CS 480

Introduction to Artificial Intelligence

November 3, 2022

Announcements / Reminders

- Follow Week 11 TO DO List
- Written Assignment #03 due on ~~Sunday (11/06/22)~~ Thursday (11/10) at 11:00 PM CST
- Programming Assignment #02 posted
- Quiz #03 posted
- Grading TA assignment:

  https://docs.google.com/spreadsheets/d/1ExS0bKnGt_fdf4LHa3YS1qRA7-Iq4xqXVjfSAPMaGVk/edit?usp=sharing
- UPDATED Final Exam date:
  - December 1st, 2022 (last week of classes!)
    - Ignore the date provided by the Registrar

Plan for Today

- **Quantifying and dealing with uncertainty**
- **Bayesian/Belief Networks**

Joint Probability

The probability of event A and event B occurring (or more than two events). It is the probability of the intersection of two or more events.

$$P(A \wedge B)$$

For any propositions $f_1, f_2, \ldots, f_n$:

$$P(f_1 \wedge f_2 \wedge \ldots \wedge f_n)$$

Chain Rule

Conditional probabilities can be used to decompose conjunctions using the chain rule. For any propositions $f_1, f_2, \ldots, f_n$:

$$P(f_1 \land f_2 \land \ldots \land f_n) =$$

$$P(f_1) *$$

$$P(f_2 \mid f_1) *$$

$$P(f_3 \mid f_1 \land f_2) *$$

$$\ldots$$

$$P(f_n \mid f_1 \land \ldots \land f_{n-1}) =$$

$$= \prod_{i=1}^{n} P(f_i \mid f_1 \land \ldots \land f_{i-1})$$

Chain Rule

Conditional probabilities can be used to decompose conjunctions using the chain rule. For any propositions $f_1, f_2, \ldots, f_n$:

$$P(f_1 = x_1 \wedge f_2 = x_2 \wedge \ldots \wedge f_n = x_n) =$$
$$P(f_1 = x_1) *$$
$$P(f_2 \mid f_1 = x_1) *$$
$$P(f_3 \mid f_1 = x_1 \wedge f_2 = x_2) *$$
$$\ldots$$
$$P(f_n = x_n \mid f_1 = x_1 \wedge \ldots \wedge f_{n-1} = x_{n-1}) =$$
$$= \prod_{i=1}^{n} P(f_i = x_i \mid f_1 = x_1 \wedge \ldots \wedge f_{i-1} = x_{i-1})$$

Bayes' Rule

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Random variables:

Toothache **- Boolean**

Cavity **- Boolean**

Catch **(dentist's probe catches tooth) - Boolean**

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Probability P(Cavity ∨ Toothache)**:**

$$P(Cavity\ = true \vee Toothache = true)\ =$$
$$= 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064$$
$$= 0.28$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Marginal probability P(Cavity)**:**

$$P(Cavity = \ true) = 0.108 + 0.012 + 0.072 + 0.008$$
$$= 0.2$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Conditional probability P(Cavity | Toothache):

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

$$= \frac{P(Cavity = true \wedge Toothache = true)}{P(Toothache = true)} =$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Conditional probability P(¬Cavity | Toothache)**:**

$$P(\neg Cavity = true \mid Toothache = true) =$$

$$= \frac{P(\neg Cavity = true \land Toothache = true)}{P(Toothache = true)} =$$

$$= \frac{0.016 + 0.164}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Note that:

$$P(Cavity \mid Toothache) = \frac{P(Cavity \wedge Toothache)}{P(Toothache)} = 0.6$$

$$P(\neg Cavity \mid Toothache) = \frac{P(\neg Cavity \wedge Toothache)}{P(Toothache)} = 0.4$$

add up to 1 and the same denominator is involved.

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Note that P() is the distribution, NOT individual probability:

$$\boldsymbol{P}(Cavity \mid Toothache) = \alpha * \boldsymbol{P}(Cavity, Toothache) =$$

$$= \alpha * [\boldsymbol{P}(Cavity, Toothache, Catch) + \boldsymbol{P}(Cavity, Toothache, \neg Catch)] =$$

$$= \alpha * [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] =$$

$$= \alpha * \langle 0.12, 0.08 \rangle =$$

$$= \langle 0.6, 0.4 \rangle$$

General Inference Procedure

Given:

- **a query involving a single variable X (in our example: Cavity),**
- **a <u>list</u> of evidence variables $\mathbf{E}$ (in our example: just Toothache),**
- **a <u>list</u> of observed values $\mathbf{e}$ for $\mathbf{E}$,**
- **a list of remaining unobserved variables $\mathbf{Y}$ (in our example: just Catch),**

where X, $\mathbf{E}$, and $\mathbf{Y}$ together are a COMPLETE set of variables for the domain, the probability $P(X \mid \mathbf{E})$ can be evaluated as:

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

where $\mathbf{y}$s are all possible values for $\mathbf{Y}$s, α - normalization constant.

$\mathbf{P}(X, \mathbf{e}, \mathbf{y})$ is a subset of probabilities from the joint distribution

Complex Joint Distributions

Consider a complex joint probability distribution involving N random variables P_1, P_2, P_3, …, P_{N-1}, Pp_N .

2^N Possible Worlds (Models)

N Random Variables						
P_1	P_2	P_3	…	P_{N-1}	P_N	Joint Probability
true	true	true	…	true	true	false
true	true	true	…	true	false	true
true	true	false	…	false	true	false
…	…	…	…	…	…	…
false	false	true	…	true	false	true
false	false	true	…	false	true	true
false	false	false	…	false	false	false

2^N values

Non-binary / Non-Boolean RVs

Some Random Variables are going to have more than two possible, discrete, values:

- height **-> short, average, tall**
- size **-> S, M, L, XL**
- streetLight **-> green, orange, red**
- vision **-> 20/20, 15/15, etc.**
- continent **-> Africa, Antarctica, Asia, Australia, Europe, North America, South America**

Non-binary RVs increase the complexity.

This May Be Impossible to Manage!

2^N Possible Worlds (Models)

N Random Variables						Joint Probability
P_1	P_2	P_3	...	P_{N-1}	P_N	
true	true	true	...	true	true	false
true	true	true	...	true	false	true
true	true	false	...	false	true	false
...	...	...	...	...	...	...
false	false	true	...	true	false	true
false	false	true	...	false	true	true
false	false	false	...	false	false	false

2^N values

Independent Variable

		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
¬Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576
		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

Let's introduce another random variable Cloudy **representing some weather conditions. It is difficult to imagine the other random variables here (**Toothache**,** Cavity**,** Catch**) being dependent on** Cloudy **and vice versa.**

Independent Variable

		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
¬Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576
		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

Let's try to calculate the following probability:

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

using the Product Rule:

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

Independent Variable

		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
¬Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576
		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

It's hard to imagine Cloudy influencing other variables, so:

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

and then:

$$P(Toothache, Catch, Cavity, Cloudy) =$$
$$= P(Cloudy) * P(Toothache, Catch, Cavity)$$

Independent Variable / Factoring

		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
¬Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

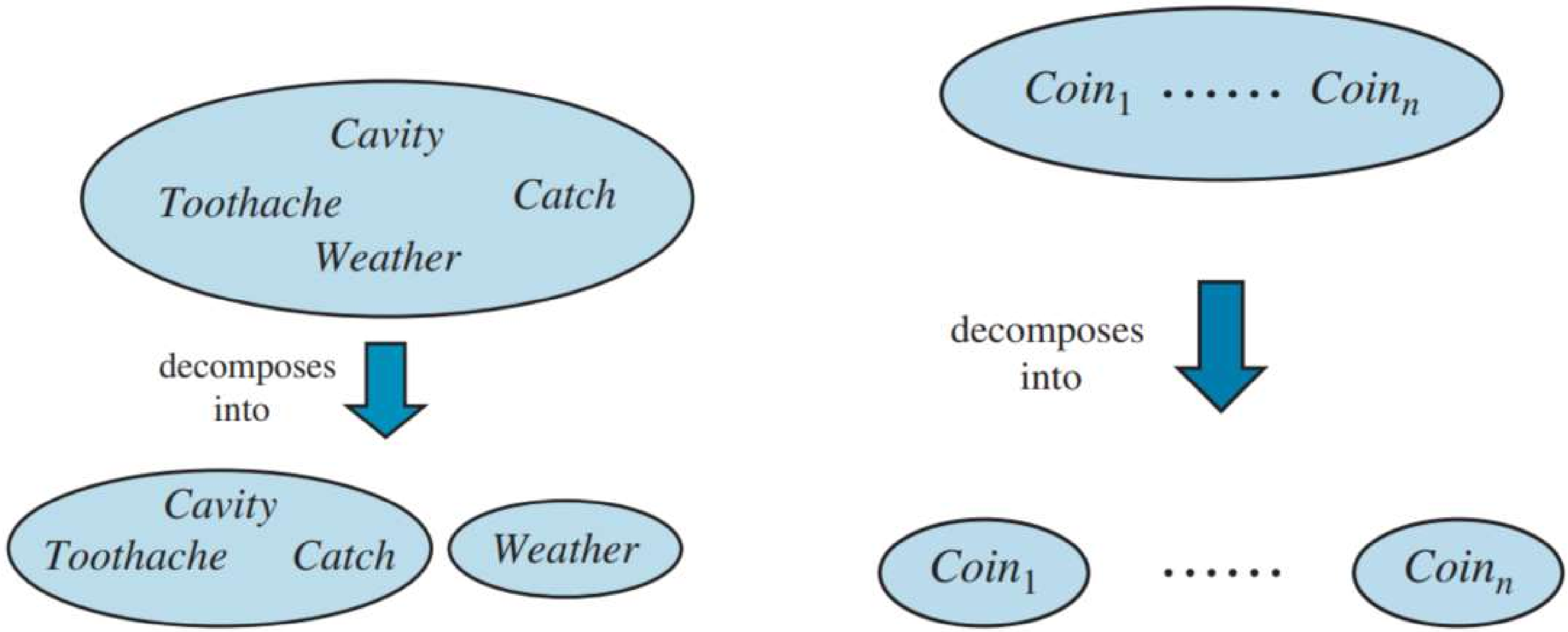
		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

It's hard to imagine Cloudy influencing other variables, so:

$$P(Toothache, Catch, Cavity, Cloudy) =$$
$$= P(Cloudy) * P(Toothache, Catch, Cavity)$$

This shows that Cloudy **is INDEPENDENT of other variables and factoring can be applied.**

Factoring / Decomposition



Use Chain Rule To Decompose

N Random Variables						Joint Probability
P_1	P_2	P_3	...	P_{N-1}	P_N	
true	true	true	...	true	true	false
true	true	true	...	true	false	true
true	true	false	...	false	true	false
...	...	...	...	...	...	...
false	false	true	...	true	false	true
false	false	true	...	false	true	true
false	false	false	...	false	false	false

Chain Rule

Conditional probabilities can be used to decompose conjunctions using the chain rule. For any propositions $f_1, f_2, \ldots, f_n$:

$$P(f_1 \wedge f_2 \wedge \ldots \wedge f_n) = \prod_{i=1}^{n} P(f_i \mid f_1 \wedge \ldots \wedge f_{i-1})$$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(grad = true \wedge female = true) = P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) \approx 0.074$
true	false	$P(grad = true \wedge female = false) = P(H, \neg e) = P(H) * P(\neg e \mid H) \approx 0.148$
false	true	$P(grad = false \wedge female = true) = P(\neg H, e) = P(\neg H) * P(e \mid \neg H) \approx 0.086$
false	false	$P(grad = false \wedge female = false) = P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) \approx 0.691$
		SUM = 1

