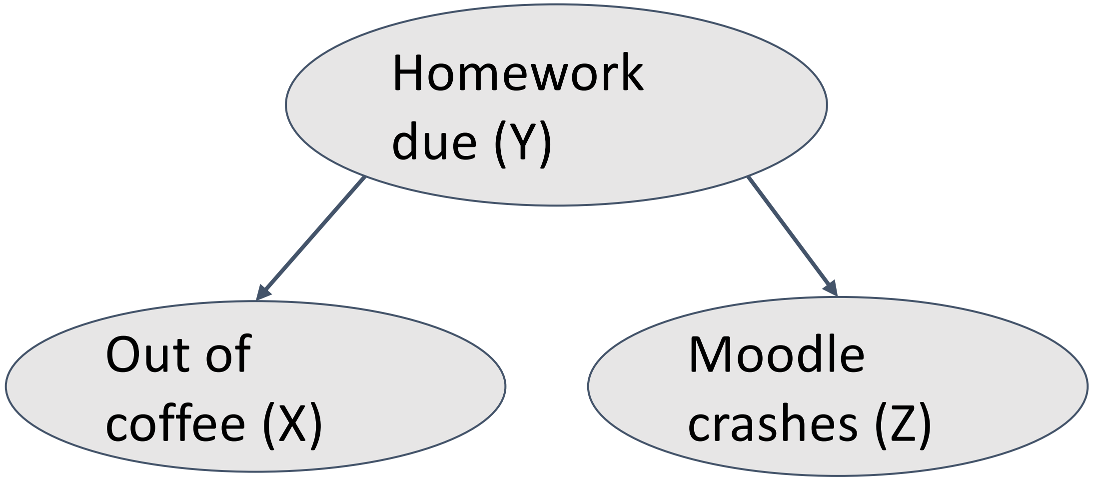
Are X and Z independent ***given*** Y?

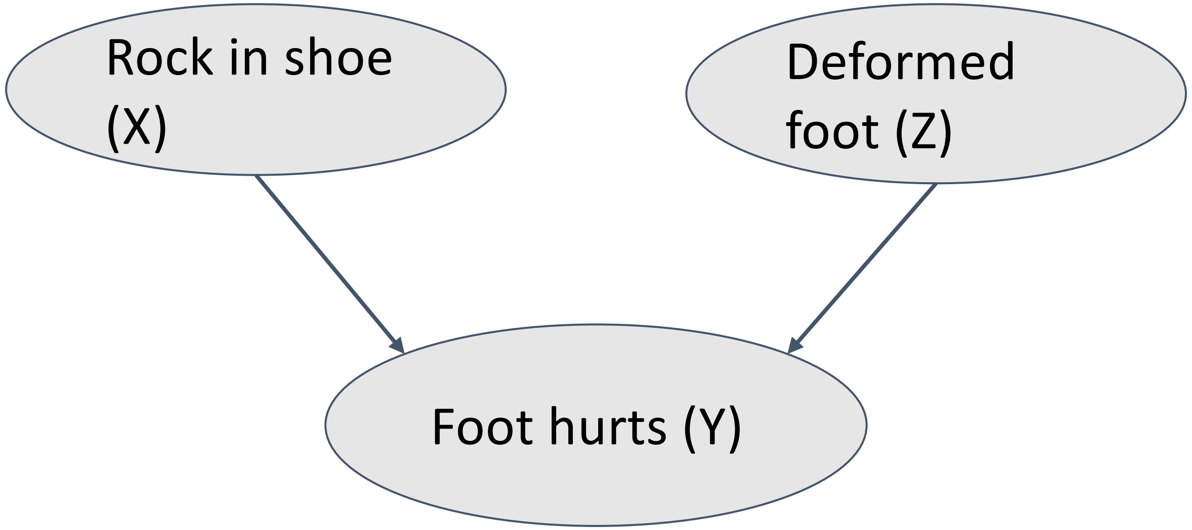
$\boxed{\text{No}}$

Causal vs Diagnostic Modeling Inference

Causal:







