Artificial Intelligence Foundation – JC3001

Lecture 21: Quantifying Uncertainty and Reasoning with Probabilities III

Prof. Aladdin Ayesh (aladdin.ayesh@abdn.ac.uk)

Dr. Binod Bhattarai (binod.bhattarai@abdn.ac.uk)

Dr. Gideon Ogunniye, (g.ogunniye@abdn.ac.uk)

September 2025

Material adapted from:

Russell and Norvig (AIMA Book): Chapters 12 and 13

Sebastian Thrun — Stanford University / Udacity

Course Progression

- Part 1: Introduction
  1. Introduction to AI ✓
  2. Agents ✓
- Part 2: Problem-solving
  1. Search 1: Uninformed Search ✓
  2. Search 2: Heuristic Search ✓
  3. Search 3: Local Search
  4. Search 4: Adversarial Search ✓
- Part 3: Reasoning and Uncertainty
  1. Reasoning 1: Constraint Satisfaction ✓
  2. Reasoning 2: Logic and Inference ✓
  3. **Probabilistic Reasoning 1: BNs**
  4. Probabilistic Reasoning 2: HMMs
- Part 4: Planning
  1. Planning 1: Intro and Formalism
  2. Planning 2: Algos and Heuristics
  3. Planning 3: Hierarchical Planning
  4. Planning 4: Stochastic Planning
- Part 5: Learning
  1. Learning 1: Intro to ML
  2. Learning 2: Regression
  3. Learning 3: Neural Networks
  4. Learning 4: Reinforcement Learning
- Part 6: Conclusion
  1. Ethical Issues in AI
  2. Conclusions and Discussion

Objectives

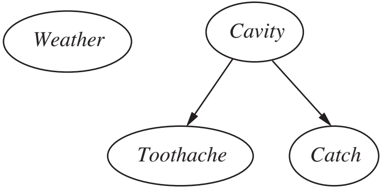
- Probability Theory
- Probabilistic Inference
- Bayes Rule
- Bayesian Networks
  - Graphical Semantics
  - Inference

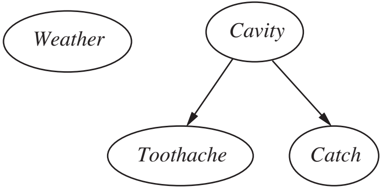
- Bayesian Networks
  - Graphical Semantics
  - Inference in Bayesian Networks

Where are We?

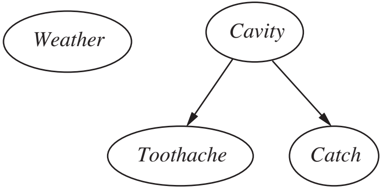
- We introduced probability theory
- We discussed how independence simplifies representation of the world
- How can we encode such independence relationships?

- A Bayesian Network (BN) represents the dependencies among variables and encodes the full joint probability distribution concisely.
- A BN is a directed graph, where each node is annotated with probability information.
  - The set of random variables makes up the nodes of the network.
  - A set of directed links connects pairs of nodes, encoding a parent-child relationship.
  - Each node X_i has a conditional probability distribution $P(X_i \mid Parents(X_i))$
  - The graph has no directed cycles.
- Intuitively, an arrow from X to Y means that X has a direct influence on Y.





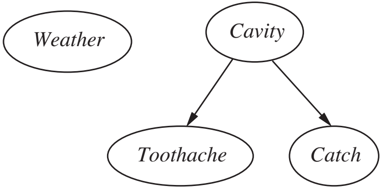
- The conditional independence of Toothache from Catch given Cavity is indicated by the absence of a link between them.
- The network indicates that Toothache is dependent on cavity (as is catch); weather is independent of all other variables in the network.
- A BN also has a conditional probability table (CPT) indicating the conditional probability for each node value for a conditioning case — the possible combination of values for its parent nodes.
- Since each row must sum to 1, if we are dealing with boolean variables, we need only show the positive case.
- If a variable has k parents, the table should have 2^k rows.



In this example, how many tables we'd have?
And how many values would we have in all tables?