Joint probabilities calculated using the Chain Rule:

$$P(f_1 \wedge f_2) = \prod_{i=1}^{2} P(f_i \mid f_1 \wedge \ldots \wedge f_{i-1})$$

$$P(f_1 \wedge f_2) = P(f_1) * P(f_2 \mid f_1)$$

so: $P(grad \wedge female) = P(H \wedge e) = P(H) * P(e \mid H)$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18 / 81 * 12 / 18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63 / 81 * 56 / 63 \approx 0.691$
		SUM = 1

Joint probabilities calculated using the Chain Rule:

$$P(f_1 \wedge f_2) = \prod_{i=1}^{2} P(f_i \mid f_1 \wedge \ldots \wedge f_{i-1})$$

$$P(f_1 \wedge f_2) = P(f_1) * P(f_2 \mid f_1)$$

so: $P(grad \wedge female) = P(H \wedge e) = P(H) * P(e \mid H)$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18 / 81 * 12 / 18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63 / 81 * 56 / 63 \approx 0.691$
		SUM = 1

Joint probabilities calculated using the Chain Rule:

$$P(f_1 \wedge f_2) = \prod_{i=1}^{2} P(f_i \mid parents(f_i))$$

$$P(f_1 \wedge f_2) = P(f_1) * P(f_2 \mid parents(f_i))$$

so: $P(H \wedge e) = P(H) * P(e \mid parents(e)) = P(H) * P(e \mid H)$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18 / 81 * 12 / 18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63 / 81 * 56 / 63 \approx 0.691$
		SUM = 1

$$P(H \wedge e) = P(H) * P(e \mid parents(e)) = \boxed{P(H)} * \boxed{P(e \mid H)}$$

H: grad	¬H: ¬grad
18 / 81 ≈ 0.22	63 / 81 ≈ 0.78

H: grad	e: female	P(e | H)
true	true	6 / 18 ≈ 0.333
true	false	12 / 18 ≈ 0.667
false	true	7 / 63 ≈ 0.111
false	false	56 / 63 ≈ 0.889

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18/81 * 6/18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18/81 * 12/18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63/81 * 7/63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63/81 * 56/63 \approx 0.691$
		SUM = 1

$$P(H \wedge e) = P(H) * P(e \mid parents(e)) = \boxed{P(H)} * \boxed{P(e \mid H)}$$

H: grad	¬H: ¬grad
18 / 81 ≈ 0.22	63 / 81 ≈ 0.78

H: grad	e: female	P(e | H)
true	true	6 / 18 ≈ 0.333
true	false	12 / 18 ≈ 0.667
false	true	7 / 63 ≈ 0.111
false	false	56 / 63 ≈ 0.889

Conditional Probability Table (CPT)

Bayesian (Belief) Network

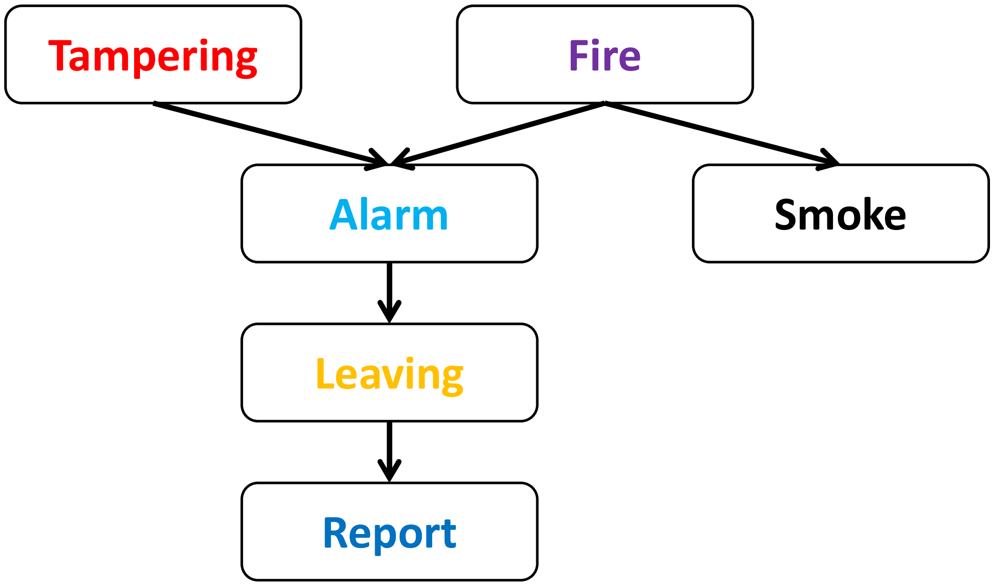
A Bayesian belief network describes the joint probability distribution for a set of variables.

A Bayesian network is an acyclic, directed graph (DAG), where the nodes are random variables (propositions). There is an edge (arc) from each elements of $\text{parents}(X_i)$ into X_i. Associated with the Bayesian network is a set of conditional probability distributions - the conditional probability of each variable given its parents (which includes the prior probabilities of those variables with no parents).

Consists of:

- **a graph (DAG) with nodes corresponding to random variables**
- **a domain for each random variable**
- **a set of conditional probability distributions $P(X_i \mid \text{parents}(X_i))$**

Bayesian (Belief) Network: Example



Random Variables (Propositions):

- Tampering: true if the alarm is tampered with
- Fire: true if there is a fire
- Alarm: true if the alarm sounds
- Smoke: true if there is smoke
- Leaving: true if people leaving the building at once
- Report: true if someone who left the building reports fire

Domain for all variables: $\{\text{true}, \text{false}\}$

NOTE: RVs don't have to be Boolean

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)	Conditional probabilities
true	true	$P(H \mid e)*P(e) \approx 0.074$	$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)} = \frac{6/18 * 18/81}{13/81} \approx 0.462$
true	false	$P(H \mid \neg e)*P(\neg e) \approx 0.148$	$P(H \mid \neg e) = \frac{P(\neg e \mid H) * P(H)}{P(\neg e)} = \frac{12/18 * 18/81}{68/81} \approx 0.176$
false	true	$P(\neg H \mid e)*P(e) \approx 0.086$	$P(\neg H \mid e) = \frac{P(e \mid \neg H) * P(\neg H)}{P(e)} = \frac{7/63 * 63/81}{13/81} \approx 0.538$
false	false	$P(\neg H \mid \neg e)*P(\neg e) \approx 0.691$	$P(\neg H \mid \neg e) = \frac{P(\neg e \mid \neg H) * P(\neg H)}{P(\neg e)} = \frac{56/63 * 63/81}{68/81} \approx 0.824$
		SUM = 1	

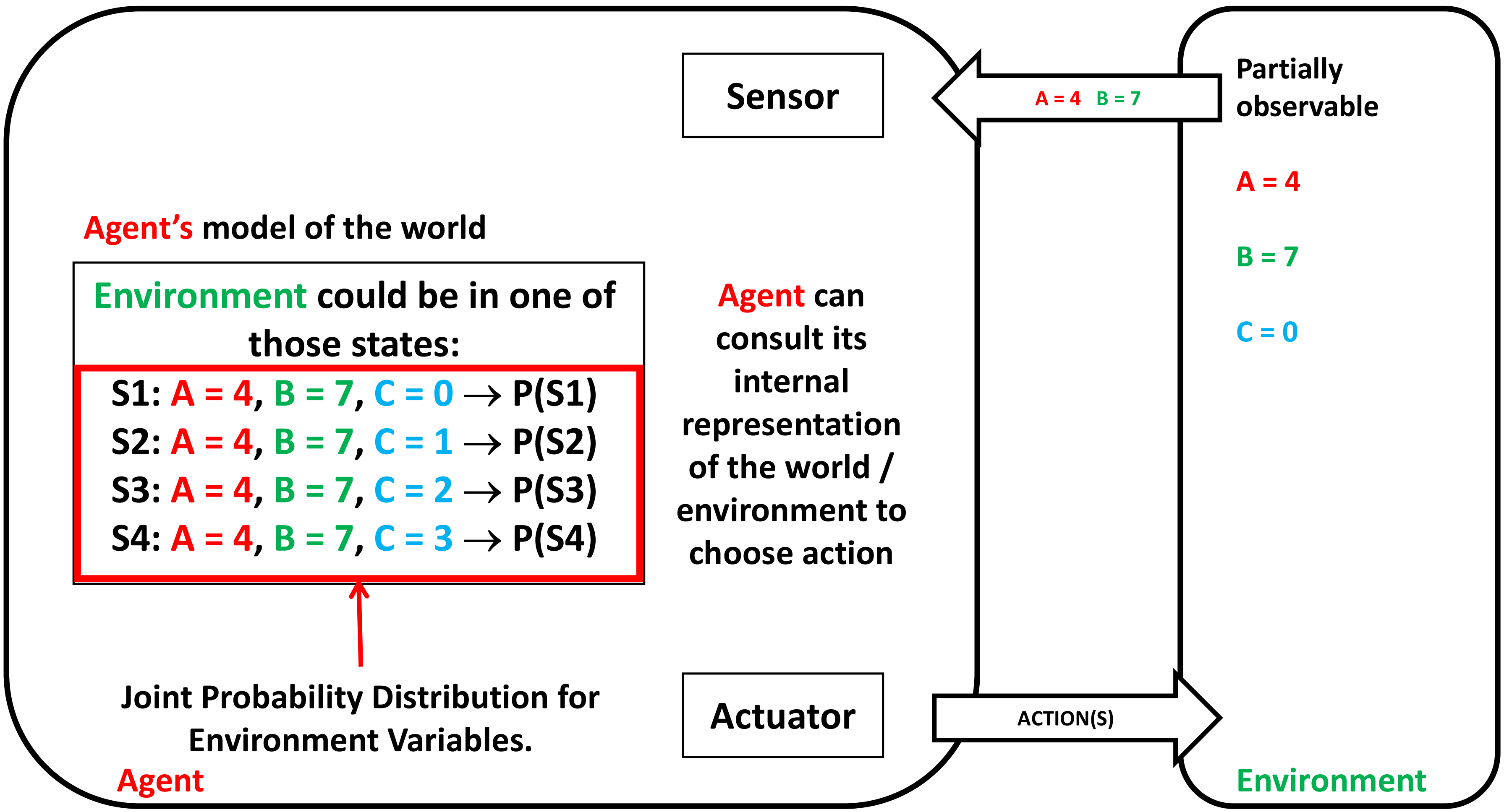
Joint probabilities calculated using the Product Rule:

$$P(A \wedge B) = P(A \mid B) * P(B)$$

Conditional probabilities calculated using Bayes' Rule:

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Agents and Belief State



Assume: $D_C = \{0,1,2,3\}$

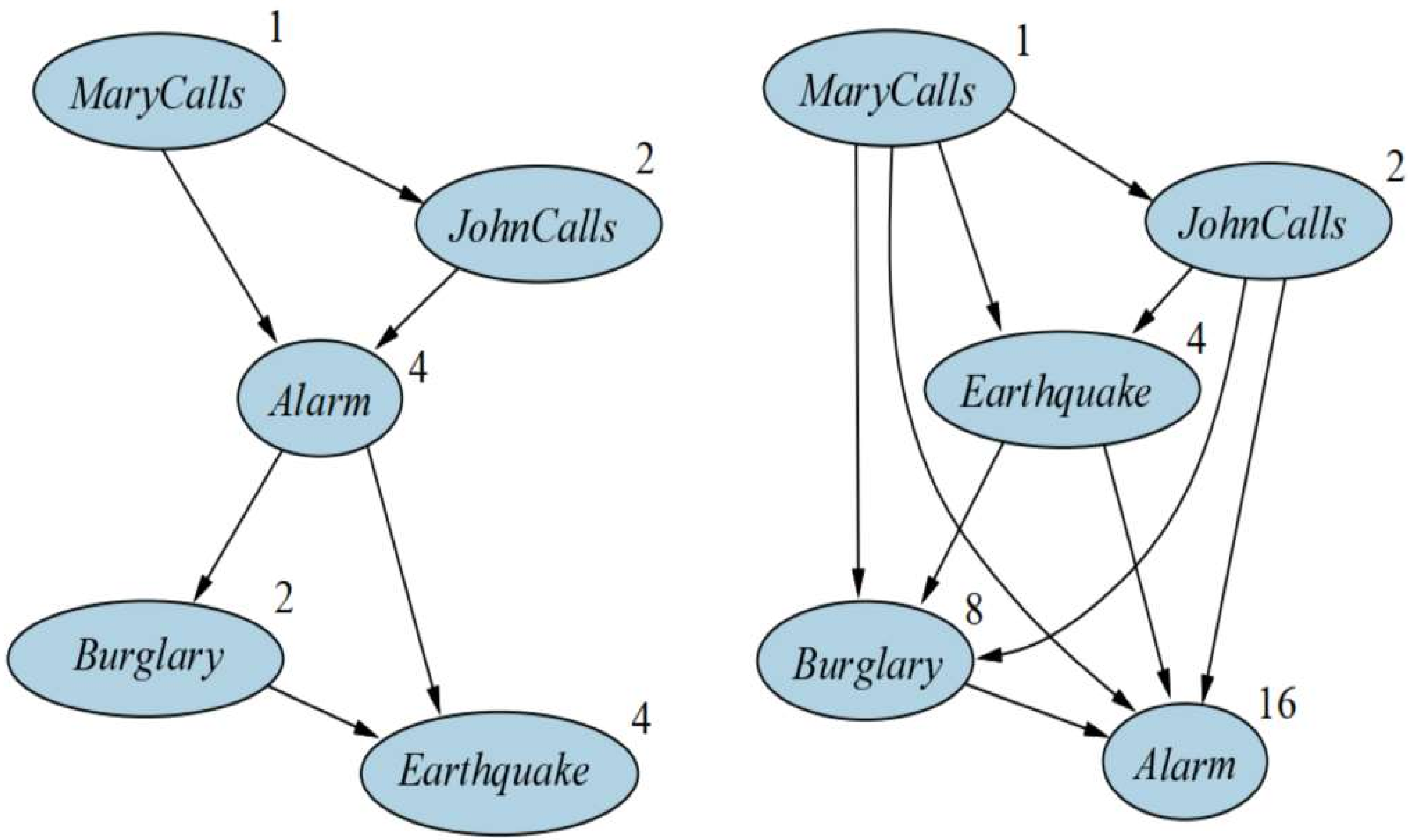
Building Bayesian (Belief) Network

1. **Order Random Variables (ordering matters!)**
2. **Create network nodes for each Random Variable**
3. **Add edges between parent nodes and children nodes**
   - **For every node node X_i:**
     - **choose a minimal set S of parents for X_i**
     - **for each parent node Y in S add an edge from Y to X_i**

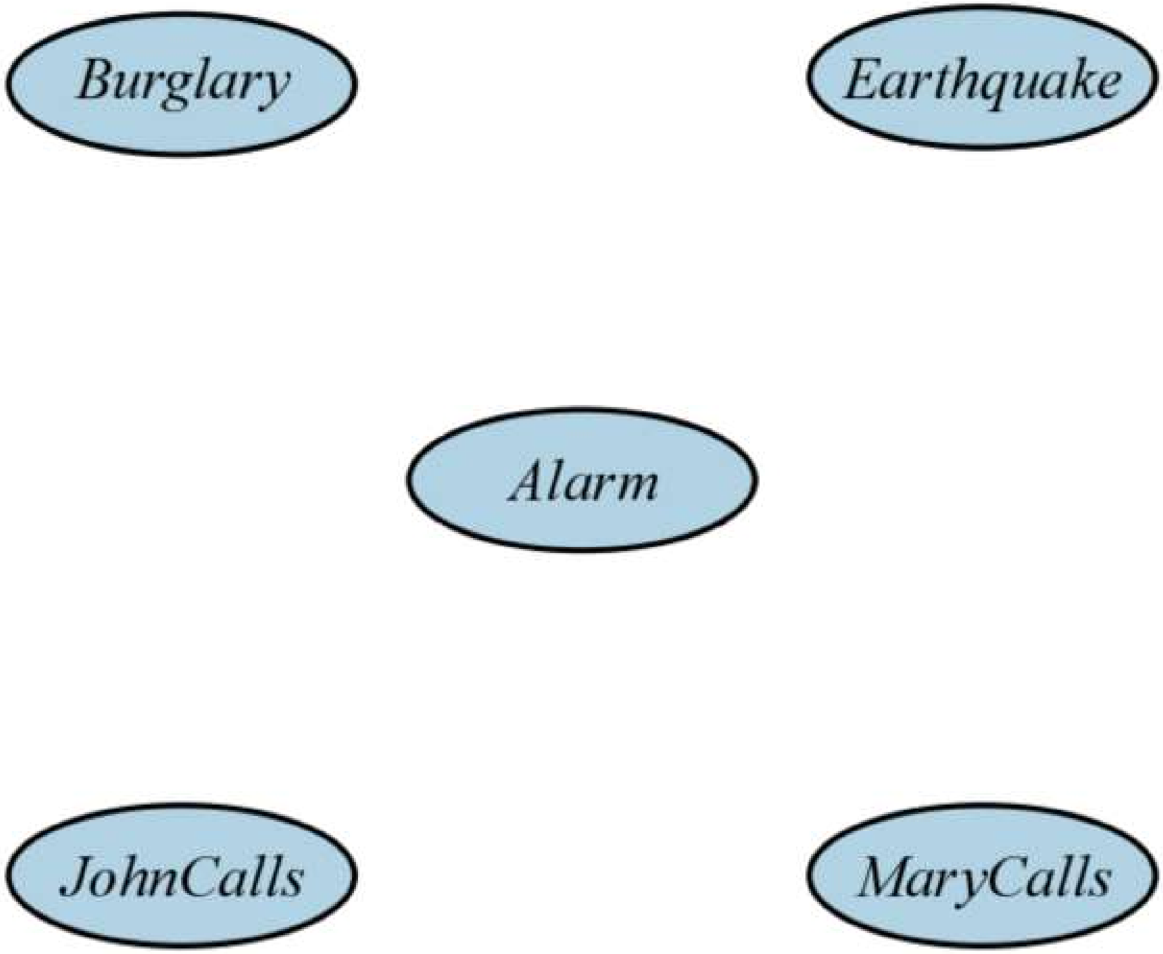
4. Add Conditional Probability Tables

Make it compact / sparse: choose your Random Variable ordering wisely.

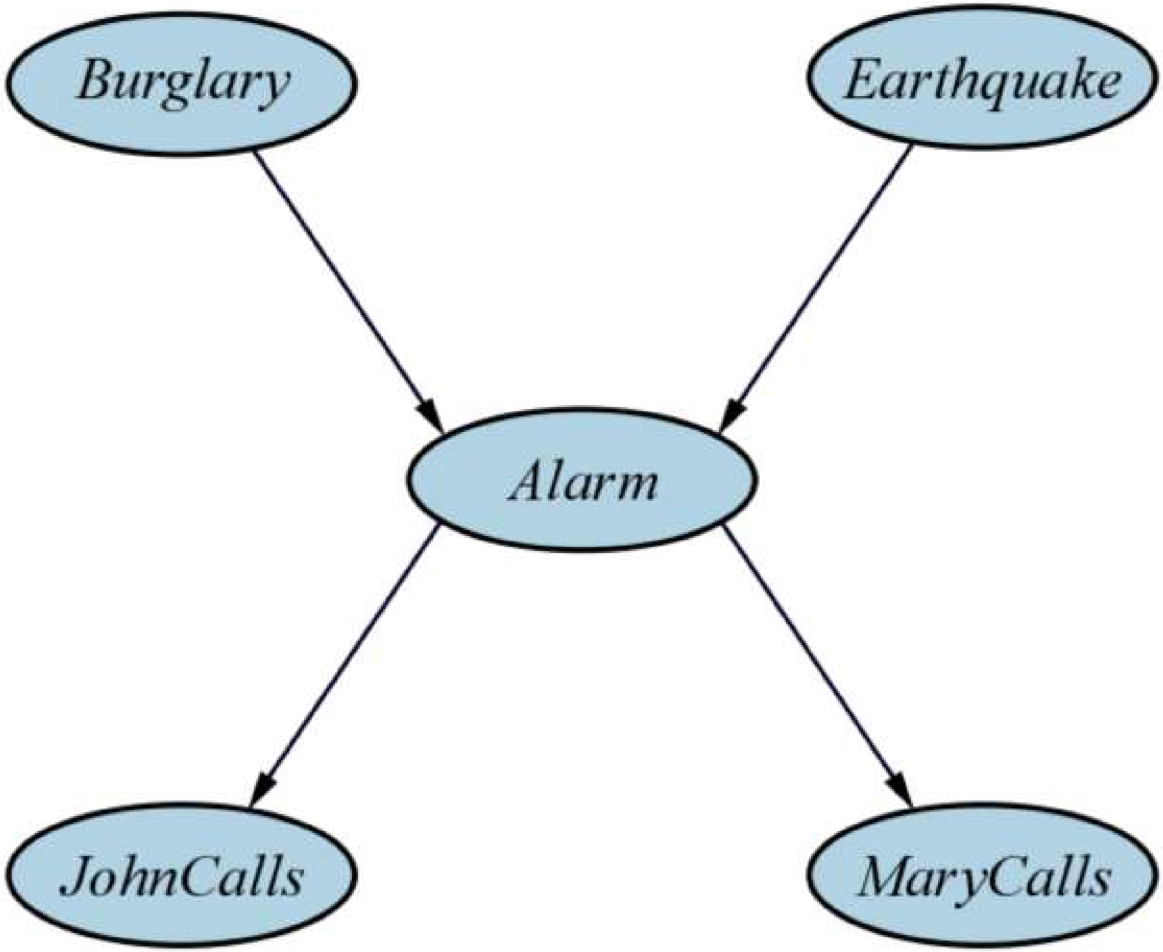
Ordering Matters!