Diagnostic: observing an effect leads to competition between possible causes

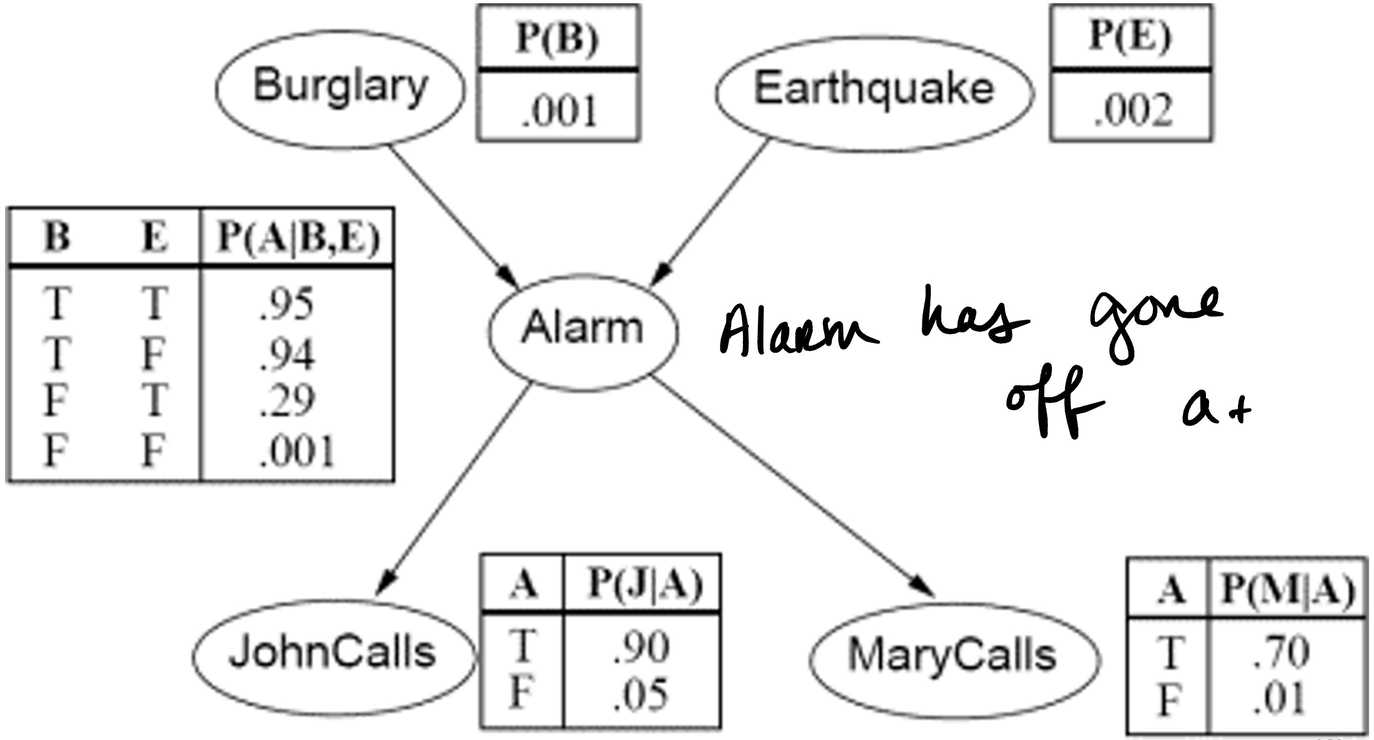
→ *diagnose* which is most likely

Bayesian Networks: “Explaining Away”

Suppose we know that the alarm has gone off.

