Introduction to Artificial Intelligence

COMP307
Introduction to Bayesian Network

Yi Mei

yi.mei@ecs.vuw.ac.nz

Outline

- Rules from previous lectures
- What is Bayesian Networks
- Why Bayesian Networks
- Cause — Effect
- Summary

Thomas Bayes (/ˈbeɪz/; c. 1701 – 7 April 1761)

Rules from Previous Lectures

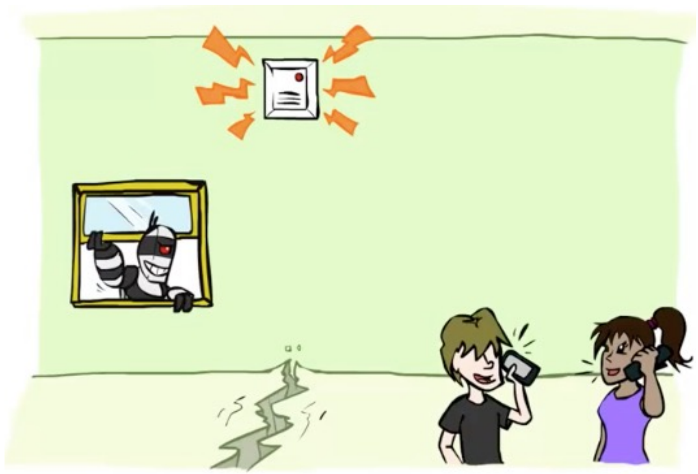
- Product Rule
  - $P(X_1, \ldots X_n, Y_1, \ldots, Y_m) = P(X_1, \ldots, X_n) * P(Y_1, \ldots, Y_m | X_1, \ldots, X_n)$
- Sum Rule:
  - $P(X_1, \ldots, X_n) = \sum_{y_1, \ldots, y_m} P(X_1, \ldots, X_n, Y_1 = y_1, \ldots, Y_m = y_m)$
- Normalisation Rule
  - $\sum_{x_1, \ldots, x_n} P(X_1 = x_1, \ldots, X_n = x_n) = 1,$
  - $\sum_{x_1, \ldots, x_n} P(X_1 = x_1, \ldots, X_n = x_n | Y_1, \ldots, Y_m) = 1$
- Independence
  - $X \perp Y$, $P(X|Y) = P(X)$, $P(X, Y) = P(X) * P(Y)$
  - $X \perp Y | Z$, $P(X| Y, Z) = P(X|Z)$, $P(X, Y|Z) = P(X|Z) * P(Y|Z)$
- Bayes Rule
  - $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
  - $P(Y|X_1, \ldots, X_n) = \frac{P(X_1|Y) * \cdots * P(X_n|Y) * P(Y)}{P(X_1, \ldots, X_n)}$ [assume conditional independence]

Alarm Network

- Your house is installed an alarm against burglary
  - The alarm will usually be set off by burglars
  - but sometimes it may also be set off by earthquakes
  - There are two neighbours, John and Mary
  - John and Mary might call you when they hear the alarm
  - They might also call you for other issues without alarm

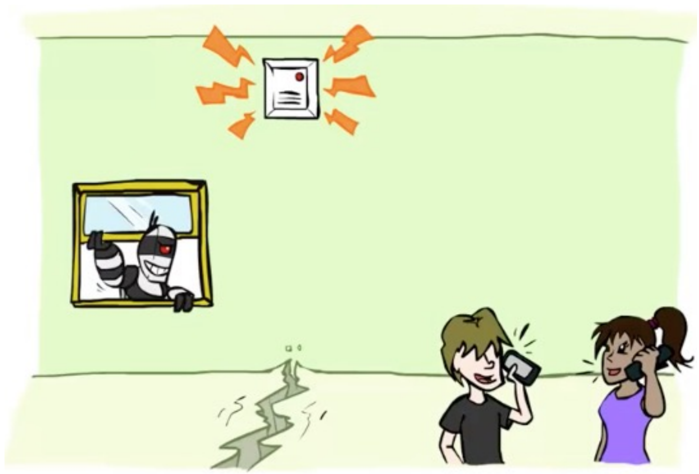
- Variables**:**
  - Burglary, Earthquake, Alarm, JohnCalls, MaryCalls
  - All binary (**true** or **false)**

- Relationship between them?
  - Cause -> Effect



Alarm Network

- Domain **causal knowledge (causes and effects)**
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call