Any fully specified question is easy to solve:

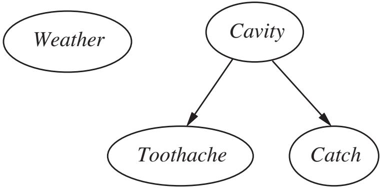
$$P(Cav \mid Cat, T) = P(T \mid Cav)P(Cat \mid Cav)P(Cav)$$
$$= 0.6 * 0.9 * 0.2 = 0.108$$



In this example, how many tables we'd have? 4
And how many values would we have in all tables?

Any fully specified question is easy to solve:

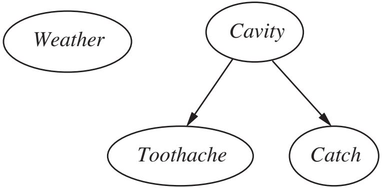
$$P(Cav \mid Cat, T) = P(T \mid Cav)P(Cat \mid Cav)P(Cav)$$
$$= 0.6 * 0.9 * 0.2 = 0.108$$



In this example, how many tables we'd have? 4
And how many values would we have in all tables? 6

Any fully specified question is easy to solve:

$$P(Cav \mid Cat, T) = P(T \mid Cav)P(Cat \mid Cav)P(Cav)$$
$$= 0.6 * 0.9 * 0.2 = 0.108$$



In this example, how many tables we'd have? 4

And how many values would we have in all tables? 6

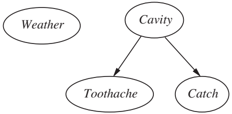
$P(Cav)$	0.2
$P(W)$	0.5

Cav	$P(T \mid Cav)$
T	0.6
F	0.1

Cav	$P(Cat \mid Cav)$
T	0.9
F	0.2

Any fully specified question is easy to solve:

$$P(Cav \mid Cat, T) = P(T \mid Cav)P(Cat \mid Cav)P(Cav)$$
$$= 0.6 * 0.9 * 0.2 = 0.108$$



$P(Cav)$	0.2
$P(W)$	0.5

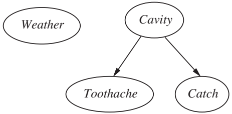
Cav	$P(T \mid Cav)$
T	0.6
F	0.1

Cav	$P(Cat \mid Cav)$
T	0.9
F	0.2

Partially specified questions are more complex:

$$\begin{aligned} P(Cav \mid Cat) =& \alpha(P(T \mid Cav)P(Cat \mid Cav)P(Cav)+ \\ & P(\neg T \mid Cav)P(Cat \mid Cav)P(Cav)) \\ =& \alpha(0.6 * 0.9 * 0.2 + 0.4 * 0.9 * 0.2) \\ =& \alpha(0.108 + 0.072) = \alpha 0.18 \end{aligned}$$

What is α?



$P(Cav)$	0.2
$P(W)$	0.5

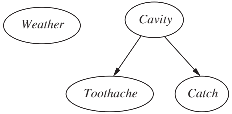
Cav	$P(T \mid Cav)$
T	0.6
F	0.1

Cav	$P(Cat \mid Cav)$
T	0.9
F	0.2

Partially specified questions are more complex:

$$\begin{aligned} P(Cav \mid Cat) =& \alpha(P(T \mid Cav)P(Cat \mid Cav)P(Cav)+ \\ & P(\neg T \mid Cav)P(Cat \mid Cav)P(Cav)) \\ =& \alpha(0.6 * 0.9 * 0.2 + 0.4 * 0.9 * 0.2) \\ =& \alpha(0.108 + 0.072) = \alpha 0.18 \end{aligned}$$

What is α? It is a **normalizing constant**



$P(Cav)$	0.2
$P(W)$	0.5

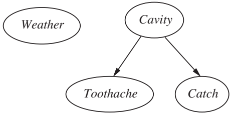
Cav	$P(T \mid Cav)$
T	0.6
F	0.1

Cav	$P(Cat \mid Cav)$
T	0.9
F	0.2

Partially specified questions are more complex:

$$\begin{aligned} P(Cav \mid Cat) =& \alpha(P(T \mid Cav)P(Cat \mid Cav)P(Cav) + \\ & P(\neg T \mid Cav)P(Cat \mid Cav)P(Cav)) \\ =& \alpha(0.6 * 0.9 * 0.2 + 0.4 * 0.9 * 0.2) \\ =& \alpha(0.108 + 0.072) = \alpha 0.18 \end{aligned}$$