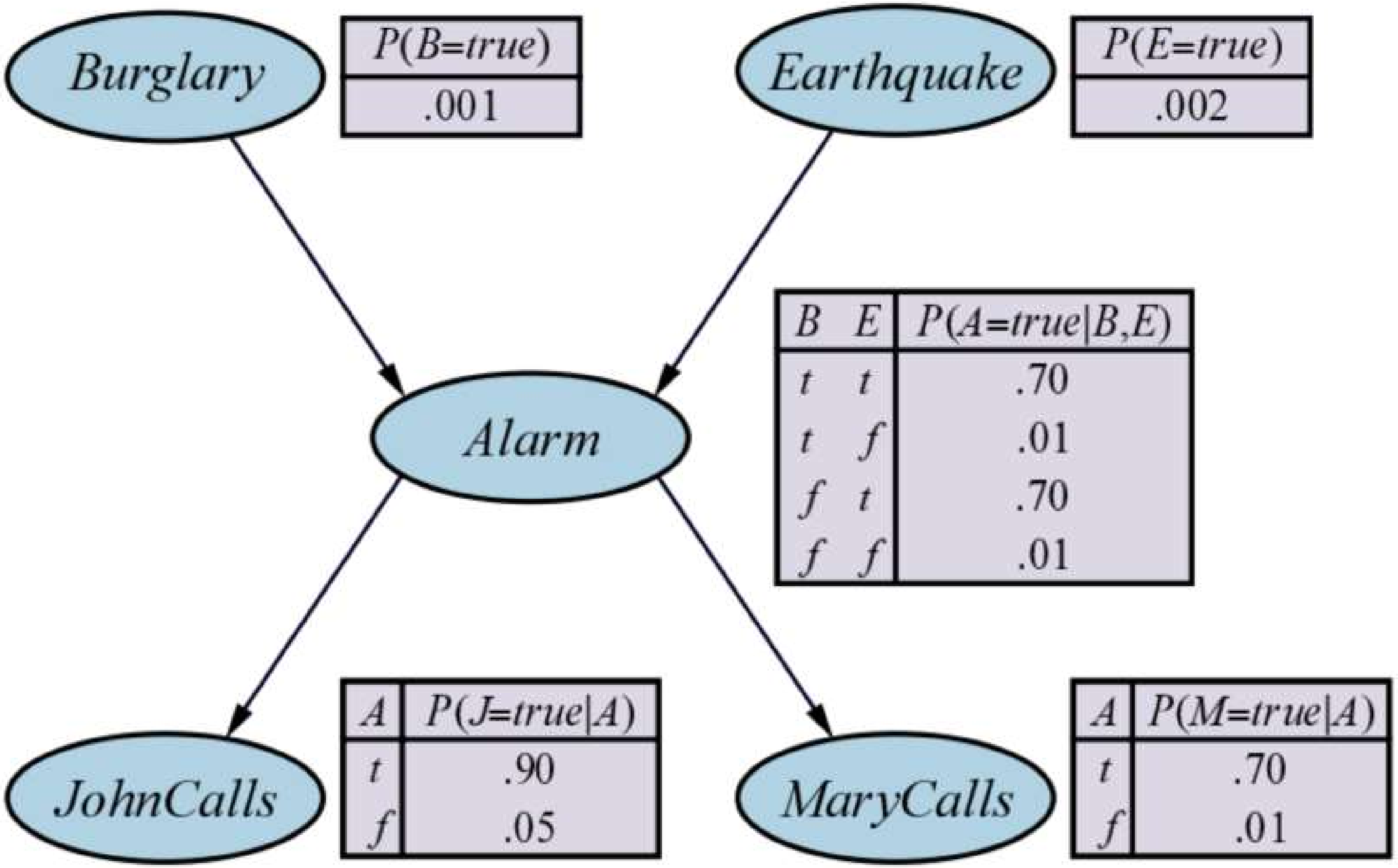
Create Vertices / Node / Random Vars



Add Edges



Add Conditional Probability Tables



P(B=true)
.001

P(E=true)
.002

B	E	P(A=true|B,E)
t	t	.70
t	f	.01
f	t	.70
f	f	.01

A	P(J=true|A)
t	.90
f	.05

A	P(M=true|A)
t	.70
f	.01

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18 / 81 * 12 / 18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63 / 81 * 56 / 63 \approx 0.691$
		SUM = 1

$$P(H \wedge e) = P(H) * P(e \mid parents(e)) = \boxed{P(H)} * \boxed{P(e \mid H)}$$

H: grad	¬H: ¬grad
18 / 81 ≈ 0.22	63 / 81 ≈ 0.78

H: grad	e: female	P(e | H)
true	true	6 / 18 ≈ 0.333
true	false	12 / 18 ≈ 0.667
false	true	7 / 63 ≈ 0.111
false	false	56 / 63 ≈ 0.889

Conditional Probability Table (CPT)

Create Vertices / Node / Random Vars

Create Vertices / Node / Random Vars

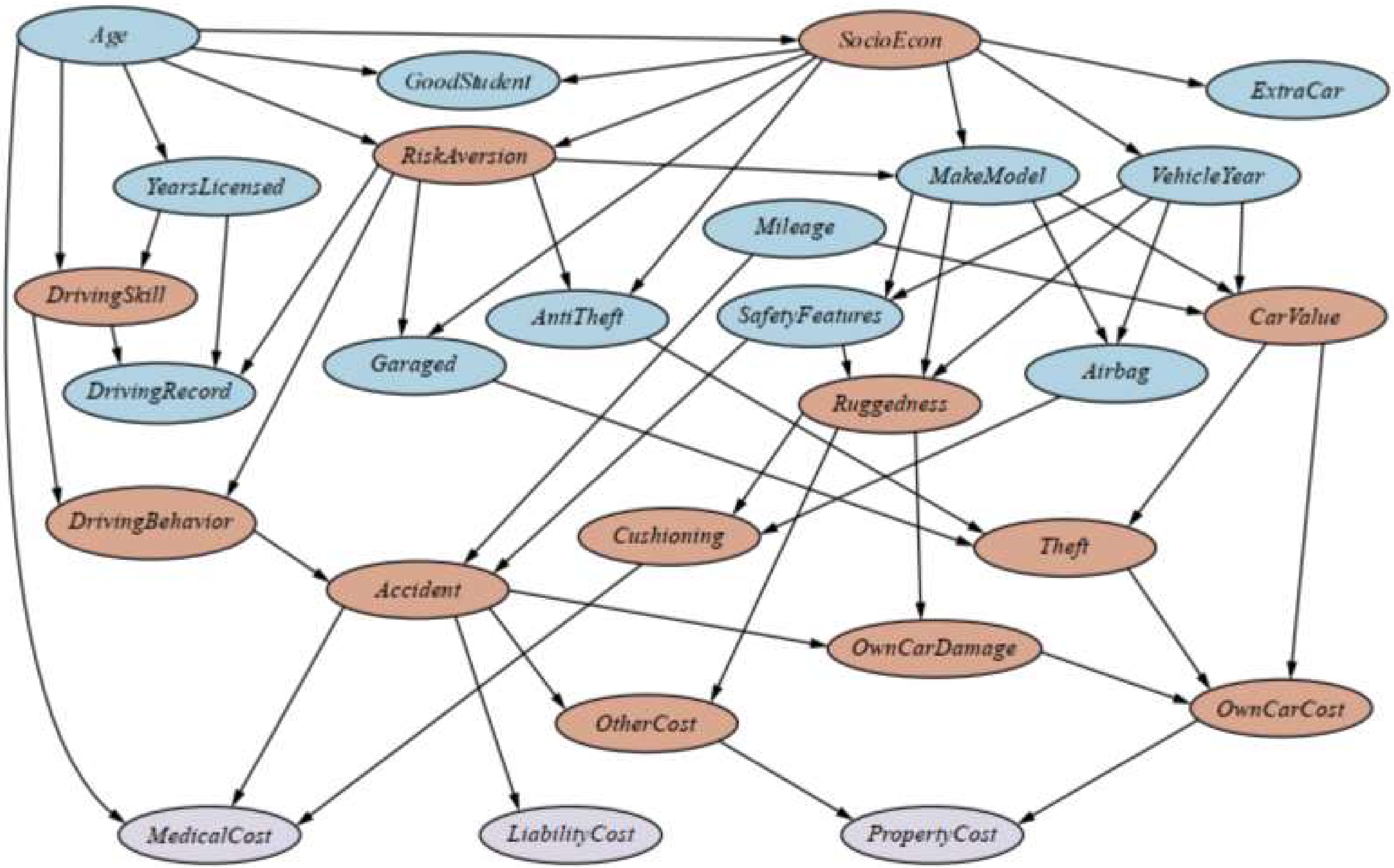
Add Edges

Add Conditional Probability Tables

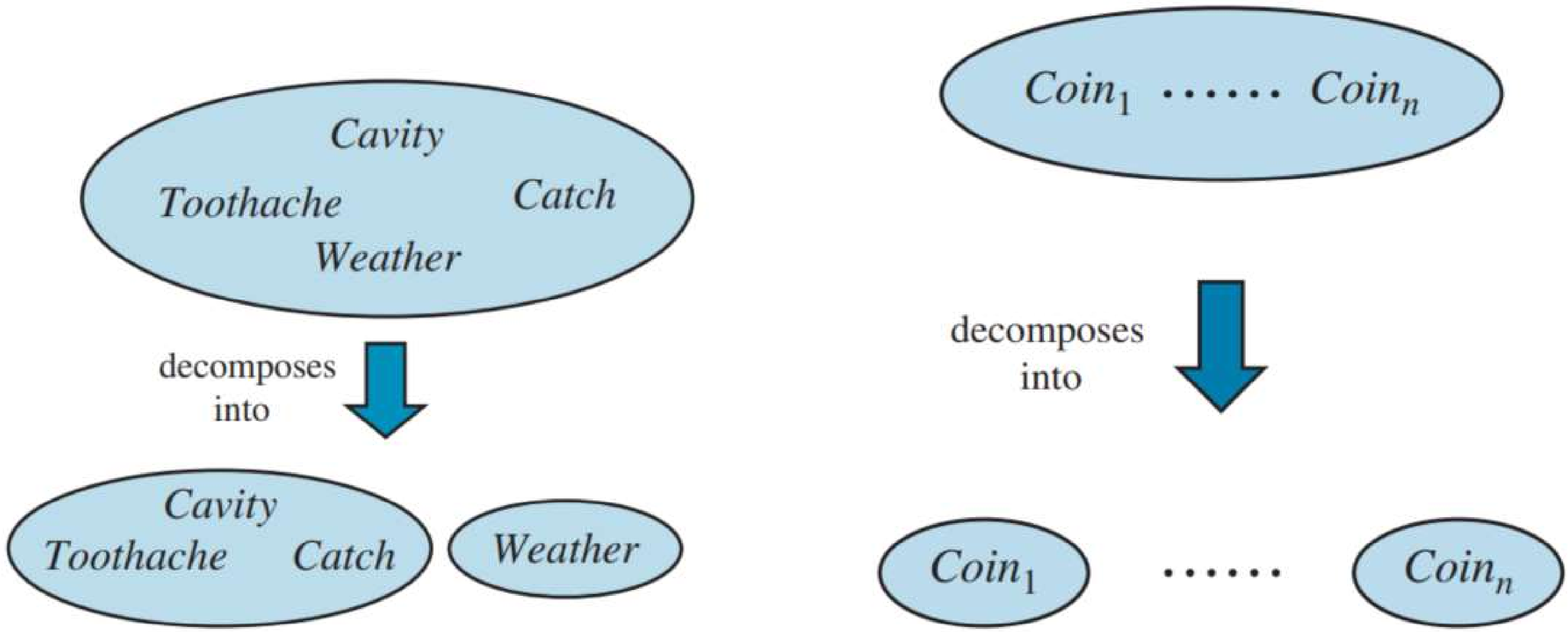
H: grad	¬H: ¬grad
$18 / 81 \approx 0.22$	$63 / 81 \approx 0.78$

H: grad	e: female	P(e | H)
true	true	$6 / 18 \approx 0.333$
true	false	$12 / 18 \approx 0.667$
false	true	$7 / 63 \approx 0.111$
false	false	$56 / 63 \approx 0.889$

Bayesian Network: Car Insurance



Factoring / Decomposition



Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18 / 81 * 12 / 18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63 / 81 * 56 / 63 \approx 0.691$
		SUM = 1

$$P(H \wedge e) = P(H) * P(e \mid parents(e)) = \boxed{P(H)} * \boxed{P(e \mid H)}$$

