CSC442: Introduction to Artificial Intelligence

Bayesian Networks

Today's Material:

AIMA 14–14.4

Representing Uncertainty

- Probability:
  - Sample space, probabilities, events
  - Random variables, domains
  - Language of probability statements
  - Probability distributions, joint distributions, full joint distribution

Probabilistic Inference

- Make implicit knowledge explicit.
- Compute what “follows” from uncertain knowledge.

Probabilistic Inference

- Given priors and evidence, compute probabilities given evidence (posteriors)
- Key — marginalization via the sum rule.

Probabilistic Inference

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\,\mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Query variable $X : Domain(X) = \{x_1, \ldots, x_m\}$

Evidence variables $\mathbf{E} : \{E_1, \ldots, E_k\}$

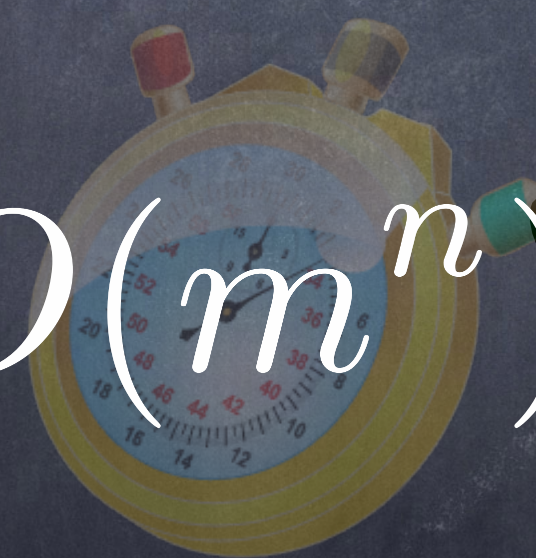
Observations $\mathbf{e} : \{e_1, \ldots, e_k\}$ s.t. $E_i = e_i$

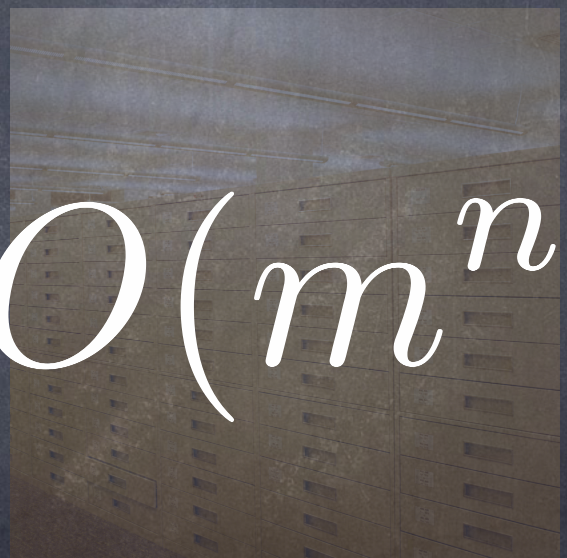
Unobserved variables $\mathbf{Y} : \{Y_1, \ldots, Y_l\}$

$Domain(Y_i) = \{y_{i,1}, \ldots, y_{i,n_i}\}$

Probabilistic Inference (Single Variable)

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$





$$O(m^n)$$

Time Complexity

$$O(m^n)$$

Space Complexity

Probabilistic Inference

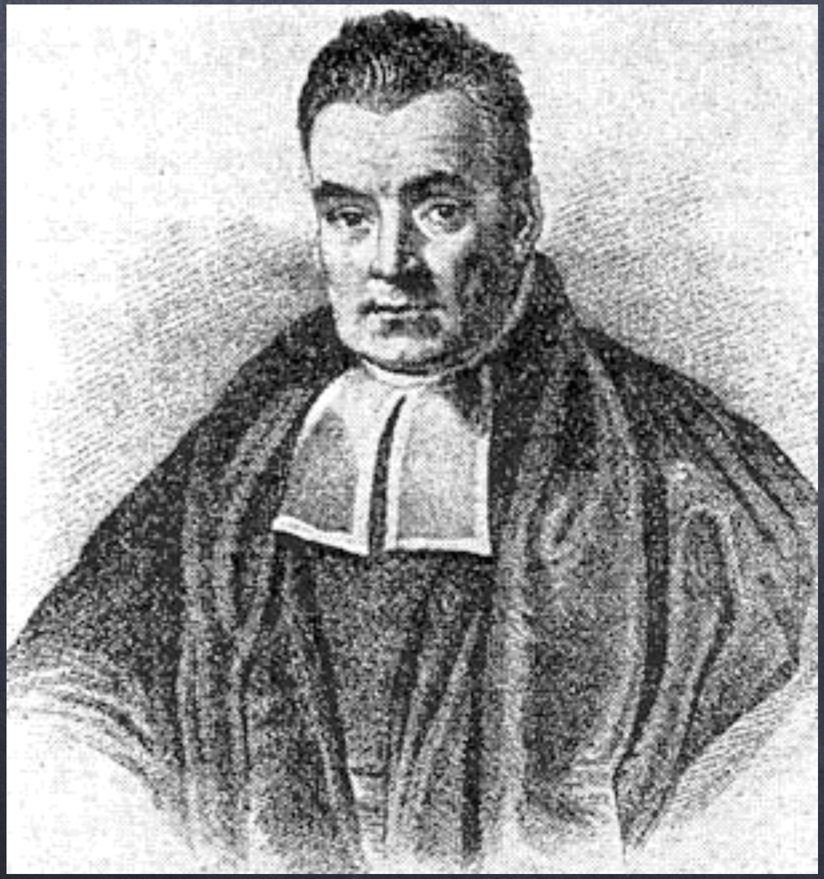
- Why is the cost so high?

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

- The sum is a **monster**, so is the **table**.
- Enter Bayesian Networks....

Bayes' Rule

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$



Thomas Bayes

(c. 1702 – 1761)

Bayesian Diagnosis

$$P(disease \mid symptom) = \frac{P(symptom \mid disease)P(disease)}{P(symptom)}$$

toothache

(Toothache = true)

catch

(Catch = true)

$P(Cavity \mid toothache \wedge catch)$

Combining Evidence

$P(Cavity \mid toothache \wedge catch)$

Combining Evidence

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$
$$= \alpha < 0.108, 0.016 > \quad \approx \quad < 0.871, 0.129 >$$

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Combining Evidence

- In general, if there are n evidence variables, then there are $O(2^n)$ possible combinations of observed values for which we would need to know the conditional probabilities

Independence

$$\mathbf{P}(X \mid Y) = \mathbf{P}(X)$$

$$\mathbf{P}(Y \mid X) = \mathbf{P}(Y)$$

$$\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$$

Can compute n×m probabilities

from n+m probabilities

(if random variables are independent)

Combining Evidence

- For n symptoms (e.g., Toothache, Catch) that are all conditionally independent given a disease (e.g., Cavity), we need $O(n)$ probabilities rather than $O(2^n)$
  - Representation scales to larger problems
  - Conditional probabilities more likely to be available than absolute independence assumptions

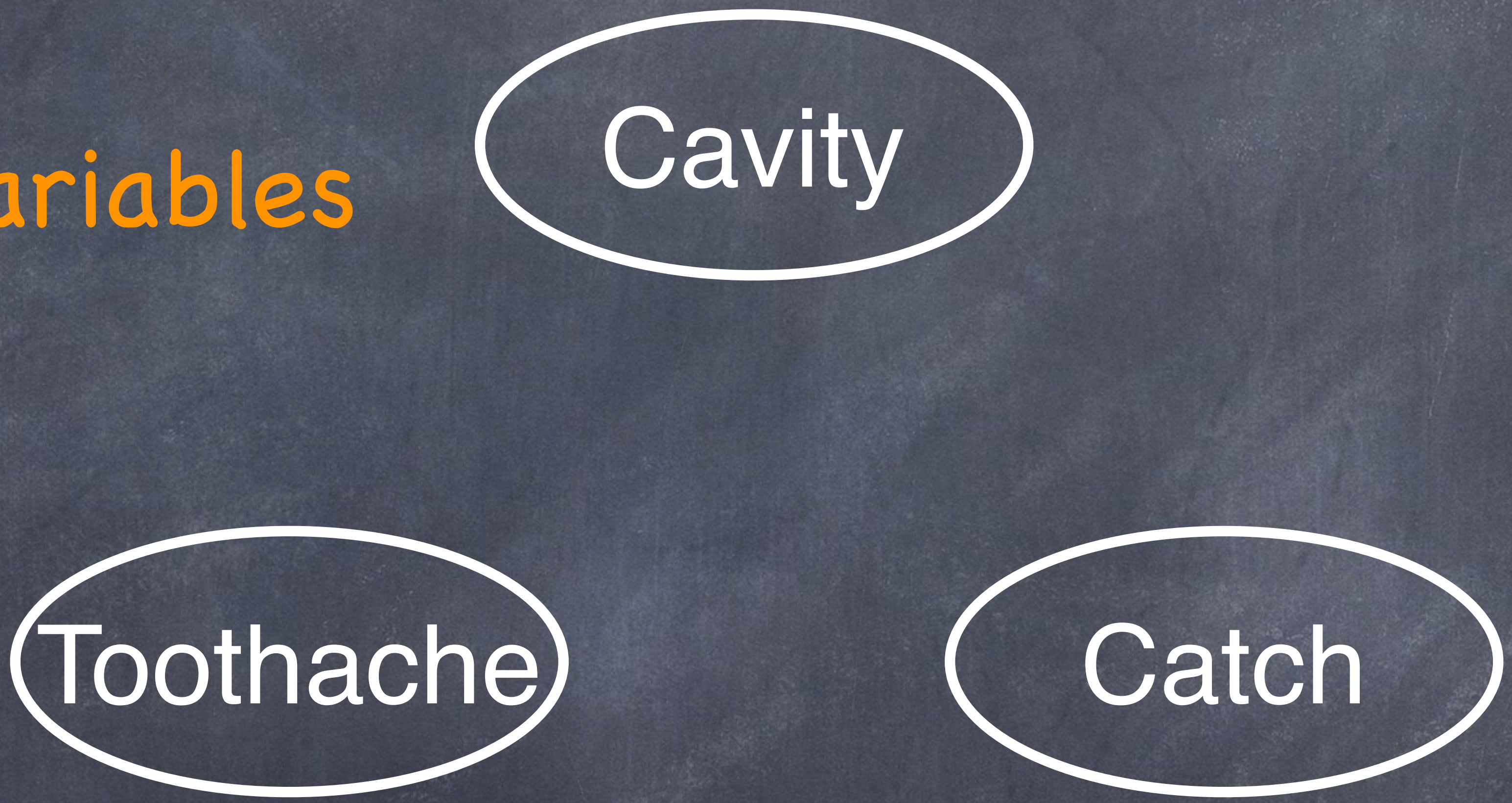
Probabilistic Inference

- Full joint distribution:
  - intractable as problem grows
- Independence assumptions reduce number of probabilities required to represent full joint distribution
- Next: Develop a data structure that represents the (in)dependencies among random variables and can be used to compute probabilities full joint distribution

Bayesian Networks

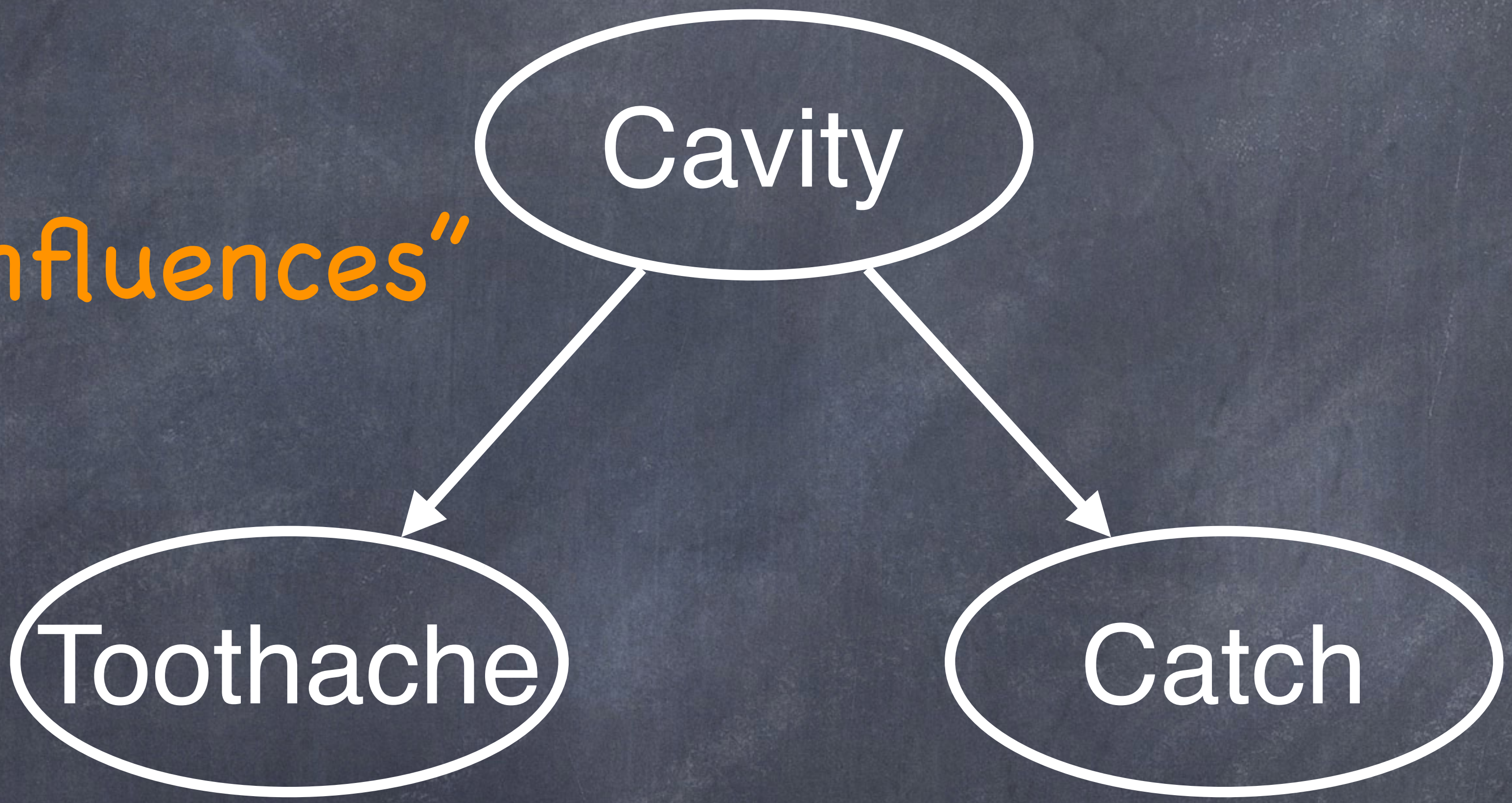
Bayesian Networks

Random Variables

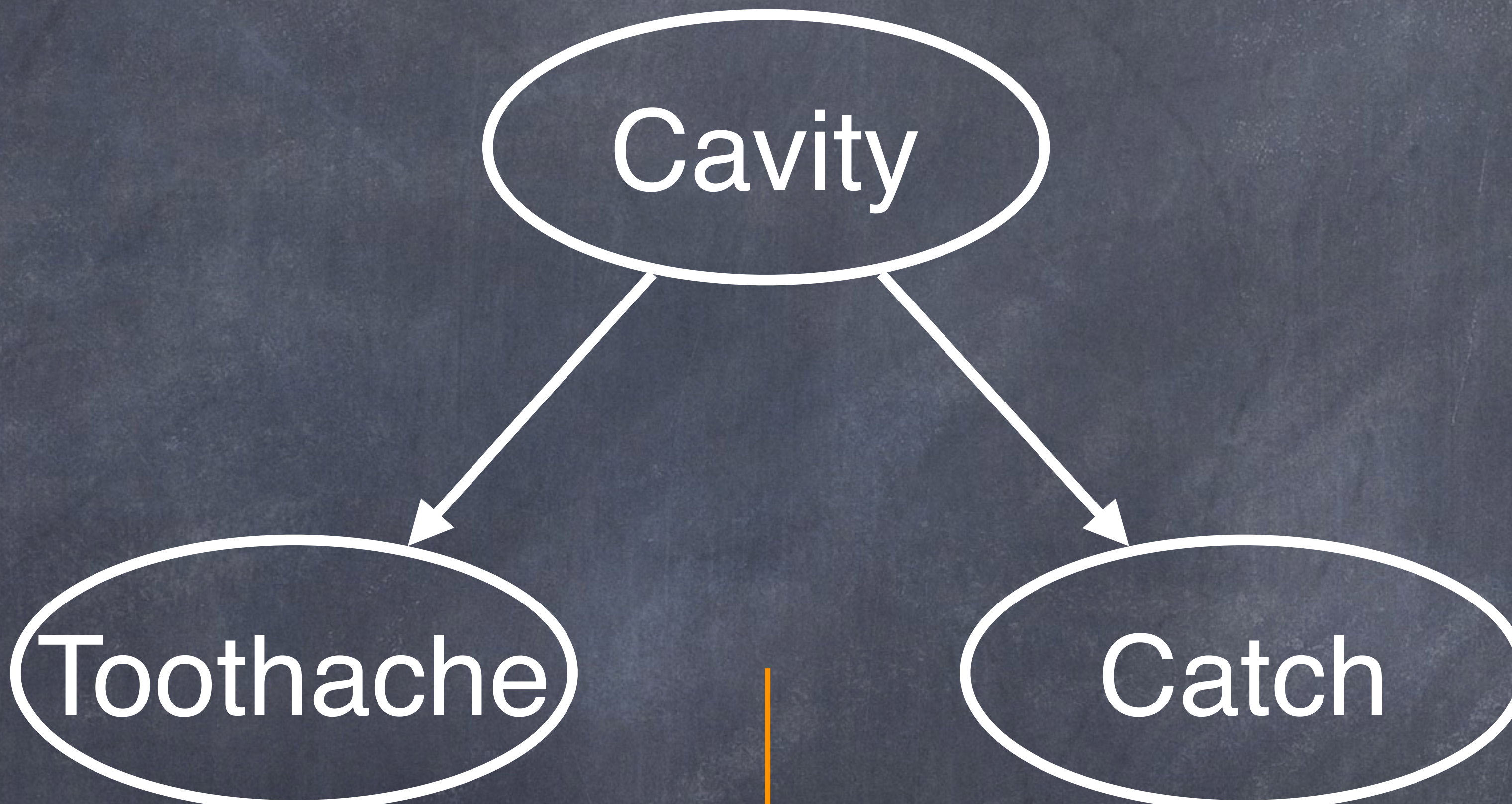


Bayesian Networks

“directly influences”