What is α? It is a **normalizing constant** $\alpha = \frac{1}{P(Cat)}$



$P(Cav)$	0.2
$P(W)$	0.5

Cav	$P(T \mid Cav)$
T	0.6
F	0.1

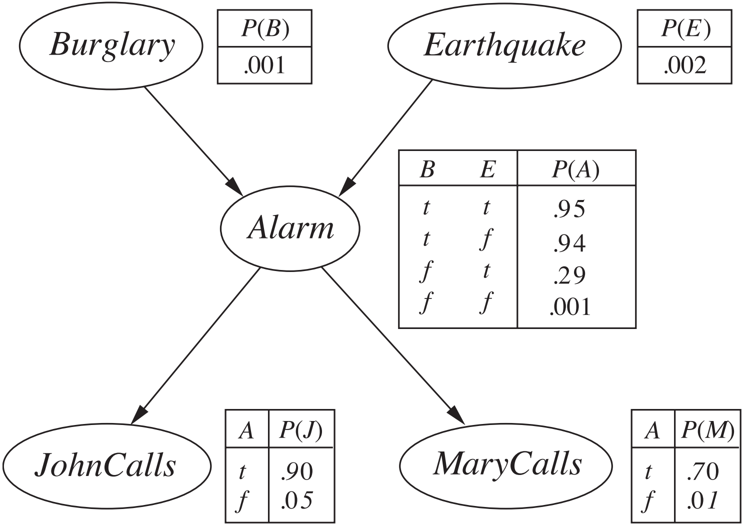
Cav	$P(Cat \mid Cav)$
T	0.9
F	0.2

Partially specified questions are more complex:

$$\begin{aligned} P(Cav \mid Cat) =& \alpha(P(T \mid Cav)P(Cat \mid Cav)P(Cav)+ \\ & P(\neg T \mid Cav)P(Cat \mid Cav)P(Cav)) \\ =& \alpha(0.6 * 0.9 * 0.2 + 0.4 * 0.9 * 0.2) \\ =& \alpha(0.108 + 0.072) = \alpha 0.18 \end{aligned}$$

What is α? It is a **normalizing constant** $\alpha = \frac{1}{P(Cat)} = P(Cat \mid Cav) + P(Cat \mid \neg Cav)$

Example II



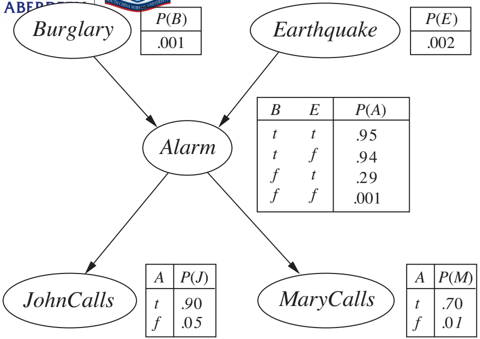
$P(B)$
.001

$P(E)$
.002

B	E	$P(A)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

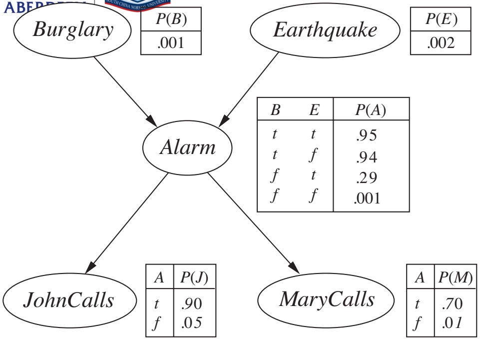
A	$P(J)$
t	.90
f	.05

A	$P(M)$
t	.70
f	.01



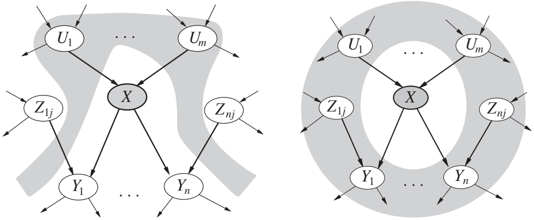
- $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$
- We can answer any query by multiplying all the relevant joint entries.