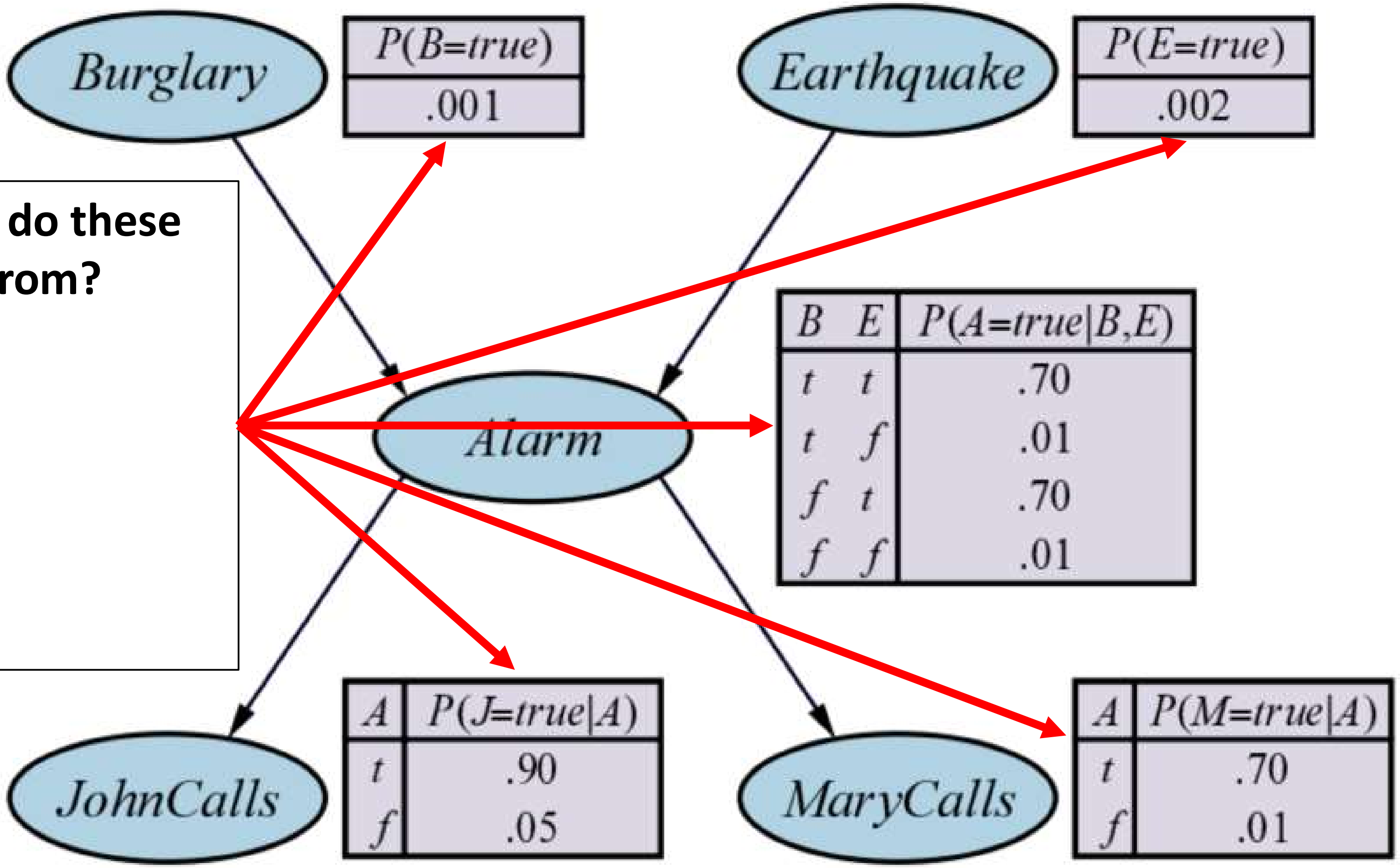
H: grad	¬H: ¬grad
18 / 81 ≈ 0.22	63 / 81 ≈ 0.78

H: grad	e: female	P(e | H)
true	true	6 / 18 ≈ 0.333
true	false	12 / 18 ≈ 0.667
false	true	7 / 63 ≈ 0.111
false	false	56 / 63 ≈ 0.889

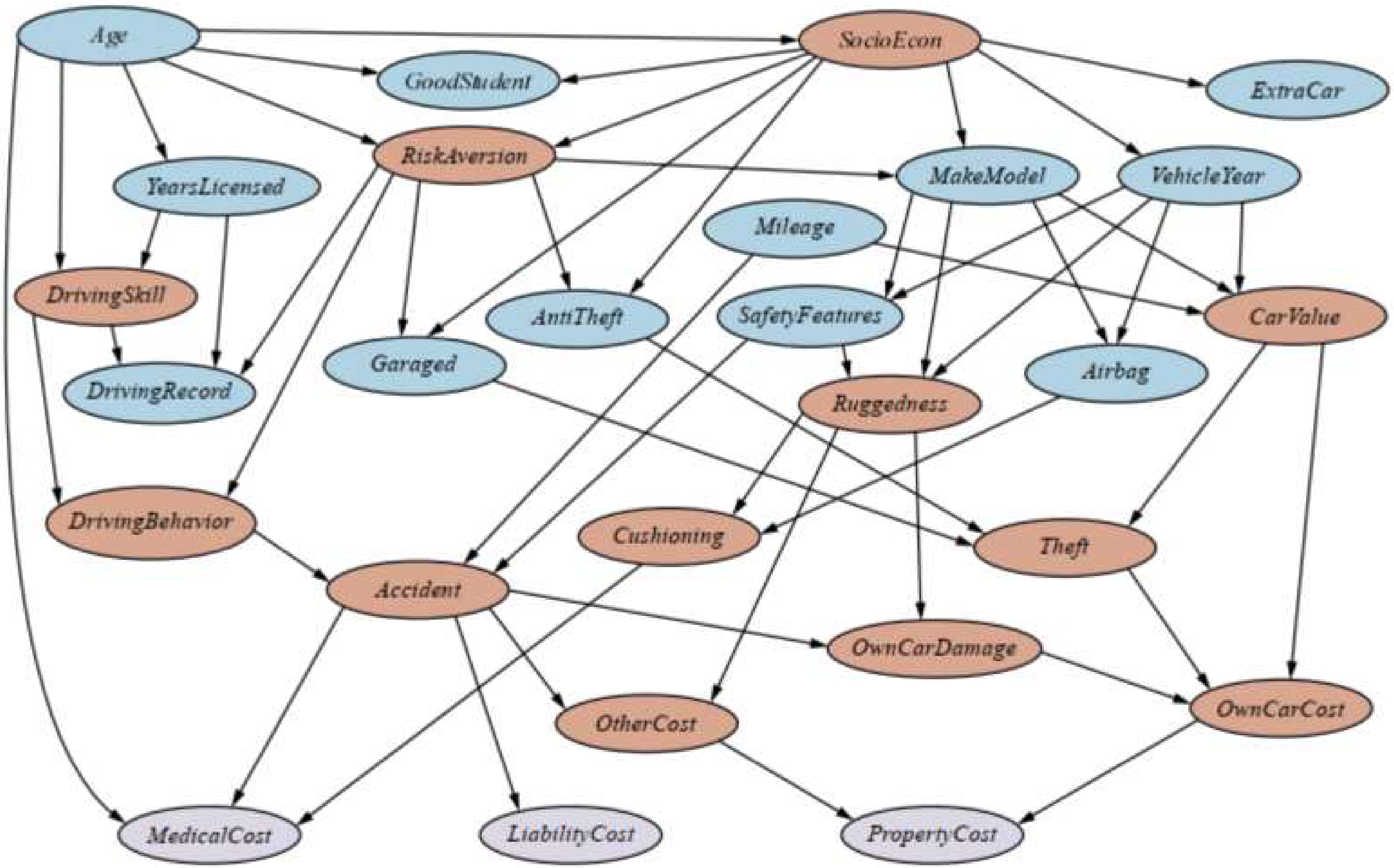
Conditional Probability Table (CPT)

Add Conditional Probability Tables

Where do these come from?



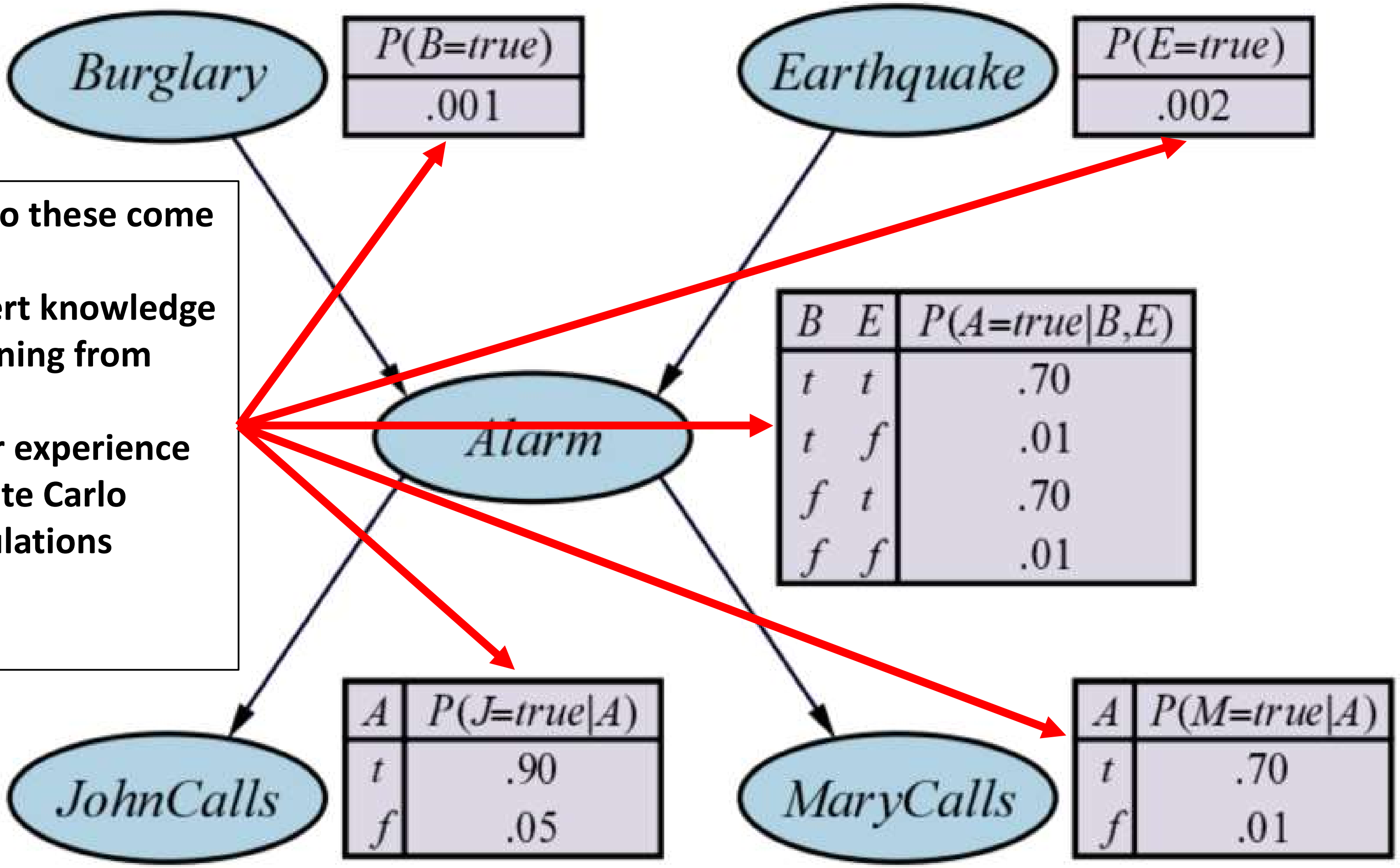
Bayesian Network: Car Insurance



Add Conditional Probability Tables

Where do these come from?

