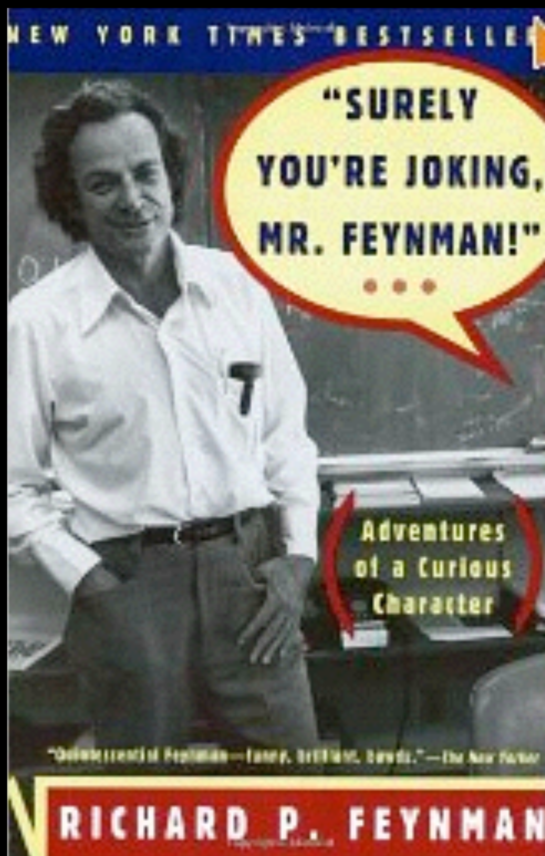
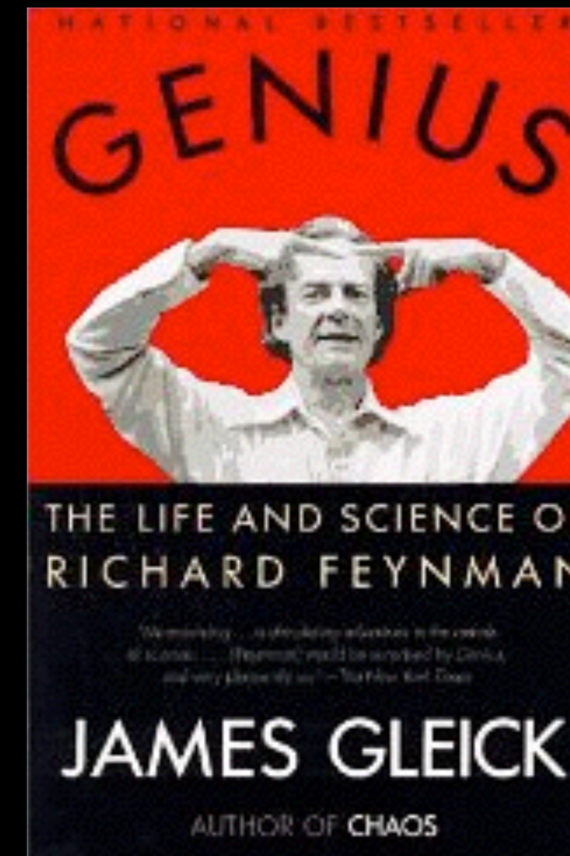
CSC242: Intro to AI

Lecture 15 Bayesian Networks





Richard Feynman
1918-1988


horizon

Othello Tournament Phase I Results

- DS-DH
- Deep-Blue
- EvanMariaPlayer
- Hyphaene-Thebaica
- KautzPlayer
- OR-KC
- a-a
- alaska-boat
- bent-paperclips
- blake-phelps
- crazy-pingpong
- delicious-fudge
- digital-teapot
- invalid-munmap
- is-rever
- jar-vis
- jesus-fish
- no-name
- othello-game
- othello-player
- problem-solved
- random-words
- robotics-anonymous
- saint-inferno
- samurai-sharks
- screaming-banjos
- spherical-cow
- spline-reticulators
- team-victory

Best Team Name

- Third Place:

0 Extra Credit Points!

- Se

  screaming-banjos

- First Place:

  spherical-cow

Defeated KautzPlayer

- DS-DH
- Hyphaene-The
- OR-
- a-a
- alaska-boat
- jar-vis
- jesus-fish
- no-name

- othello-game
- othello-player
- lved
- rds
- no
- spherical-cow
- spline-reticulators
- team-victory

10 Extra Credit Points!

Best Performance

- 2nd Place:

20 Extra Credit Points!

94 a-a

Best Performance

- 1st

30 Extra Credit Points!

Calendar

- April 1: ULW First Draft Due
- April 8: Project 2: Planning Due
- April 8: Exam 3: Probability
- April 29: Project 3: Neural Networks Due
- April 29: ULW Final Draft Due
- May 9: Final Exam

Bayesian Diagnosis

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

Cavity

Toothache

Catch

Combining Evidence

$$\mathbf{P}(Cavity \mid toothache \land catch)$$
$$= \alpha \langle 0.180, 0.016 \rangle \approx \langle 0.871, 0.129 \rangle$$

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Exponential Growth of Combinations of Evidence

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

	$toothache$				$\neg toothache$			
	$catch$		$\neg catch$		$catch$		$\neg catch$	
	$bleeding$	$\neg bleeding$	$bleeding$	$\neg bleeding$	$bleeding$	$\neg bleeding$	$bleeding$	$\neg bleeding$
$cavity$	?	?	?	?	?	?	?	?
$\neg cavity$	?	?	?	?	?	?	?	?

Conditional Independence

- Both $toothache$ and $catch$ are caused by a cavity, but neither has a direct effect on the other
- The variables are independent given the presence or absence of a cavity
- Notation: $Toothache \perp\!\!\!\perp Catch \mid Cavity$

Benefit of Conditional Independence Assumptions

$$\mathbf{P}(Cavity \mid toothache \wedge catch) =$$

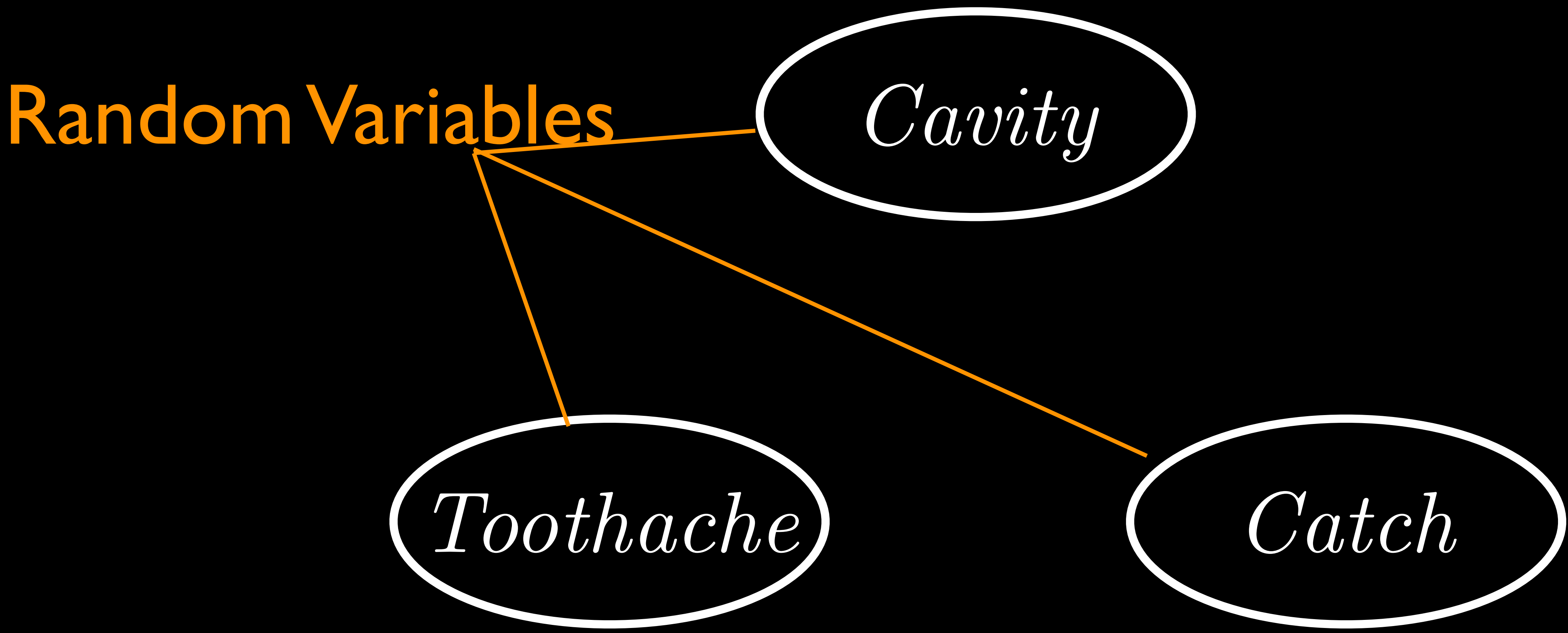
$$\alpha\, \mathbf{P}(toothache \mid Cavity)\mathbf{P}(catch \mid Cavity)\mathbf{P}(Cavity)$$

Only need these probabilities -
linear in the number of evidence variables!

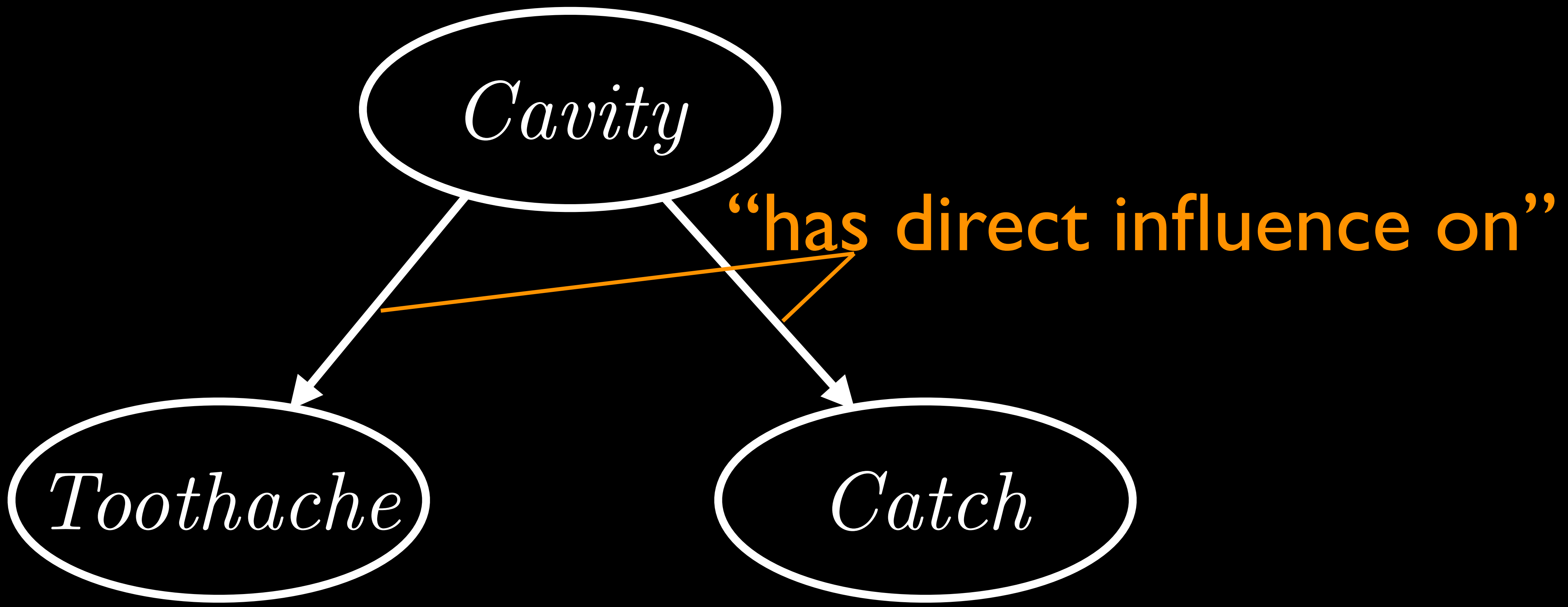
Bayesian Network

- Data structure for compactly representing a joint probability distributions
- Leverages (conditional) independencies between variables
- Can be exponentially smaller than explicit tabular representation of the joint distribution
- Supports many algorithms for inference and learning