Suppose we find out later that we have been robbed

$P(E \mid +a, +b) < P(E)$



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

Alarm has gone off a+

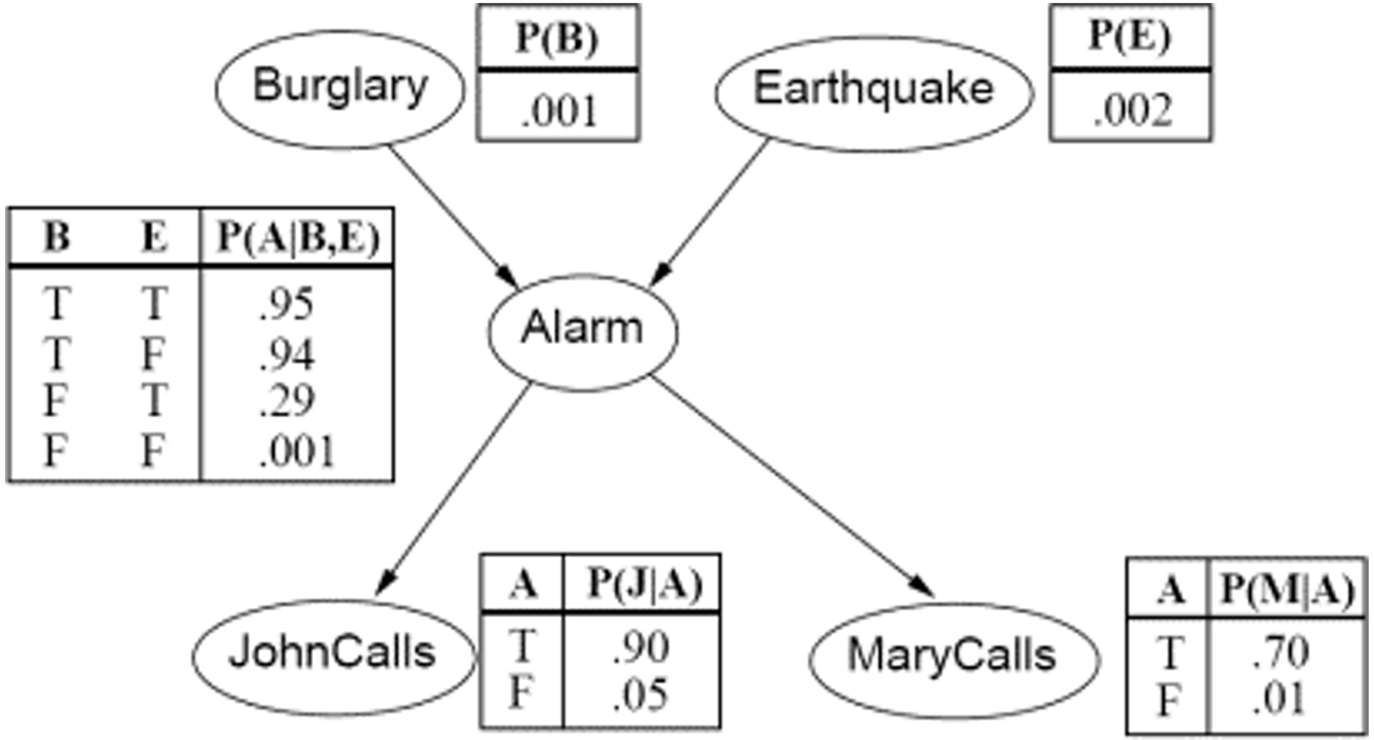
Bayesian Networks

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



P(B)
.001

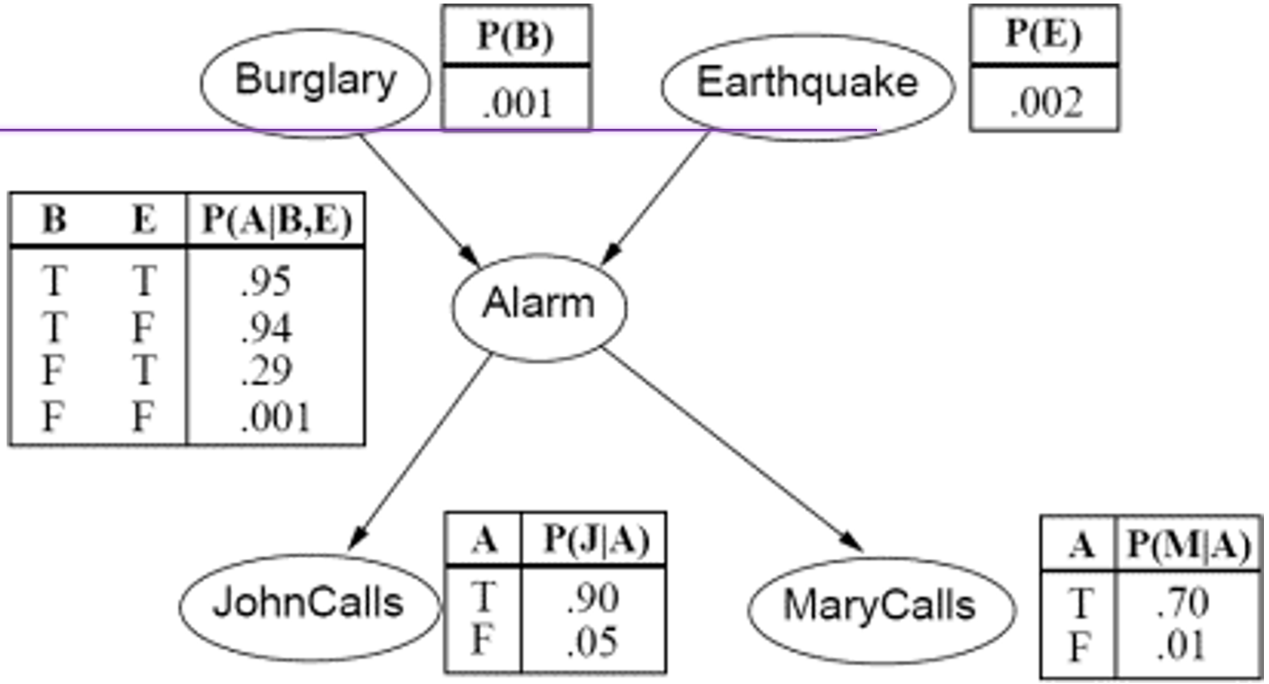
P(E)
.002

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



Example: S'pose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ $P(+b \mid +j, +m) = ?$

Query variables: what we want the **posterior** probability of, **given** some **evidence**

X = B

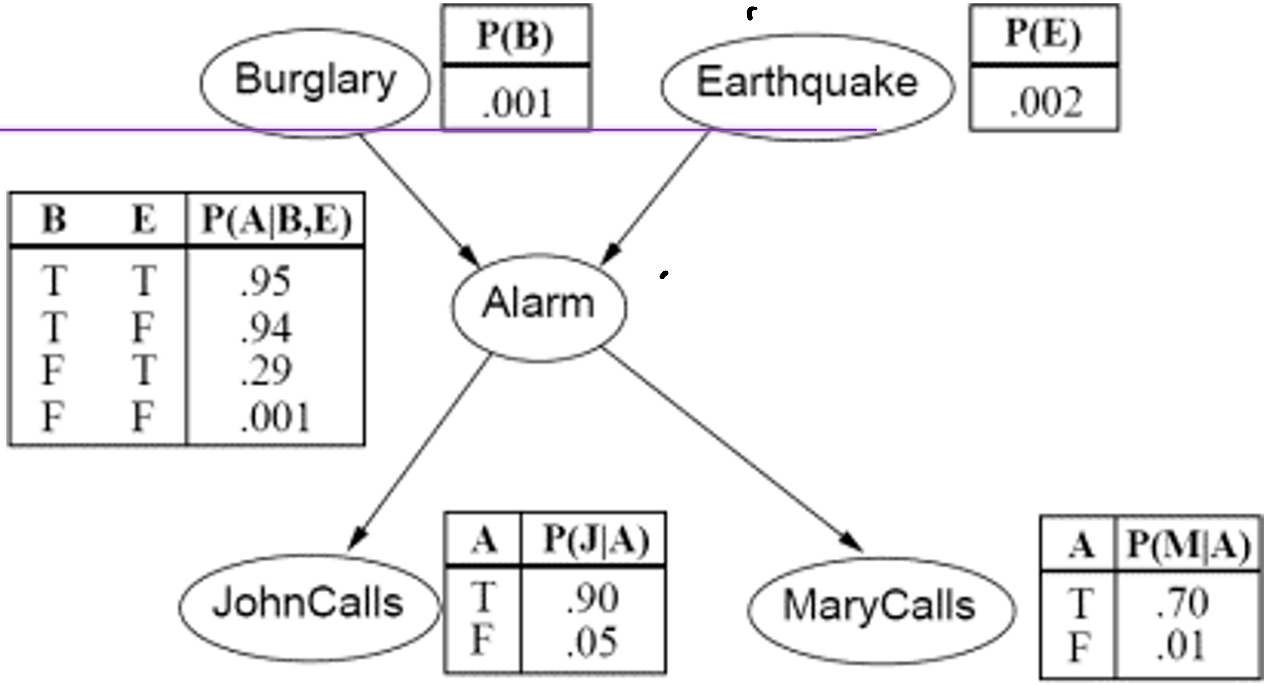
Evidence variables: the variables we are given an assignment of (the **data**)

E = [+j, +m]

Hidden variables: the non-evidence, non-query variables

y = [E, A]