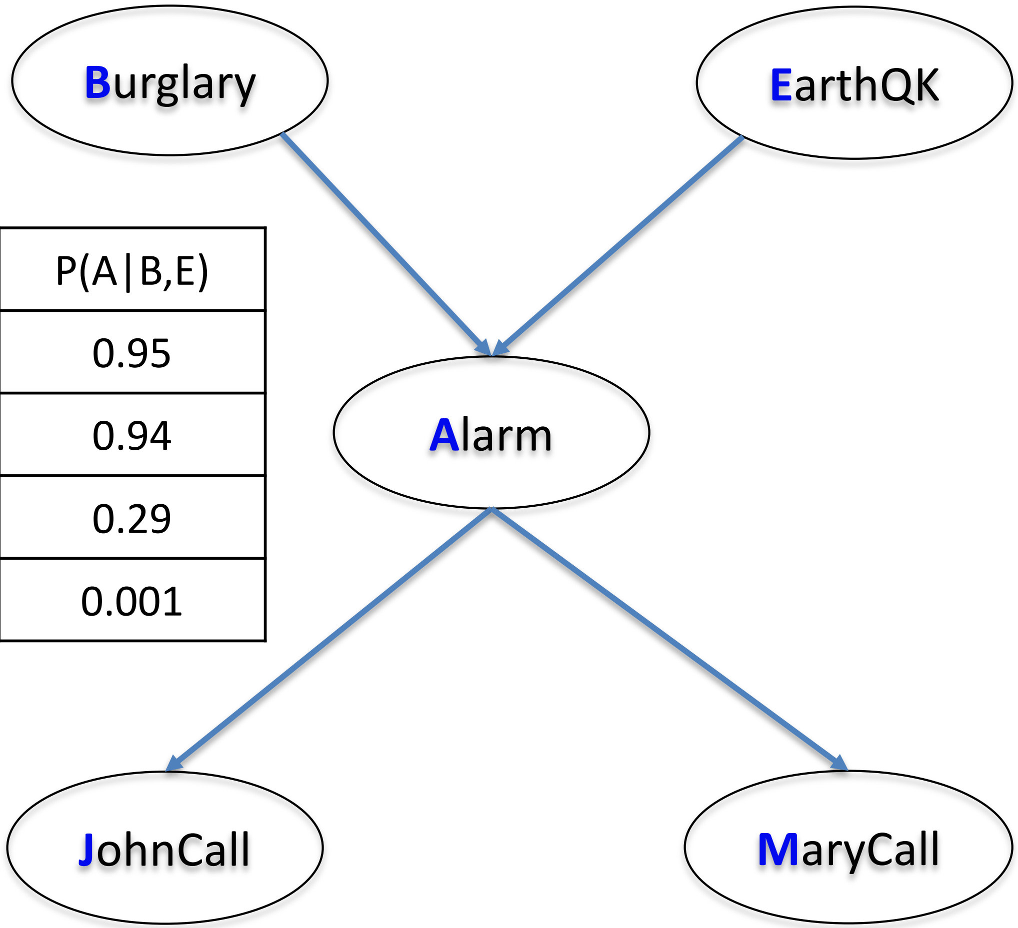
Alarm Network

- **Conditional Probability Tables**
  - Quantify likelihood under different conditions/situations