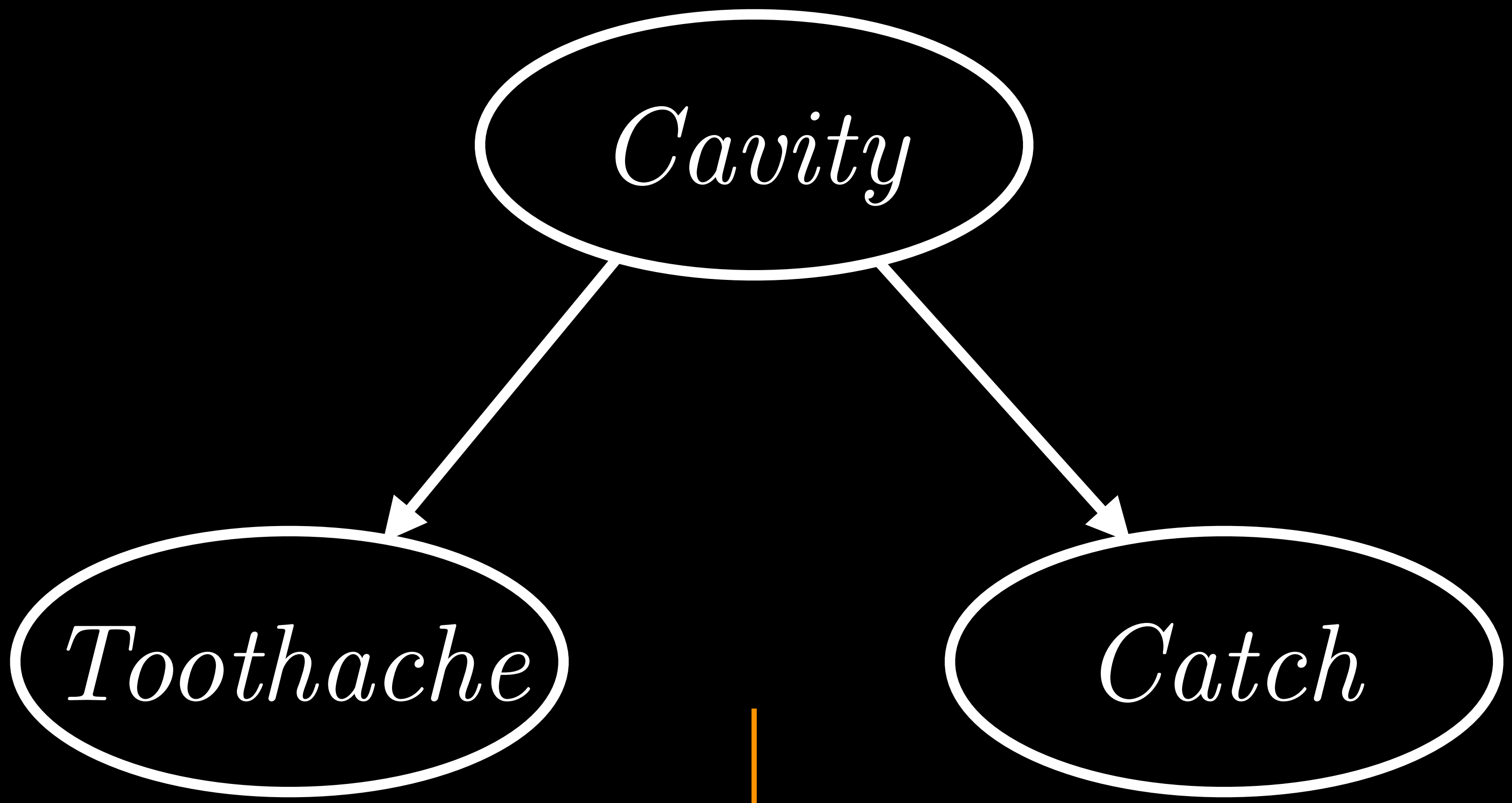
Bayesian Networks



Bayesian Networks

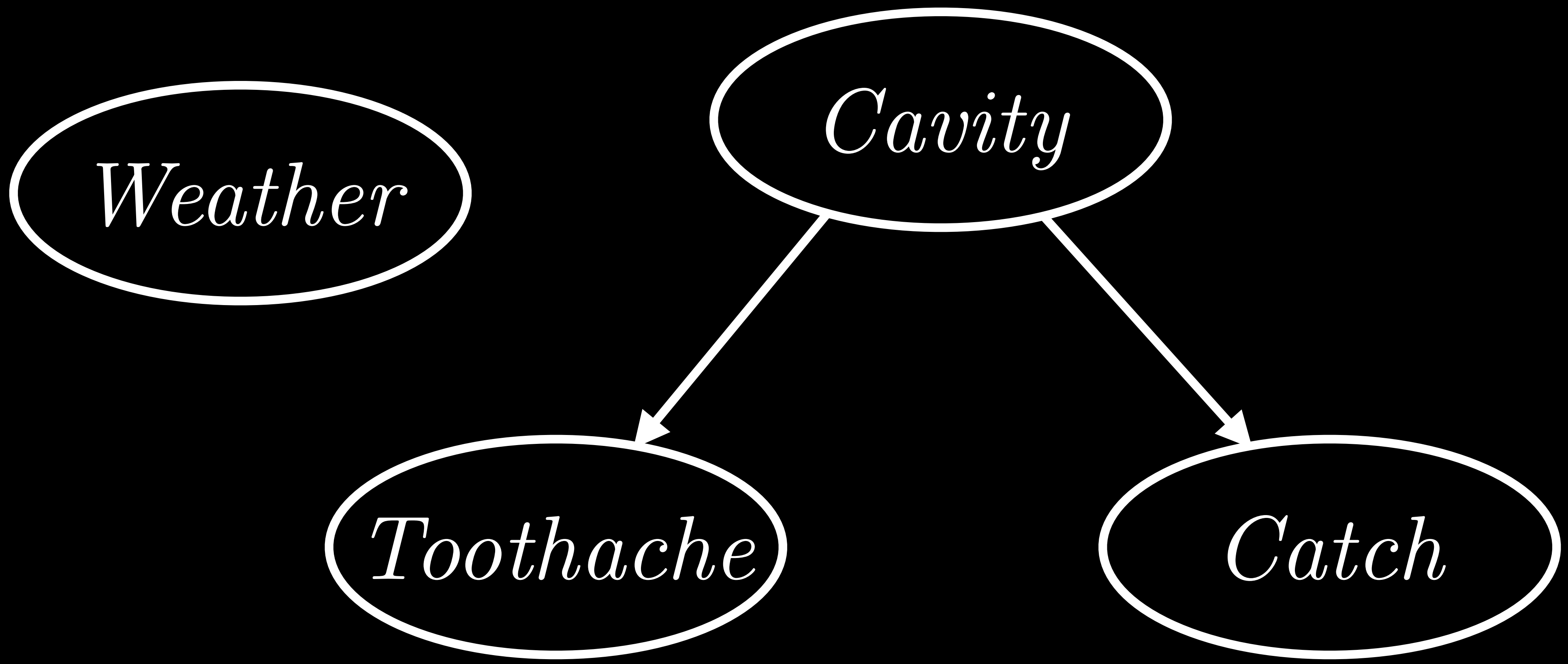


Bayesian Networks

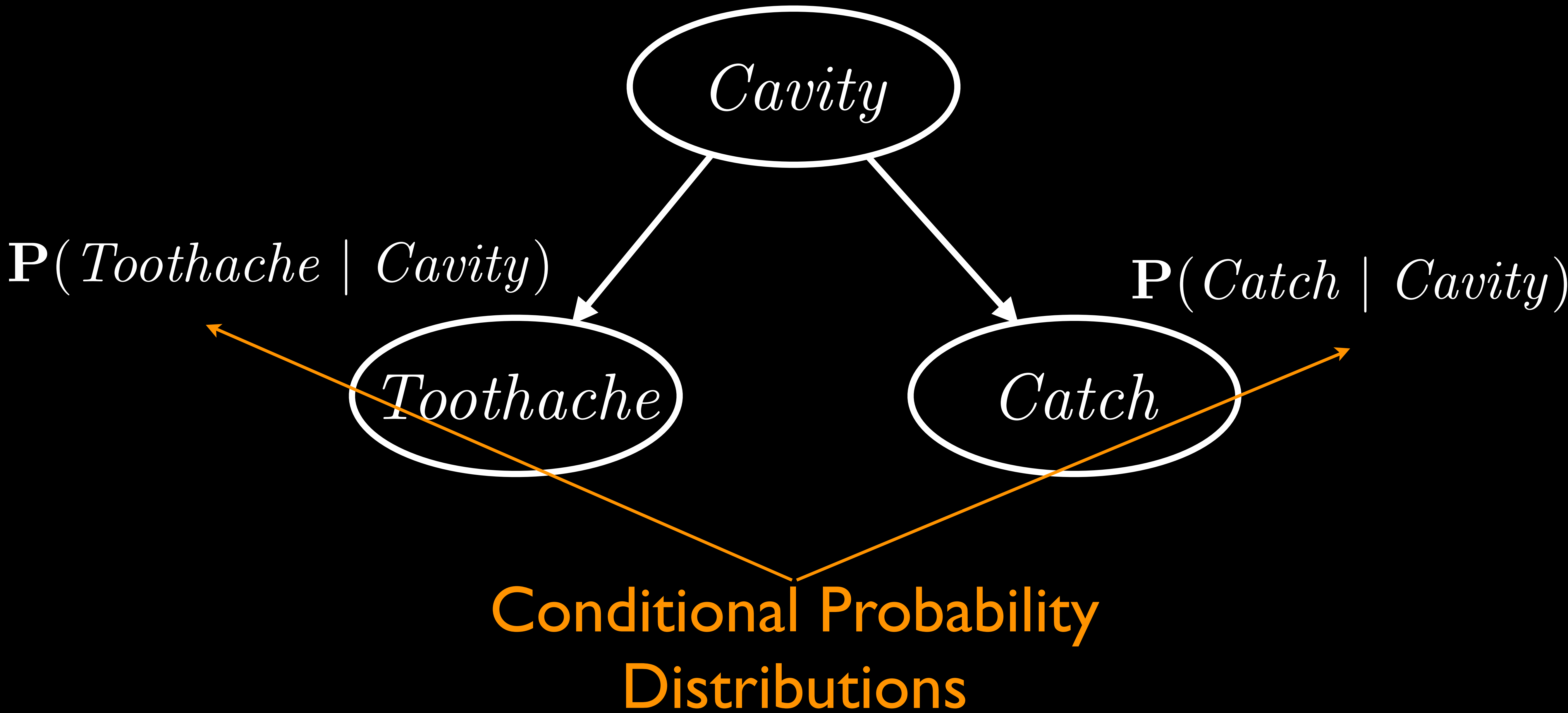


conditionally independent given parents

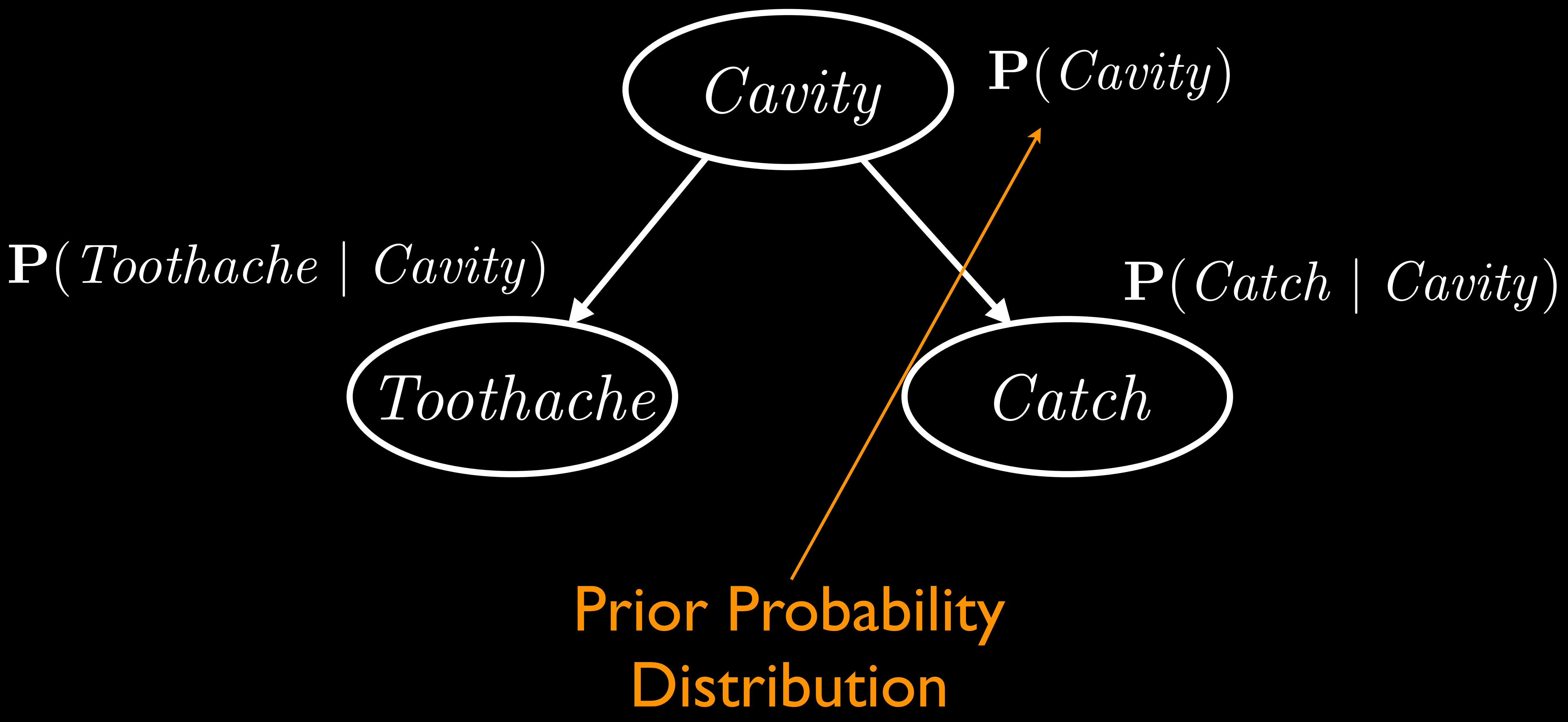
Bayesian Networks



Bayesian Networks



Bayesian Networks



Bayesian Networks

- Each node corresponds to a random variable
- There is a link from X to Y if X has a direct influence on Y (no cycles; DAG)
- The node for X_i stores the conditional distribution $\mathbf{P}(X_i \mid Parents(X_i))$
- Root nodes store the priors $\mathbf{P}(X_i)$

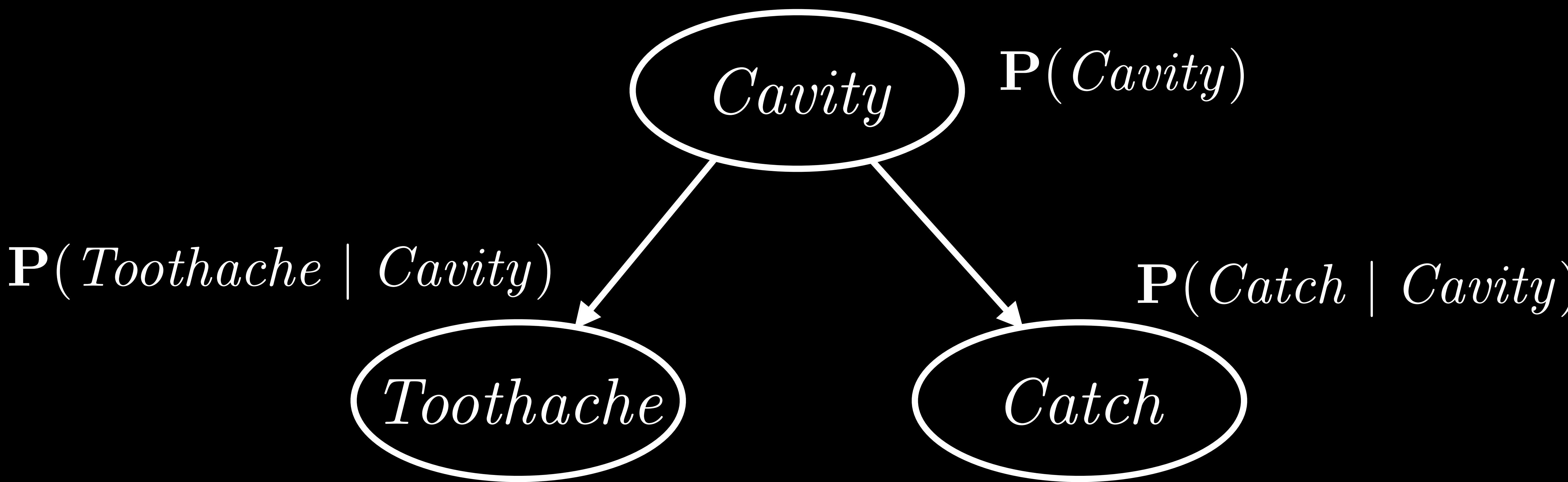
Bayesian Networks How-To

- Select random variables required to model the domain
- Add links from causes to effects
  - “Directly influences”
  - No cycles
- Write down (conditional) probability distributions for each node

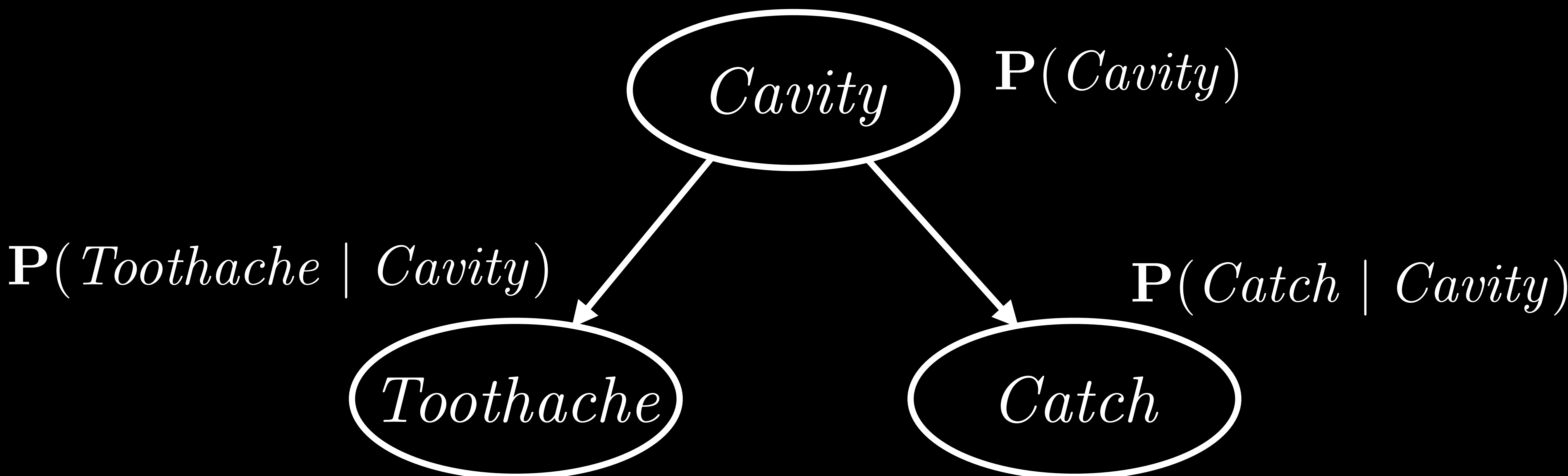
Semantics of Bayesian Networks

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