- Expert knowledge
- Learning from data
- Prior experience
- Monte Carlo simulations
- etc.



P(B=true)
.001

P(E=true)
.002

B	E	P(A=true|B,E)
t	t	.70
t	f	.01
f	t	.70
f	f	.01

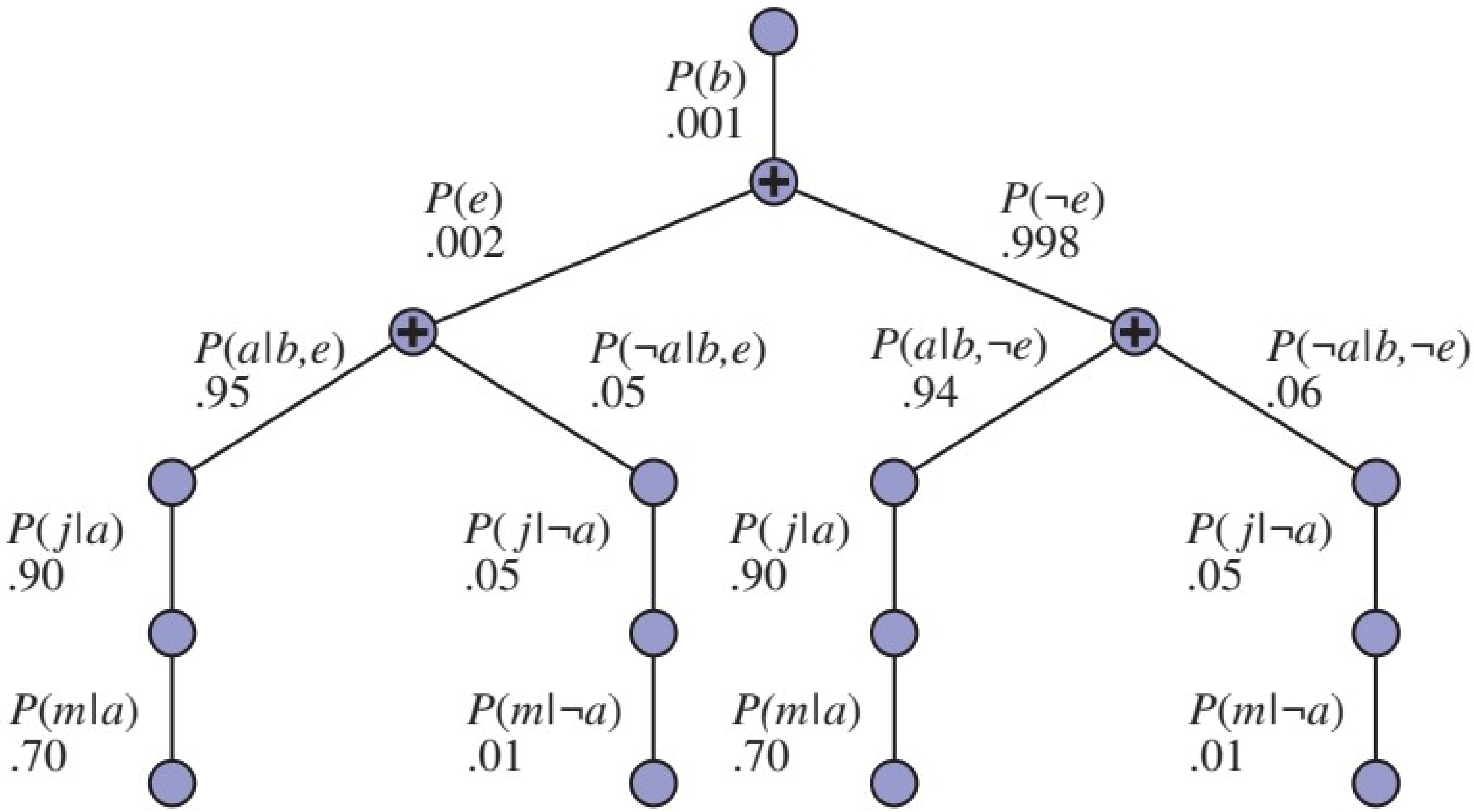
A	P(J=true|A)
t	.90
f	.05

A	P(M=true|A)
t	.70
f	.01

Inference by Enumeration: Example

Query:

$$\boldsymbol{P}(Burglary \mid JohnCalls = true \wedge MaryCalls = true)$$



General Inference Procedure

Given:

- **a query involving a single variable X (in our example: Cavity),**
- **a list of evidence variables E (in our example: just Toothache),**
- **a list of observed values e for E,**
- **a list of remaining unobserved variables Y (in our example: just Catch),**

where X, E, and Y together are a COMPLETE set of variables for the domain, the probability P(X | E) can be evaluated as:

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

where ys are all possible values for Ys, α - normalization constant.

P(X, e, y) is a subset of probabilities from the joint distribution

Inference

Query:

$$P(Burglary \mid JohnCalls = true \wedge MaryCalls = true)$$

Given:

- **a query involving a single variable** X
- **a list of evidence variables** K,
- **a list of observed values** k **for** K,
- **a list of remaining unobserved variables** Y

the probability P(X | K) **can be evaluated as:**

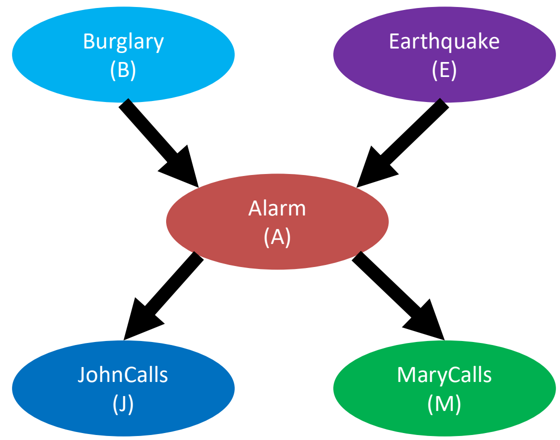
$$\boldsymbol{P}(X \mid \boldsymbol{k}) = \alpha * \boldsymbol{P}(X, \boldsymbol{k})$$

$$= \alpha * \sum_{\boldsymbol{y}} \boldsymbol{P}(X, \boldsymbol{k}, \boldsymbol{y})$$

where ys **are all possible values for** Ys, α **- normalization constant.**

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query:

$$P(Burglary \mid JohnCalls = true \wedge MaryCalls = true)$$

Given:

- **a query involving a single variable** X**:** $Burglary$
- **a list of evidence variables** $\mathbf{K}$**:** $JohnCalls$, $MaryCalls$
- **a list of observed values** k **for** $\mathbf{K}$**:** $johnCalls$, $maryCalls$
- **a list of remaining unobserved variables** $\mathbf{Y}$**:** $Earthquake$, $Alarm$

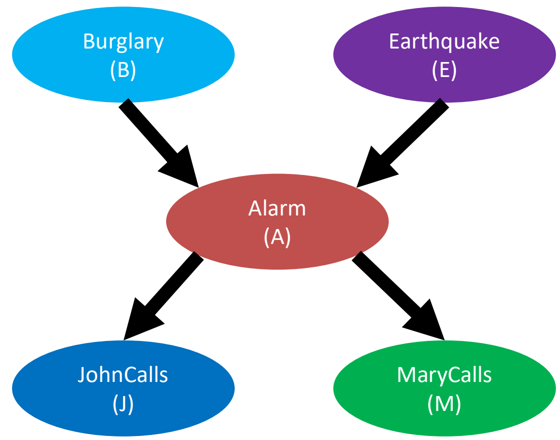
the probability $\mathrm{P(X \mid \mathbf{K})}$ **can be evaluated as:**

$$\boldsymbol{P}(X \mid \boldsymbol{k}) = \alpha * \sum_{\boldsymbol{y}} \boldsymbol{P}(X, \boldsymbol{k}, \boldsymbol{y})$$

where y**s are all possible values for** Y**s,** α **- normalization constant.**

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query:

$$P(Burglary \mid JohnCalls = true \wedge MaryCalls = true)$$

Given:

- a query involving a single variable X: B
- a list of **evidence** variables **K**: J, M
- a list of **observed** values k for **K**: j, m
- a list of remaining **unobserved** variables **Y**: E, A

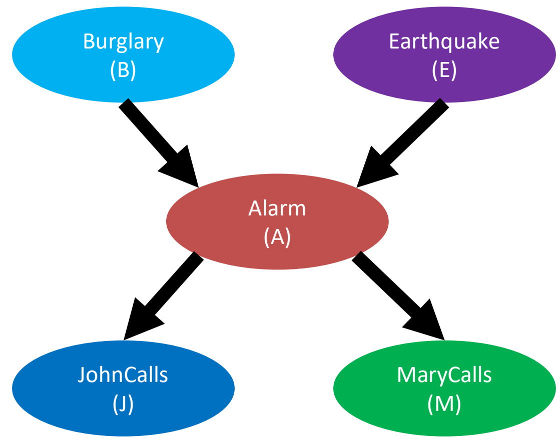
the probability P(X | **K**) can be evaluated as:

$$P(X \mid \boldsymbol{k}) = \alpha * \sum_{\boldsymbol{y}} P(X, \boldsymbol{k}, \boldsymbol{y})$$

where ys are all possible values for Ys, α - normalization constant.

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query:

$$P(Burglary \mid JohnCalls = true \wedge MaryCalls = true)$$

Given:

- **a query involving a single variable B**
- **a list of evidence variables K: J, M**
- **a list of observed values k for K: j, m**
- **a list of remaining unobserved variables Y: E, A**

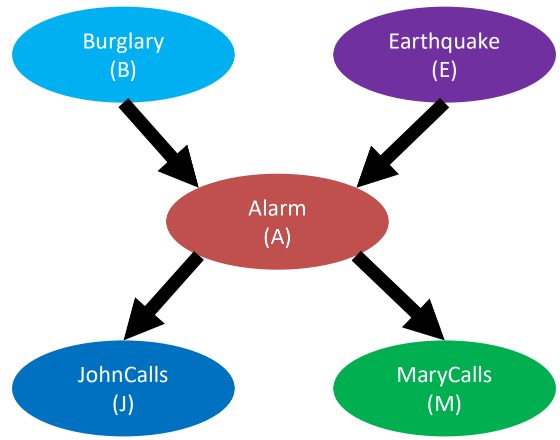
the probability P(B | J, M) can be evaluated as:

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

where ys are all possible values for Ys, α - normalization constant.