Bayesian Networks



Example: S'pose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

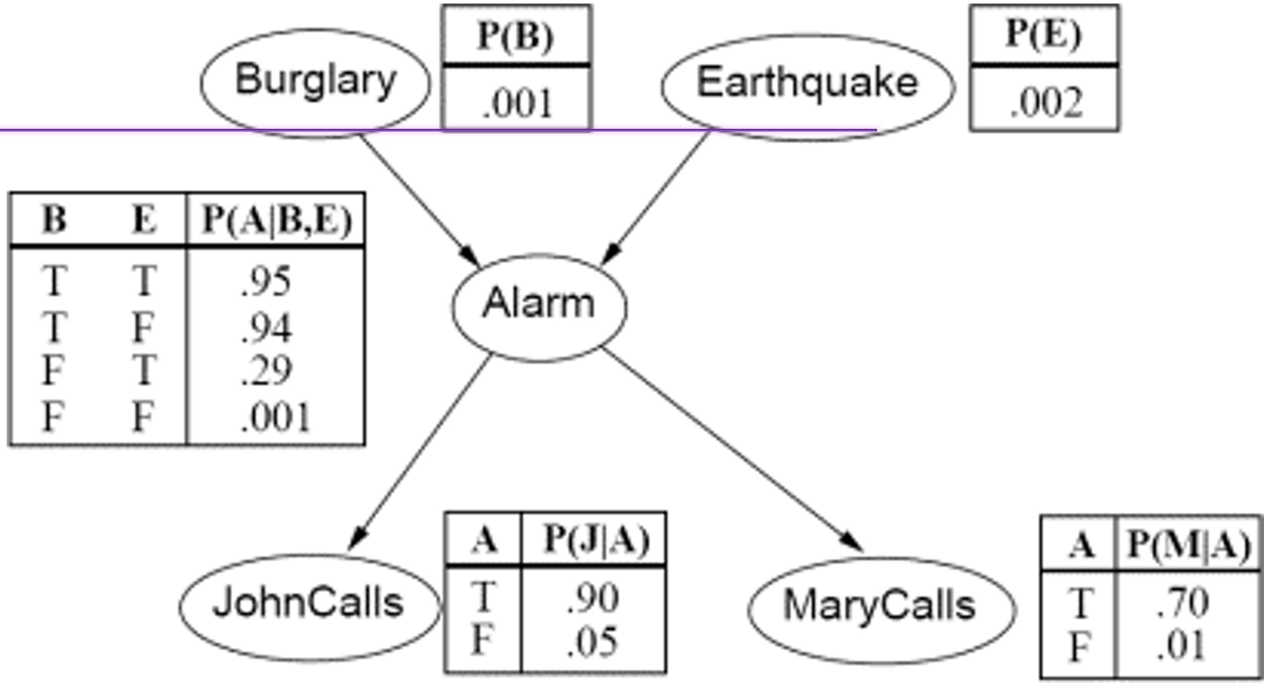
→ *P*(+b | +j, +m) = ?

Calculation by **enumeration:**

$$P(B,+j,+m,A,E) = \prod_{i=1}^{5} P(X_i \mid \text{Parents}(X_i))$$

$$P(B \mid +j,+m) = \frac{P(+j,+m,B)}{P(+j,+m)} = \underbrace{\boxed{\frac{1}{P(+j,+m)}}}_{\alpha} \cdot P(+j,+m,B)$$

Bayesian Networks



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

Example: S'pose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

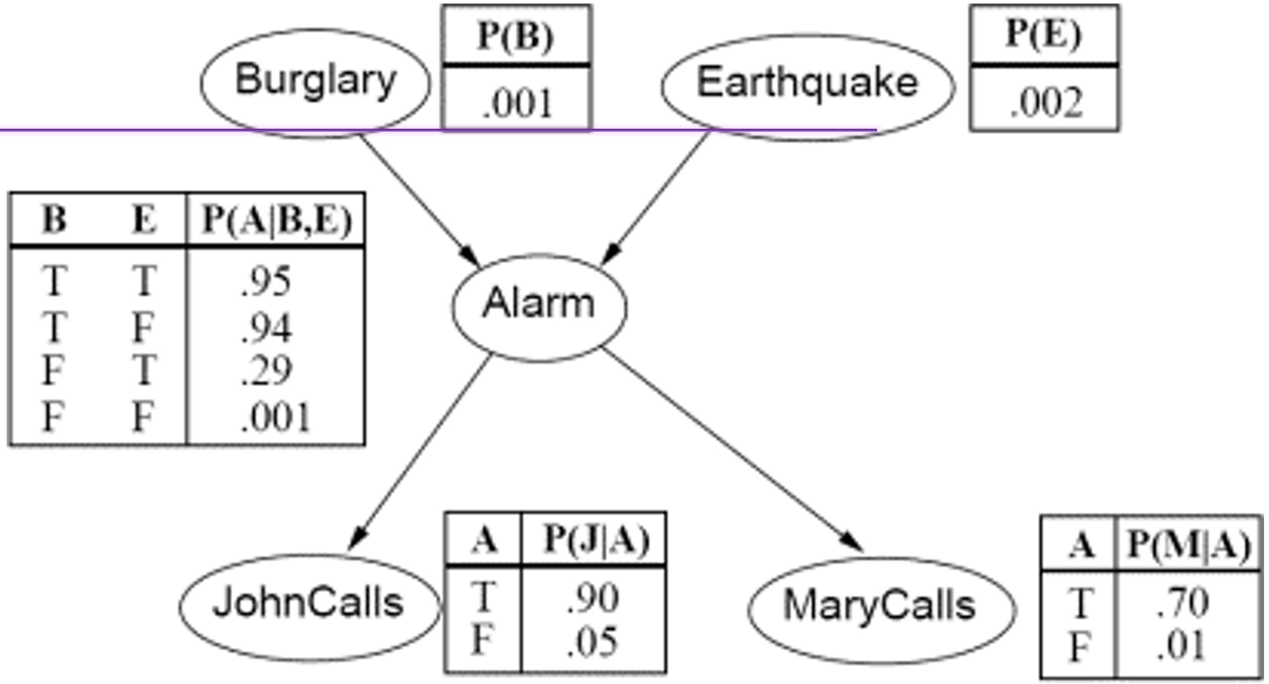
→ $P(+b \mid +j, +m) = ?$

Calculation by **enumeration:**

$$P(B \mid j, m) = \frac{P(B, j, m)}{P(j, m)} \quad \Rightarrow \quad \alpha = \frac{1}{P(j, m)} \quad \Rightarrow \quad P(B \mid j, m) = \alpha P(B, j, m)$$