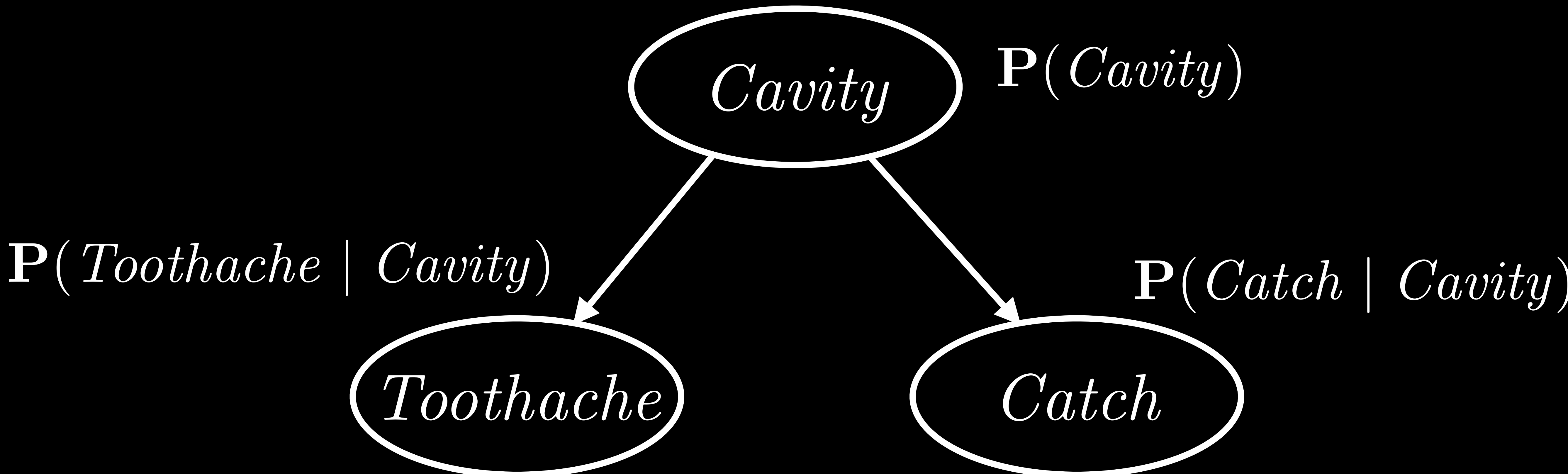
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$



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



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



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

Inference in Bayesian Networks

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

Inference in Bayesian Networks

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

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

Inference in Bayesian Networks

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

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

Inference in Bayesian Networks

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

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

Inference in Bayesian Networks

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

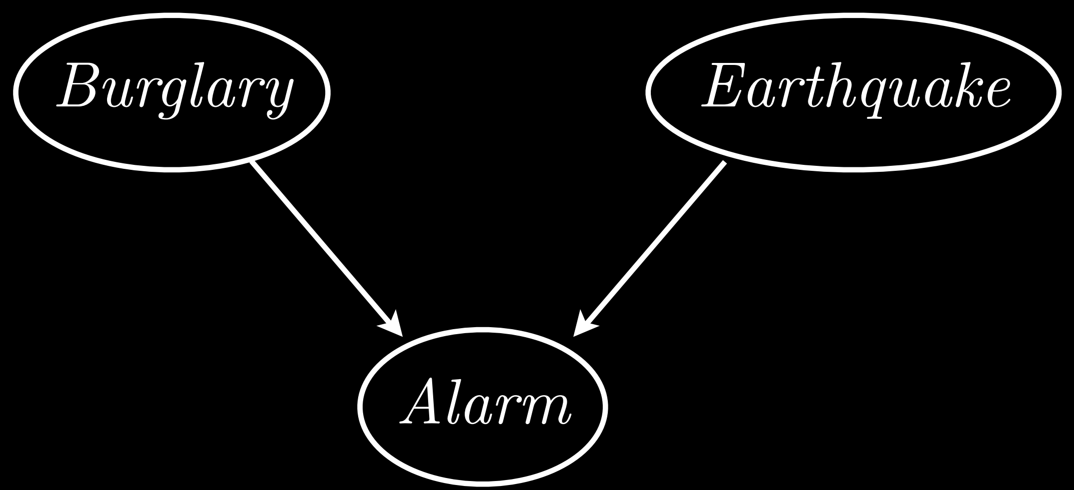
$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