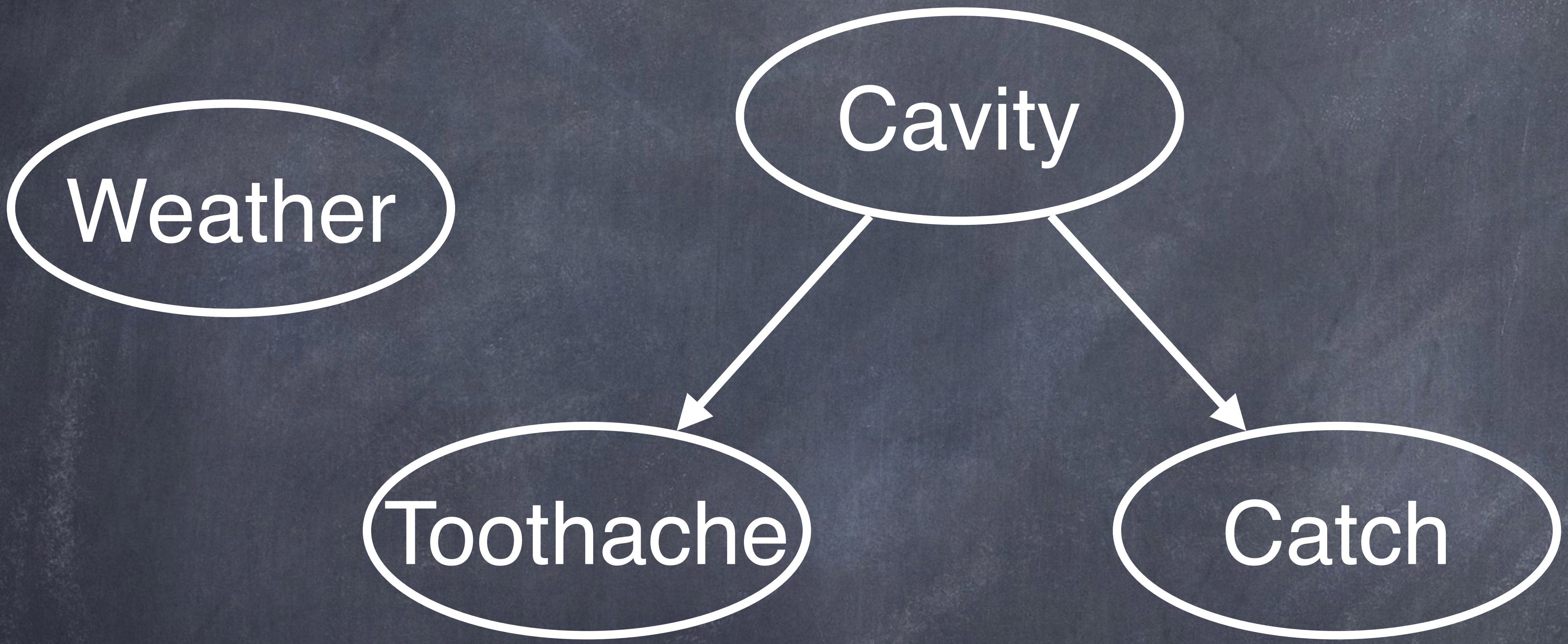


Bayesian Networks

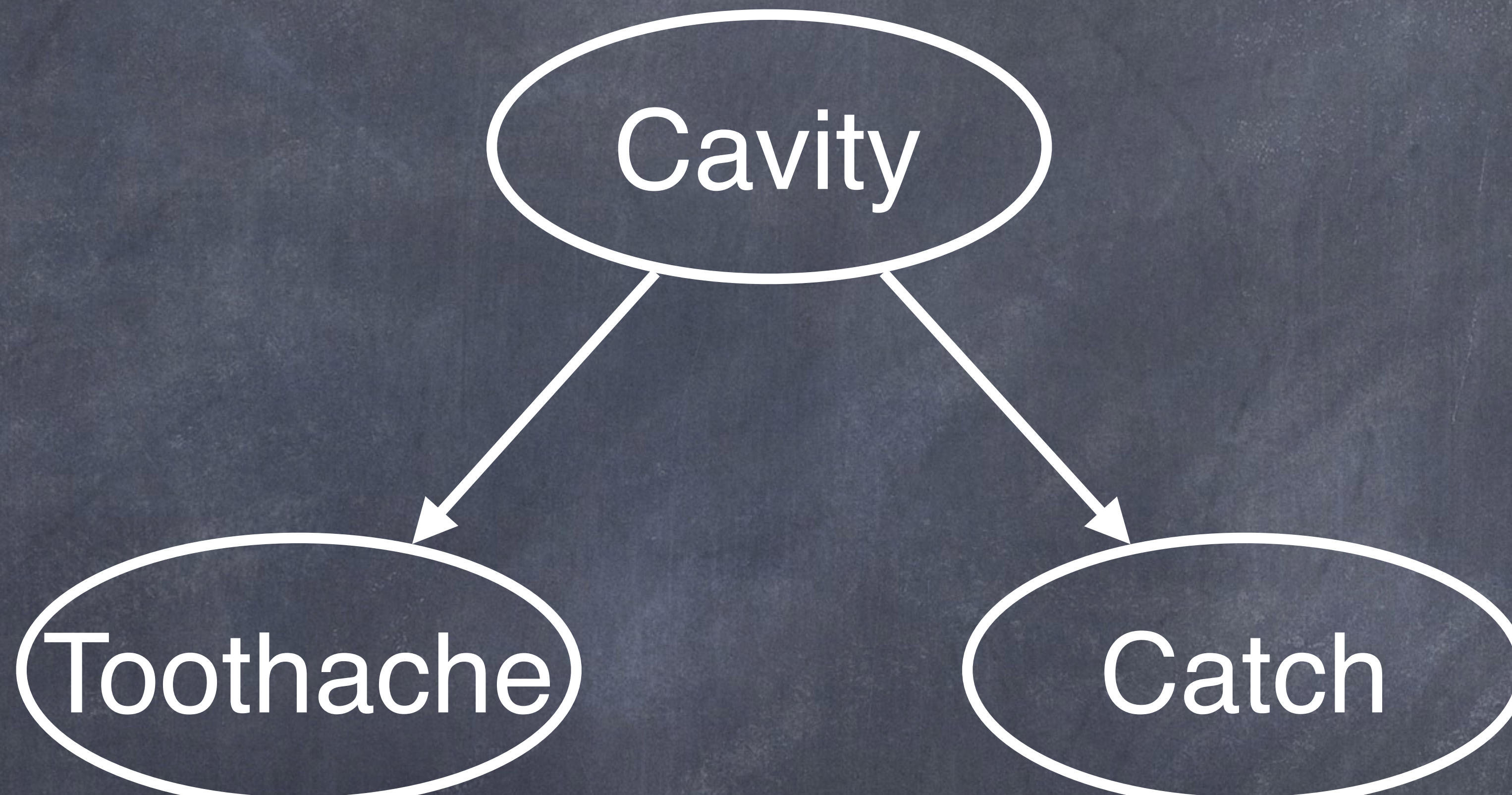


no “direct influence”