We'll do our thing, then figure out the normalizing constant α later

Bayesian Networks



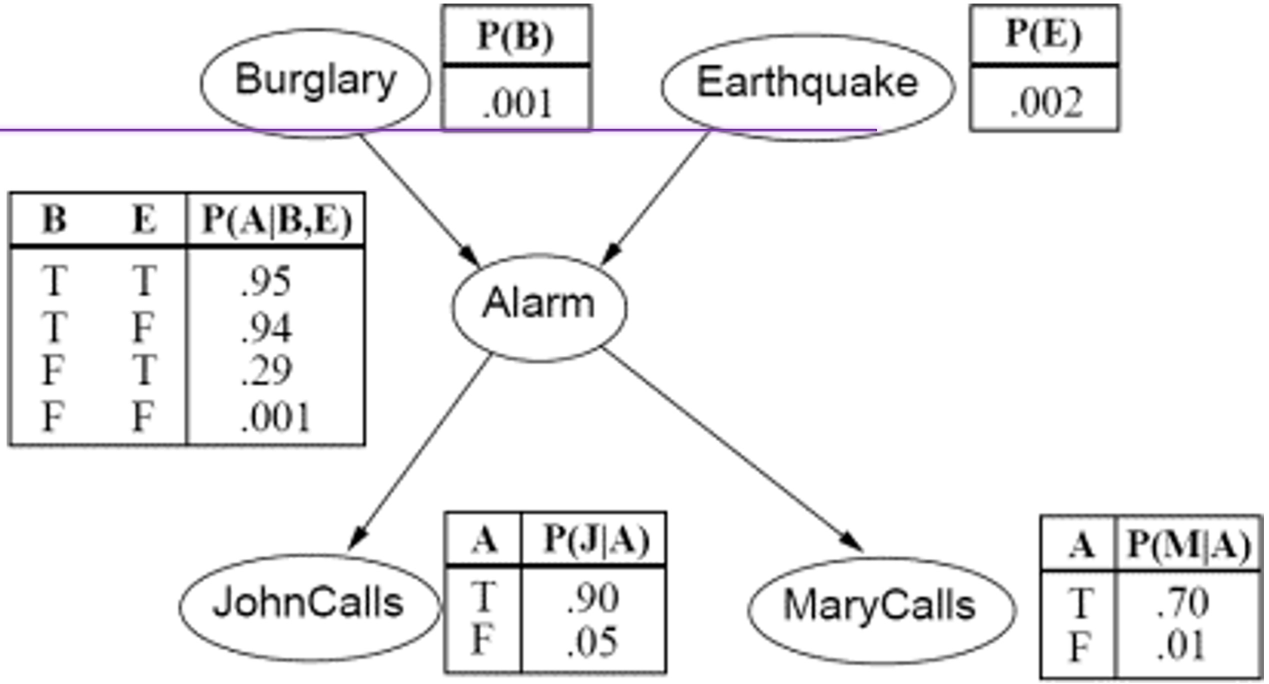
Example: Suppose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ *P*(+b | +j, +m) = ?

Calculation by **enumeration:**

$$P(B \mid j, m) = \alpha P(B, j, m)$$
$$= \alpha \sum_a P(B, j, m \mid a) P(a)$$
$$= \alpha \sum_a P(B, j, m, a)$$

Bayesian Networks



Example: Suppose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ *P*(+b | +j, +m) = ?