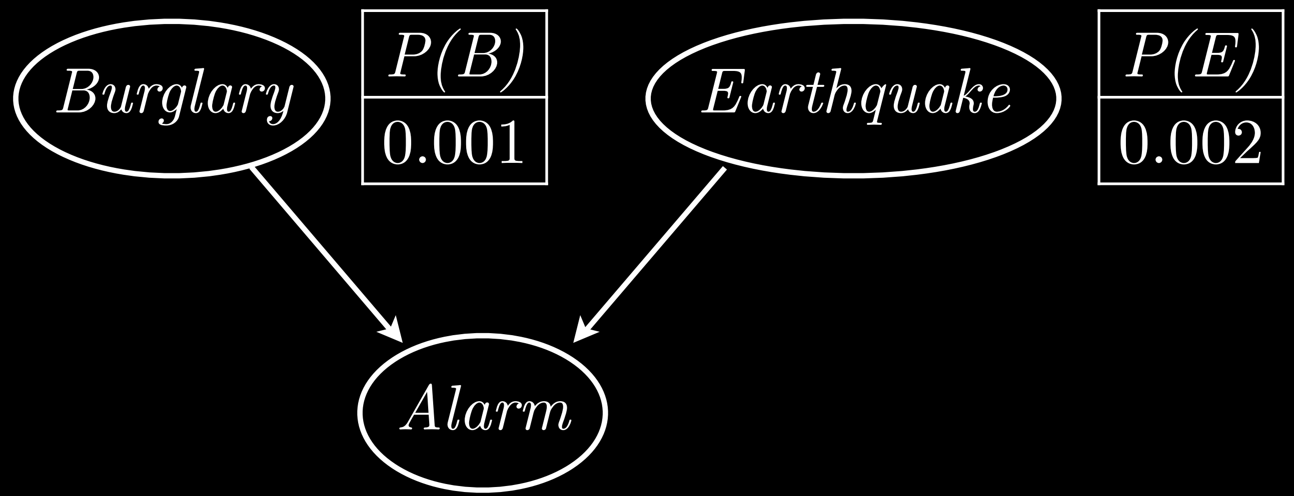
$$= \alpha \sum_{\mathbf{y}} \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$



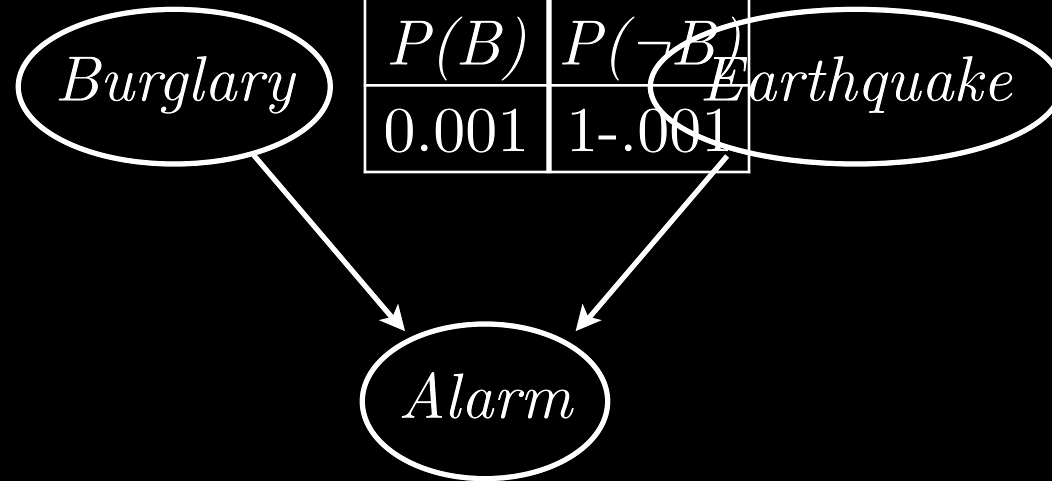

HOME
ALARM


Alarm


Burglary
Earthquake
Alarm

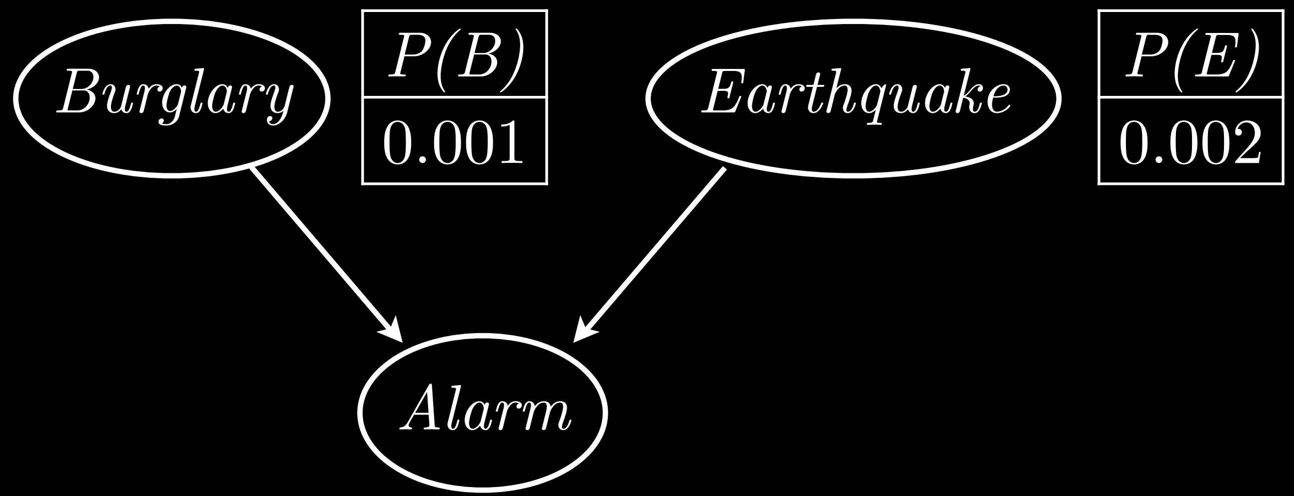


Burglary
P(B)
0.001
Earthquake
P(E)
0.002
Alarm

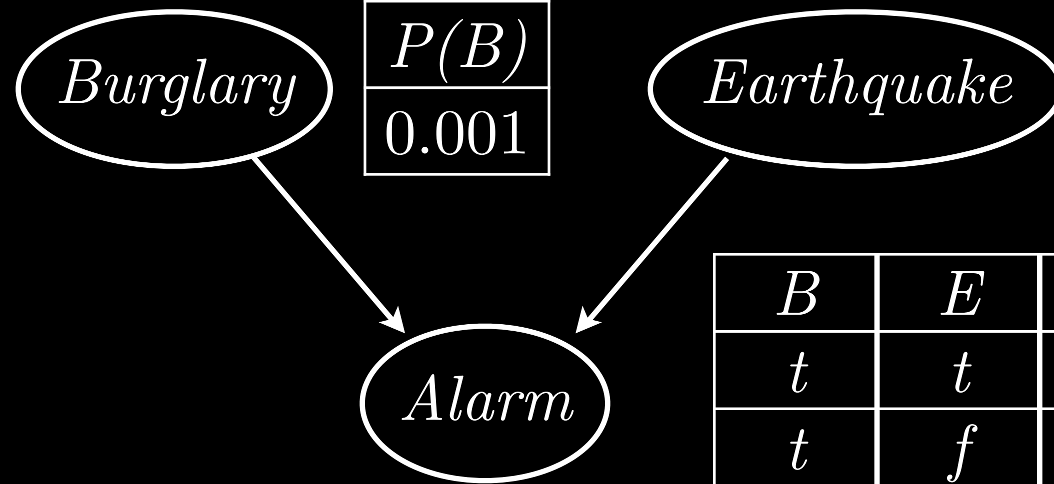


$P(B)$	$P(\neg B)$
0.001	1-.001

$P(E)$	$P(\neg E)$
0.002	1-.002



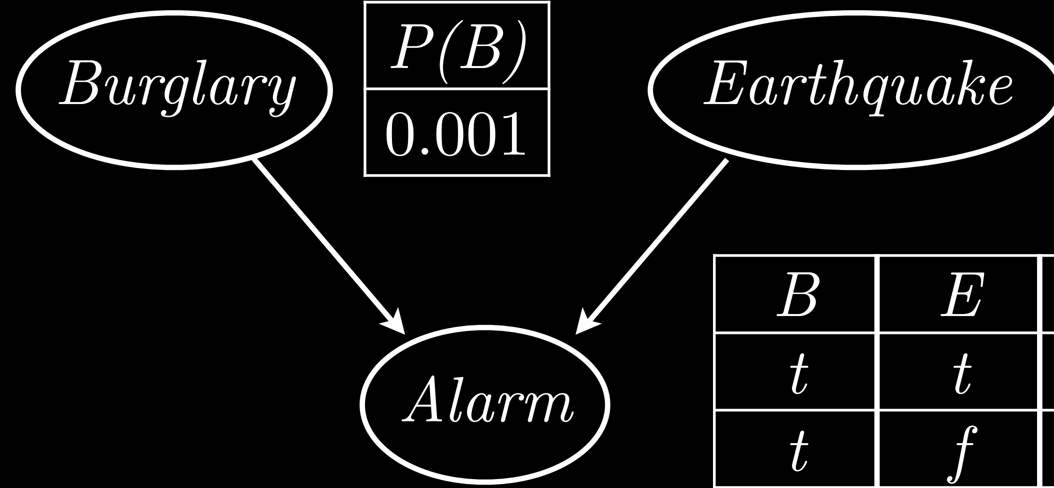
Burglary
P(B)
0.001
Earthquake
P(E)
0.002
Alarm



$P(B)$
0.001

$P(E)$
0.002

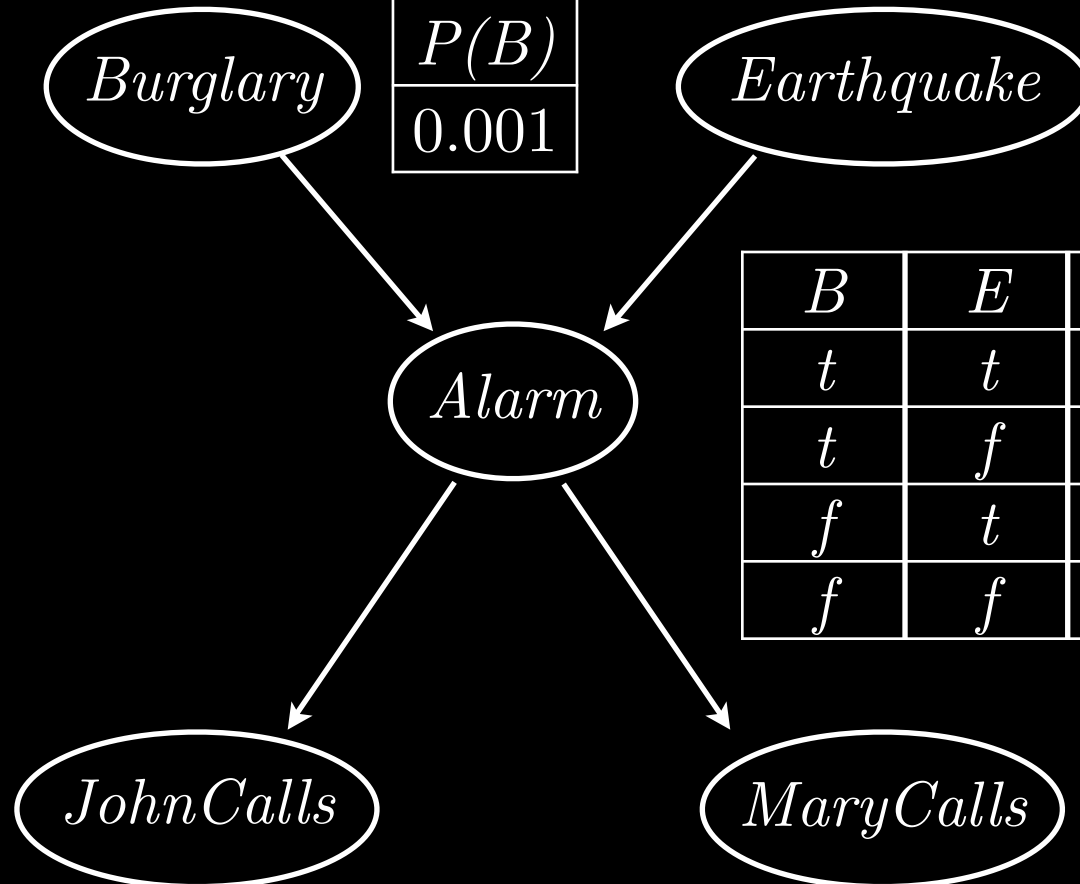
B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001