Bayesian Networks



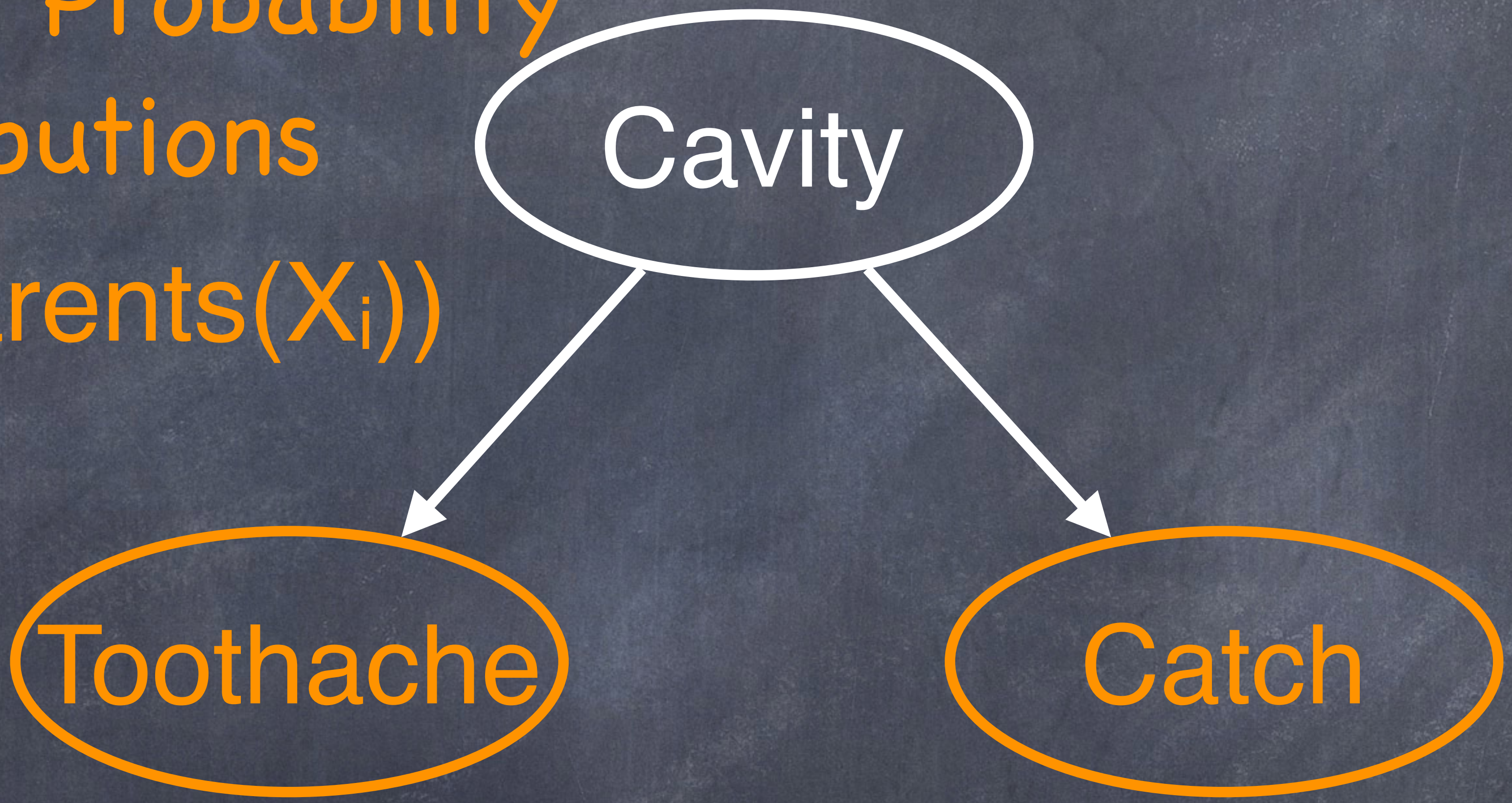
Bayesian Networks



Bayesian Networks

Conditional Probability Distributions

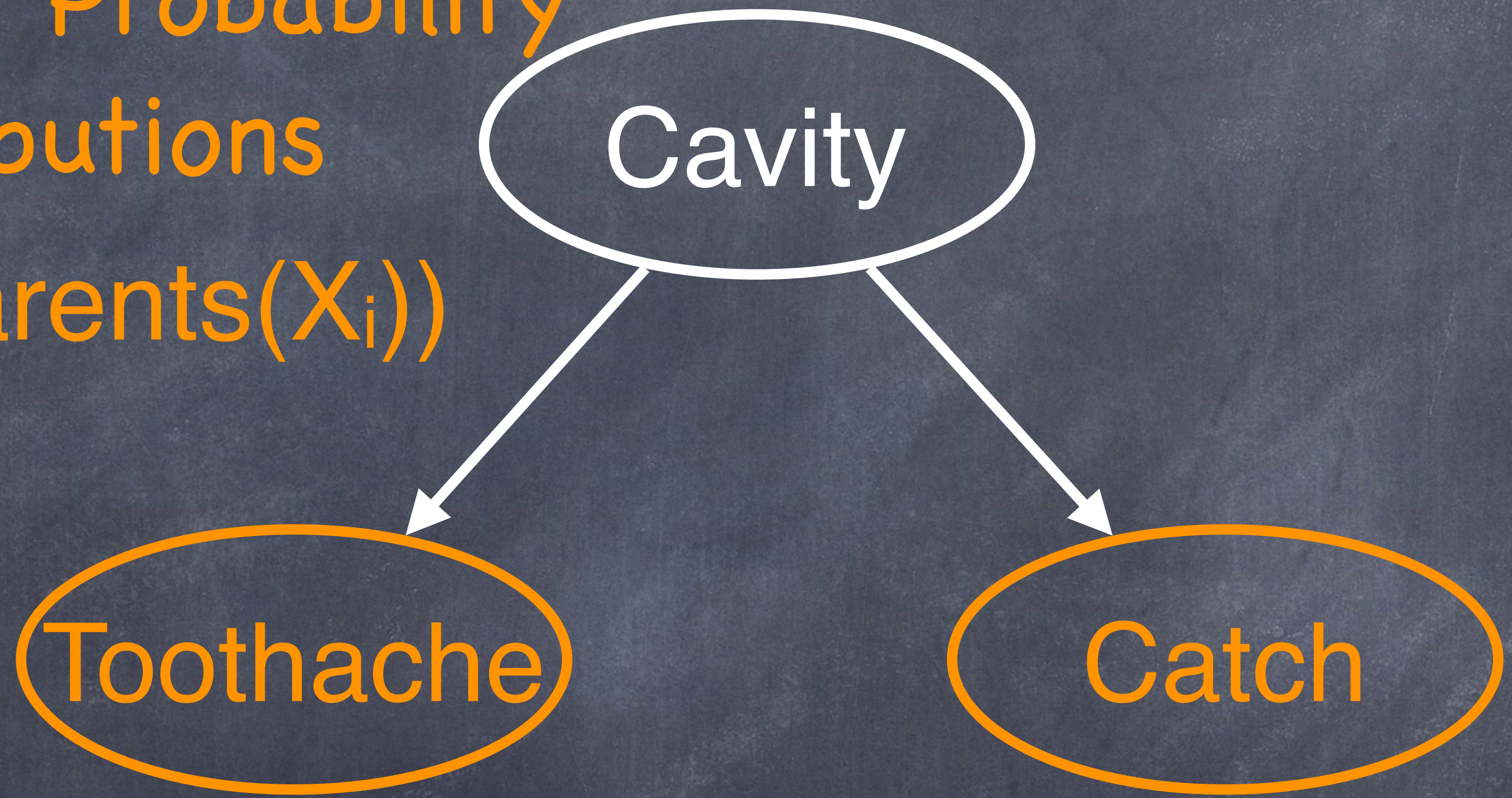
$P(X_i \mid \text{Parents}(X_i))$



Bayesian Networks

Conditional Probability Distributions

$P(X_i \mid Parents(X_i))$



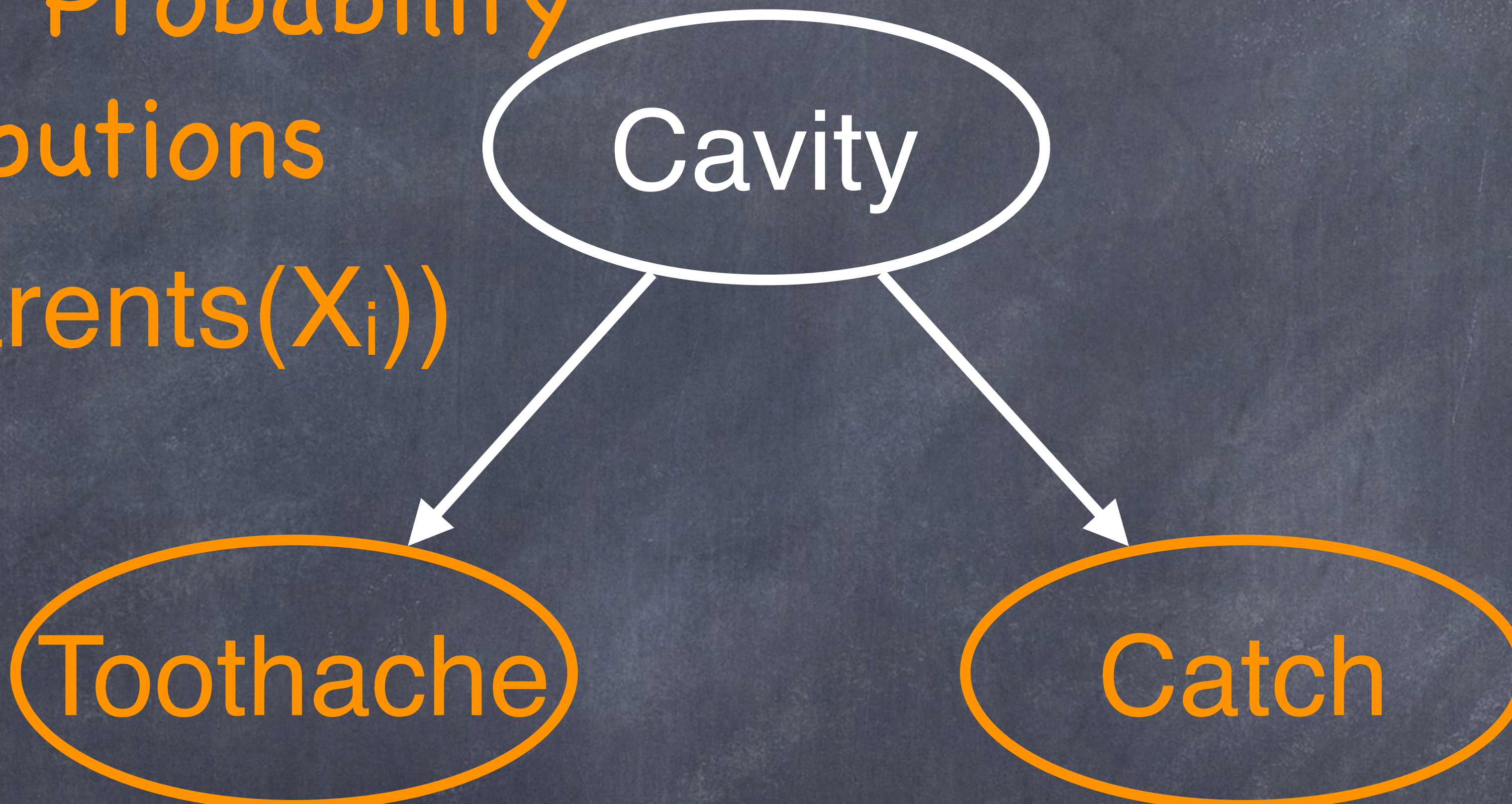
$P(Toothache \mid Cavity)$

Cavity	toothache	¬toothache
cavity		
¬cavity		

Bayesian Networks

Conditional Probability Distributions

$P(X_i \mid Parents(X_i))$



$P(Toothache \mid Cavity)$

Cavity	toothache	¬toothache
cavity		
¬cavity		

$P(Catch \mid Cavity)$

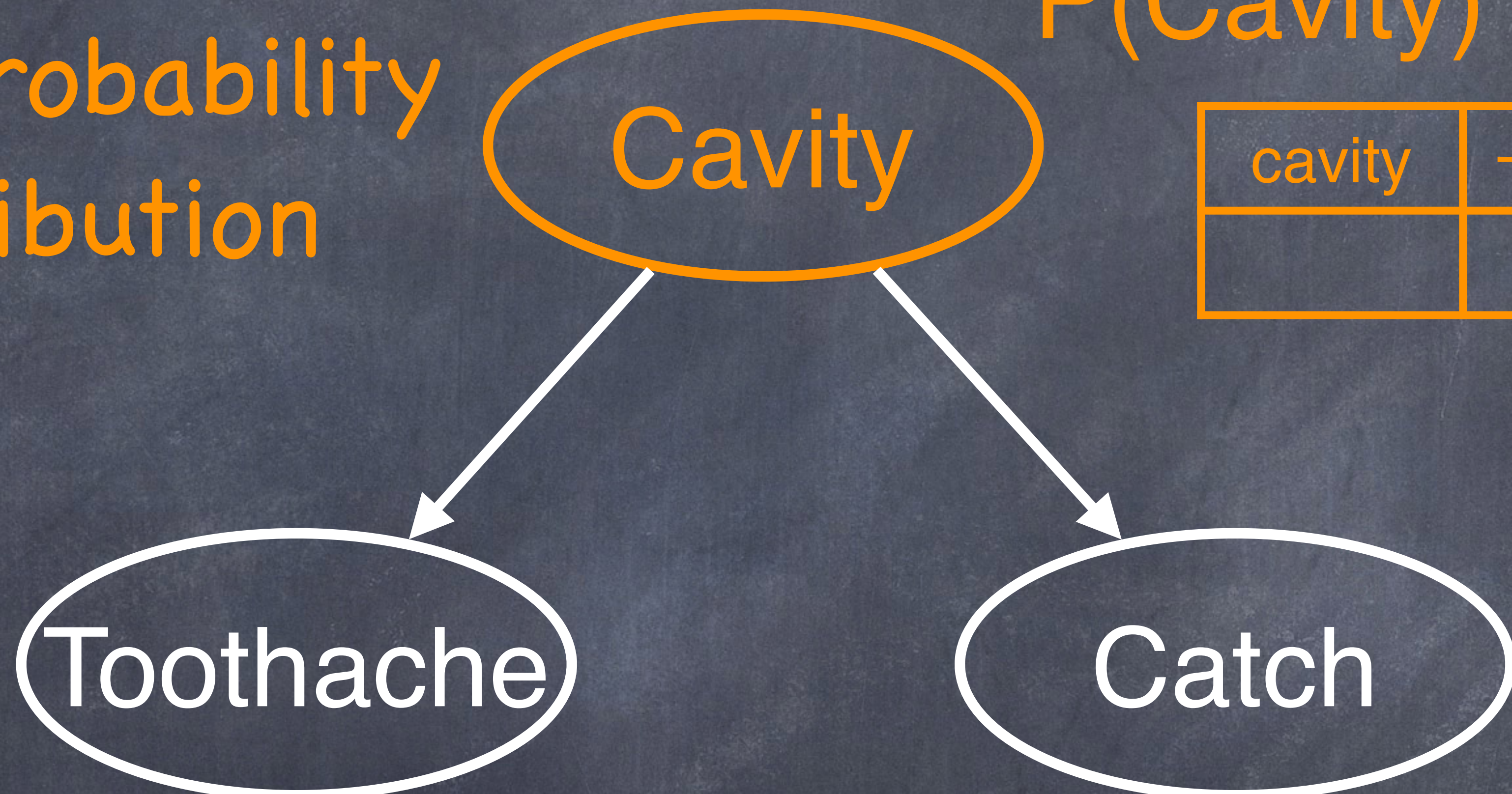
Cavity	catch	¬catch
cavity		
¬cavity		

Bayesian Networks

Prior Probability Distribution

P(Cavity)

cavity	¬cavity



P(Toothache|Cavity)

Cavity	toothache	¬toothache
cavity		
¬cavity		

P(Catch|Cavity)

Cavity	catch	¬catch
cavity		
¬cavity		

Bayesian Networks

- Nodes correspond to a random variables
- Link from X to Y iff X "directly influences" Y
  - No link: no "direct influence"
- Non-root nodes store the conditional distribution: $P(X_i \mid Parents(X_i))$
- Root nodes store their prior $P(X_i)$

Bayesian Networks How-To

- Select random variables required to model the domain
- Add links from causes to effects ("directly influences")
  - No cycles allowed. :(
- Add conditional probability distributions for $P(X_i \mid Parents(X_i))$ and priors $P(X_i)$

Probabilistic Inference

- Computing posterior probabilities for statements given observed evidence and probabilistic background knowledge

Probabilistic Inference (Single Variable)

Full Joint Prob. Dist.

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Query variable $X : Domain(X) = \{x_1, \ldots, x_m\}$

Evidence variables $\mathbf{E} : \{E_1, \ldots, E_k\}$

Observations $\mathbf{e} : \{e_1, \ldots, e_k\}$ s.t. $E_i = e_i$

Unobserved variables $\mathbf{Y} : \{Y_1, \ldots, Y_l\}$

$Domain(Y_i) = \{y_{i,1}, \ldots, y_{i,n_i}\}$

Semantics of Bayesian Networks

- Full joint distribution can be computed as the product of the separate conditional probabilities stored in the network

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

$$P(Toothache, Catch, Cavity)$$

Product Rule

$$= P(Toothache, Catch \mid Cavity)\ P(Cavity)$$

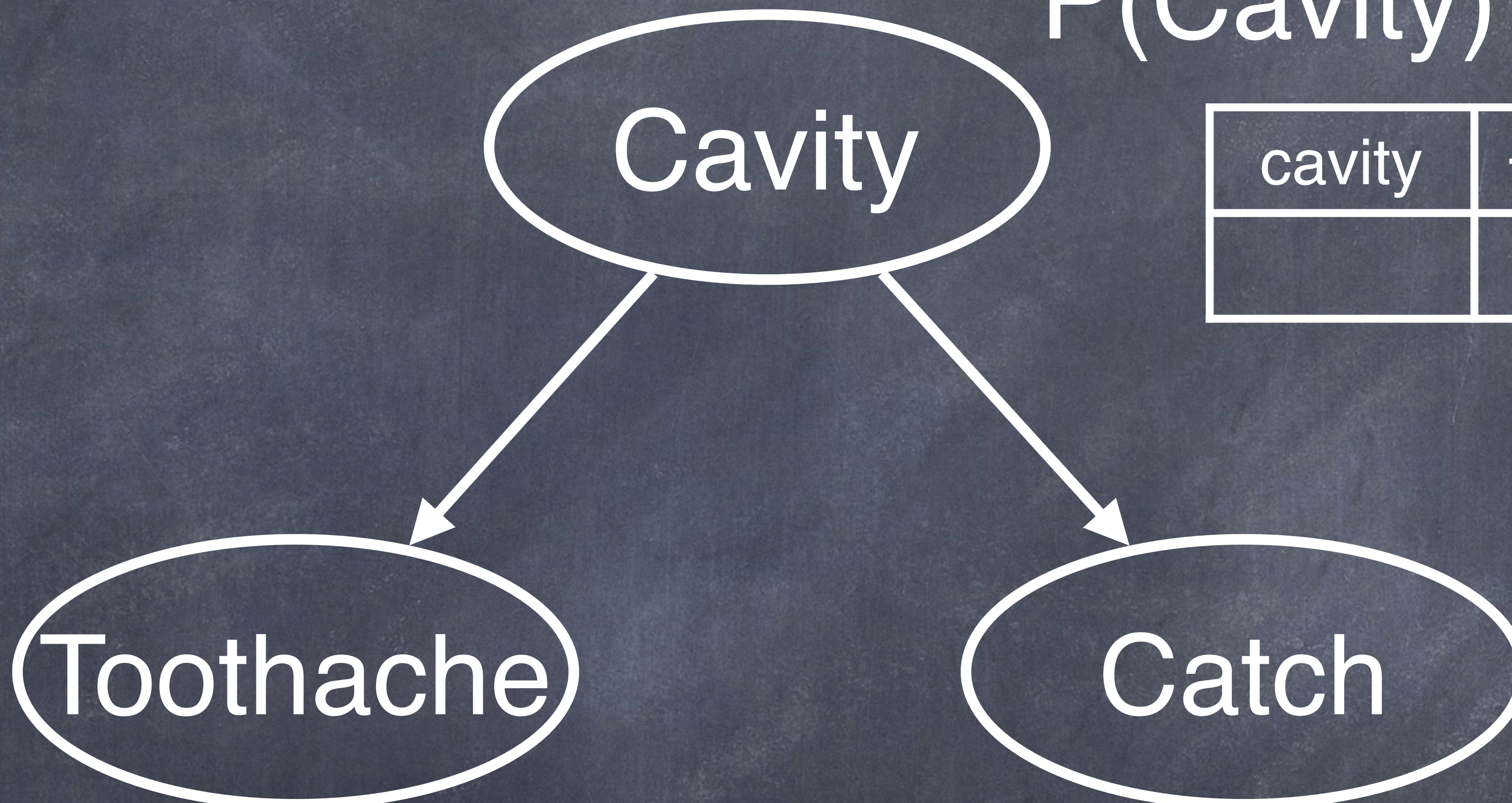
$$= P(Toothache \mid Cavity)\ P(Catch \mid Cavity)\ P(Cavity)$$

Conditional Independence

Bayesian Networks

P(Cavity)

cavity	¬cavity



P(Toothache|Cavity)

Cavity	toothache	¬toothache
cavity		
¬cavity		

P(Catch|Cavity)

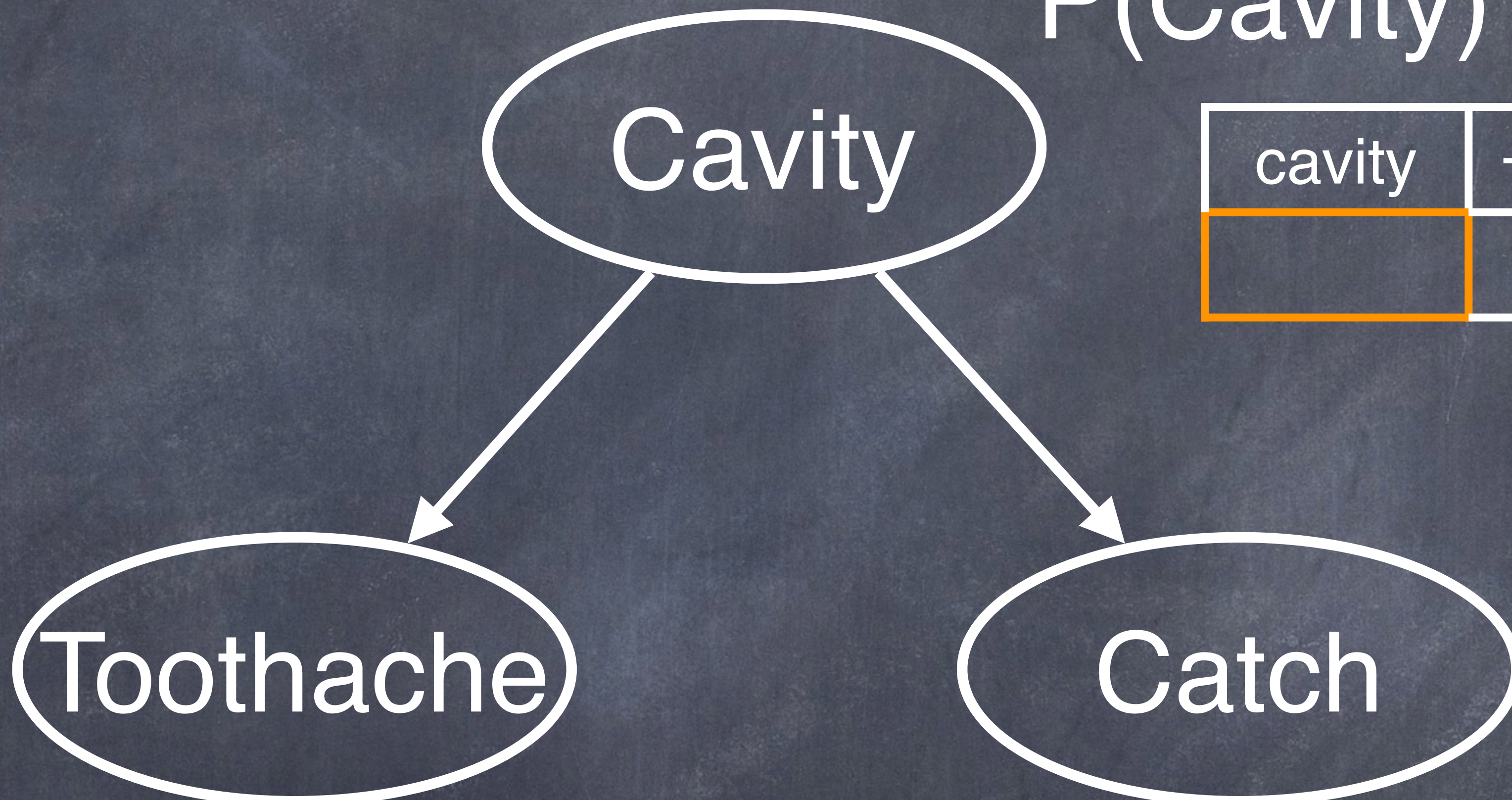
Cavity	catch	¬catch
cavity		
¬cavity		

Semantics of Bayesian Networks

- Full joint distribution can be computed as the product of the separate conditional probabilities stored in the network

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

$$P(\text{toothache}, \text{cavity}, \text{catch}) = P(\text{toothache} \mid \text{cavity}) P(\text{catch} \mid \text{cavity}) P(\text{cavity})$$



P(Cavity)

cavity	¬cavity

P(Toothache | Cavity)

Cavity	toothache	¬toothache
cavity		
¬cavity		

P(Catch | Cavity)

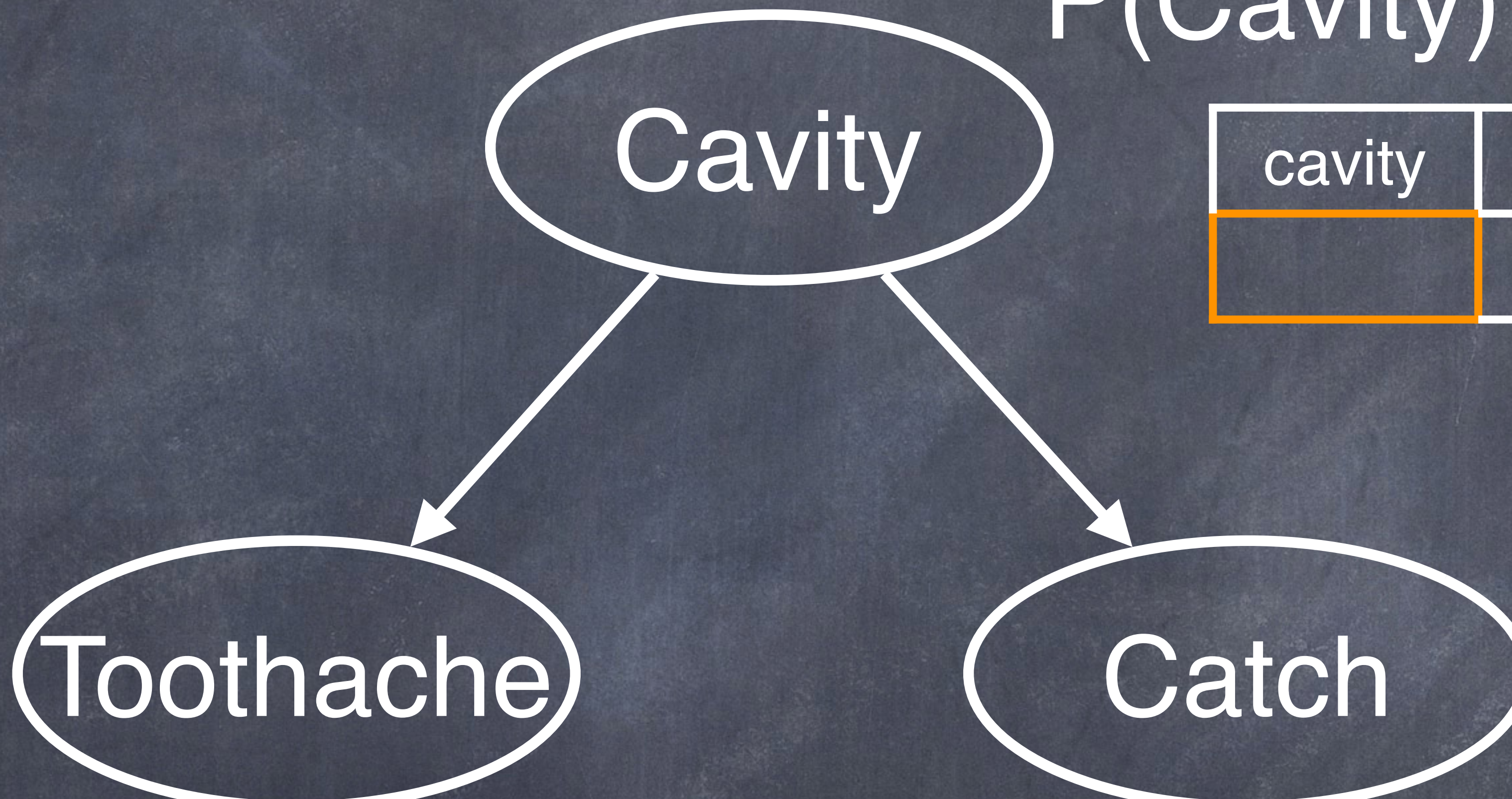
Cavity	catch	¬catch
cavity		
¬cavity		

$$P(\neg toothache, cavity, catch) =$$

$$P(\neg toothache | cavity) P(catch | cavity) P(cavity)$$

P(Cavity)

cavity	¬cavity



P(Toothache|Cavity)