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$$



- $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$
- We can answer any query by multiplying all the relevant joint entries.

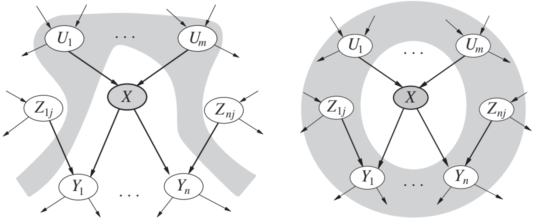
$$
\begin{aligned}
&P(j \wedge m \wedge a \wedge \neg b \wedge \neg e) \\
&\quad = P(j \mid a)P(m \mid a)P(a \mid \neg b \wedge \neg e)P(\neg b)P(\neg e) \\
&\quad = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062
\end{aligned}
$$



(a) (b)

- We just saw the numerical semantics of BNs, but we also have a “topological” semantics:

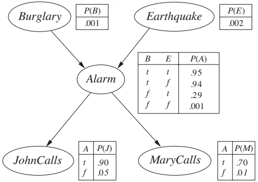
Conditional Independence in BNs



(a) (b)

- We just saw the numerical semantics of BNs, but we also have a “topological” semantics:
- In a BN, each variable is conditionally independent of:
  - (a) Non-descendants given its parents
    E.g. *JohnCalls* is independent of *Burglary*, *Earthquake*, and *MaryCalls* given *Alarm*
  - (b) All other nodes in the network given parents, children, and children’s parents (its Markov Blanket)
    E.g. *Burglary* is independent of *JohnCalls* and *MaryCalls*, given *Alarm* and *Earthquake*

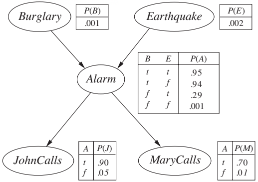
Bayesian Networks—bit by bit



$$P(j \land m \land a \land \neg b \land \neg e)$$

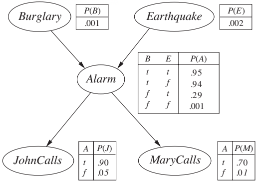
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

Bayesian Networks—bit by bit



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

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$$
$= P(j \mid m, a, \neg b, \neg e) P(m, a, \neg b, \neg e)$ by the chain rule



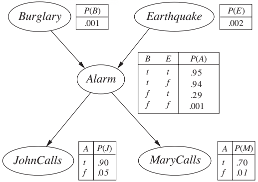
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

$P(j \land m \land a \land \neg b \land \neg e)$

$= P(j \mid m, a, \neg b, \neg e) P(m, a, \neg b, \neg e)$ by the chain rule

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a, \neg b, \neg e)$ chain rule again

Bayesian Networks—bit by bit



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

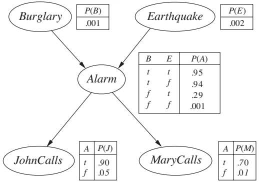
$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$= P(j \mid m, a, \neg b, \neg e) P(m, a, \neg b, \neg e)$ by the chain rule

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a, \neg b, \neg e)$ chain rule again

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a \mid \neg b, \neg e) P(\neg b, \neg e)$ CR...

Bayesian Networks—bit by bit



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

$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

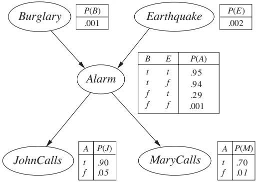
$= P(j \mid m, a, \neg b, \neg e) P(m, a, \neg b, \neg e)$ by the chain rule

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a, \neg b, \neg e)$ chain rule again

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a \mid \neg b, \neg e) P(\neg b, \neg e)$ CR...

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a \mid \neg b, \neg e) P(\neg b \mid \neg e) P(\neg e)$ CR...