$P(B)$
0.001

$P(E)$
0.002

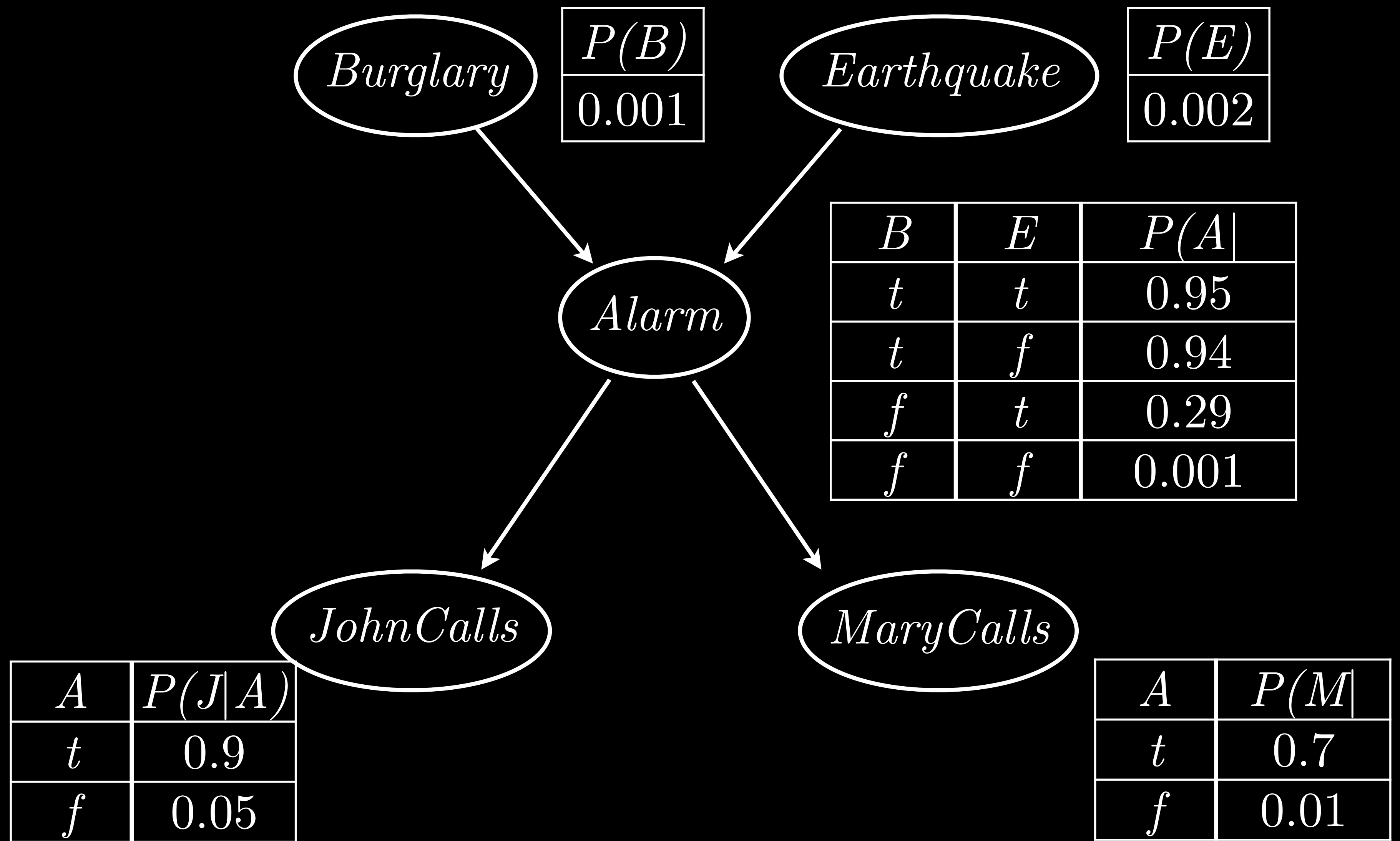
B	E	$P(A\|$	$P(\neg A\|$
t	t	0.95	1-.95
t	f	0.94	1-.94
f	t	0.29	1-.29
f	f	0.001	1-.001



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001



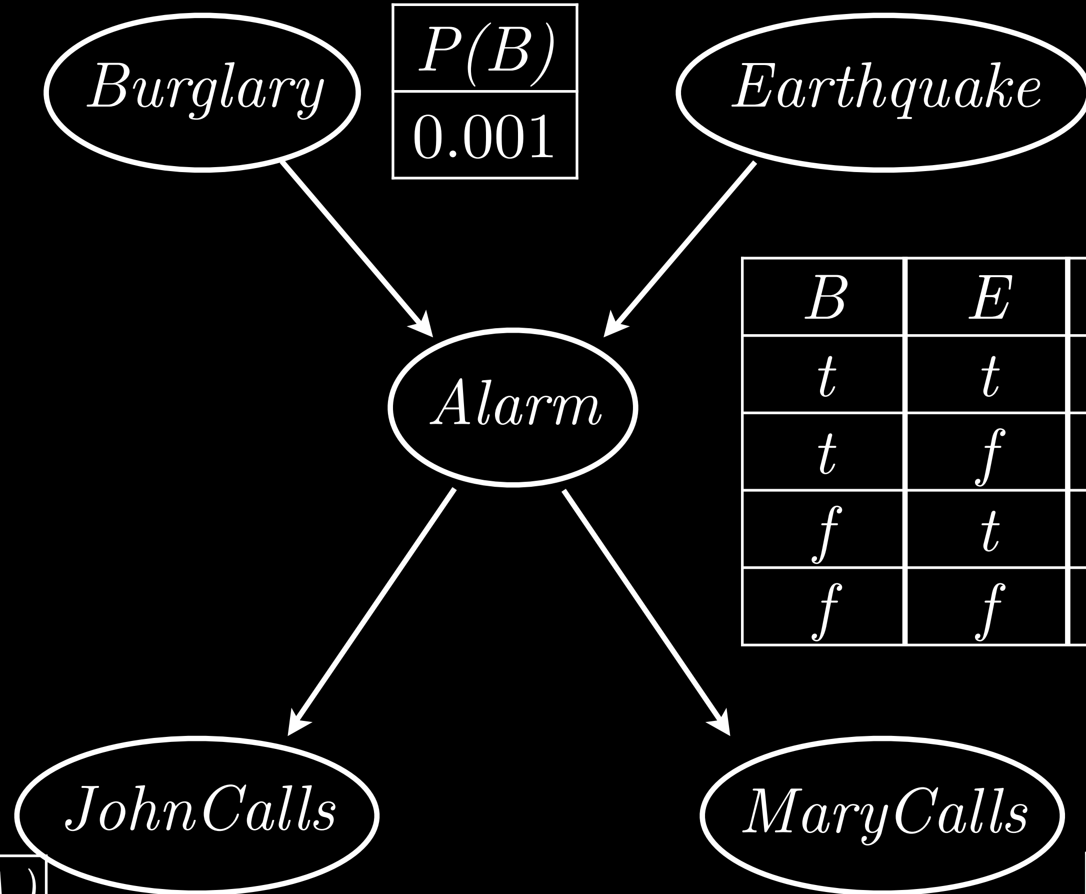
$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

A	$P(M\|$
t	0.7
f	0.01



$P(B)$
0.001

$P(E)$
0.002

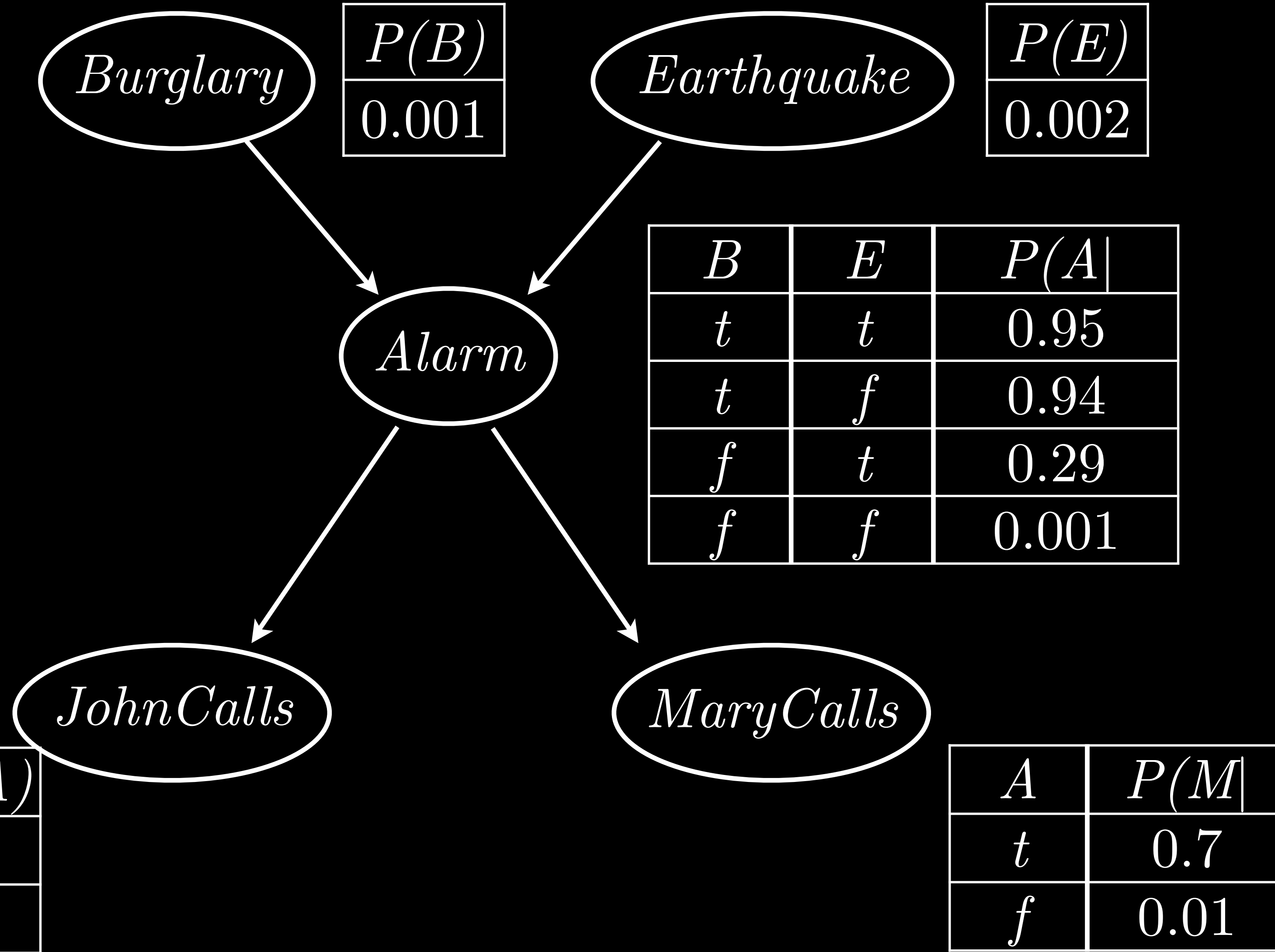
B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