B	P(B)
T	0.001

E	P(E)
T	0.002

B	E	A	P(A|B,E)
T	T	T	0.95
T	F	T	0.94
F	T	T	0.29
F	F	T	0.001



A	J	P(J|A)
T	T	0.9
F	T	0.05

A	M	P(M|A)
T	T	0.7
F	T	0.01

Bayesian Networks

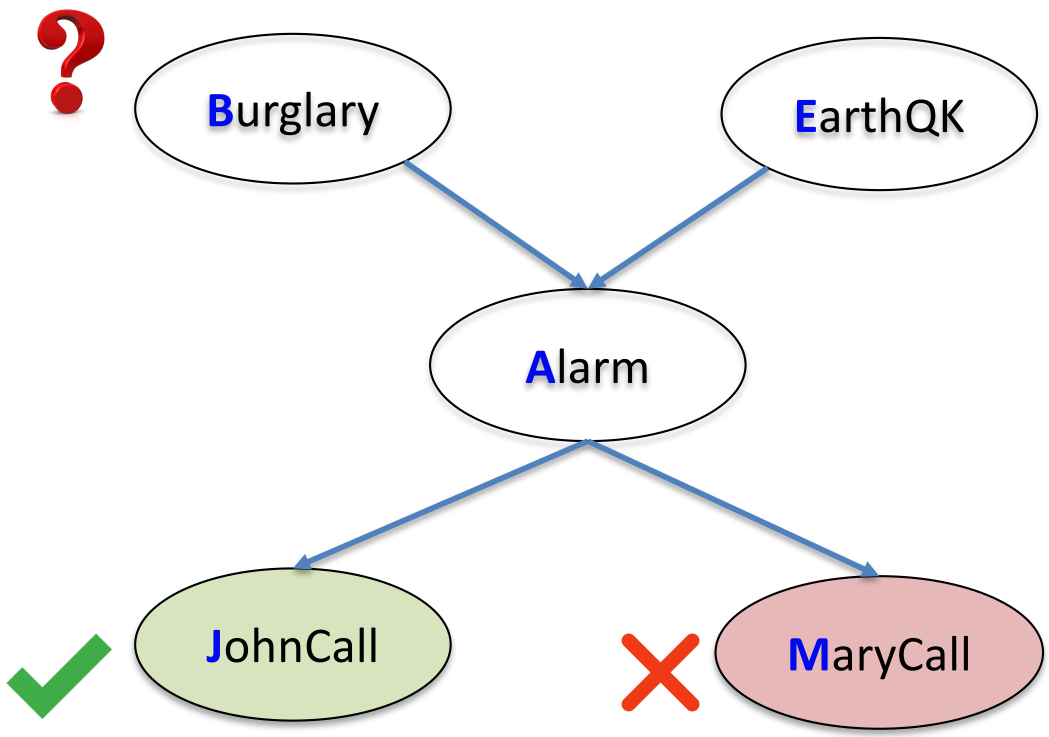
- Bayesian networks (BNs): a graphical representation of a probabilistic dependency model
  - also known as Belief networks (or Bayes nets for short)
  - Belong to the family of probabilistic graphical models (GMs).
    - Other GMs: Markov network …

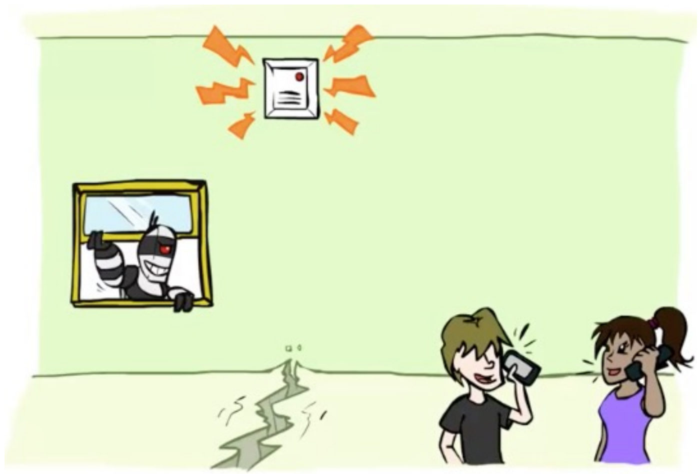
- These graphical structures are used to represent knowledge about an uncertain domain.
  - each node in the graph represents a random variable,
  - the edges between the nodes represent probabilistic dependencies among the corresponding random variables.
  - The conditional dependencies in the graph are often estimated by using known statistical and computational methods.

- BNs combine principles from graph theory, probability theory, computer science, and statistics.

Bayesian Networks

- Each node or variable may take one of a number of possible states or values.
- The belief each of these values is determined from the belief in each possible value of every node directly connected to it and its relationship with each of these nodes.
- The belief in each state of a node is updated whenever the belief in each state of any directly connected node changes.