Cavity	toothache	¬toothache
cavity		
¬cavity		

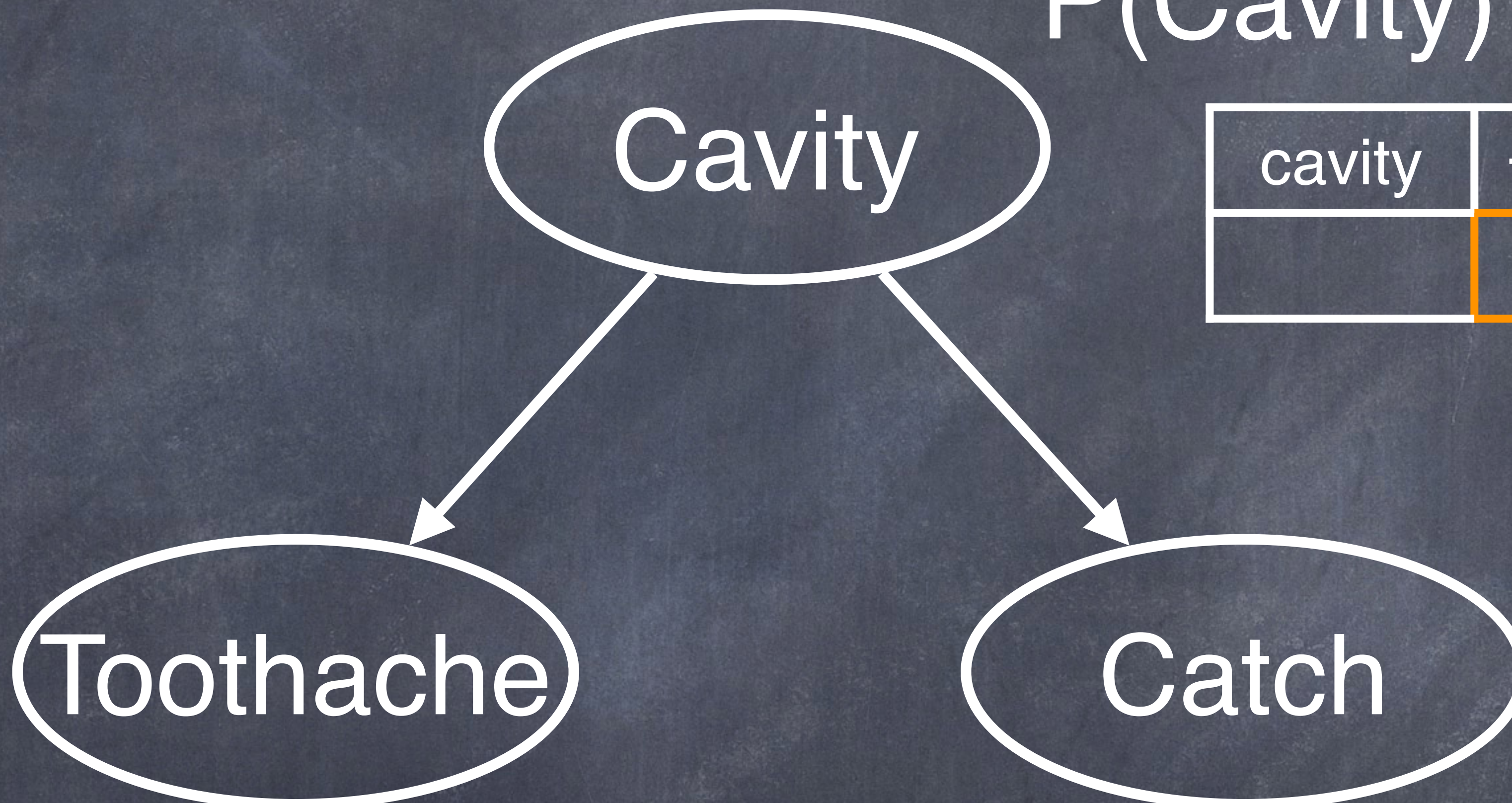
P(Catch|Cavity)

Cavity	catch	¬catch
cavity		
¬cavity		

$$P(\neg toothache, \neg cavity, \neg catch) =$$
$$P(\neg toothache \mid \neg cavity) P(\neg catch \mid \neg cavity) P(\neg cavity)$$

P(Cavity)

cavity	¬cavity



P(Toothache|Cavity)

Cavity	toothache	¬toothache
cavity		
¬cavity		

P(Catch|Cavity)

Cavity	catch	¬catch
cavity		
¬cavity		

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Semantics of Bayesian Networks

- Full joint distribution can be computed as the product of the separate conditional probabilities stored in the network

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

Inference in Bayesian Networks

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

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

Inference in Bayesian Networks

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

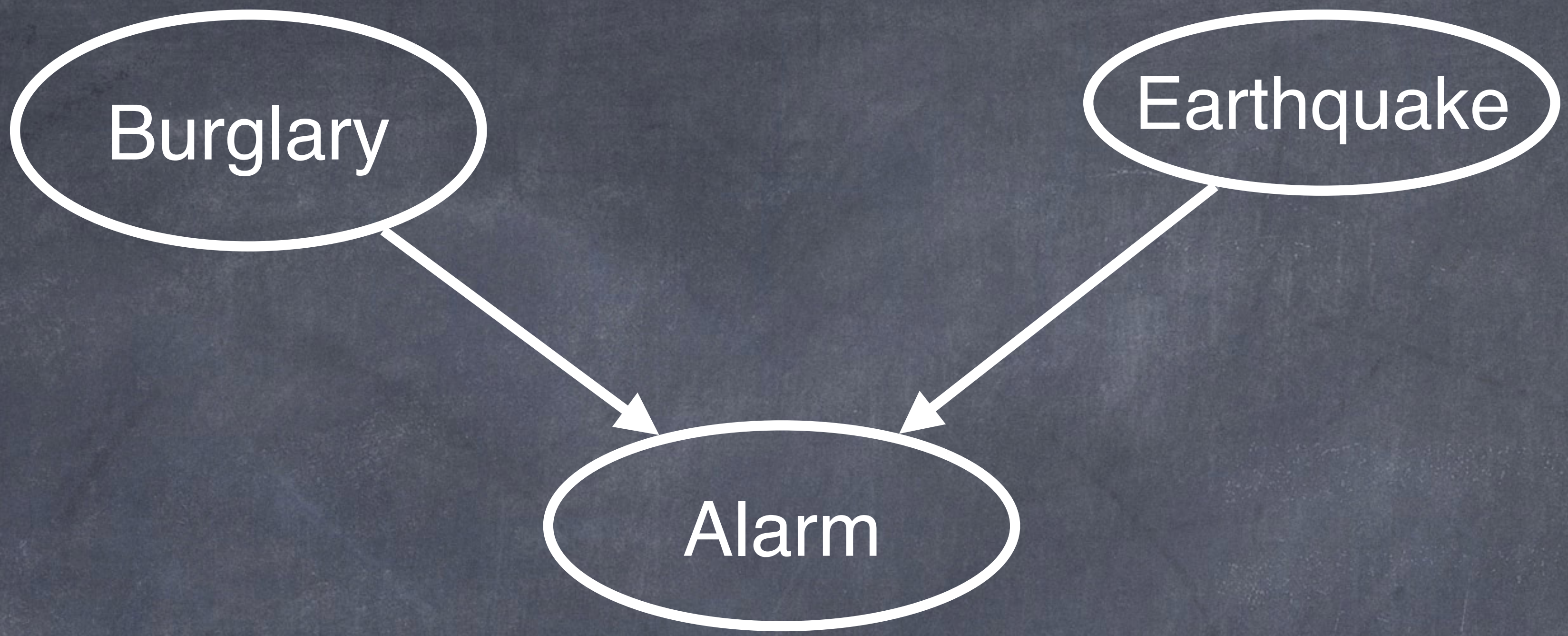
$$= \alpha \sum_{\mathbf{y}} \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- "A query can be answered from a Bayesian Network by computing sums of products of conditional probabilities from the network."

Example:
Judea Pearl's Alarm


HOME
ALARM

Alarm


Burglary
Earthquake
Alarm


Burglary
Earthquake
Alarm
JohnCalls
MaryCalls

P(B)

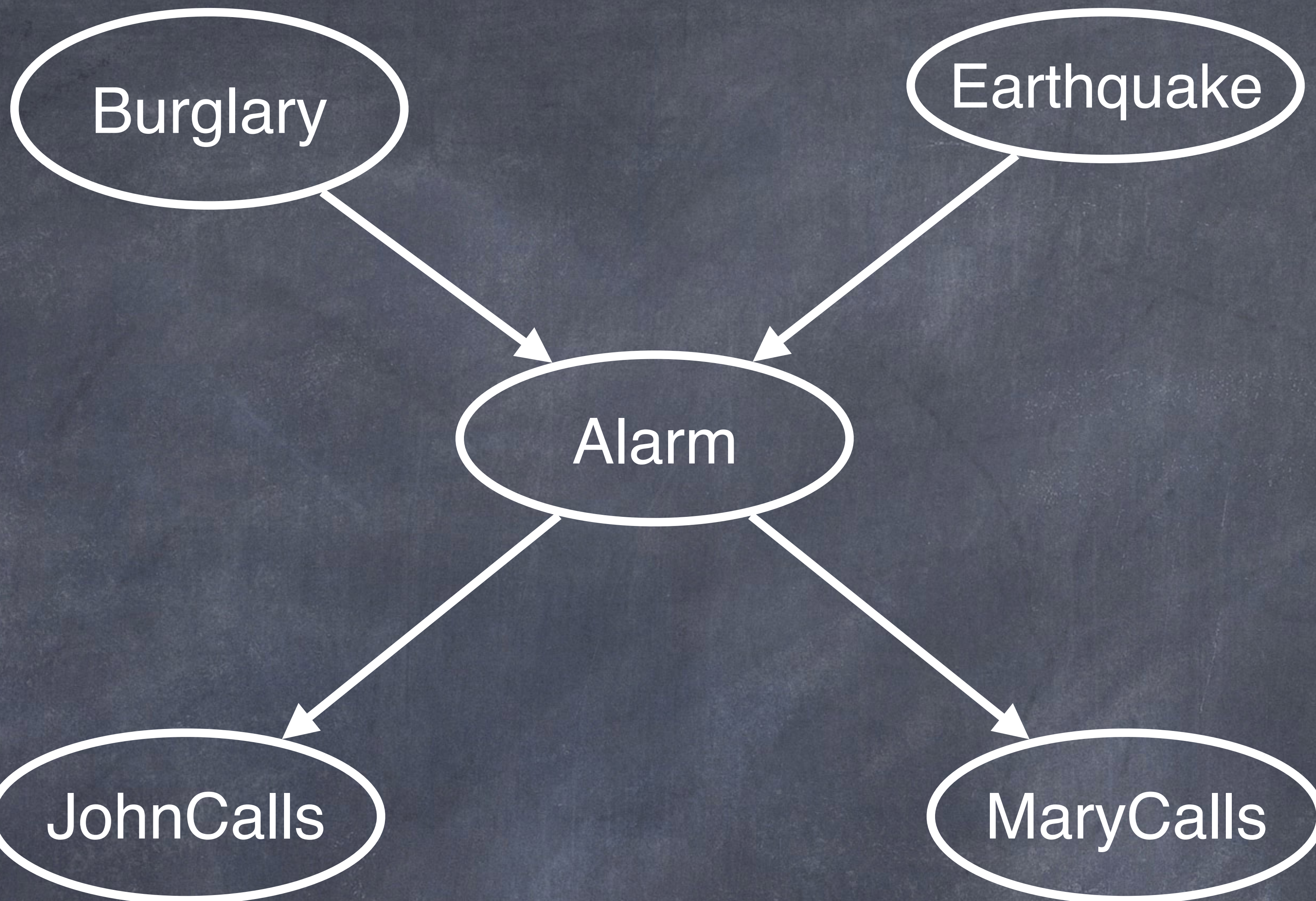
P(b)
0.001

P(E)

P(e)
0.002

P(B)

P(b)	P(¬b)
0.001	1-0.001



P(E)

P(e)	P(¬e)
0.002	1-0.002

P(B)

P(b)
0.001

P(E)

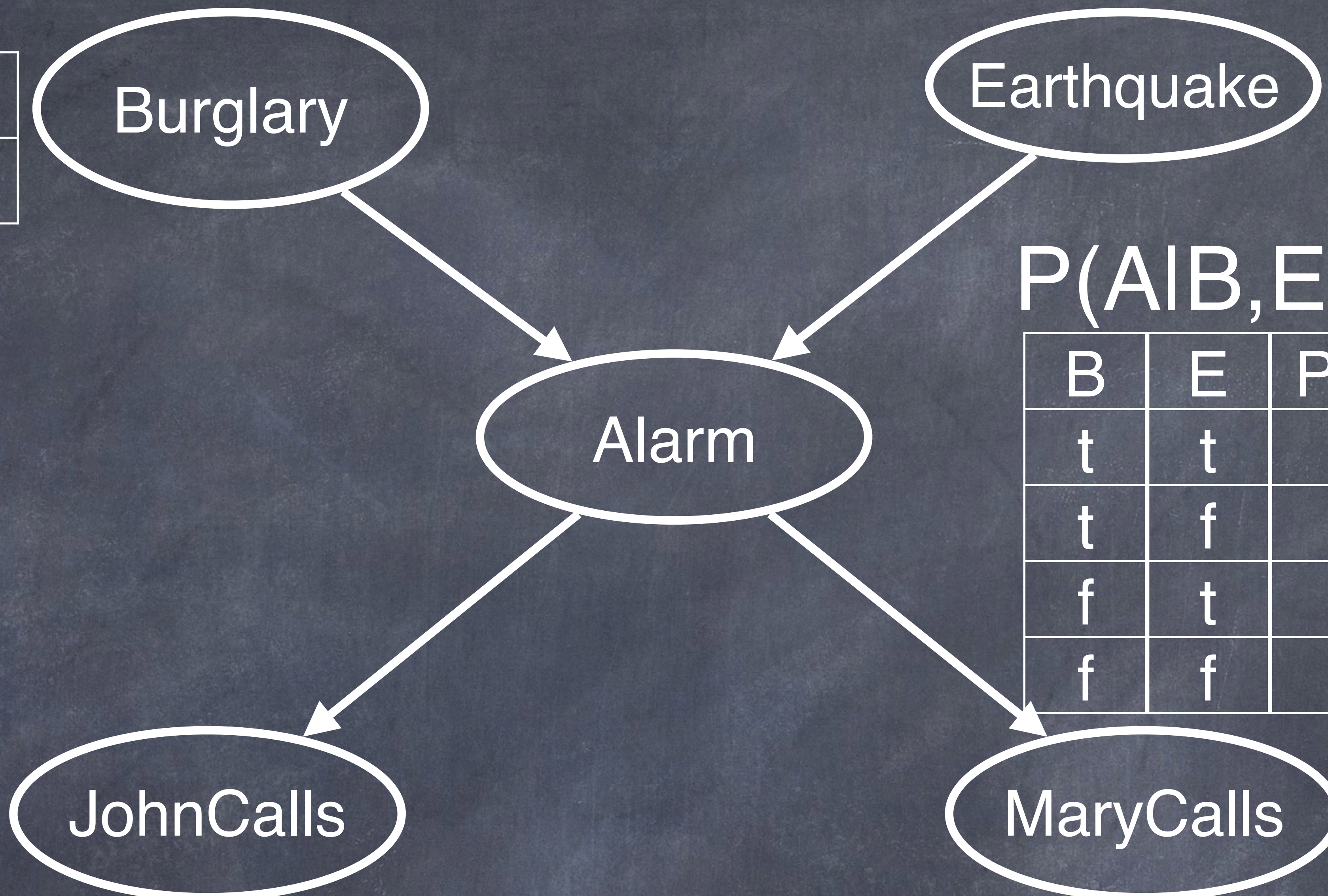
P(e)
0.002

P(B)

P(b)
0.001

P(E)