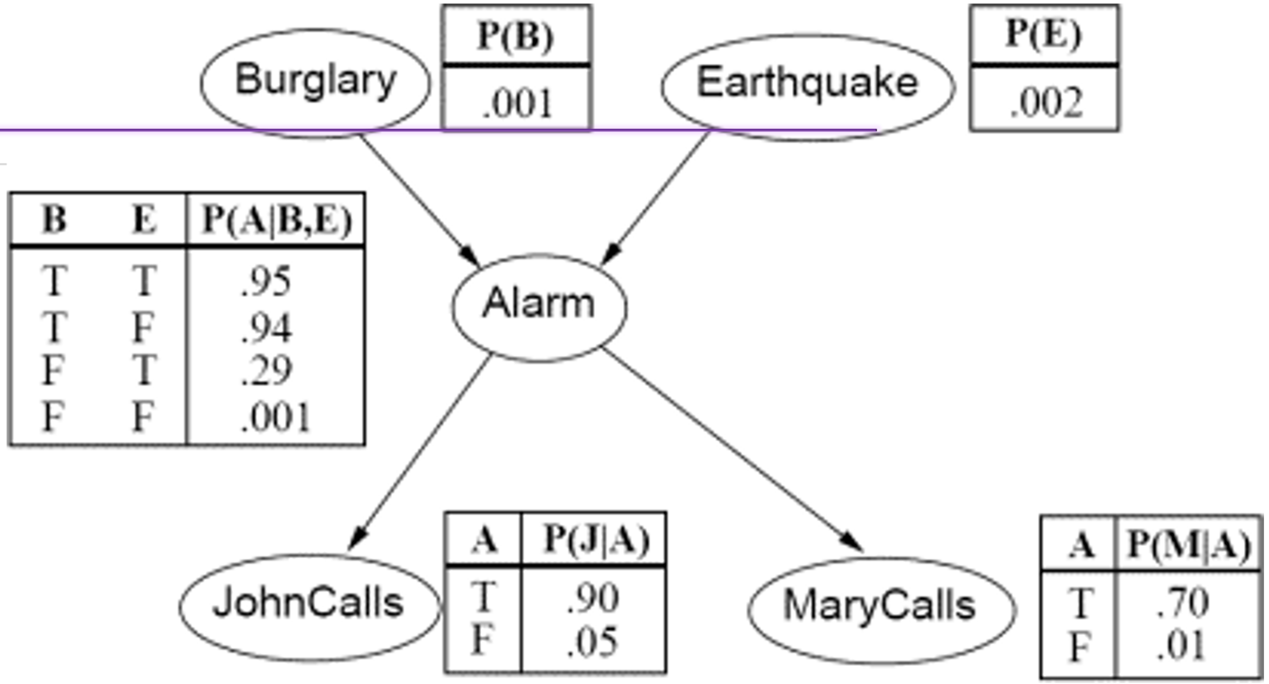
Calculation by **enumeration:**

$$P(B \mid j, m) = \alpha \sum_{a} P(B, j, m, a)$$

$$= \alpha \sum_{e} \sum_{a} P(B, j, m, a \mid e) P(e)$$

$$= \alpha \sum_{e} \sum_{a} P(B, j, m, a, e)$$

Bayesian Networks



Example: Suppose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ $P(+b \mid +j, +m) = ?$

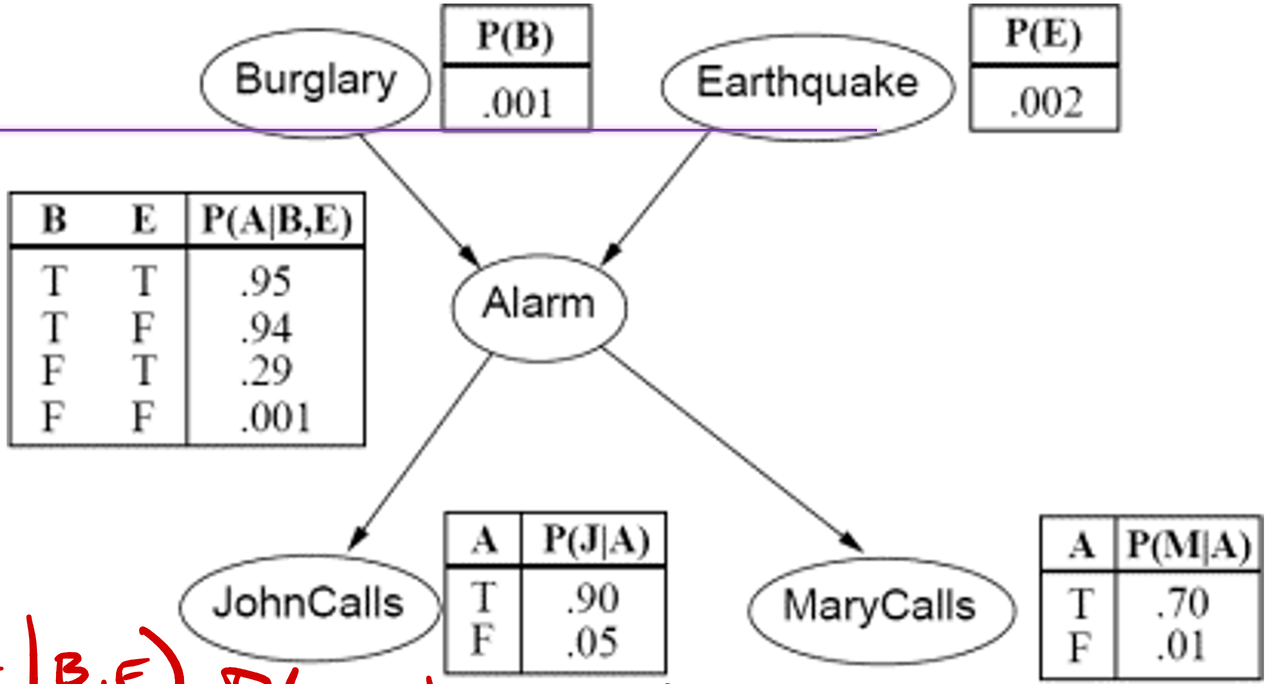
FINALLY we have:

$$P(B \mid j, m) = \alpha \sum_e \sum_a P(B, j, m, a, e)$$

From the conditional independence of the Bayes net:

$$P(B \mid j, m) = \alpha \sum_e \sum_a \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

Bayesian Networks



Example: ~~Suppose we know the alarm has gone off,~~ and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ $P(+b \mid +j, +m) = ?$

P(B)P(E)P(A | B,E) P(J|A) P(M|A)