Semantics of Bayesian Networks

- A set of nodes, one for a variable X
- A directed, acyclic graph
  - Each edge shows the direct influence between parent and child
  - A child depends on its parents
- A conditional probability table for each node
  - a collection of distributions over X, one for each combination of parents values

$$P(X \mid a_1, \ldots, a_n)$$

  - (usually) description of a "causal" process

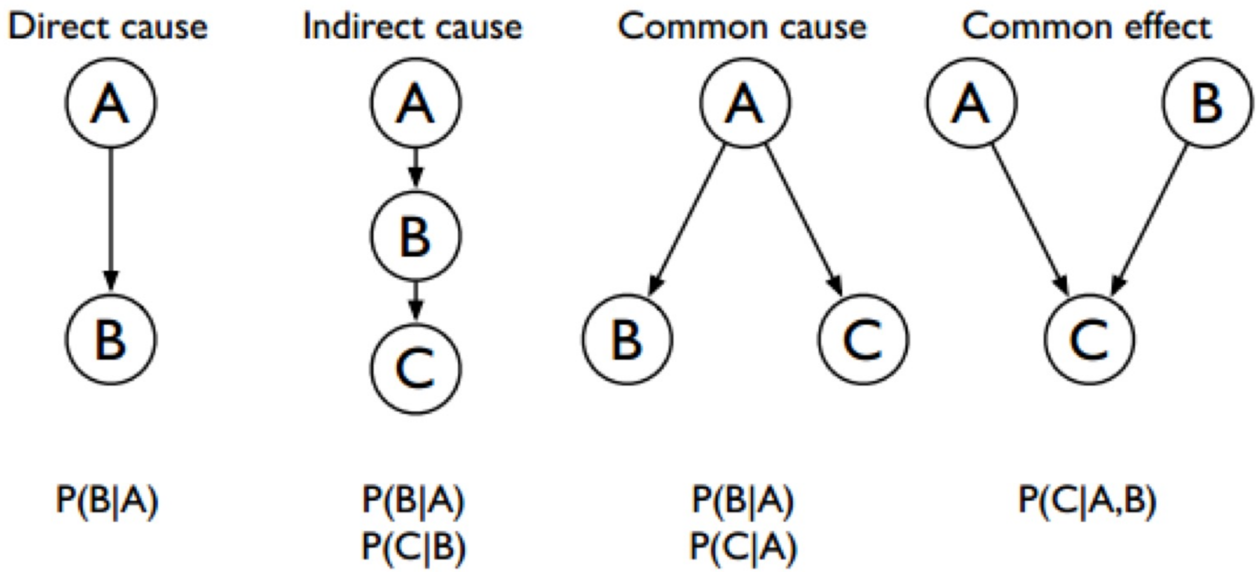
A Bayes Net = Topology (graph) + Local Conditional Probabilities

Why Bayesian Networks

- Several advantages for data analysis:
  - the model encodes dependencies among all variables, it readily handles situations where some data entries are missing.
  - a Bayesian network can be used to learn causal relationships, and hence can be used to gain understanding about a problem domain and to predict the consequences of intervention.
  - the model has both a causal and probabilistic semantics, it is an ideal representation for combining prior knowledge (which often comes in causal form) and data.
  - Bayesian statistical methods in conjunction with Bayesian networks offer an efficient and principled approach for avoiding the overfitting of data.