P(e)
0.002



$P(A \mid B,E)$

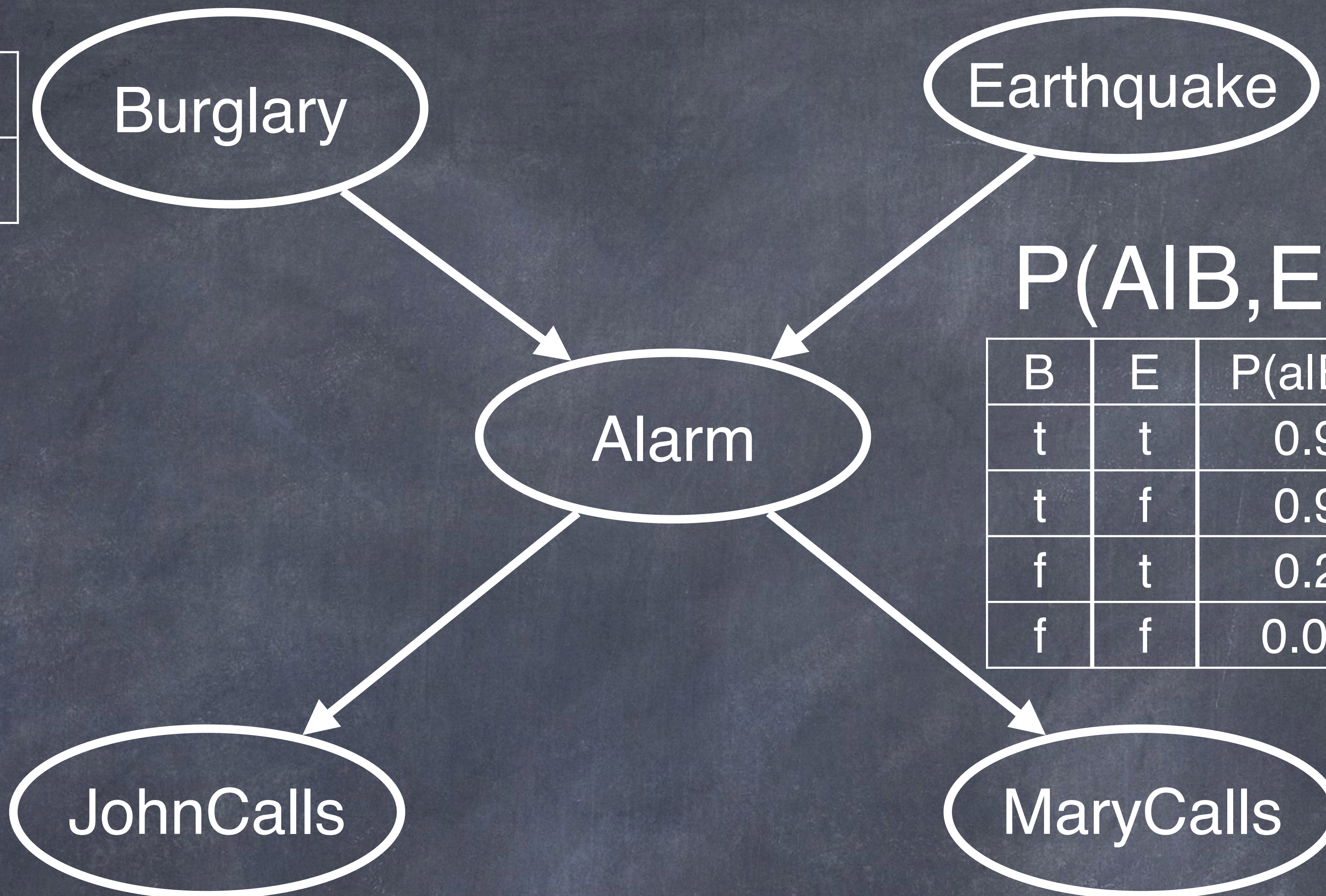
B	E	$P(a \mid B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(B)

P(b)
0.001

P(E)

P(e)
0.002



P(A|B,E)

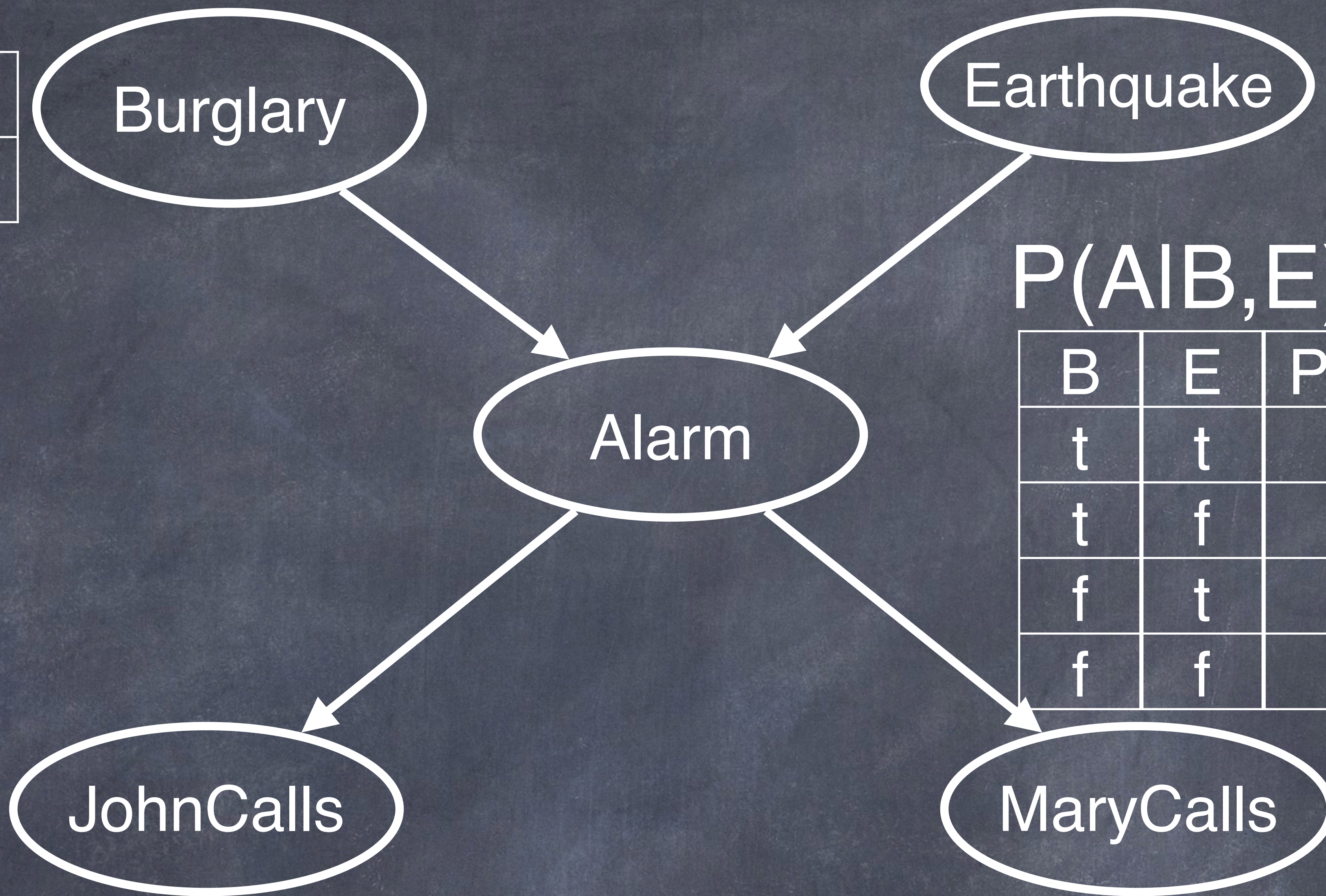
B	E	P(a|B,E)	P(¬a|B,E)
t	t	0.95	1-0.95
t	f	0.94	1-0.94
f	t	0.29	1-0.29
f	f	0.001	1-0.001

P(B)

P(b)
0.001

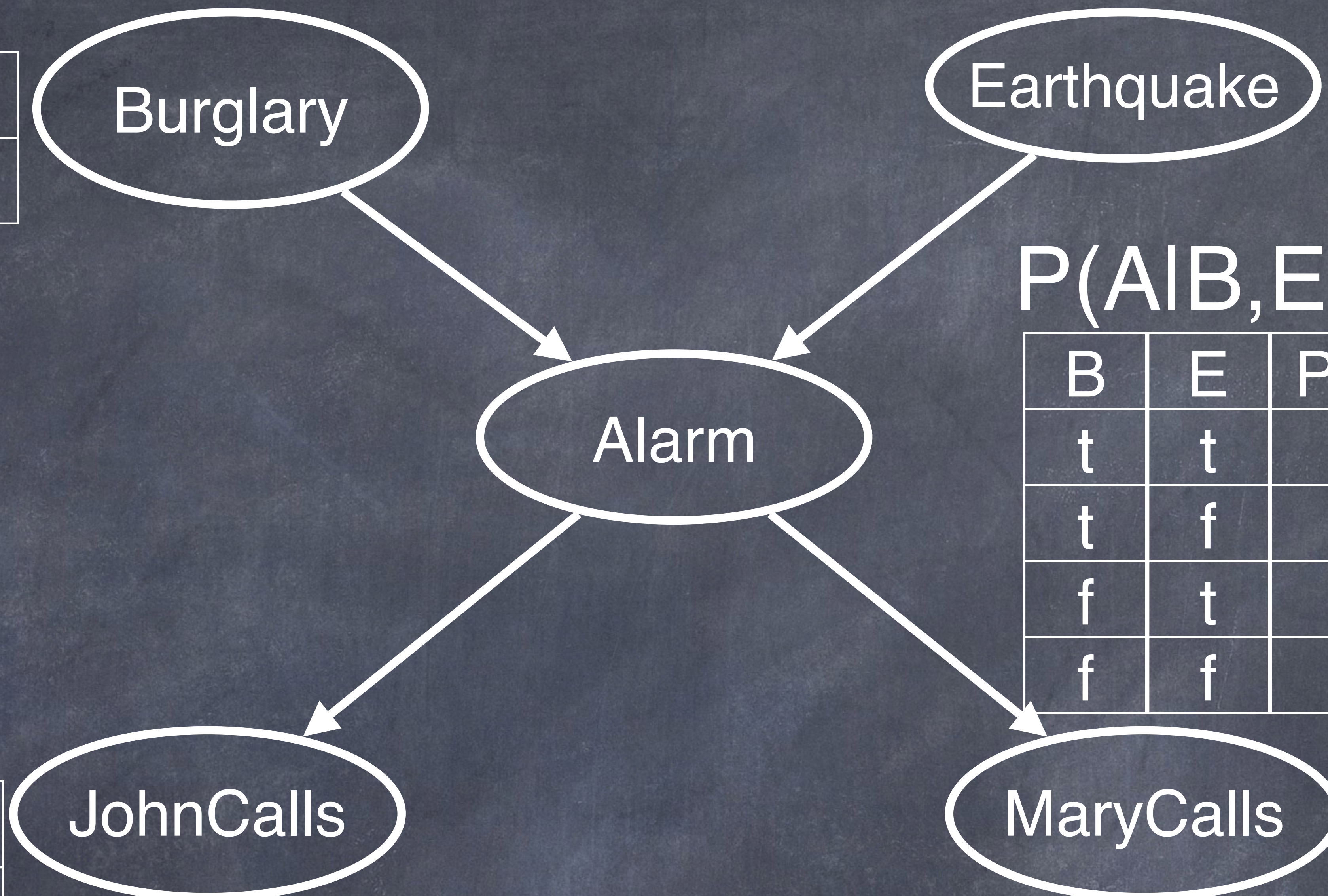
P(E)

P(e)
0.002



P(A∣B,E)

B	E	P(a∣B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001



$P(B)$

$P(b)$
0.001

$P(E)$

$P(e)$
0.002

$P(A \mid B,E)$

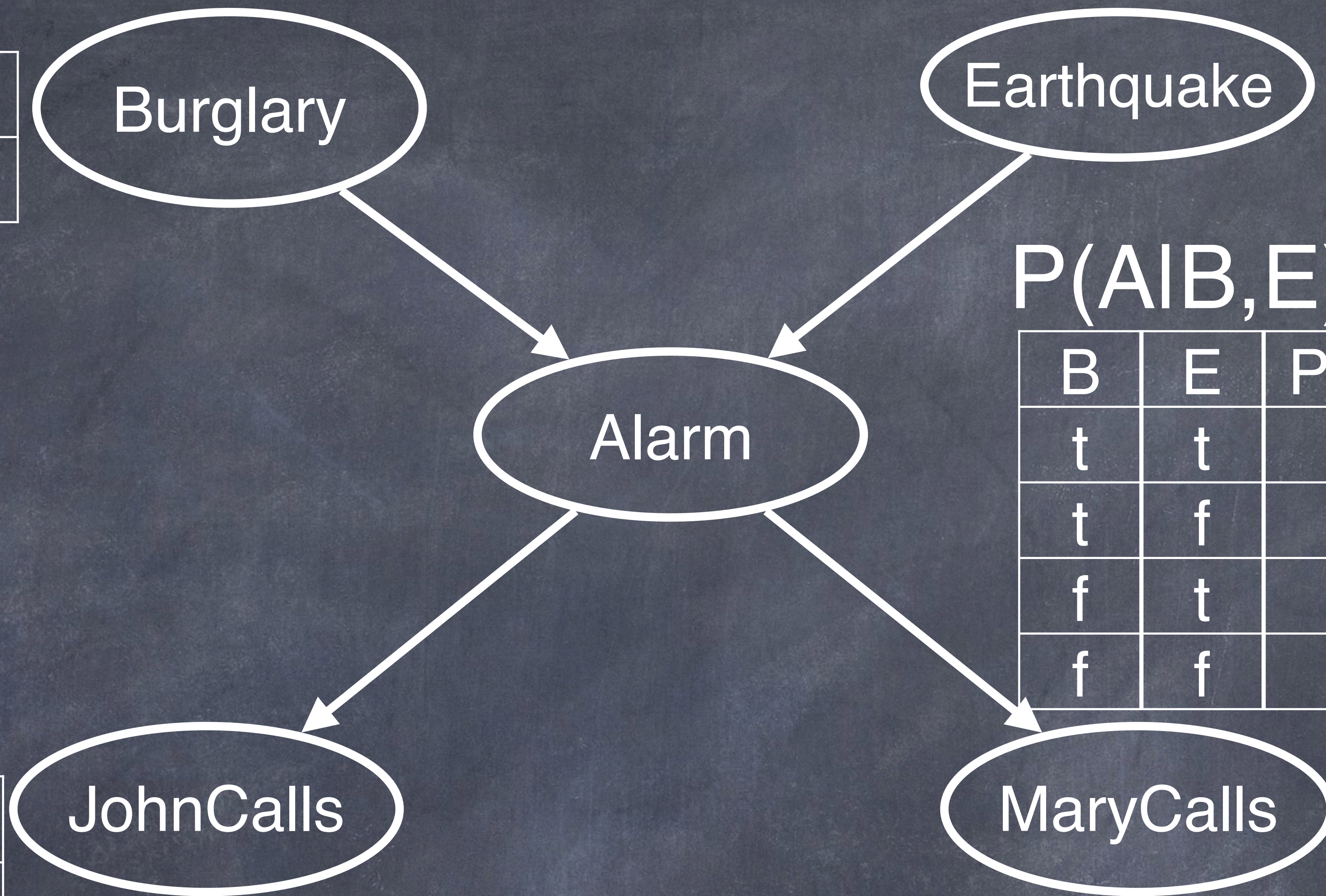
B	E	$P(a \mid B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$P(J \mid A)$

A	$P(j \mid A)$
t	0.9
f	0.05

$P(M \mid A)$

A	$P(m \mid A)$
t	0.7
f	0.01



P(B)

P(b)
0.001

P(E)

P(e)
0.002

P(A|B,E)

B	E	P(a|B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(J |A)

A	P(j |A)
t	0.9
f	0.05

P(M |A)

A	P(m|A)
t	0.7
f	0.01

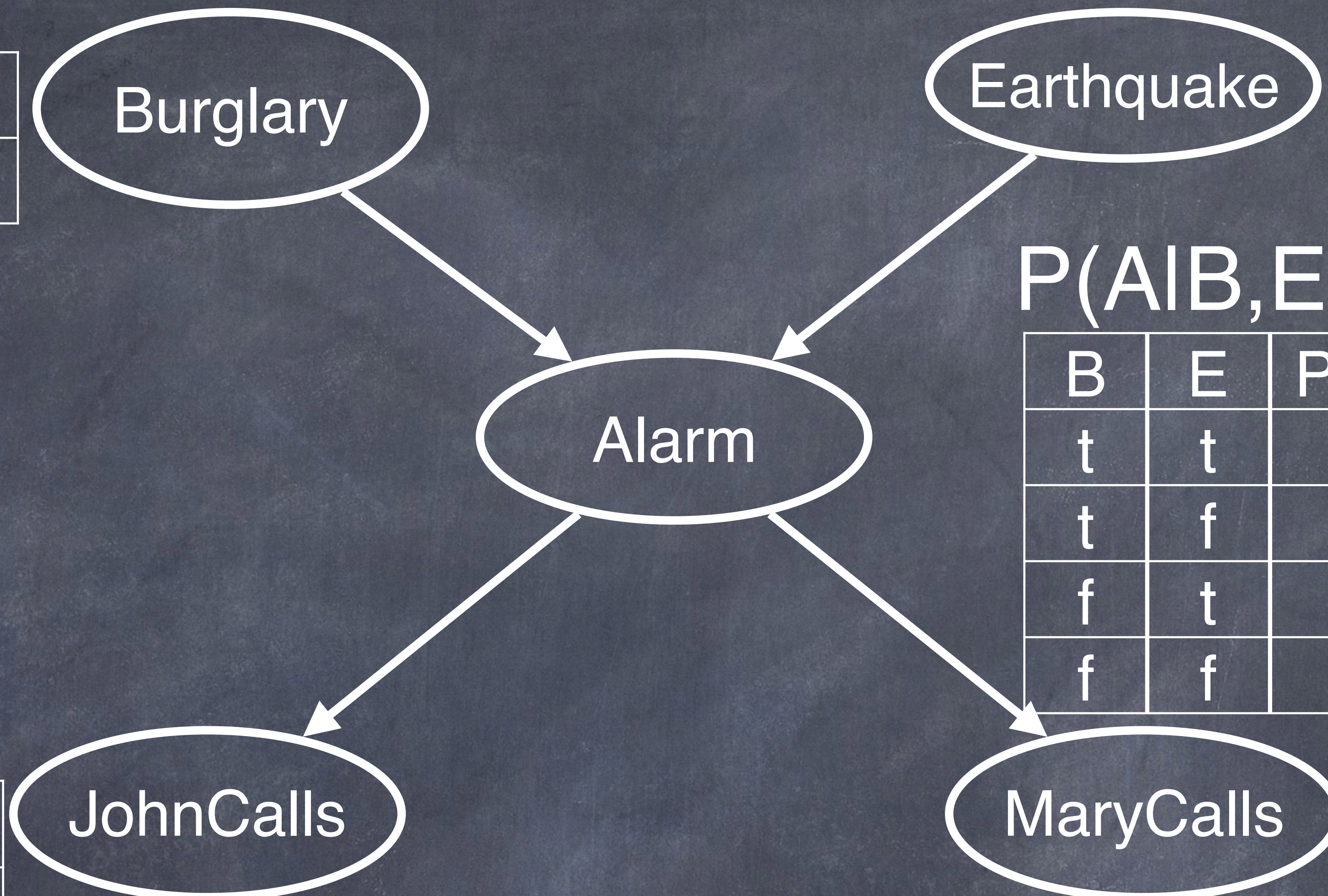
Full joint distribution: 2^5=32 entries

Bayesian network: 10 entries

Assuming conditional independences encoded in the network

Call from: John

Call from: Mary



P(B)