Bayesian Networks—bit by bit

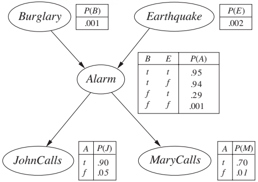


$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$
$= P(j \mid m, a, \neg b, \neg e)P(m, a, \neg b, \neg e)$ by the chain rule
$= P(j \mid m, a, \neg b, \neg e)P(m \mid a, \neg b, \neg e)P(a, \neg b, \neg e)$ chain rule again
$= P(j \mid m, a, \neg b, \neg e)P(m \mid a, \neg b, \neg e)P(a \mid \neg b, \neg e)P(\neg b, \neg e)$ CR...
$= P(j \mid m, a, \neg b, \neg e)P(m \mid a, \neg b, \neg e)P(a \mid \neg b, \neg e)P(\neg b \mid \neg e)P(\neg e)$ CR...

We can now take into consideration conditional independence (CI)

$= P(j \mid m, a, \neg b, \neg e)P(m \mid a, \neg b, \neg e)P(a \mid \neg b, \neg e)P(\neg b)P(\neg e)$ CI



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

$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$= P(j \mid m, a, \neg b, \neg e) P(m, a, \neg b, \neg e)$ by the chain rule

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a, \neg b, \neg e)$ chain rule again

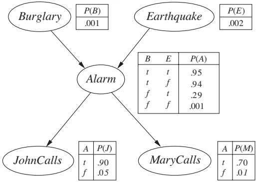
$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a \mid \neg b, \neg e) P(\neg b, \neg e)$ CR...

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a \mid \neg b, \neg e) P(\neg b \mid \neg e) P(\neg e)$ CR...

We can now take into consideration conditional independence (CI)

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a \mid \neg b, \neg e) P(\neg b) P(\neg e)$ CI

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a) P(a \mid \neg b, \neg e) P(\neg b) P(\neg e)$ CI



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

$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$= P(j \mid m, a, \neg b, \neg e)P(m, a, \neg b, \neg e)$ by the chain rule

$= P(j \mid m, a, \neg b, \neg e)P(m \mid a, \neg b, \neg e)P(a, \neg b, \neg e)$ chain rule again

$= P(j \mid m, a, \neg b, \neg e)P(m \mid a, \neg b, \neg e)P(a \mid \neg b, \neg e)P(\neg b, \neg e)$ CR...

$= P(j \mid m, a, \neg b, \neg e)P(m \mid a, \neg b, \neg e)P(a \mid \neg b, \neg e)P(\neg b \mid \neg e)P(\neg e)$ CR...

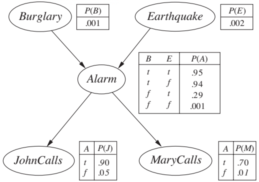
We can now take into consideration conditional independence (CI)

$= P(j \mid m, a, \neg b, \neg e)P(m \mid a, \neg b, \neg e)P(a \mid \neg b, \neg e)P(\neg b)P(\neg e)$ CI

$= P(j \mid m, a, \neg b, \neg e)P(m \mid a)P(a \mid \neg b, \neg e)P(\neg b)P(\neg e)$ CI

$= P(j \mid a)P(m \mid a)P(a \mid \neg b, \neg e)P(\neg b)P(\neg e)$ CI

Bayesian Networks—bit by bit



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

$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$= P(j \mid m, a, \neg b, \neg e) P(m, a, \neg b, \neg e)$ by the chain rule

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a, \neg b, \neg e)$ chain rule again

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a \mid \neg b, \neg e) P(\neg b, \neg e)$ CR...

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a \mid \neg b, \neg e) P(\neg b \mid \neg e) P(\neg e)$ CR...

We can now take into consideration conditional independence (CI)

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a, \neg b, \neg e) P(a \mid \neg b, \neg e) P(\neg b) P(\neg e)$ CI

$= P(j \mid m, a, \neg b, \neg e) P(m \mid a) P(a \mid \neg b, \neg e) P(\neg b) P(\neg e)$ CI

$= P(j \mid a) P(m \mid a) P(a \mid \neg b, \neg e) P(\neg b) P(\neg e)$ CI

$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062$

- That still sounds really complicated!!

- That still sounds really complicated!!
  Will I need to redo all of those simplifications?