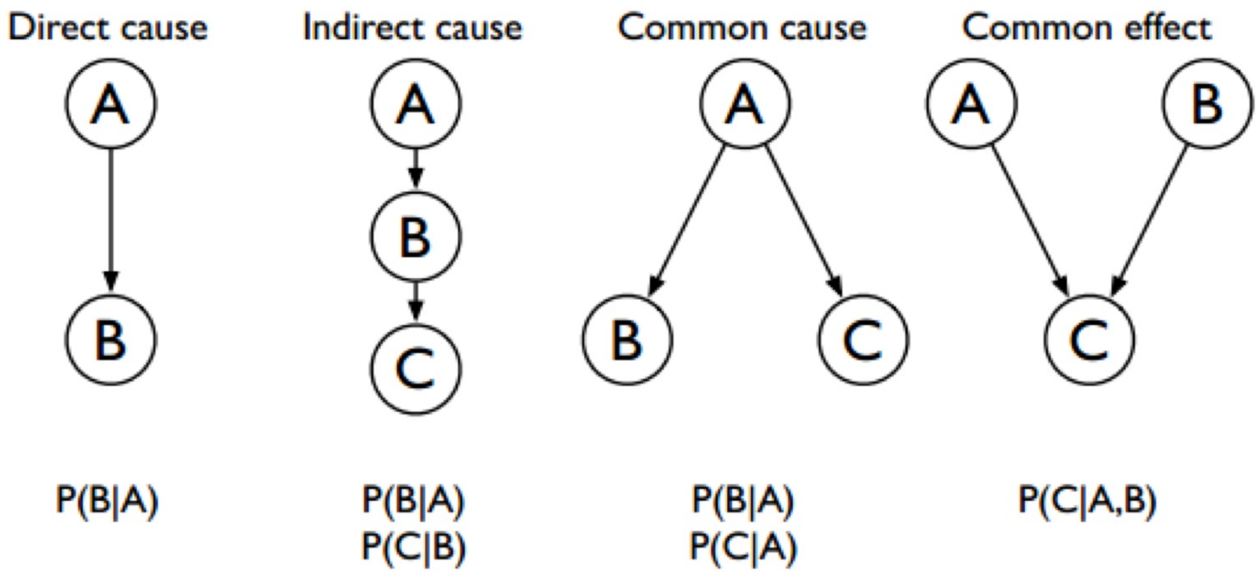
(In)Dependencies in BN

- [**Direct cause**]: A is a direct cause of B
  - A and B are dependent
- [**Indirect cause**]: A is a direct cause of B, B is a direct cause of C
  - A and C are independent given B



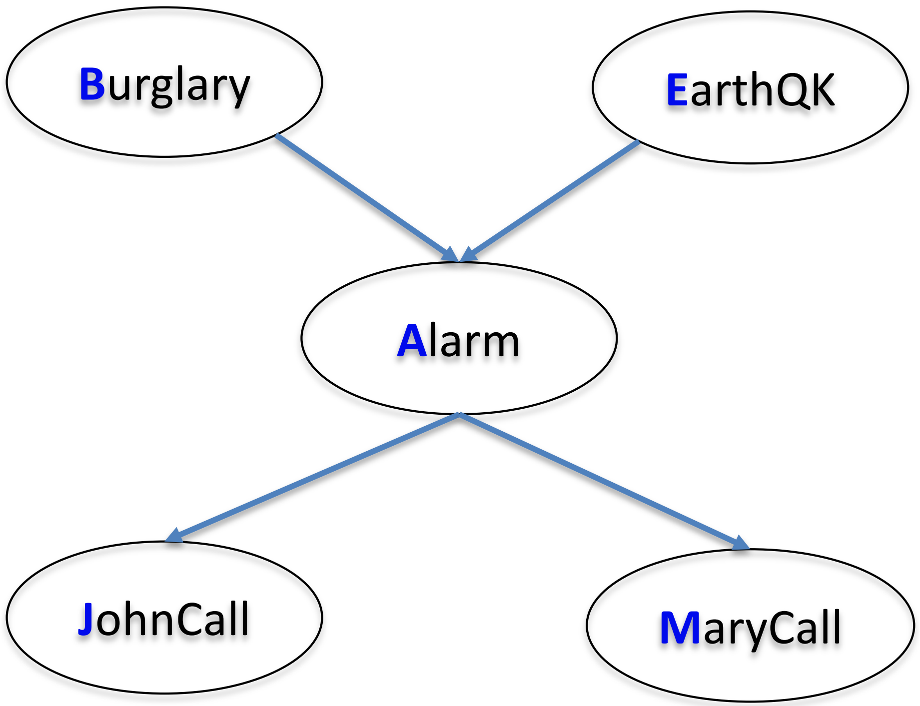
(In)Dependencies in BN

- [**Common Cause**]: A is a **common direct cause** of B and C
  - B and C are dependent **(if A is not given)**
  - B and C are independent **given A**
- [**Common Effect**]: C is a **common direct effect** of A and B
  - A and B are independent **(if C is not given)**
  - A and B are dependent **given C** ("explaining away")



(In)Dependencies in BN

- Which are true?
  - B and E are independent
  - B and E are independent given A
  - B and M are independent
  - B and M are independent given A
  - J and M are independent
  - J and M are independent given A