P(b)
0.001

P(E)

P(e)
0.002

P(A|B,E)

B	E	P(a|B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(J |A)

A	P(j |A)
t	0.9
f	0.05

P(M |A)

A	P(m |A)
t	0.7
f	0.01

$$\mathbf{P}(Burglary \mid JohnCalls = True, MaryCalls = True)$$

$$\mathbf{P}(B \mid j, m)$$

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

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

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

P(B)

P(b)
0.001

P(E)

P(e)
0.002

P(A lB,E)

B	E	P(alB,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

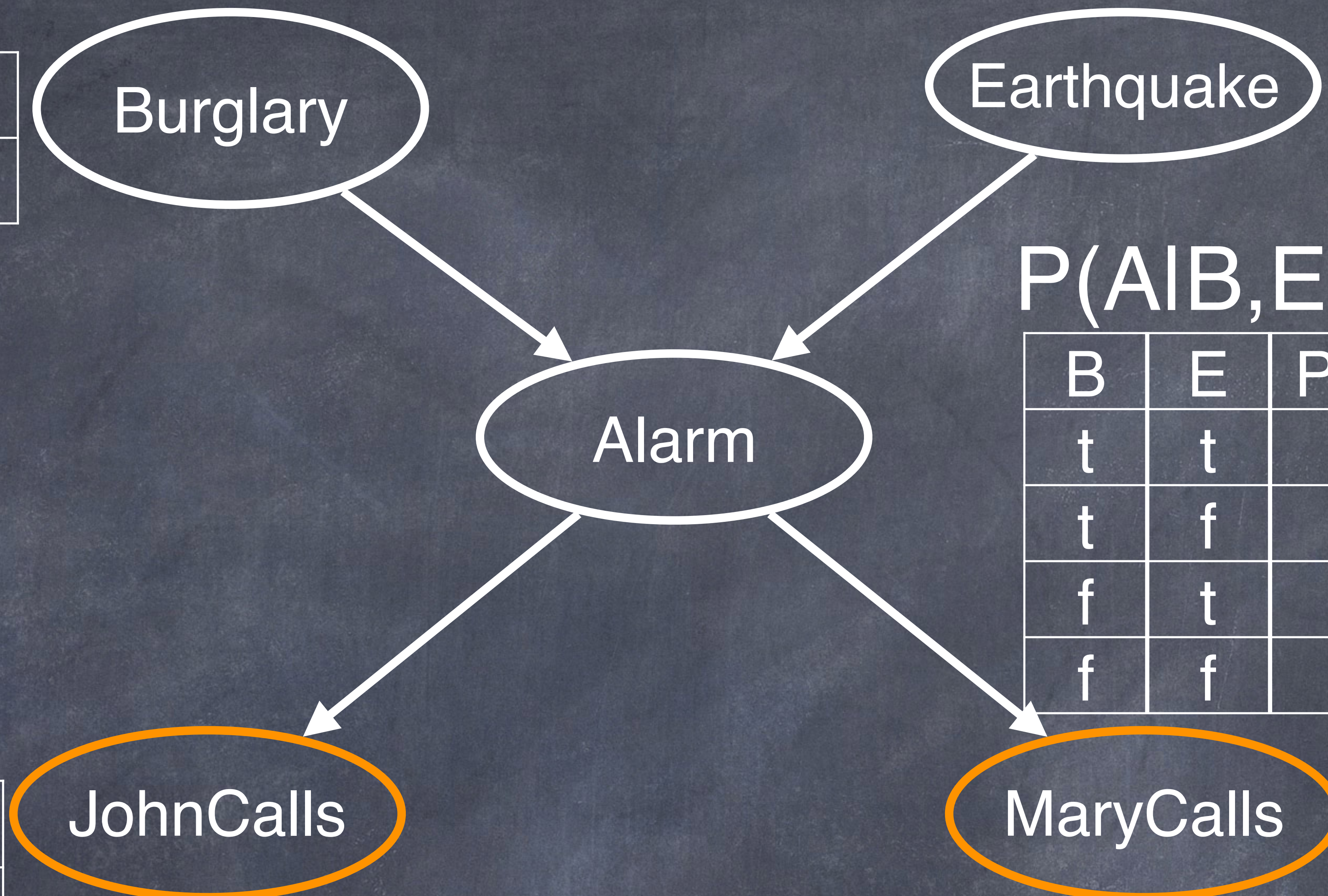
P(J l A)

A	P(jlA)
t	0.9
f	0.05

P(M l A)

A	P(mlA)
t	0.7
f	0.01

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



P(B)

P(b)
0.001

P(E)

P(e)
0.002

P(AIB,E)

B	E	P(aIB,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(JIA)

A	P(jIA)
t	0.9
f	0.05

P(MIA)

A	P(mIA)
t	0.7
f	0.01

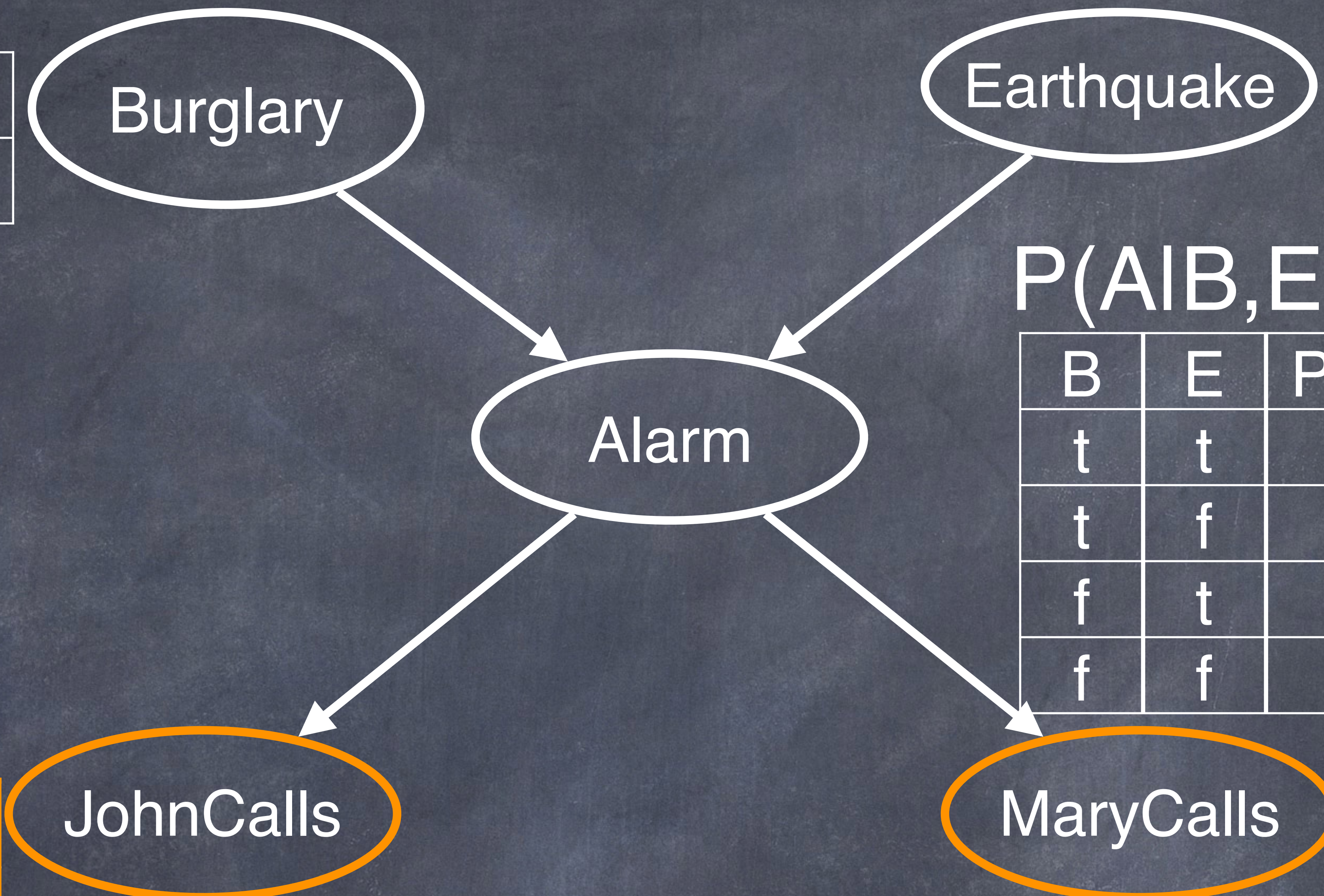
$$\mathbf{P}(B, E, A, \underline{J, M}) = \alpha\, \mathbf{P}(B)\, \mathbf{P}(E)\, \mathbf{P}(A \mid B, E)\, \mathbf{P}(\underline{J} \mid A)\, \mathbf{P}(\underline{M} \mid A)$$

P(B)

P(b)
0.001

P(E)

P(e)
0.002



$P(A \mid B,E)$

B	E	$P(a \mid B,E)$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

$P(j \mid A)$

A	$P(j \mid A)$
t	0.9
f	0.05

$P(m \mid A)$

A	$P(m \mid A)$
t	0.7
f	0.01

$$\mathbf{P}(B, E, A, \underline{j, m}) = \alpha\, \mathbf{P}(B)\, \mathbf{P}(E)\, \mathbf{P}(A \mid B, E)\, \mathbf{P}(\underline{j} \mid A)\, \mathbf{P}(\underline{m} \mid A)$$

P(B)

P(b)
0.001

P(E)

P(e)
0.002

P(A|B,E)

B	E	P(a|B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(j|A)

A	P(j|A)
t	0.9
f	0.05

P(m|A)

A	P(m|A)
t	0.7
f	0.01

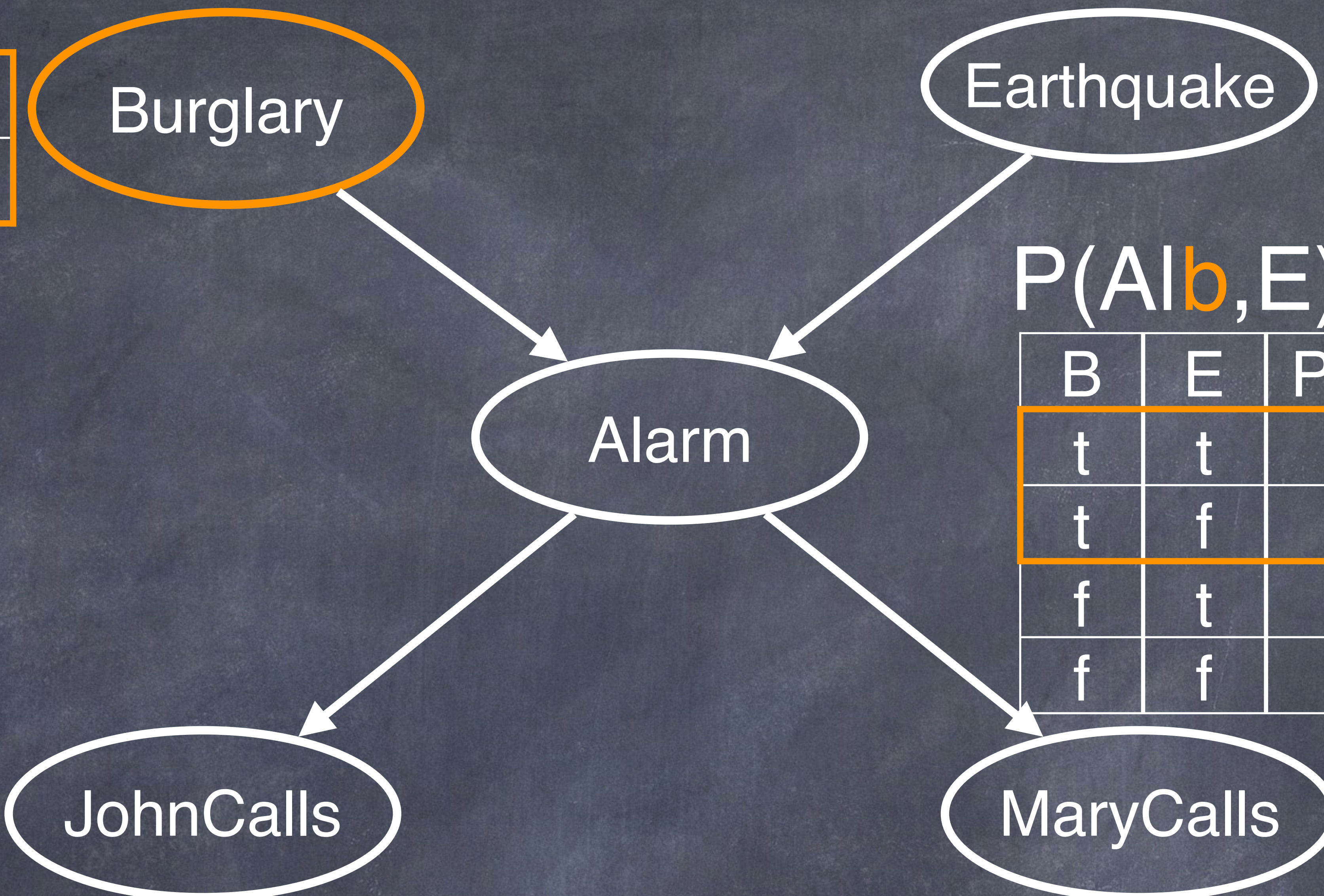
$$\mathbf{P}(\underline{B}, E, A, j, m) = \alpha\, \mathbf{P}(\underline{B})\, \mathbf{P}(E)\, \mathbf{P}(A \mid \underline{B}, E)\, \mathbf{P}(j \mid A)\, \mathbf{P}(m \mid A)$$

P(b)

P(b)
0.001

P(E)

P(e)
0.002



P(A|b,E)

B	E	P(a|B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

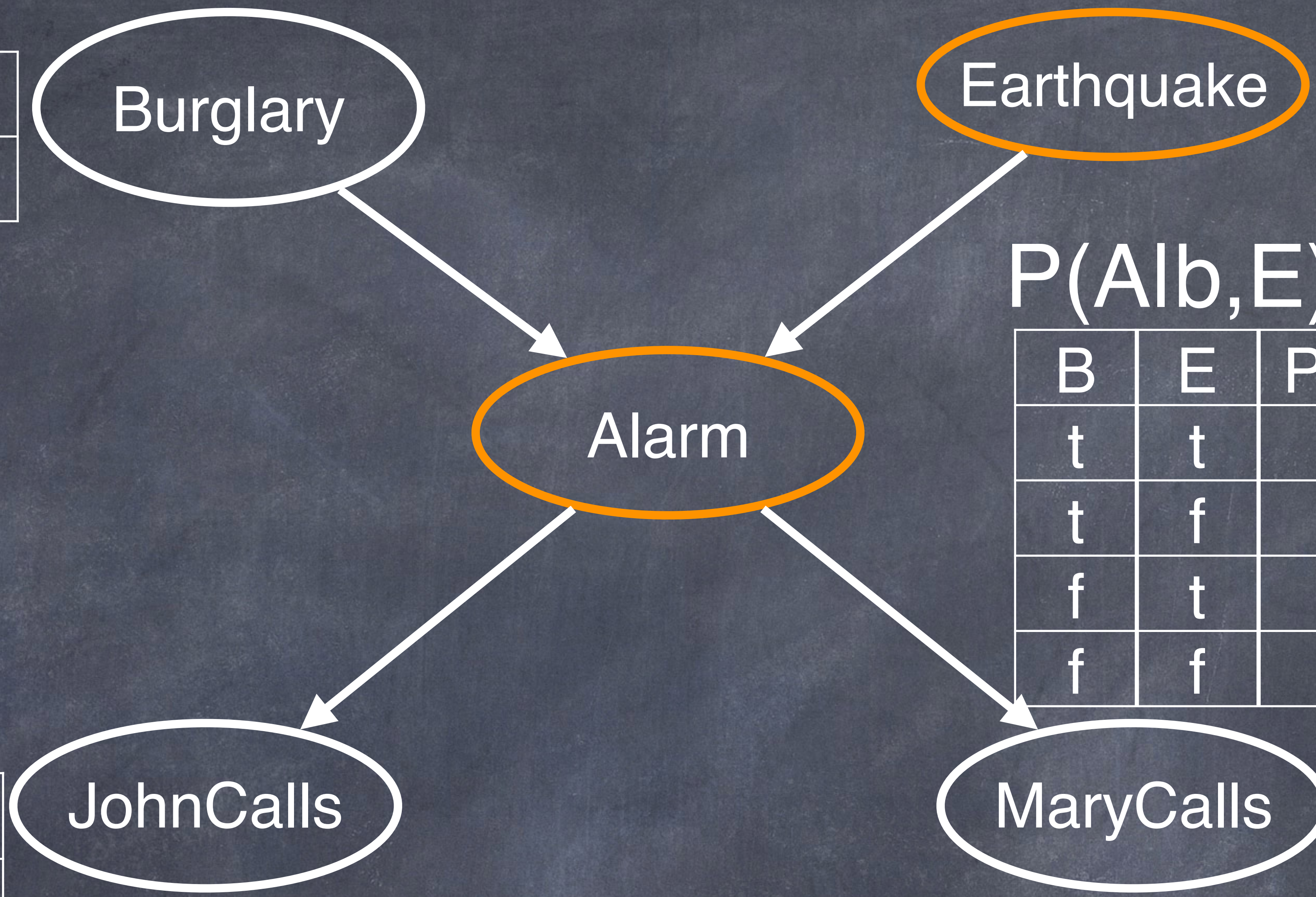
P(j|A)