A	$P(M\|$
t	0.7
f	0.01

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

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



$P(B)$
0.001

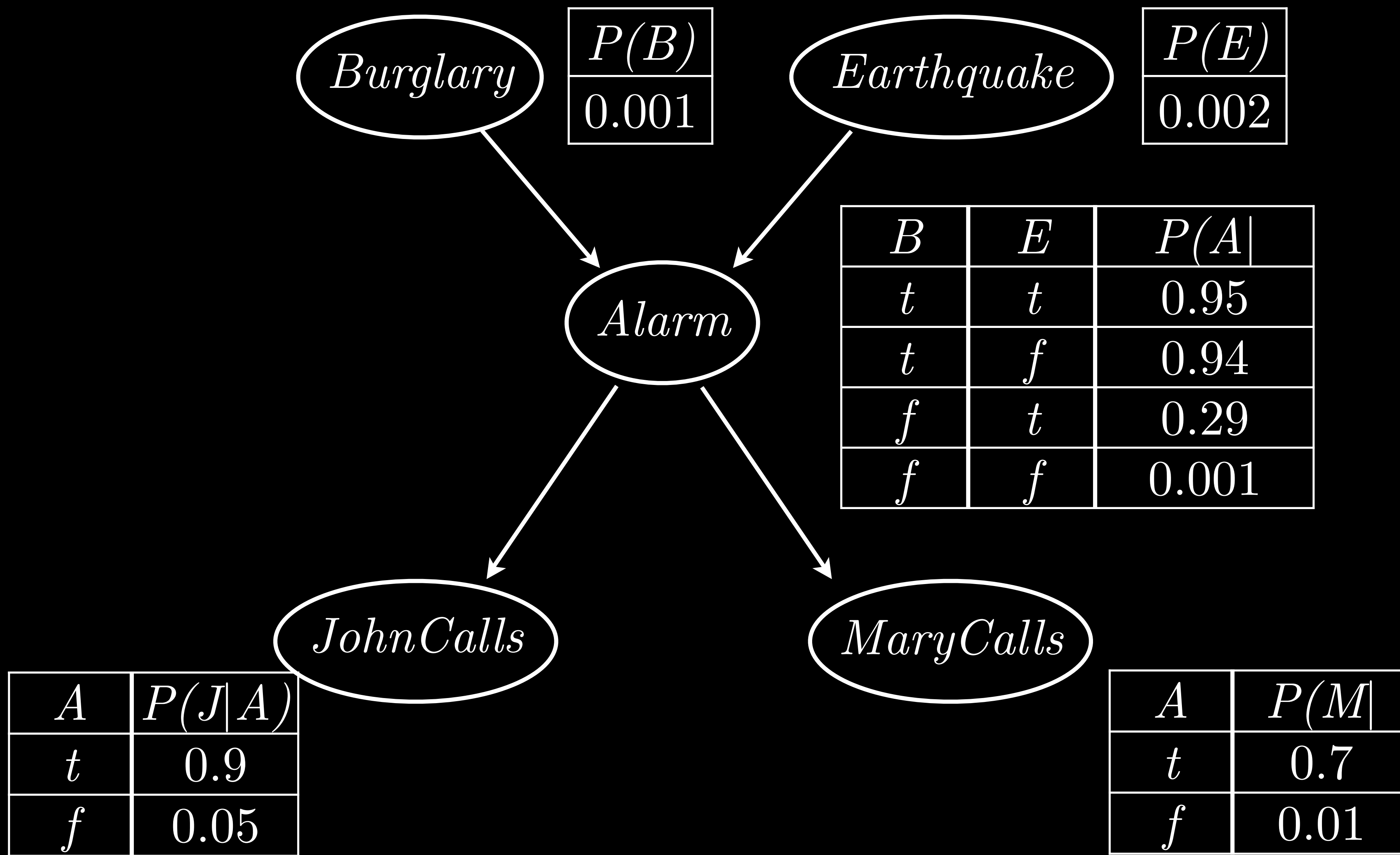
$P(E)$
0.002

B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

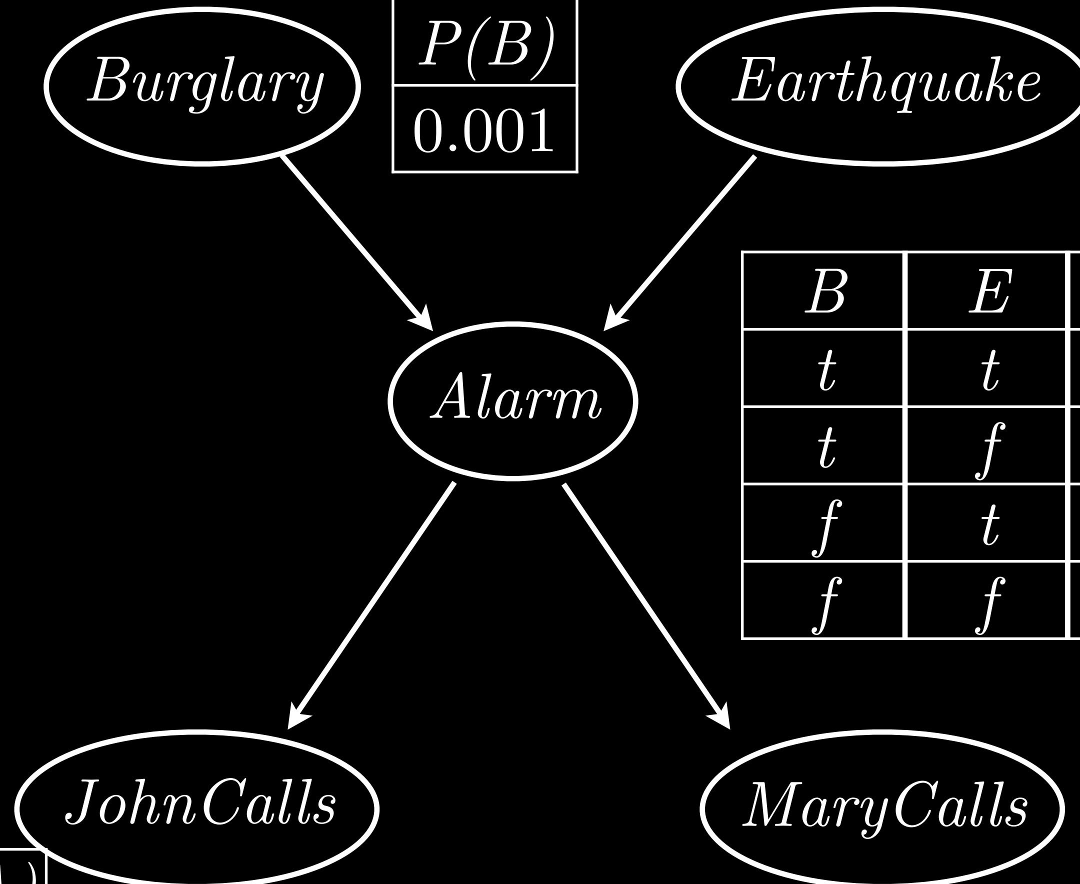
A	$P(M\|$
t	0.7
f	0.01

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



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

WHY?



$P(B)$
0.001

$P(E)$
0.002

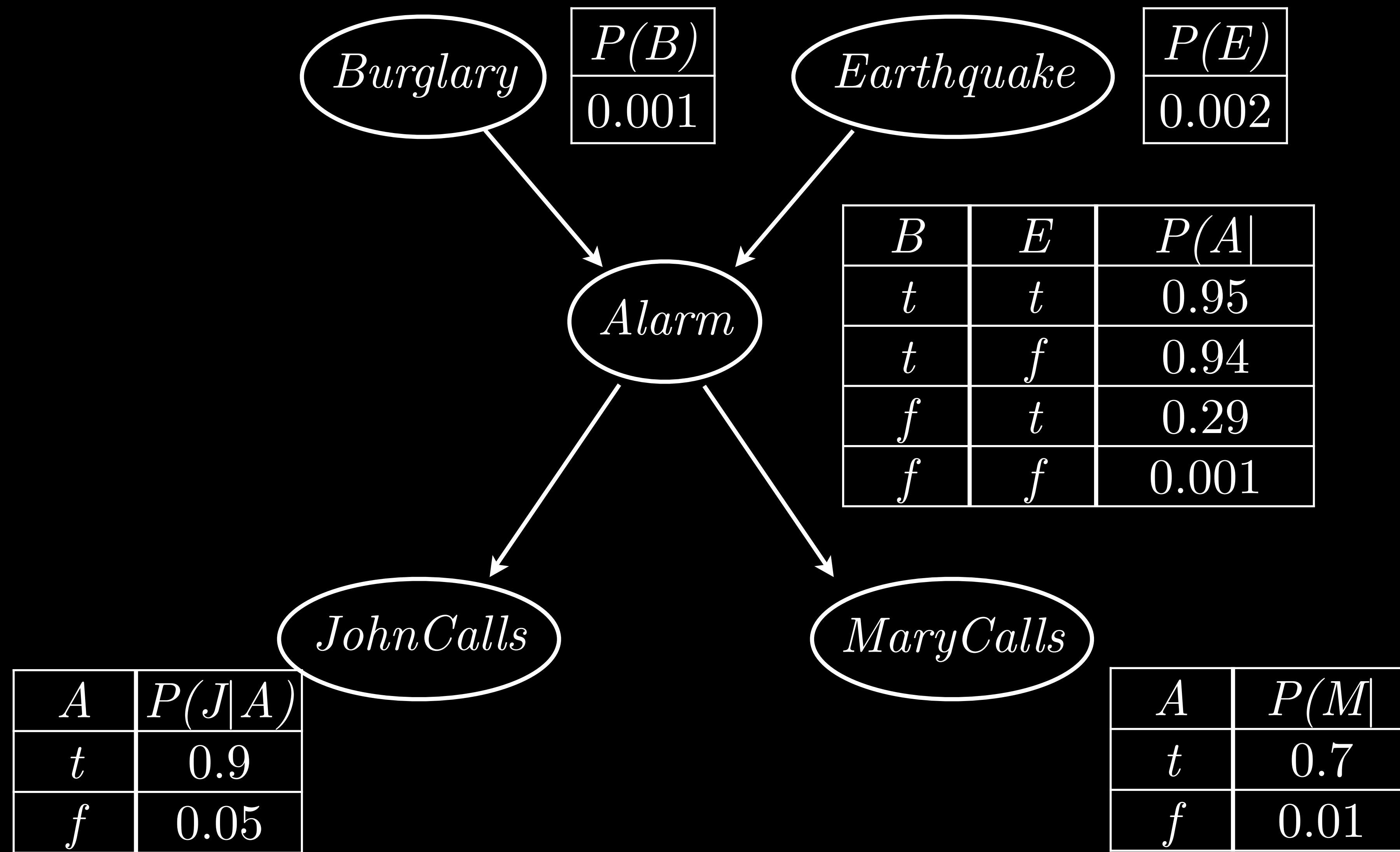
B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

A	$P(M\|$
t	0.7
f	0.01

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

Bayes Rule +
Normalization Trick!