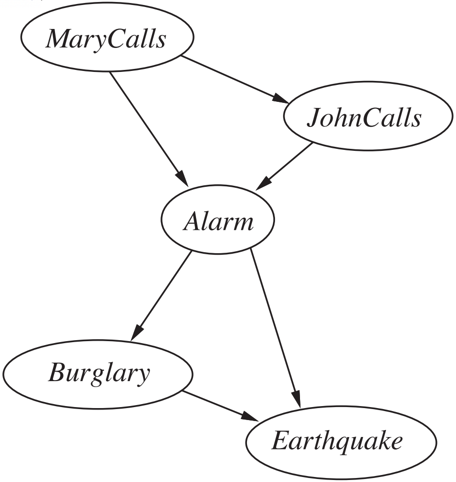
- That still sounds really complicated!!
  Will I need to redo all of those simplifications?
- Luckily, no! (But still need to multiply a lot, for a human anyway)
- The inference algorithm needs you to separate variables into three groups
  - Query variables (this is what you *want to know*)
  - Evidence variables (this is what you *know*)
  - Hidden variables (this is what you *don't care*)
- In a naïve algorithm, you just pick all relevant lines of the CPTs and mix them up!
  - Sum out all hidden variables (**marginalisation**)
  - Fix the numbers for the lines with evidence variables (**conditioning**)
  - Compute both true and false probabilities of the query for **normalization**
- Conditional independence will allow us to ignore many of these numbers

- Our goal is to compute the posterior probability distribution for a set of query variables X given some observed event – assignment of values e to a set of evidence variables E, i.e. $P(X \mid e)$
- E.g. How likely is it that a burglary occurred given that both John and Mary called? $P(Burglary \mid JohnCalls = true, MaryCalls = true)$
- Query Variable: $Burglary$, hidden variables are $Earthquake$ and $Alarm$

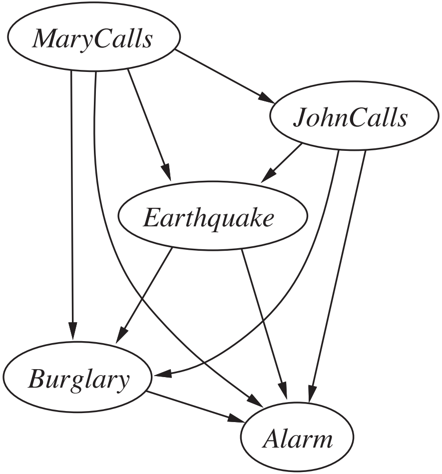
$$P(B \mid j, m) = \alpha P(B, j, m) = \alpha \sum_e \sum_a P(B, e, a, j, m)$$

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

- α is a normalising factor to make sure that all the probabilities sum to one, we can compute it from $P(b \mid j, m)$ and $P(\neg b \mid j, m)$. From this we see that the probability of a burglary is around 28%.


MaryCalls
JohnCalls
Alarm
Burglary
Earthquake

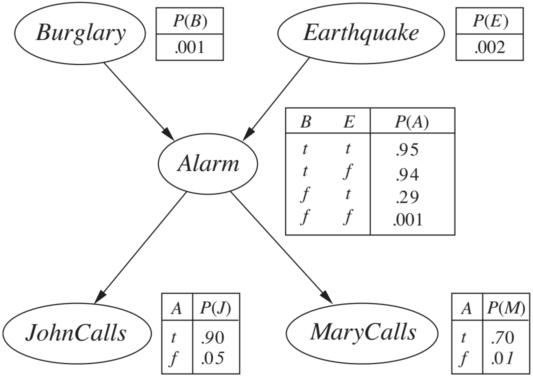
(a)


MaryCalls
JohnCalls
Earthquake
Burglary
Alarm

(b)

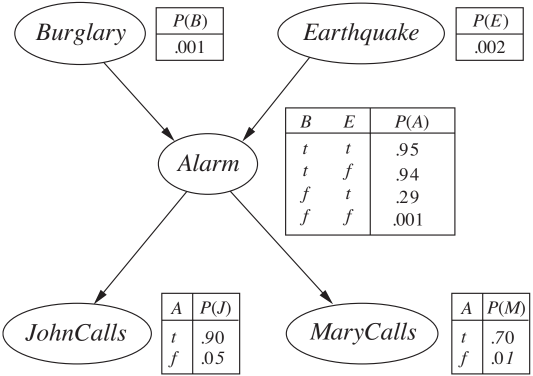
Given $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$

Compute the probability that I will get warned of a burglary when there is no earthquake (the alarm rang)