A	P(j|A)
t	0.9
f	0.05

P(m|A)

A	P(m|A)
t	0.7
f	0.01

$$\mathbf{P}(\underline{b,} E, A, j, m) = \alpha\, \mathbf{P}\underline{(b)}\, \mathbf{P}(E)\, \mathbf{P}(A \mid \underline{b,} E)\, \mathbf{P}(j \mid A)\, \mathbf{P}(m \mid A)$$



P(b)

P(b)
0.001

P(E)

P(e)
0.002

P(A|b,E)

B	E	P(a|B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(j|A)

A	P(j|A)
t	0.9
f	0.05

P(m|A)

A	P(m|A)
t	0.7
f	0.01

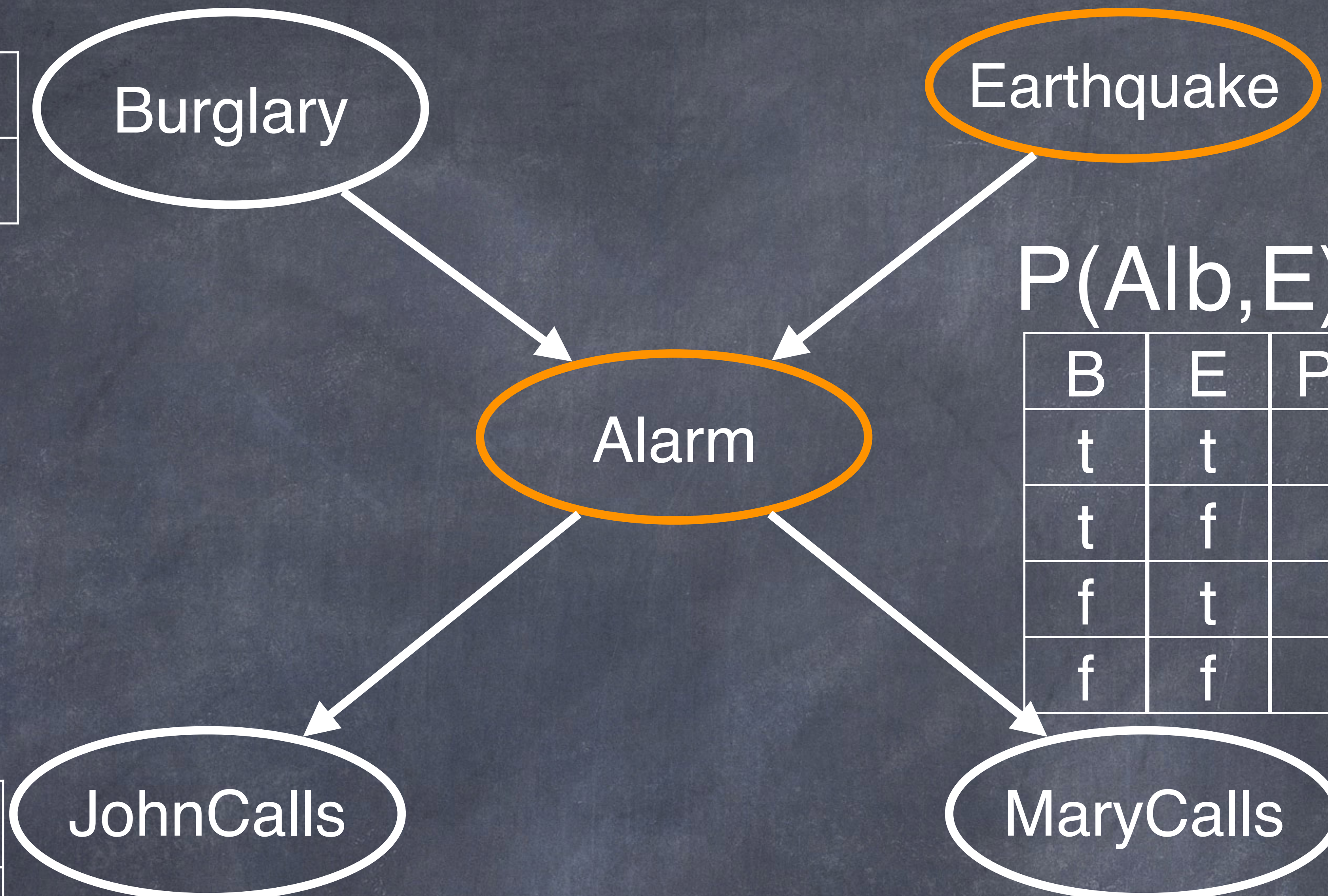
$$\mathbf{P}(b, \underline{E, A}, j, m) = \alpha\, \mathbf{P}(b)\, \mathbf{P}(\underline{E})\, \mathbf{P}(\underline{A} \mid b, \underline{E})\, \mathbf{P}(j \mid \underline{A})\, \mathbf{P}(m \mid \underline{A})$$

P(b)

P(b)
0.001

P(E)

P(e)
0.002



P(A|b,E)

B	E	P(a|B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(j|A)

A	P(j|A)
t	0.9
f	0.05

P(m|A)

A	P(m|A)
t	0.7
f	0.01

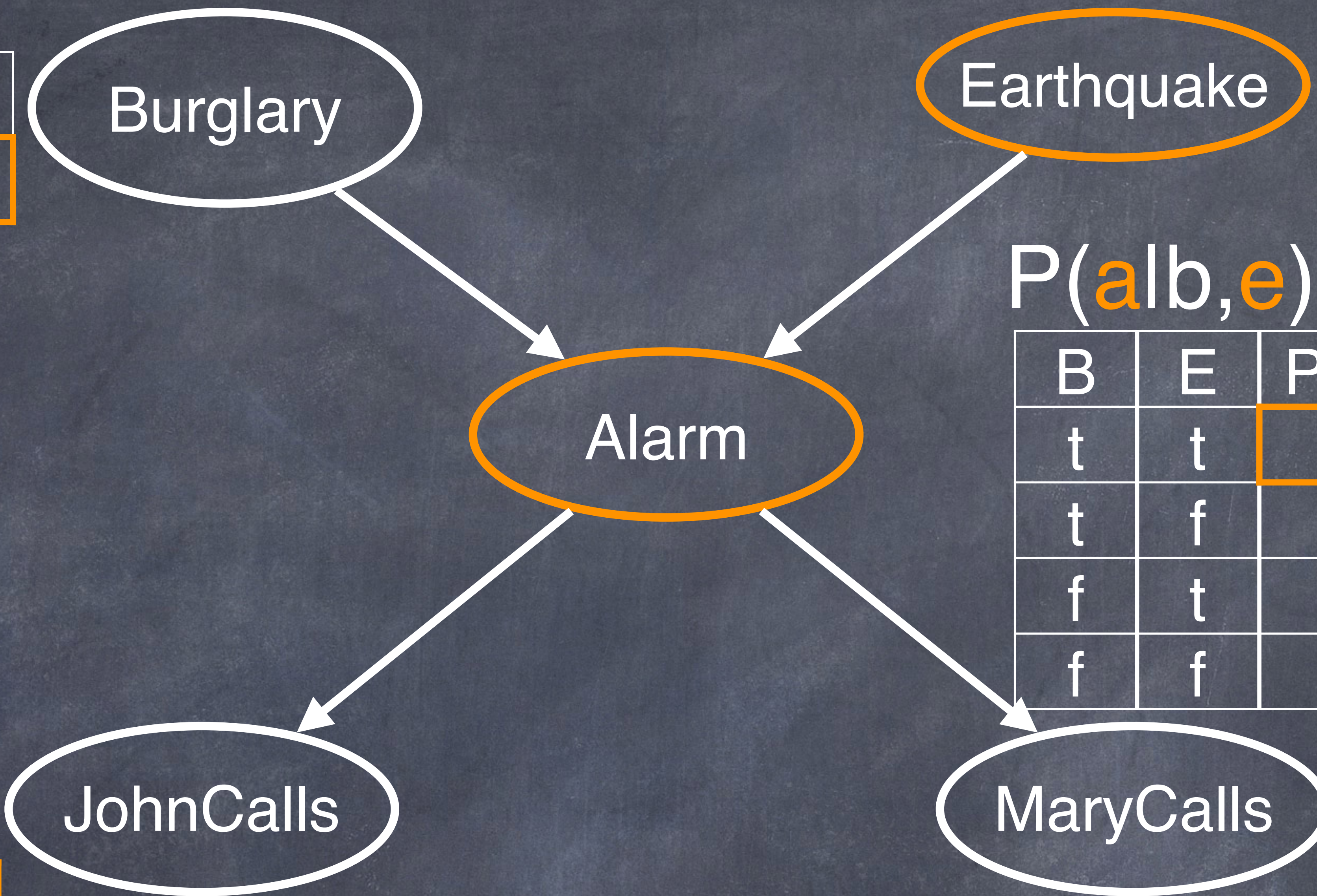
$$\mathbf{P}(b, j, m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(a \mid b, e)\, \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a)$$

P(b)

P(b)
0.001

P(e)

P(e)
0.002



P(a|b,e)

B	E	P(a|B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(j|a)

A	P(j|A)
t	0.9
f	0.05

P(m|a)

A	P(m|A)
t	0.7
f	0.01

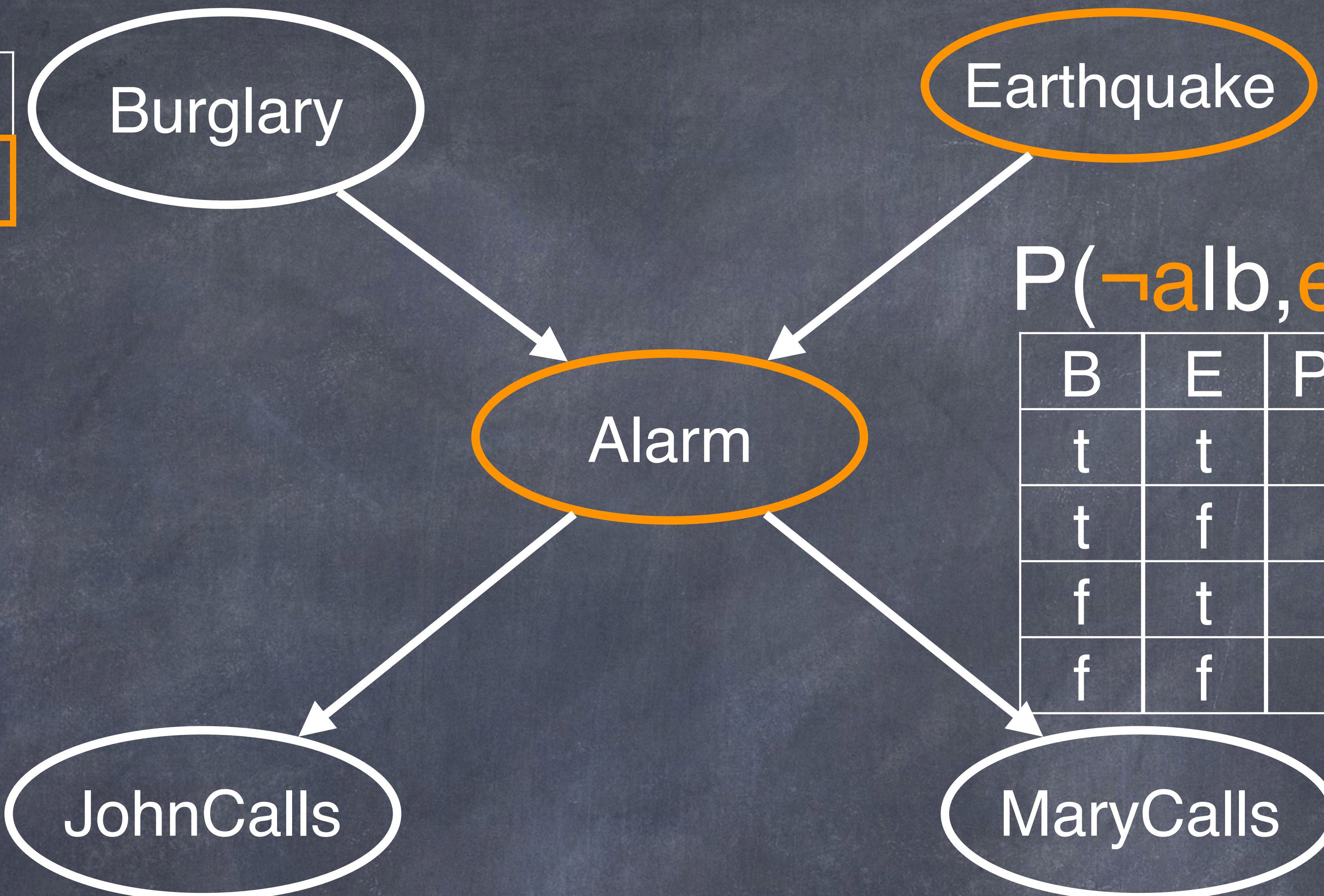
$$\mathbf{P}(b, j, m) = \alpha \, \mathbf{P}(b) \, \mathbf{P}(e) \, \mathbf{P}(a \mid b, e) \, \mathbf{P}(j \mid a) \, \mathbf{P}(m \mid a) +$$

P(b)

P(b)
0.001

P(e)

P(e)
0.002



P(¬a|b,e)

B	E	P(a|B,E)	
t	t	0.95	1-0.95
t	f	0.94	
f	t	0.29	
f	f	0.001	

P(j|¬a)

A	P(j|A)
t	0.9
f	0.05

P(m|¬a)

A	P(m|A)
t	0.7
f	0.01

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

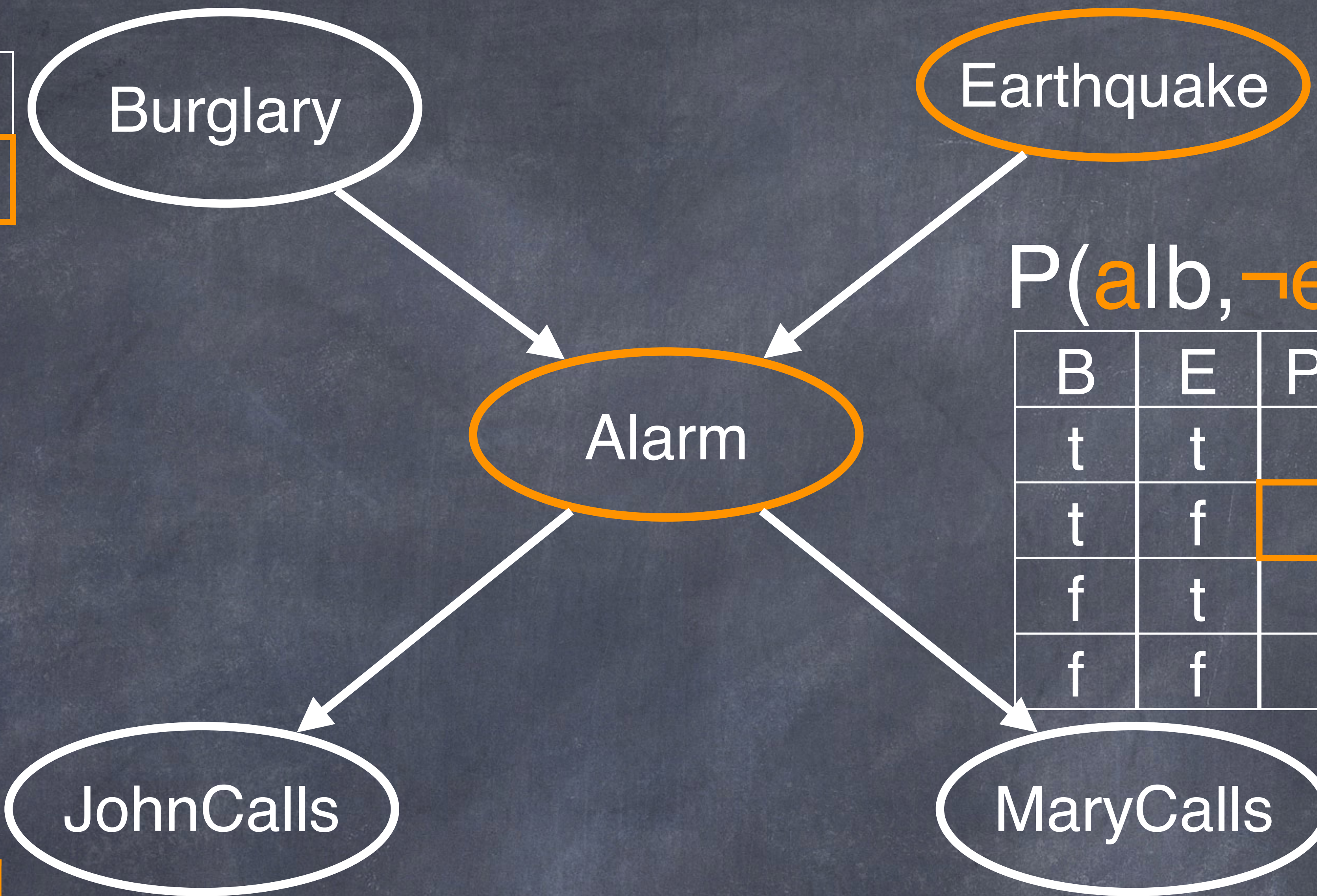
$$\mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(\neg a \mid b,e)\, \mathbf{P}(j \mid \neg a)\, \mathbf{P}(m \mid \neg a) +$$