$P(B)$
0.001

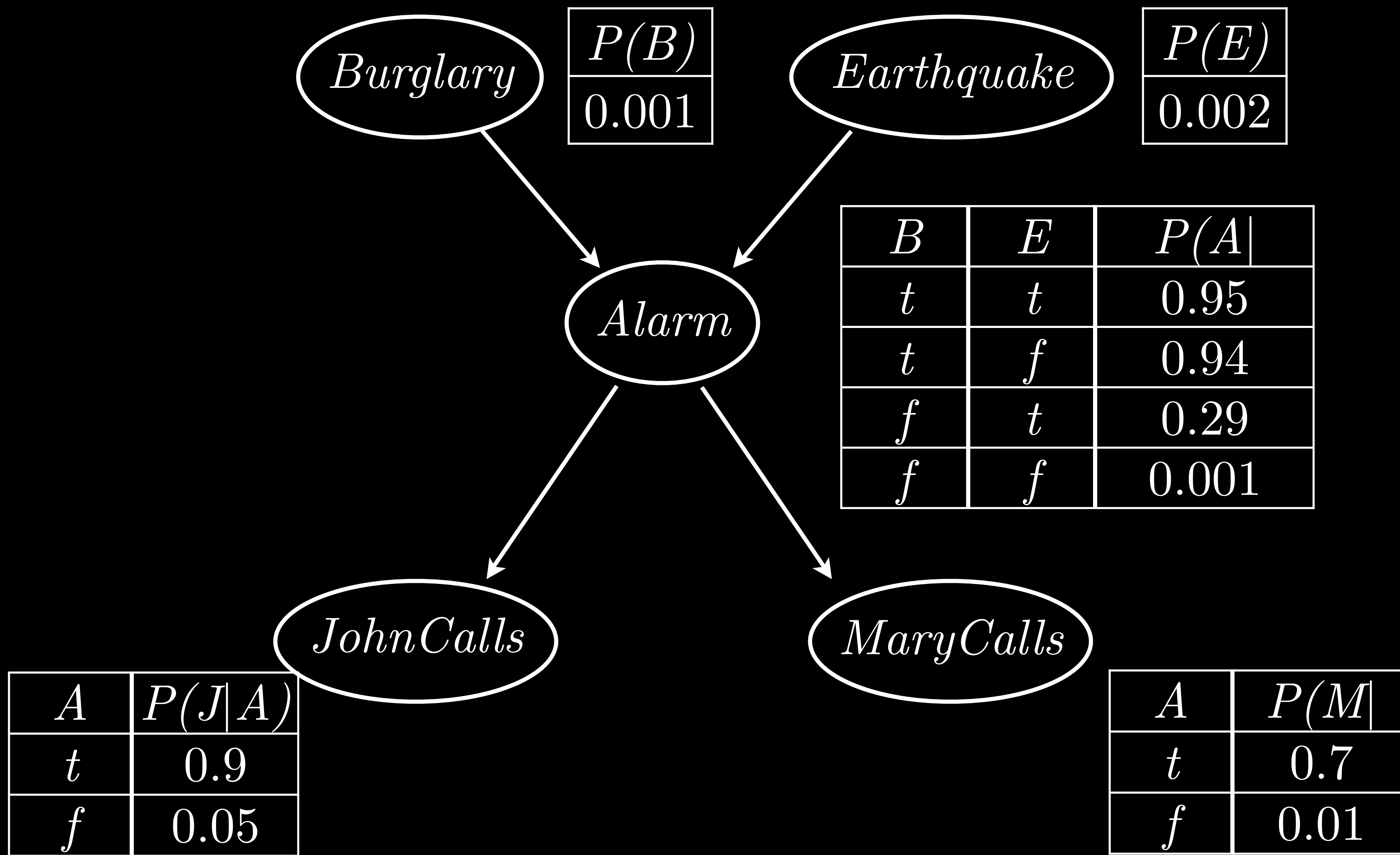
$P(E)$
0.002

B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

A	$P(M\|$
t	0.7
f	0.01

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



$P(B)$
0.001

$P(E)$
0.002

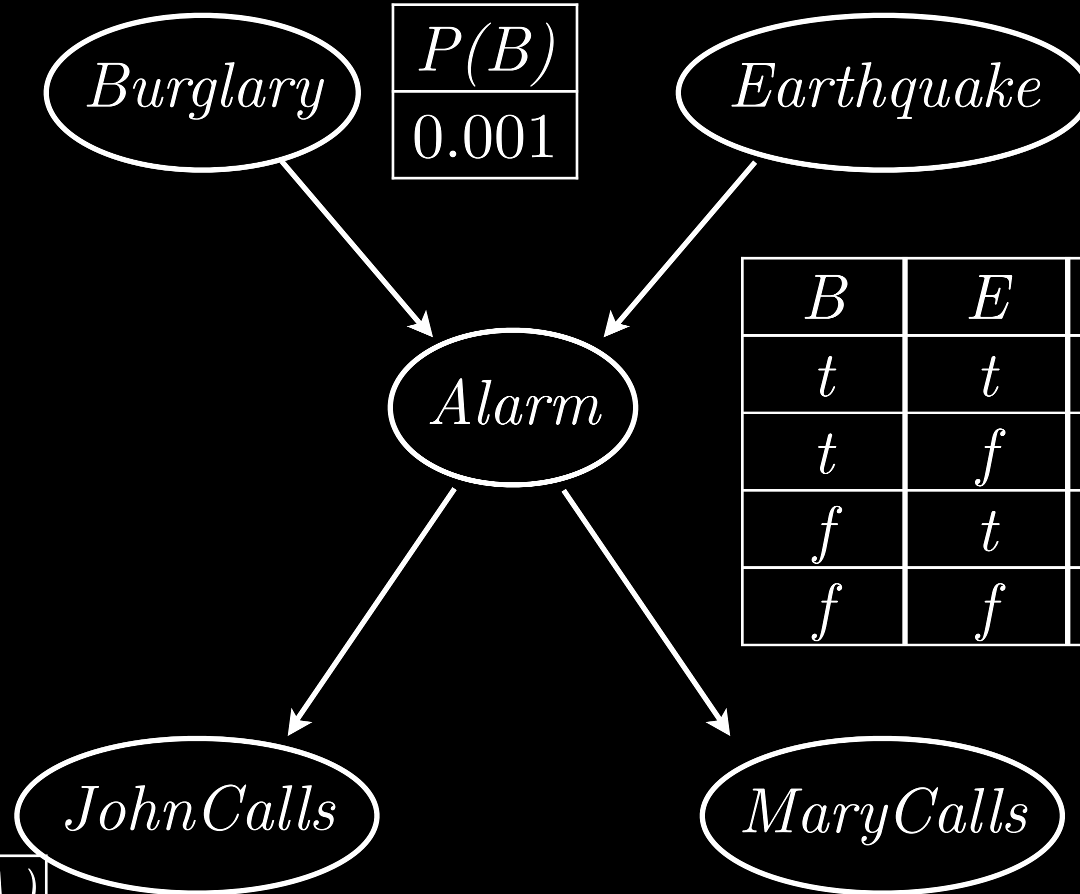
B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

A	$P(M\|$
t	0.7
f	0.01

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

WHY?



$P(B)$
0.001

$P(E)$
0.002

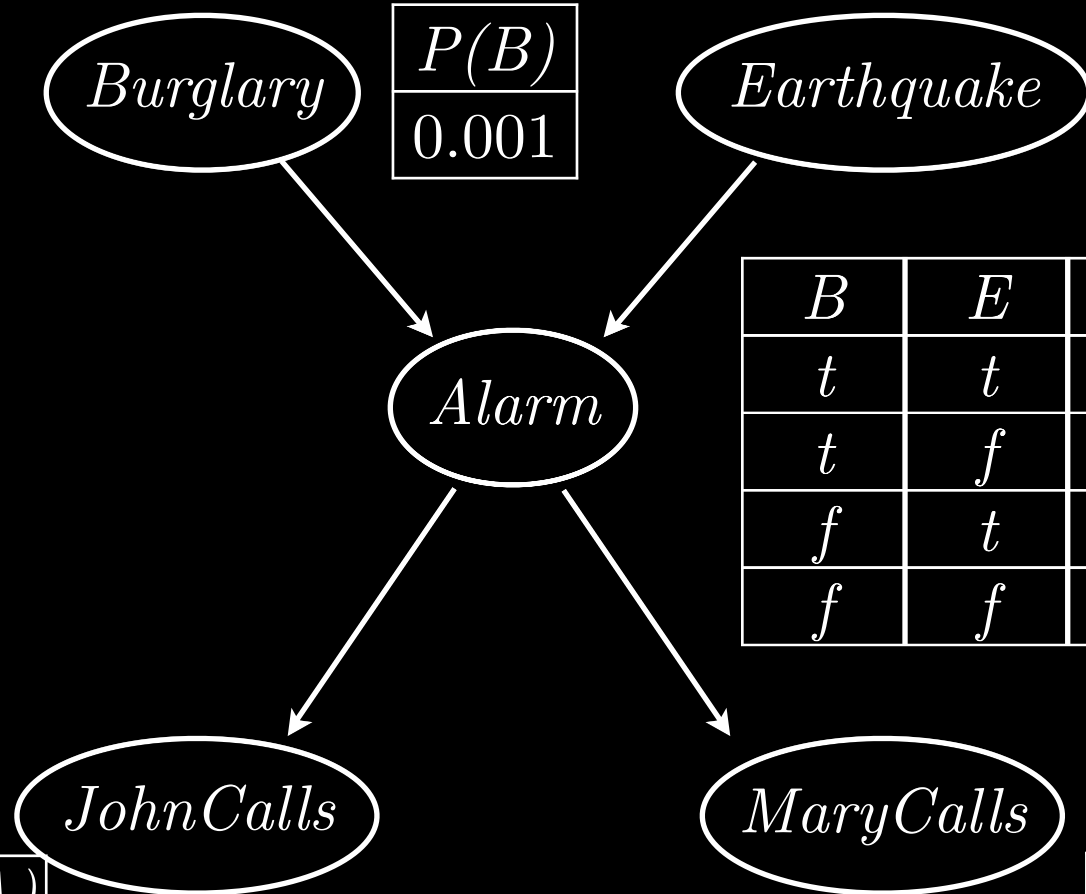
B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

A	$P(M\|$
t	0.7
f	0.01

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

Marginalizing Joint Distribution!



$P(B)$
0.001

$P(E)$
0.002

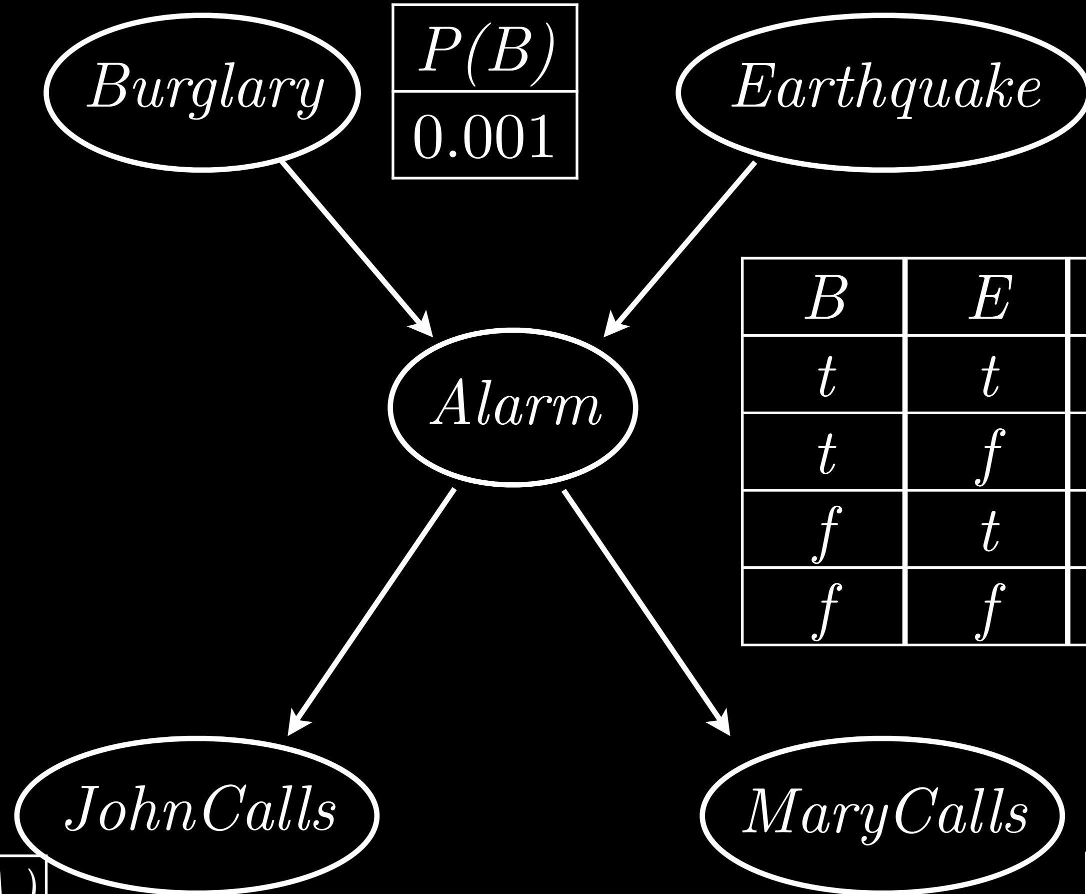
B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