Exercise

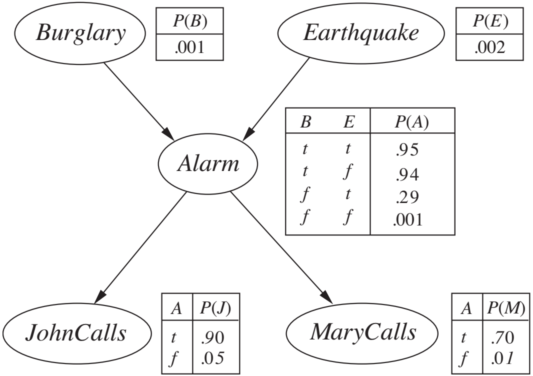
Given $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$



Compute the probability that I will get warned of a burglary when there is no earthquake (the alarm rang)

$$
\begin{aligned}
&P(j \mid a)P(m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
&+P(j \mid a)P(\neg m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
&+P(\neg j \mid a)P(m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
&= .00091
\end{aligned}
$$

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



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

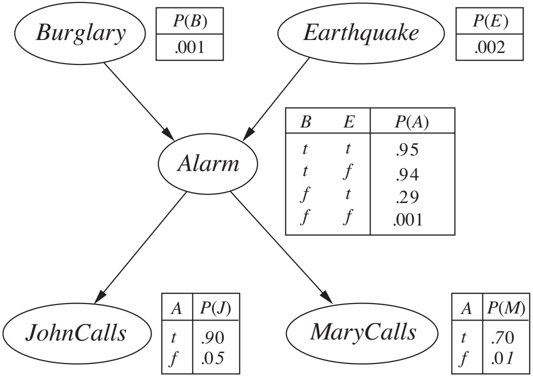
A	P(M)
t	.70
f	.01

Compute the probability that I will get warned of a burglary when there is no earthquake (the alarm rang)

$$
\begin{aligned}
& P(j \mid a)P(m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
& +P(j \mid a)P(\neg m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
& +P(\neg j \mid a)P(m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
& = .00091
\end{aligned}
$$

Is this correct?

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



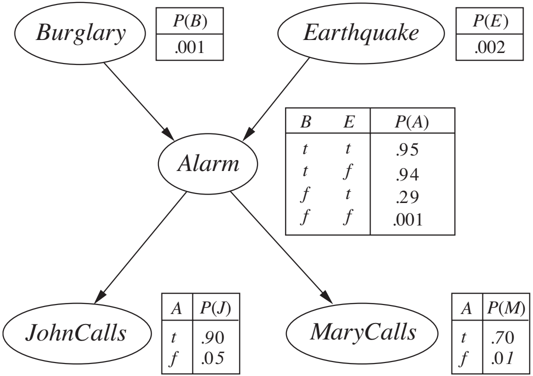
Compute the probability that I will get warned of a burglary when there is no earthquake (the alarm rang)

$$
\begin{aligned}
& P(j \mid a)P(m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
& + P(j \mid a)P(\neg m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
& + P(\neg j \mid a)P(m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
& = .00091
\end{aligned}
$$

Is this correct? Let us see the complementary case:

$$
\begin{aligned}
& P(\neg j \mid a)P(\neg m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
& = .00003
\end{aligned}
$$

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



Compute the probability that I will get warned of a burglary when there is no earthquake (the alarm rang)

$$
\begin{aligned}
&P(j \mid a)P(m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
&+P(j \mid a)P(\neg m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
&+P(\neg j \mid a)P(m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
&= .00091
\end{aligned}
$$

Is this correct? Let us see the complementary case:

$$
\begin{aligned}
&P(\neg j \mid a)P(\neg m \mid a)P(a \mid b \wedge \neg e)P(b)P(\neg e) \\
&= .00003
\end{aligned}
$$

$$
\begin{aligned}
P(A \mid b, \neg e) &= \alpha\langle .00091, .00003\rangle \\
&= \frac{1}{.00094}\langle .00091, .00003\rangle \\
&\approx \langle .97, .03\rangle
\end{aligned}
$$