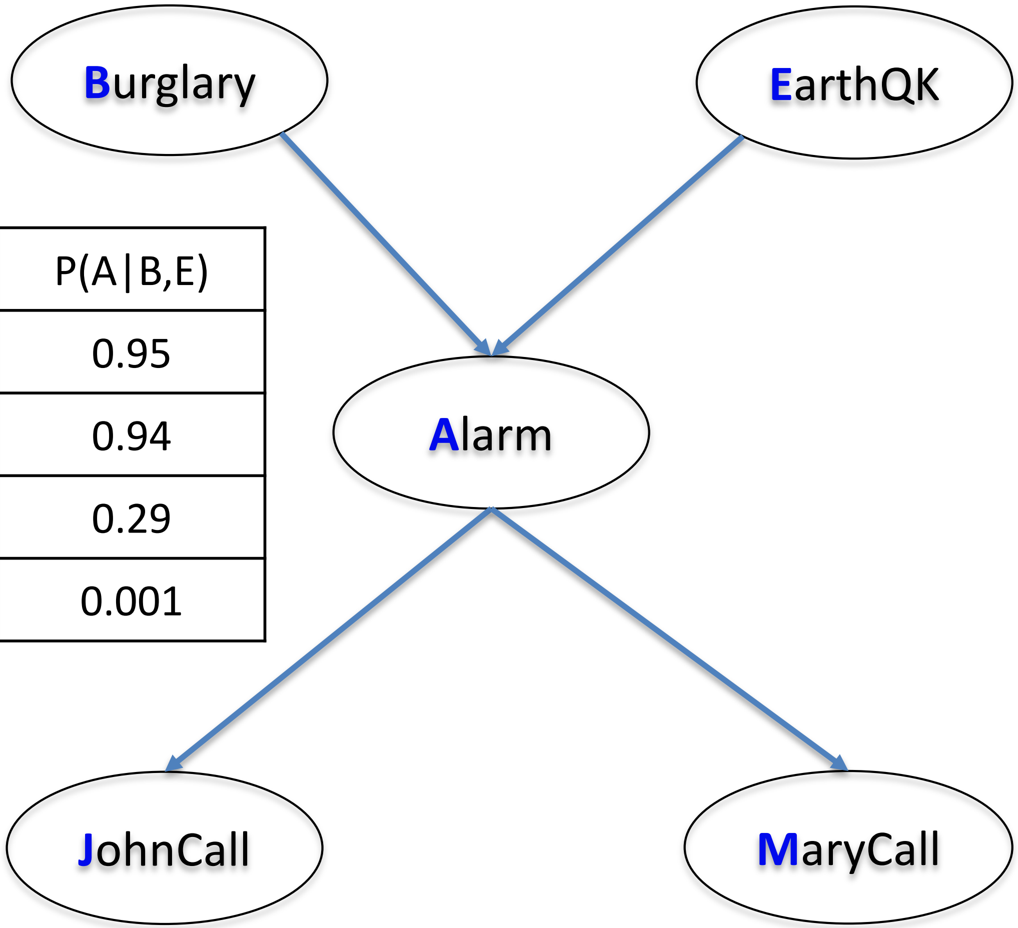


Factorisation

- In a Bayesian network, a node is conditionally independent from **all the nodes except its direct effects**, if the **direct causes are all given**, and **no direct effect is given**.

B	P(B)
T	0.001

E	P(E)
T	0.002



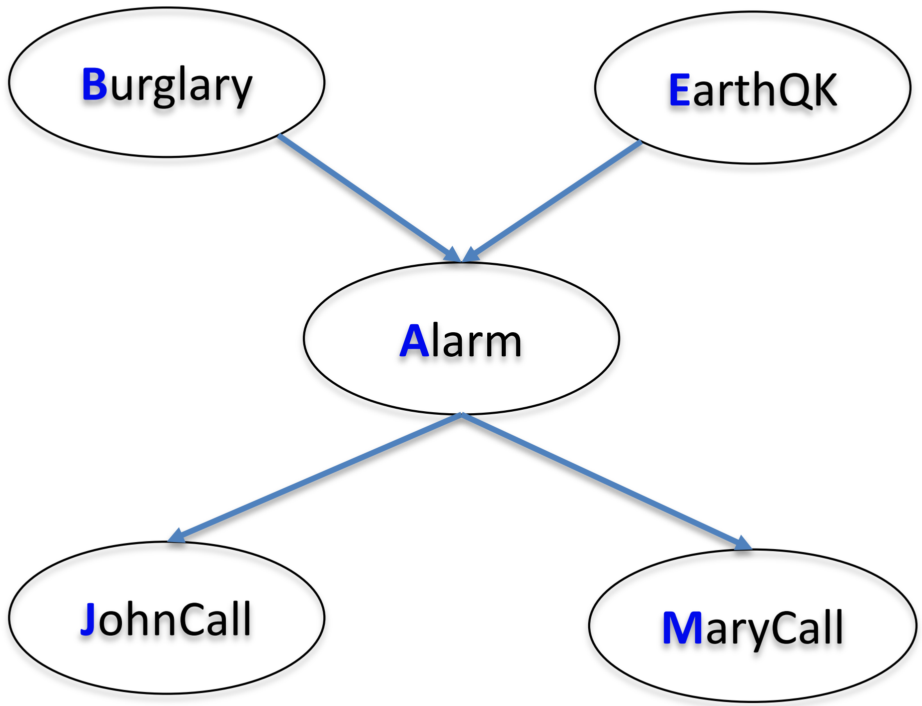
B	E	A	P(A|B,E)
T	T	T	0.95
T	F	T	0.94
F	T	T	0.29
F	F	T	0.001