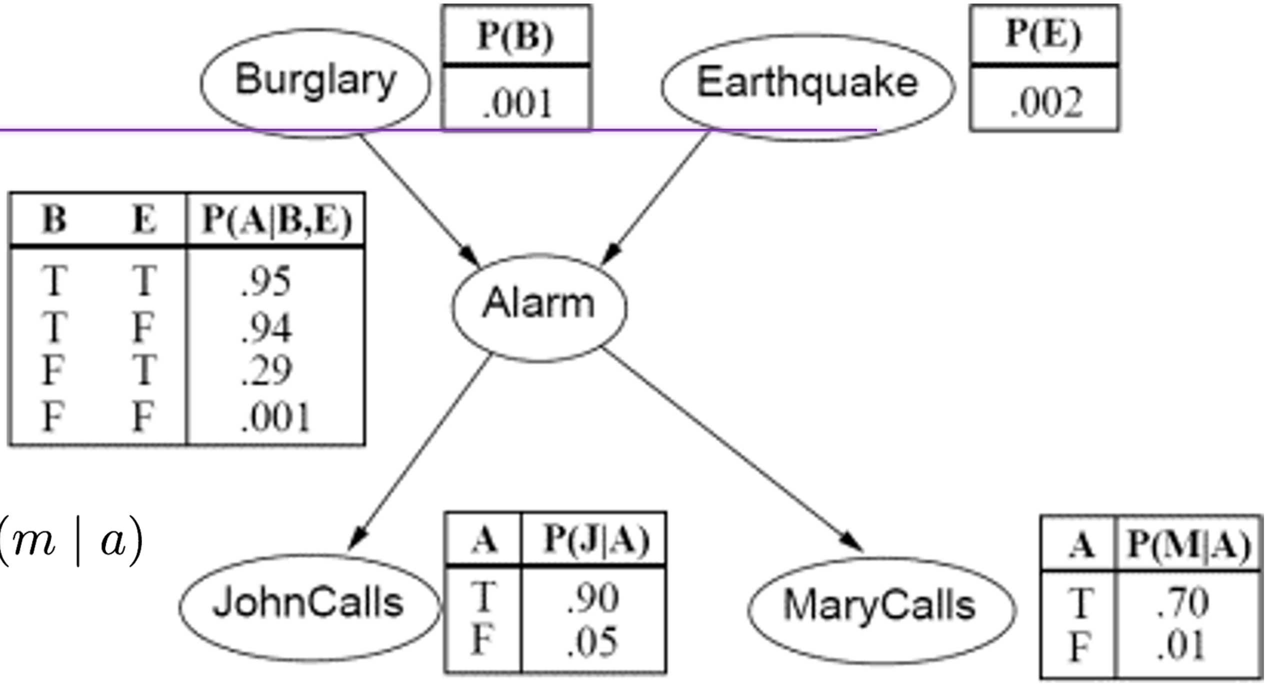
So for this problem…

$$P(B \mid +j, +m) = \alpha \sum_e \sum_a \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

$$= \alpha \sum_e \sum_a P(B)P(e)P(a \mid B, e)P(j \mid a)P(m \mid a)$$

$$= \alpha P(B) \sum_e P(e) \sum_a P(a \mid B, e)P(j \mid a)P(m \mid a)$$

Bayesian Networks



Example: Are we ever going to actually calculate the probability that we have been burgled???

$$P(B \mid j,m) = \alpha P(B) \sum_{e} P(e) \sum_{a} P(a \mid B,e) P(j \mid a) P(m \mid a)$$