Normalizer in Bayes Network

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

$$P(a \mid b) + P(\neg a \mid b) = 1$$

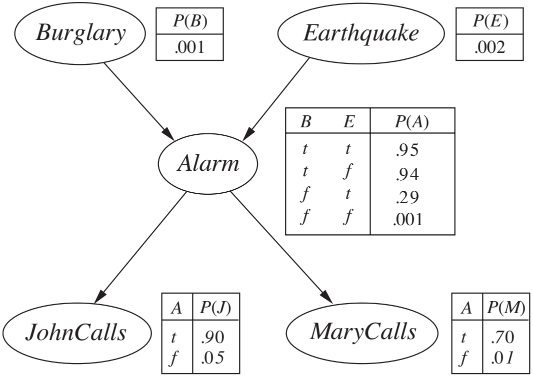
$$P'(a \mid b) = P(b \mid a)P(a)$$
$$P'(\neg a \mid b) = P(b \mid \neg a)P(\neg a)$$

$$P(a \mid b) = \alpha P'(a \mid b)$$
$$P(\neg a \mid b) = \alpha P'(\neg a \mid b)$$

$$\alpha = (P'(a \mid b) + P'(\neg a \mid b))^{-1}$$

Conditional Independence in BNs - simplification in action

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



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

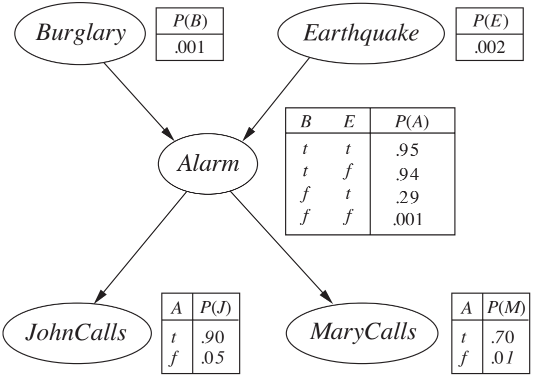
A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Re-compute the probability that I will get warned of a burglary when there is no earthquake (the alarm rang)

Conditional Independence in BNs - simplification in action

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



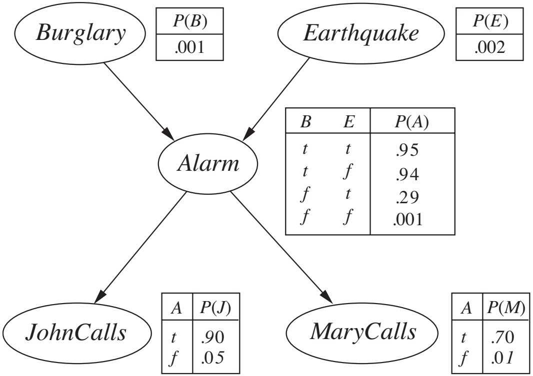
Re-compute the probability that I will get warned of a burglary when there is no earthquake (the alarm rang)

$$P(j \mid a)P(m \mid a)P(a \mid b \land \neg e)$$
$$+P(j \mid a)P(\neg m \mid a)P(a \mid b \land \neg e)$$
$$+P(\neg j \mid a)P(m \mid a)P(a \mid b \land \neg e)$$
$$= .9215$$

$$P(\neg j \mid a)P(\neg m \mid a)P(a \mid b \land \neg e)$$
$$= .0285$$

Conditional Independence in BNs - simplification in action

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



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Re-compute the probability that I will get warned of a burglary when there is no earthquake (the alarm rang)

$$P(j \mid a)P(m \mid a)P(a \mid b \wedge \neg e) \\ +P(j \mid a)P(\neg m \mid a)P(a \mid b \wedge \neg e) \\ +P(\neg j \mid a)P(m \mid a)P(a \mid b \wedge \neg e) \\ = .9215$$

$$P(\neg j \mid a)P(\neg m \mid a)P(a \mid b \wedge \neg e) \\ = .0285$$

$$\begin{aligned} P(A \mid b, \neg e) &= \alpha\langle .9215, .0285\rangle \\ &= \tfrac{1}{.95}\langle .9215, .0285\rangle \\ &\approx \langle .97, .03\rangle \end{aligned}$$

- Probability Theory (quick recap)
- Bayes Rule
- Bayesian Networks
  - Graphical Semantics
  - Inference