P(b)

P(b)
0.001

P(¬e)

P(e)	
0.002	1-0.00



P(a|b,¬e)

B	E	P(a|B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(j|a)

A	P(j|A)
t	0.9
f	0.05

P(m|a)

A	P(m|A)
t	0.7
f	0.01

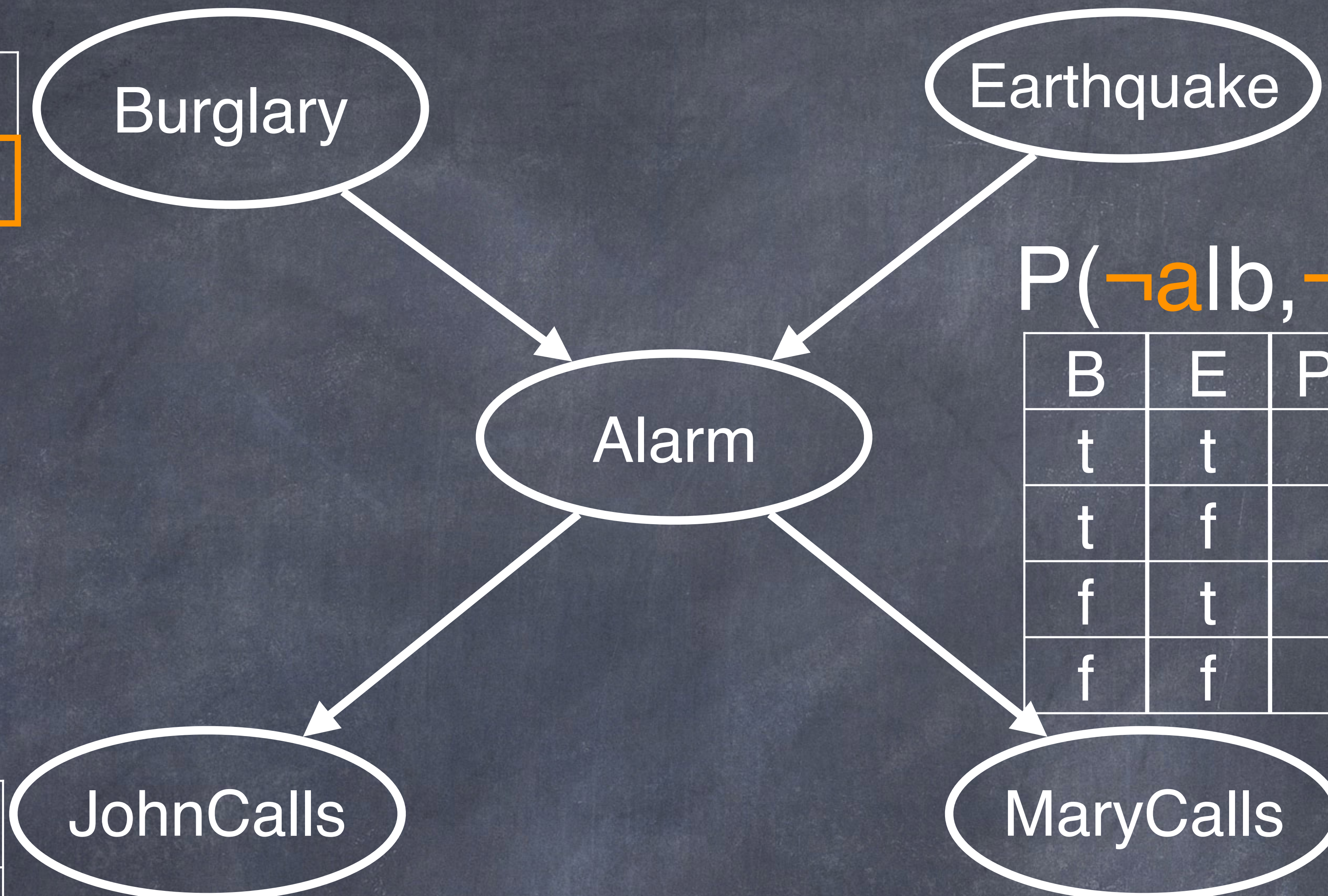
$$\mathbf{P}(b, j, m) = \alpha\, \mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(a \mid b, e)\, \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a) +$$
$$\mathbf{P}(b)\, \mathbf{P}(e)\, \mathbf{P}(\neg a \mid b, e)\, \mathbf{P}(j \mid \neg a)\, \mathbf{P}(m \mid \neg a) +$$
$$\mathbf{P}(b)\, \mathbf{P}(\neg e)\,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

P(b)

P(b)
0.001

P(¬e)

P(e)	
0.002	1-.002



P(¬a|b,¬e)

B	E	P(a|B,E)	
t	t	0.95	
t	f	0.94	1-.94
f	t	0.29	
f	f	0.001	

P(j|¬a)

A	P(j|A)
t	0.9
f	0.05

P(m|¬a)

A	P(m|A)
t	0.7
f	0.01

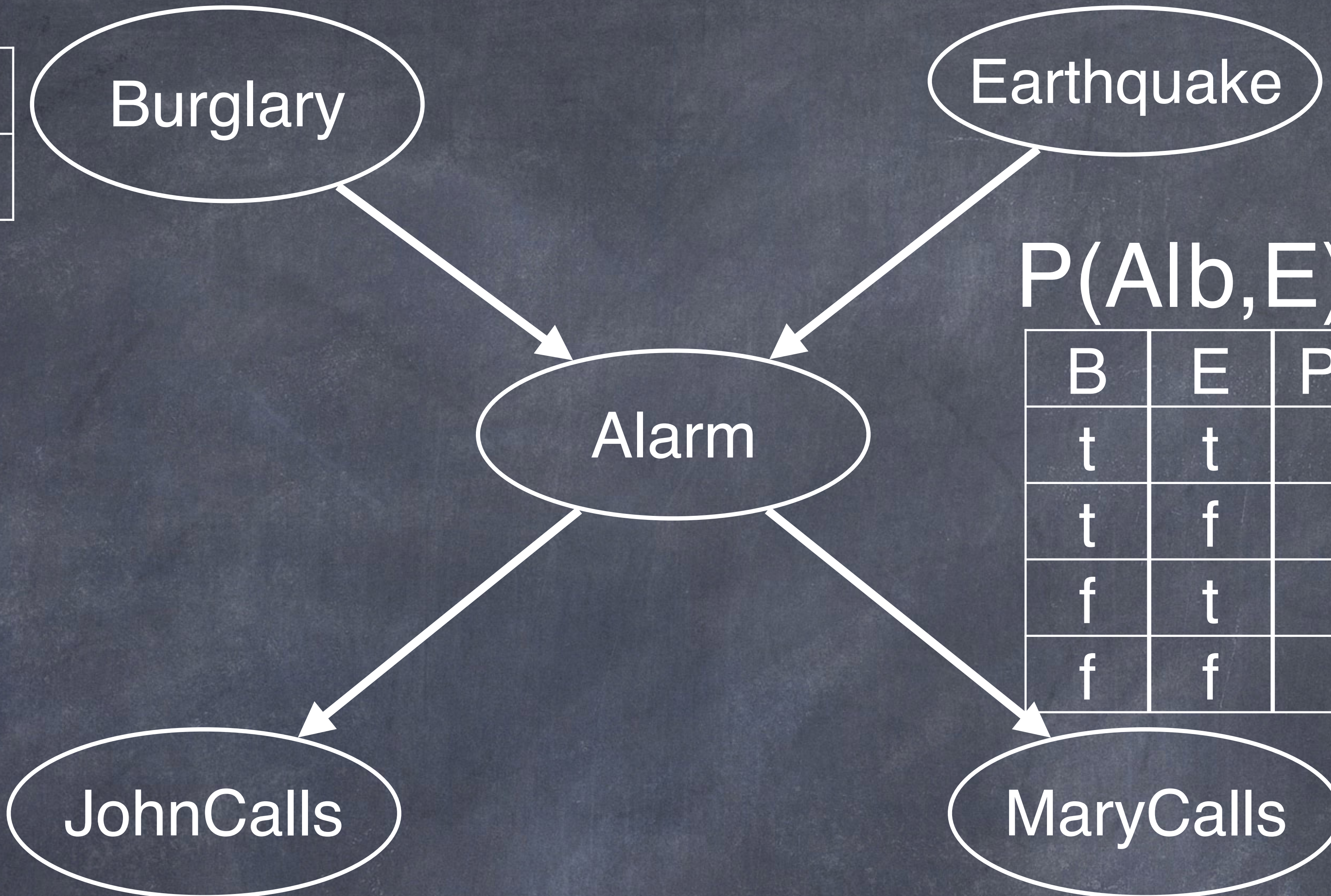
$$\begin{aligned}\mathbf{P}(b,j,m) = \alpha\Big(&\mathbf{P}(b)\,\mathbf{P}(e)\,\mathbf{P}(a \mid b,e)\,\mathbf{P}(j \mid a)\,\mathbf{P}(m \mid a)+\\ &\mathbf{P}(b)\,\mathbf{P}(e)\,\mathbf{P}(\neg a \mid b,e)\,\mathbf{P}(j \mid \neg a)\,\mathbf{P}(m \mid \neg a)+\\ &\mathbf{P}(b)\,\mathbf{P}(\neg e)\,\mathbf{P}(a \mid b,\neg e)\,\mathbf{P}(j \mid a)\,\mathbf{P}(m \mid a)+\\ &\mathbf{P}(b)\,\mathbf{P}(\neg e)\,\mathbf{P}(\neg a \mid b,\neg e)\,\mathbf{P}(j \mid \neg a)\,\mathbf{P}(m \mid \neg a)\Big)\end{aligned}$$

P(b)

P(b)
0.001

P(E)

P(e)
0.002



P(A|b,E)

B	E	P(a|B,E)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

P(j|A)

A	P(j|A)
t	0.9
f	0.05

P(m|A)

A	P(m|A)
t	0.7
f	0.01

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

Factored full joint dist.

Marginalization

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

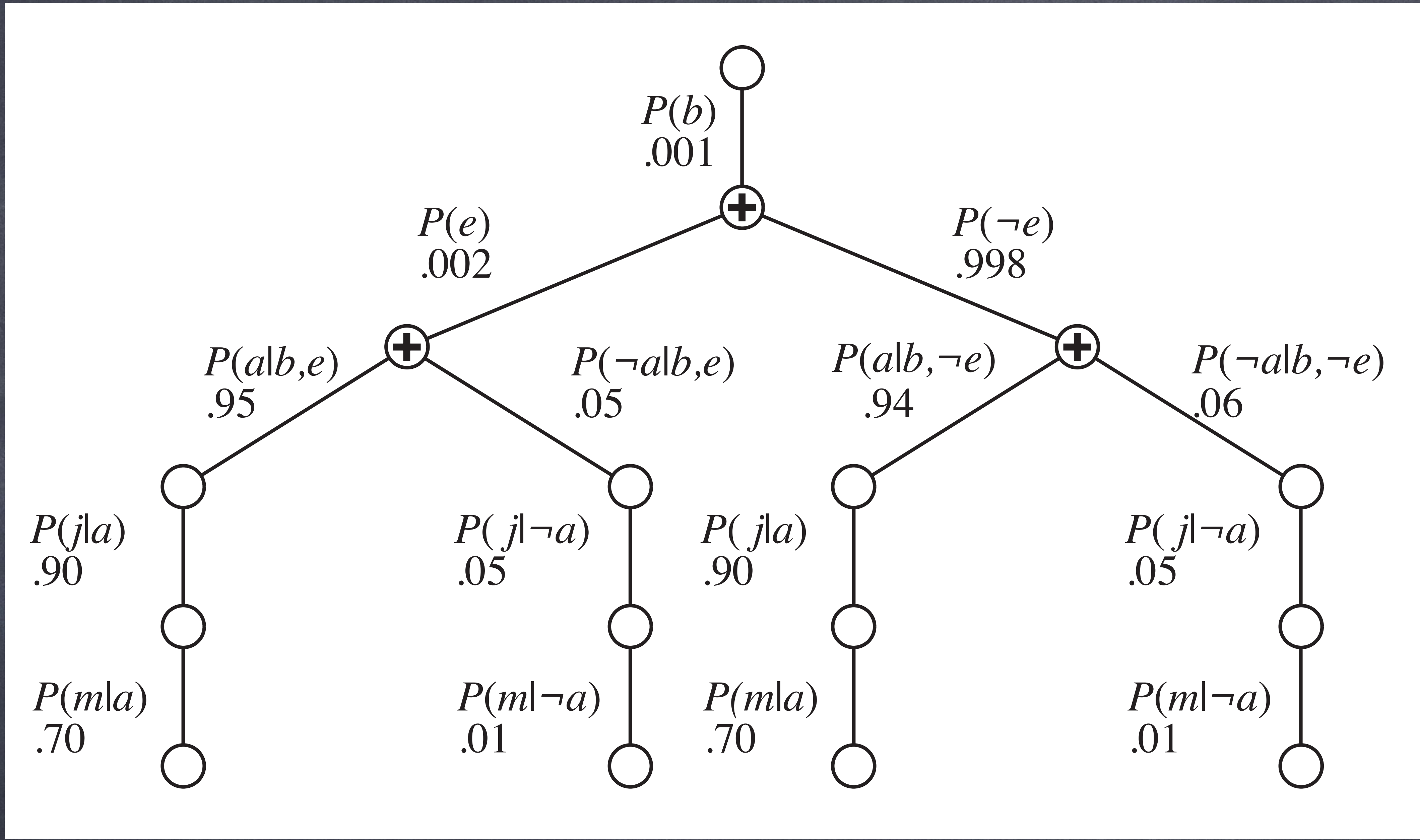
$$P(b \mid j, m) = \alpha \sum_{e} \sum_{a} P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a)$$

$$O(n2^n)$$

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

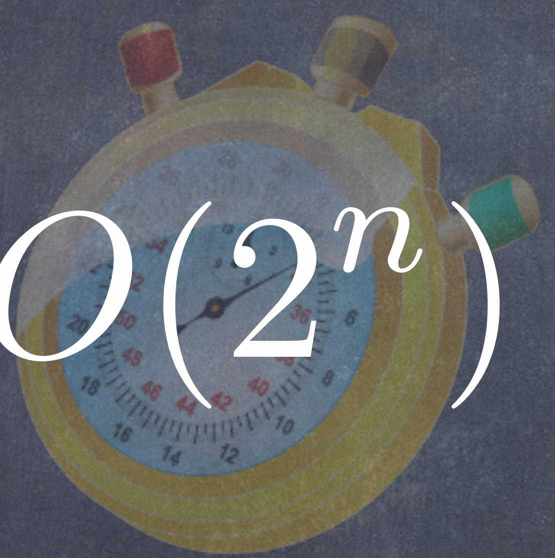
$$P(b \mid j, m) = \alpha \sum_{e} \sum_{a} P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a)$$

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



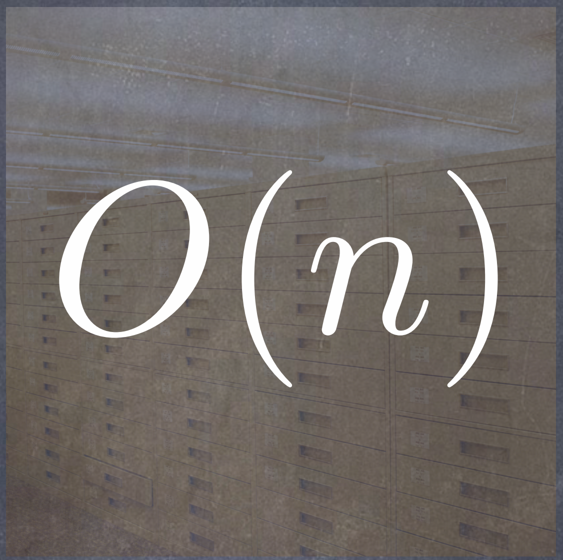
$$\mathbf{P}(B \mid j, m) = \alpha \langle 0.00059224, 0.0014919 \rangle \approx \langle 0.284, 0.716 \rangle$$

Exact Inference in BNs



$O(2^n)$

Time Complexity



$O(n)$

Space Complexity

Bayesian Networks Summary

- Independence assumptions make probabilistic inference easier
  - By factoring the joint distribution
- Bayesian Networks encode conditional independence assumptions among random variables
  - And store conditional probabilities

Exact Inference in BNs

- Exact inference in BNs is NP-hard
- Can be shown to be as hard as computing the number of satisfying assignments of a propositional logic formula => #P-hard

Inference in Bayesian Networks

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\,\mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

$$= \alpha \sum_{\mathbf{y}} \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

Inference in Bayesian Networks

- Exact inference with BNs is still hard
  - Dynamic programming can sometimes help (next time)
  - But we can do approximate inference efficiently (next time)
  - We can learn the conditional probabilities required to do inference from data (in a few weeks)

Project 3

- Implement Bayesian Network inference.
- Stages:
  - Read a BN from an XMLBIF file.
  - Implement a query/evidence API.
  - Implement exact and approximate inference.
  - Compare your results by resource requirements (time and memory) and accuracy.
- Formal specification will be posted Friday.

For Next Time:

AIMA 14.4, 14.5; 14.7 fyi