Suppose we do this double sum and we find

$$P(+b \mid e) = 0.02 \cdot \alpha = \frac{.02}{.06} = \frac{1}{3}$$

$$P(-b \mid e) = 0.04 \cdot \alpha = \frac{.04}{.06} = \frac{2}{3}$$

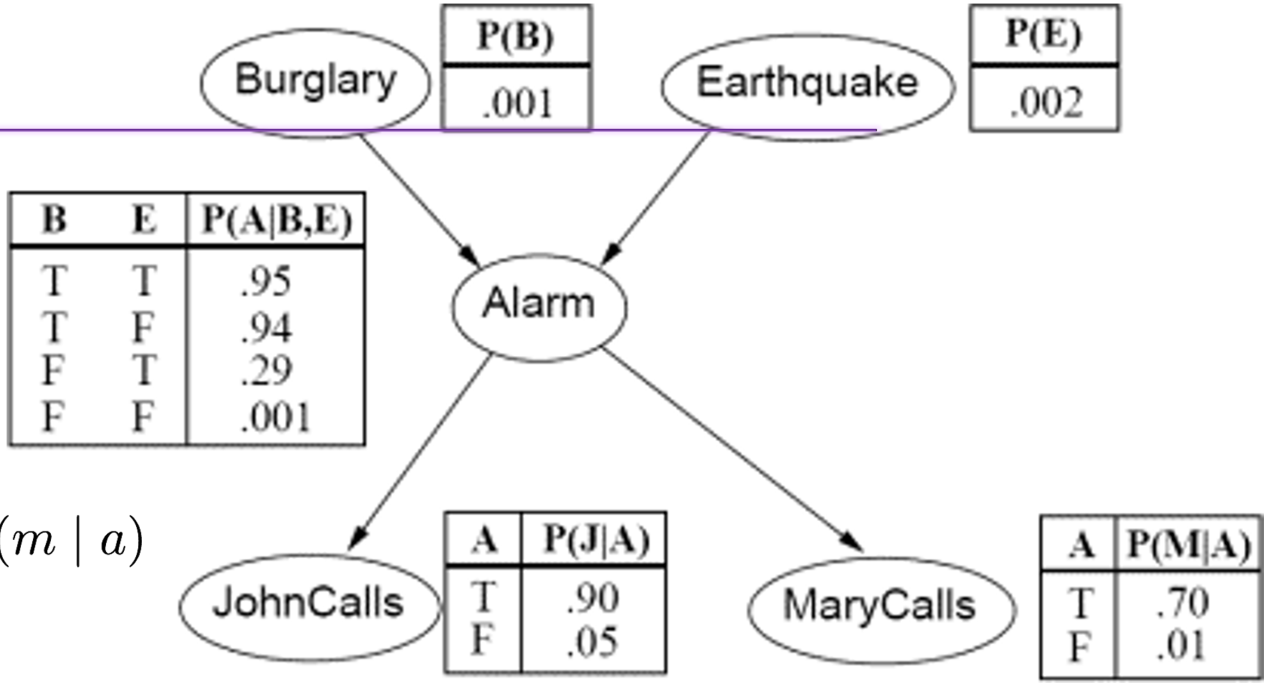
How do we normalize

$$1 = 0.02\alpha + 0.04\alpha$$

$$1 = 0.06\alpha$$

$$1/0.06 = \alpha$$

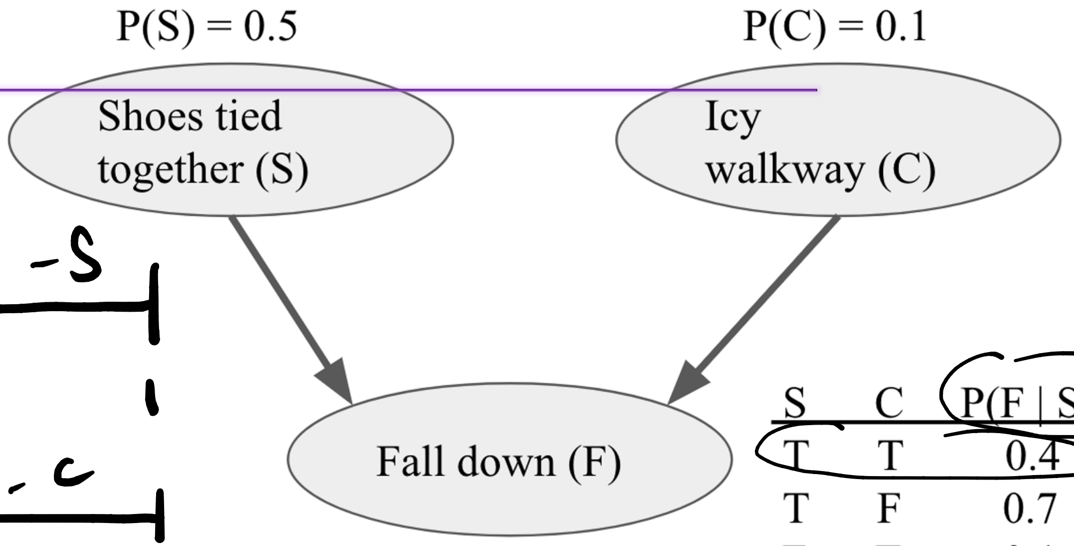
Bayesian Networks



Example: Graphical representation

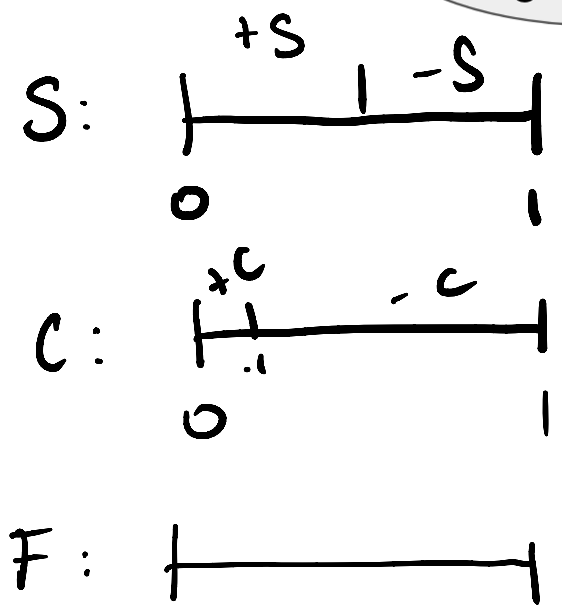
$$P(B \mid j, m) = \alpha P(B) \sum_e P(e) \sum_a P(a \mid B, e) P(j \mid a) P(m \mid a)$$

Prior Sampling



S	C	P(F | S, C)
T	T	0.4
T	F	0.7
F	T	0.1
F	F	0.01

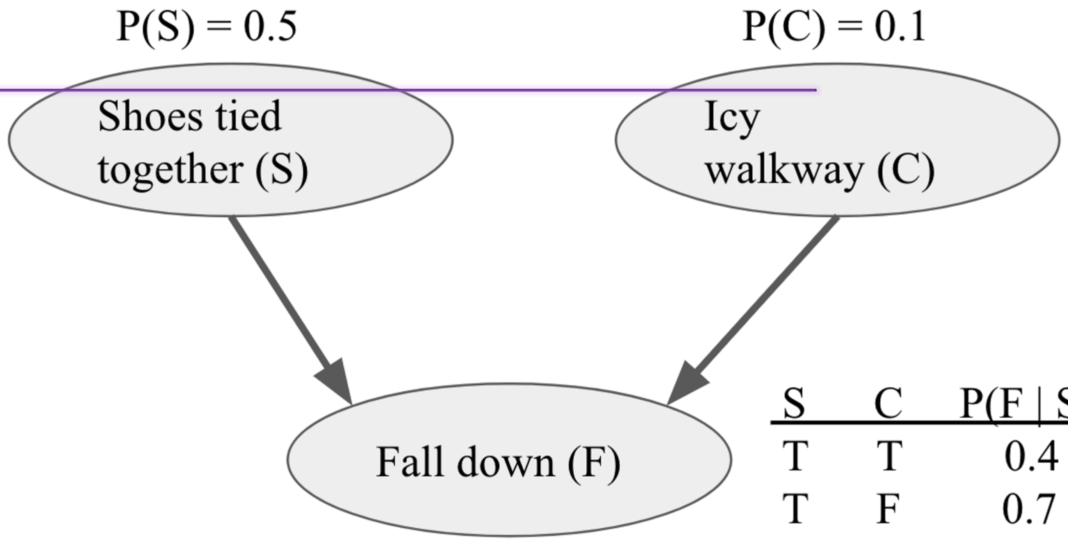
Calculation: P(+s, -f, -c) = ?



S	C	F
T	T	T
T	T	
F	F	
T	F	
F	F	
F	F	
T	T	

$$\frac{\text{counting \# True F}}{\text{num samples}}$$

Rejection Sampling



S	C	P(F | S, C)
T	T	0.4
T	F	0.7
F	T	0.1
F	F	0.01

Calculation: P(+s | -f, -i) = ?

Next Time

- ***Markov Models***