A	J	P(J|A)
T	T	0.9
F	T	0.05

A	M	P(M|A)
T	T	0.7
F	T	0.01

Factorisation

- If we sort the variables so that the causes are always before the effects, e.g., [B, E, A, J, M], then we have
  - $P(X_i \mid X_1, \ldots, X_{i-1}) = P(X_i \,\big|\, parents(X_i))$
  - $P(E \mid B) = P(E)$
  - $P(J \mid A, B, E) = P(J \mid A)$
  - …



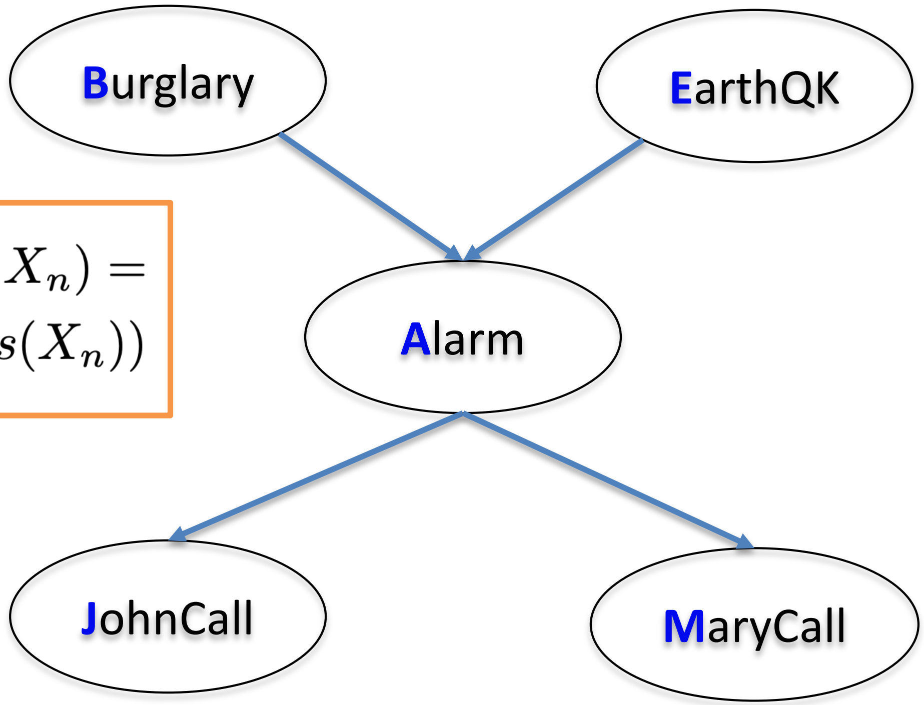
Factorisation

- Combine the chain (product) rule

$$
\begin{aligned}
P(B,E,A,J,M) &= P(B) * P(E|B) * P(A|B,E) * P(J|B,E,A) * P(M|B,E,A,J) \\
&= P(B) * P(E) * P(A|B,E) * P(J|A) * P(M|A)
\end{aligned}
$$

- Bayesian network is a **factorization** of the joint probability distribution of the variables

$$
\boxed{\begin{array}{r} P(X_1, \ldots, X_n) = \\ P(X_1 \mid parents(X_1)) * \ldots * P(X_n \mid parents(X_n)) \end{array}}
$$



Summary

- Bayesian networks
  - A directed acyclic graph
  - Represent conditional dependencies between variables
  - Conditional distribution tables
- Independencies between variables in BN
- Factorisation

- Next lecture: how to build a BN