A	$P(M\|$
t	0.7
f	0.01

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

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



$P(B)$
0.001

$P(E)$
0.002

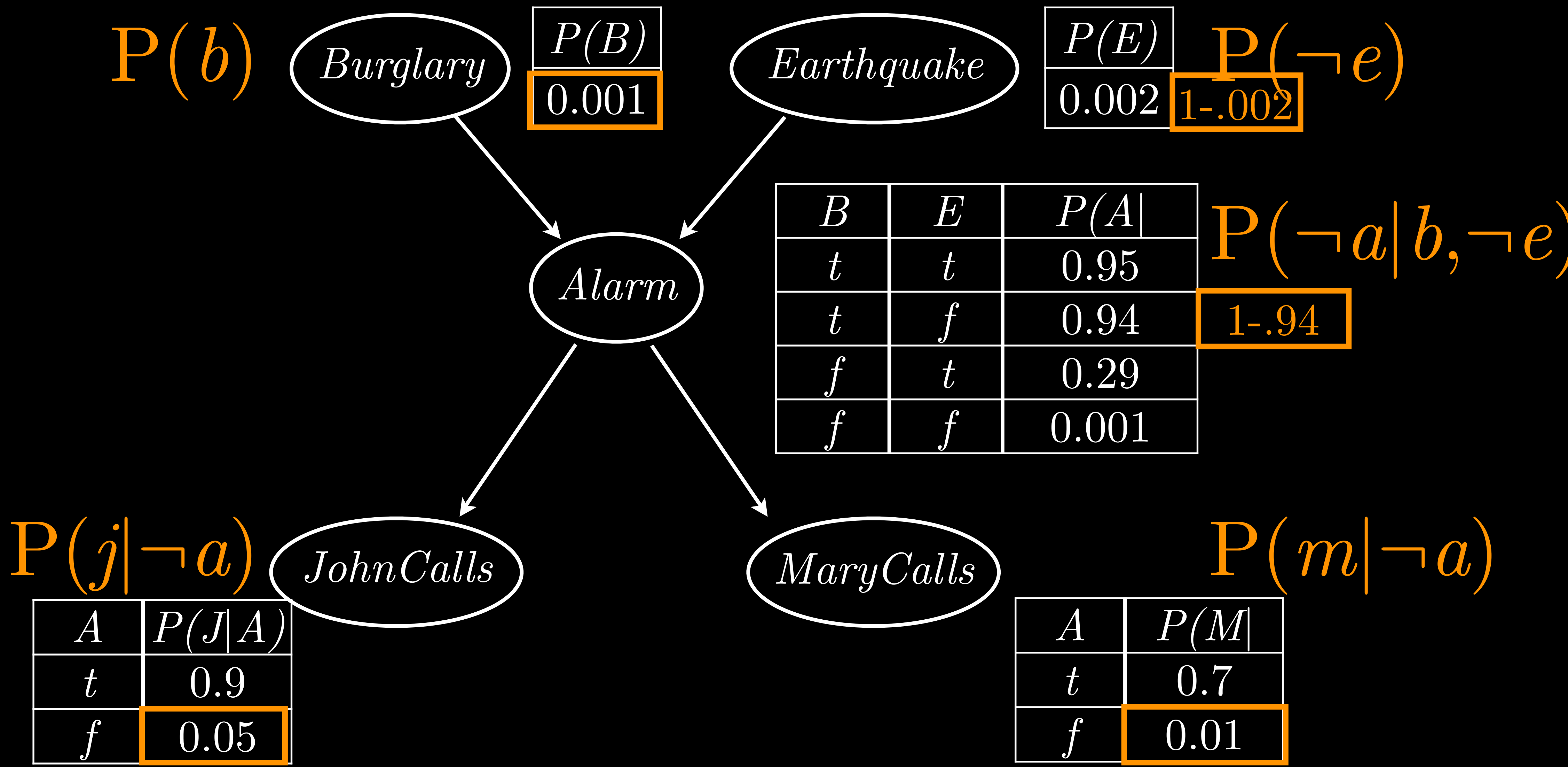
B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

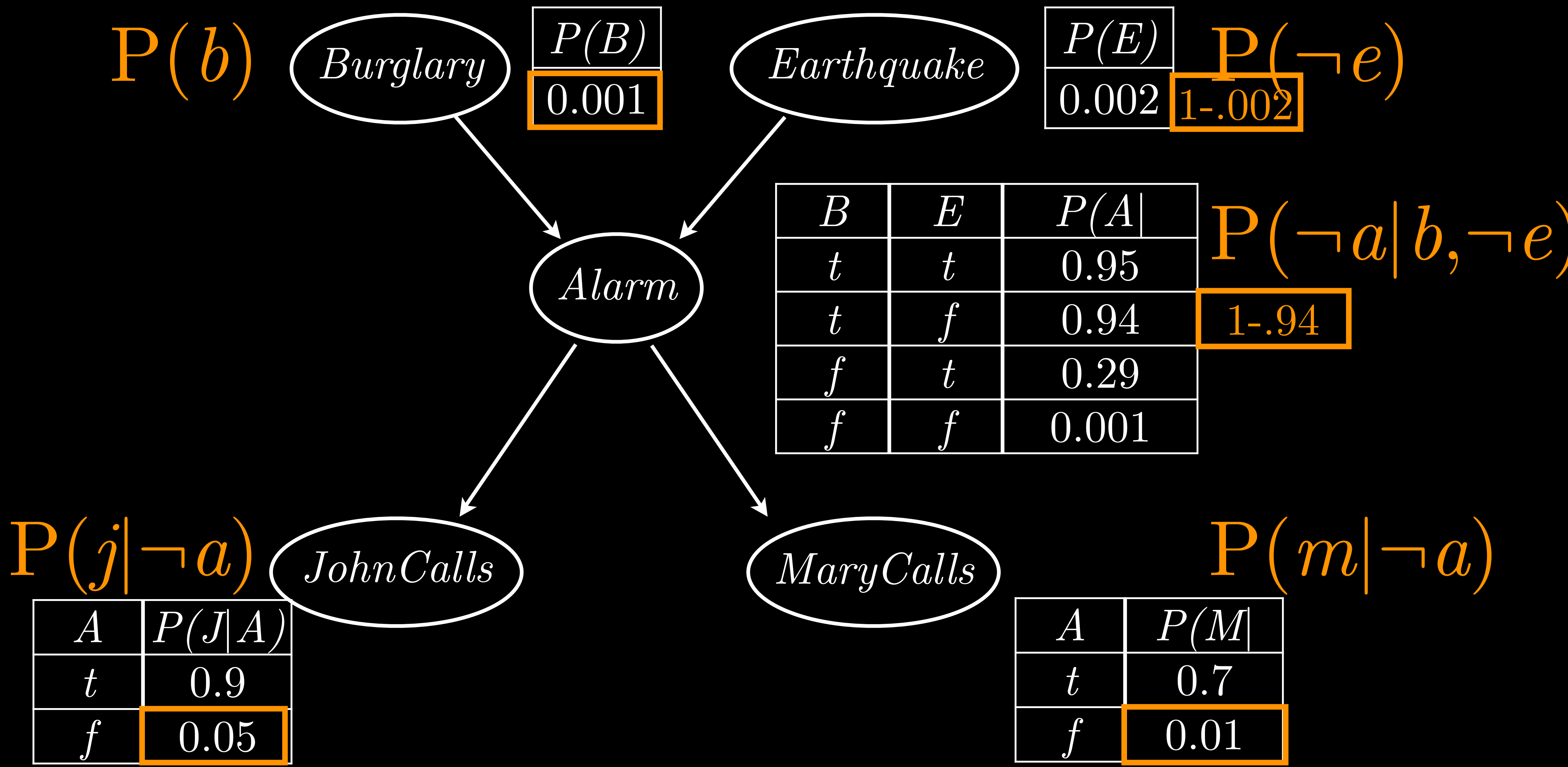
A	$P(M\|$
t	0.7
f	0.01

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

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



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



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A\|$
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	$P(J\|A)$
t	0.9
f	0.05

A	$P(M\|$
t	0.7
f	0.01

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

Optimizing Bayesian Network Inference

- It is often possible to optimize a query to a Bayesian Network
- Idea: rearrange terms, so that each is evaluated as few times as possible

Example: Optimizing Inference

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

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

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

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

Example: Optimizing Inference

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

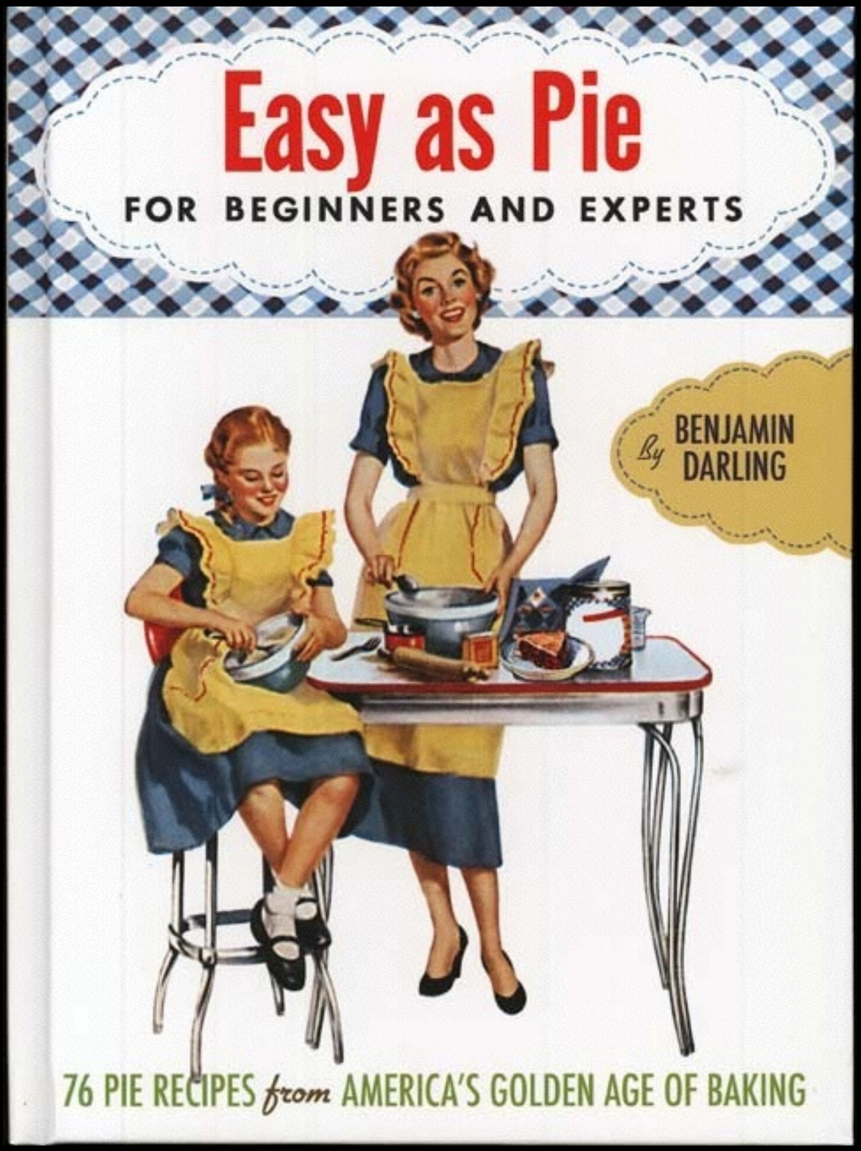
Before optimization: 2 x 2 x 5 = 20 multiplies

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

After optimization: 1 + 2 x 3 = 7 multiplies

Bayes Net Toolkits

- Many Bayesian Network tools are available
- Variety of built-in optimization routines
- Just input the network and let the system do the work!



Worst-Case Complexity

- Exact inference in Bayesian Networks can be shown to be as hard as computing the number of satisfying assignments of a propositional logic formula
- #P-complete (harder than NP-complete)

Next Questions

- How do we learn the (conditional) probabilities for a Bayesian Network from a set of data?
- How can be we do even faster approximate probabilistic inference?