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query (let's change it a bit for simplicity):

$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$

Given:

- a query involving a single variable B
- a list of evidence variables $\mathbf{K}$: J, M
- a list of observed values $\mathbf{k}$ for $\mathbf{K}$: j, m
- a list of remaining unobserved variables $\mathbf{Y}$: E, A

the query can be evaluated as:

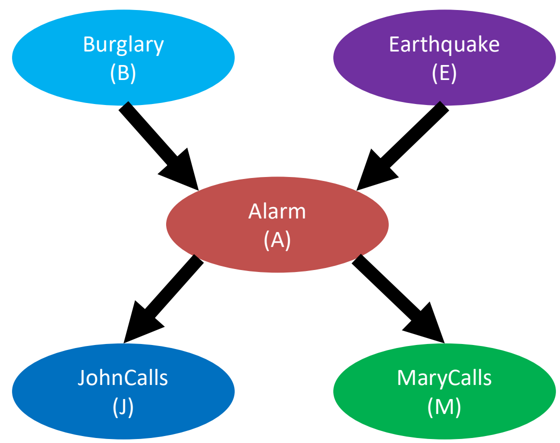
$$P(b \mid j, m) = \alpha * \sum_{e} \sum_{a} P(b, j, m, e, a)$$

By Chain rule:

$$P(b, j, m, e, a) = P(b) * P(e) * P(a|b,e) * P(j|a) * P(m|a)$$

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query (let's change it a bit for simplicity):

$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$

Given:

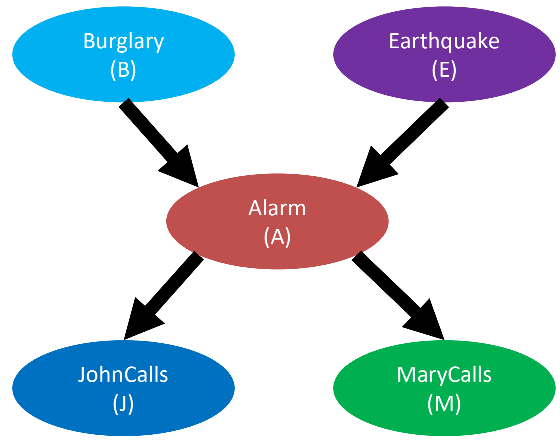
- **a query involving a single variable** B
- **a list of evidence variables** $\mathbf{K}$**:** J, M
- **a list of observed values** $\mathbf{k}$ **for** $\mathbf{K}$**:** j, m
- **a list of remaining unobserved variables** $\mathbf{Y}$**:** E, A

the query can be evaluated as:

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

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query (let's change it a bit for simplicity):

$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$

Given:

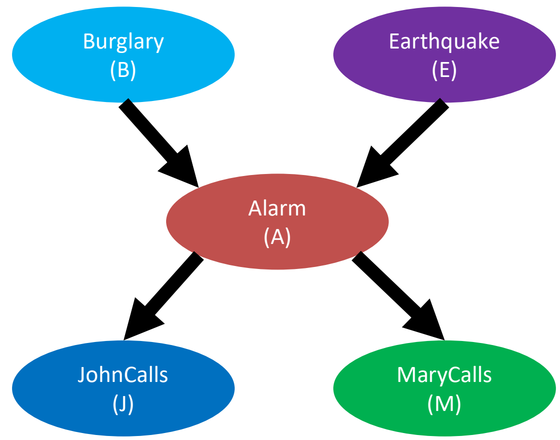
- **a query involving a single variable B**
- **a list of evidence variables $\mathbf{K}$: J, M**
- **a list of observed values $\mathbf{k}$ for $\mathbf{K}$: j, m**
- **a list of remaining unobserved variables $\mathbf{Y}$: E, A**

the query can be evaluated as:

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

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query (let's change it a bit for simplicity):

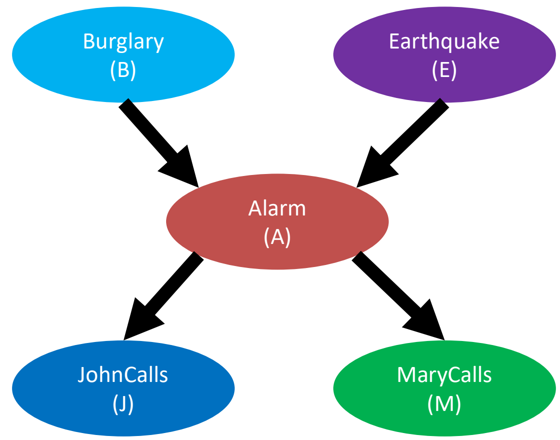
$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$

Query rewritten:

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

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query (let's change it a bit for simplicity):

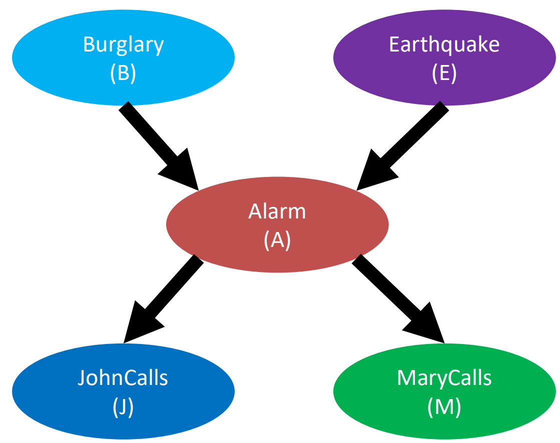
$P(Burglary = true \mid JohnCalls = true \land MaryCalls = true)$

Query rewritten:

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

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

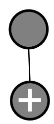
Query (let's change it a bit for simplicity):

$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$

Query rewritten:

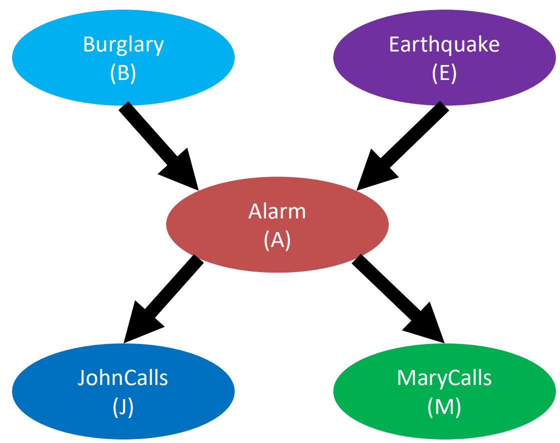
$$P(b \mid j, m)$$
$$= \alpha * \sum_e \sum_a P(b) * P(e) * P(a|b,e) * P(j|a) * P(m|a)$$
$$= \alpha * P(b) * \boxed{\sum_e} P(e) * \sum_a P(a|b,e) * P(j|a) * P(m|a)$$

$P(b)$



P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

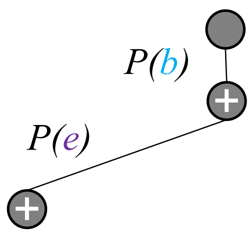
Inference

Query (let's change it a bit for simplicity):

$$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$$

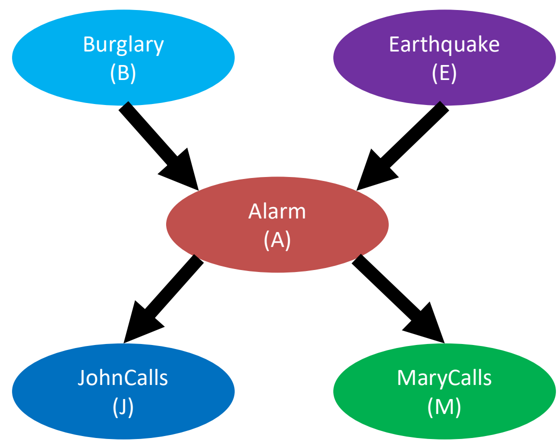
Query rewritten:

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



P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

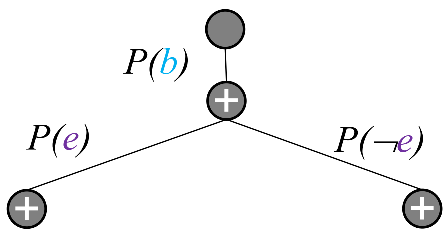
Inference

Query (let's change it a bit for simplicity):

$$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$$

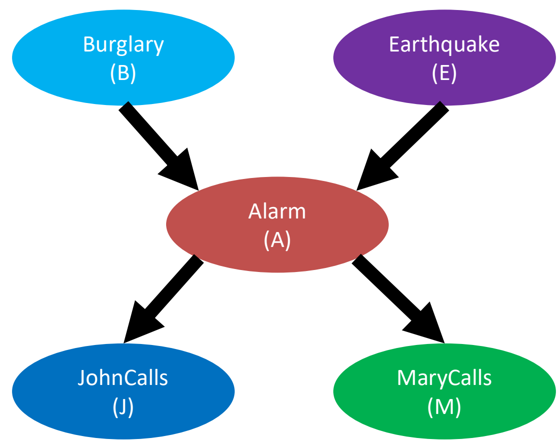
Query rewritten:

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



P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

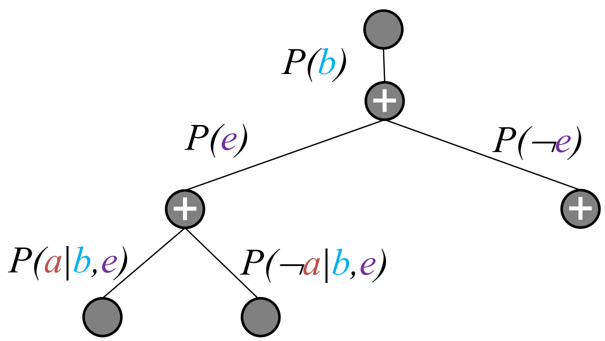
Inference

Query (let's change it a bit for simplicity):

$$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$$

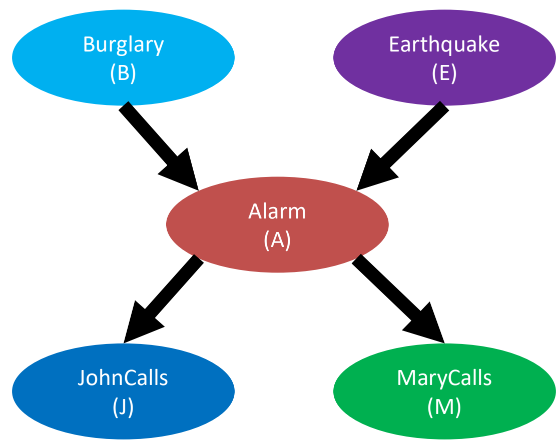
Query rewritten:

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



P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

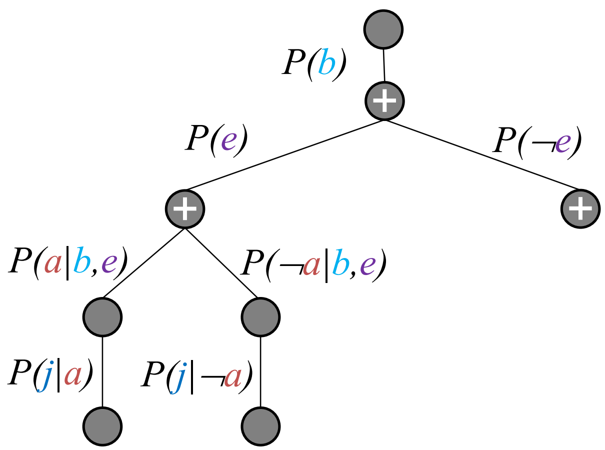
Inference

Query (let's change it a bit for simplicity):

$$P(Burglary = true \mid JohnCalls = true \land MaryCalls = true)$$

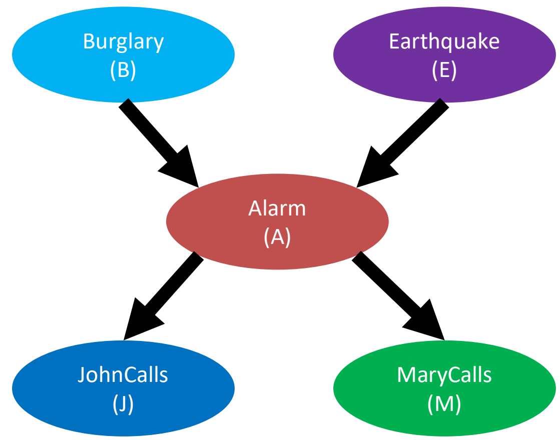
Query rewritten:

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



P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

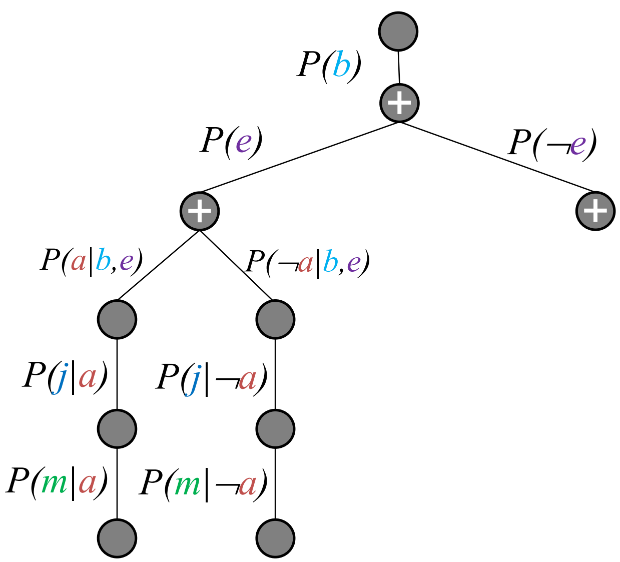
Inference

Query (let's change it a bit for simplicity):

$$P(Burglary = true \mid JohnCalls = true \land MaryCalls = true)$$

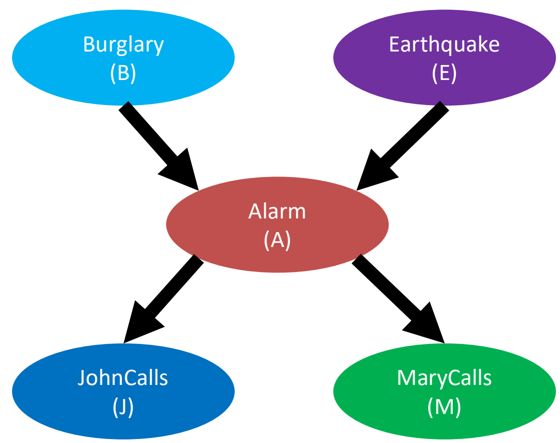
Query rewritten:

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



P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

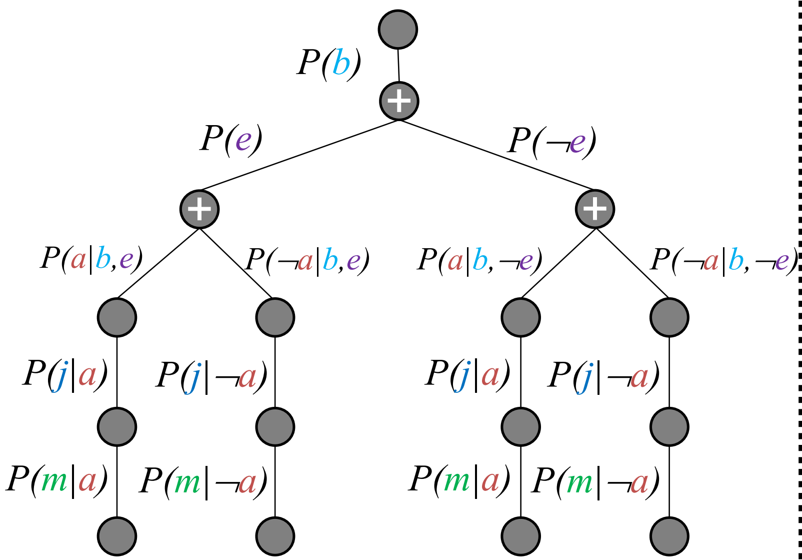
Inference

Query (let's change it a bit for simplicity):

$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$

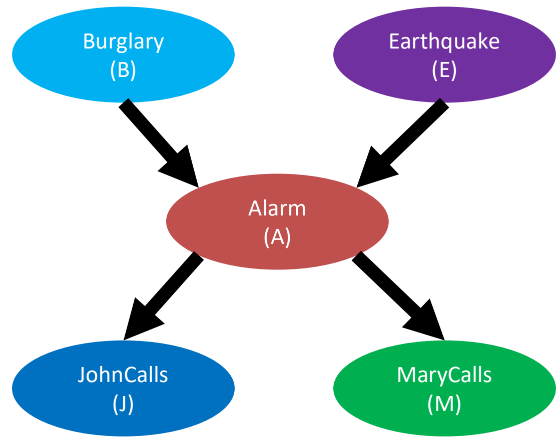
Query rewritten:

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



P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query (let's change it a bit for simplicity):

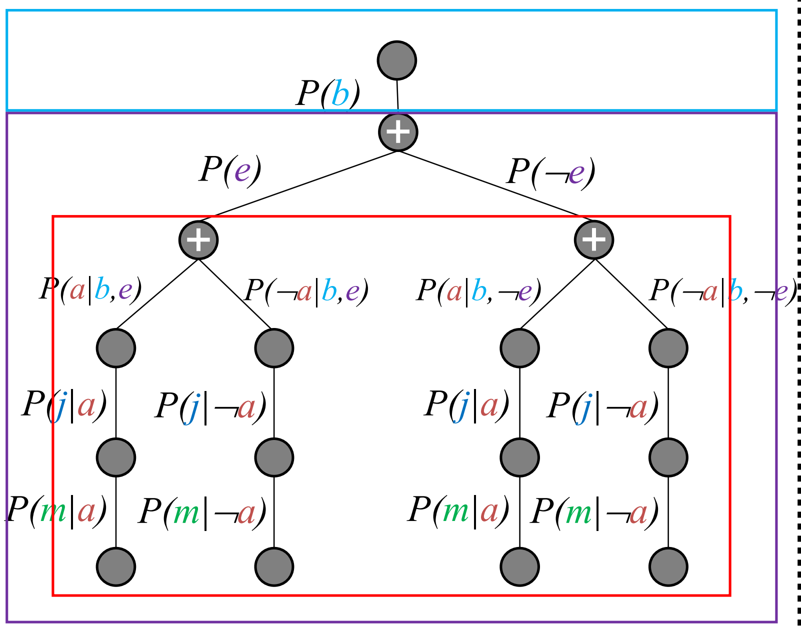
$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$

Query rewritten:

$$P(b \mid j, m)$$

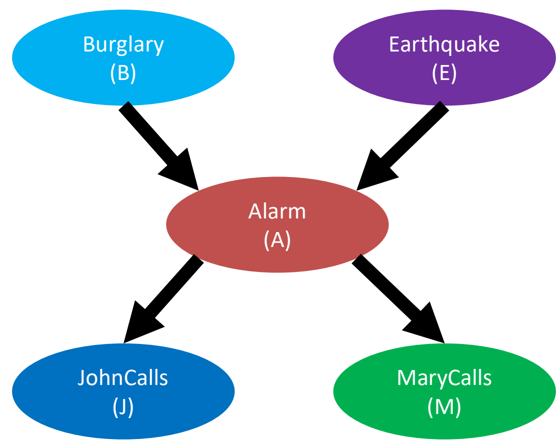
$$= \alpha * \sum_{e}\sum_{a} P(b) * P(e) * P(a|b,e) * P(j|a) * P(m|a)$$

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



P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

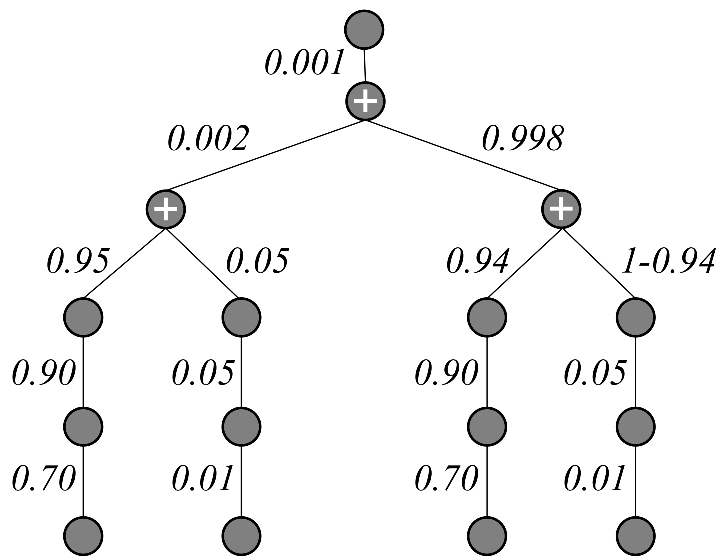
Inference

Query (let's change it a bit for simplicity):

$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$

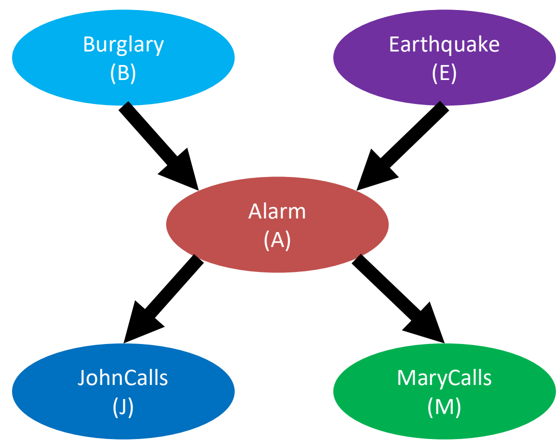
Query rewritten:

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



P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query (let's change it a bit for simplicity):

$P(Burglary = true \mid JohnCalls = true \wedge MaryCalls = true)$

We can now calculate:

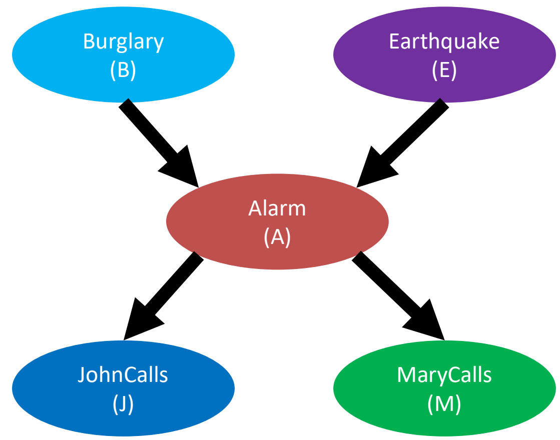
$$P(b \mid j, m) = \alpha * 0.00059224$$

And through similar process (not shown):

$$P(\neg b \mid j, m) = \alpha * 0.0014919$$

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01

Inference

Query (now we can get joint distribution):

$$\boldsymbol{P}(Burglary \mid JohnCalls = true \wedge MaryCalls = true)$$

We can now calculate:

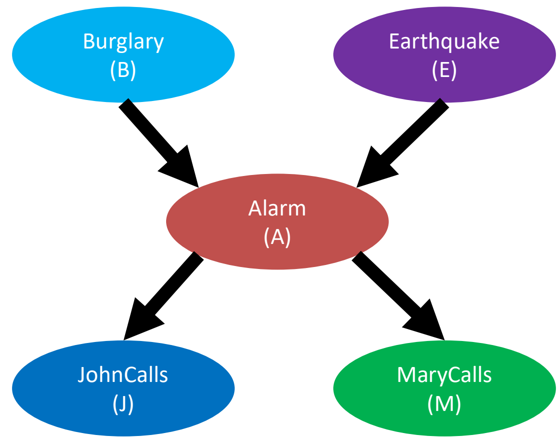
$$\boldsymbol{P}(B \mid j, m) = \alpha * \ < 0.00059224, 0.0014919 >$$

And after normalization:

$$\boldsymbol{P}(B \mid j, m) \approx < 0.284, 0.716 >$$

P(B)	P(¬B)
0.001	0.999

P(E)	P(¬E)
0.002	0.998



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

A	P(J|A)
t	0.90
f	0.05

A	P(M|A)
t	0.70
f	0.01