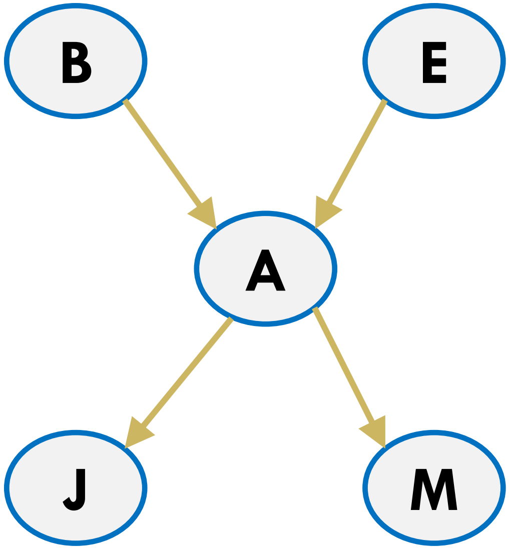
CS 4104
APPLIED MACHINE LEARNING

Dr. Hashim Yasin

National University of Computer and Emerging Sciences,

Faisalabad, Pakistan.



BAYESIAN BELIEF NETWORKS

Bayesian Belief Networks

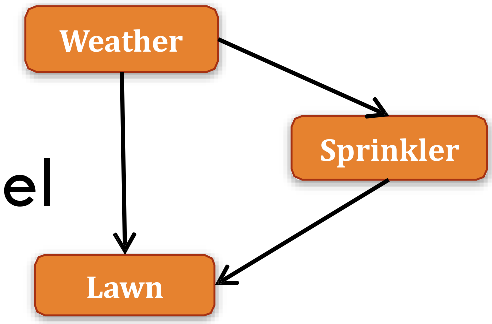
Naive Bayes classifier

- which assumes that ***all*** the variables are conditionally independent given the value of the target variable,

Bayesian belief networks

- which allows stating conditional independence assumptions that apply to ***subsets*** of the variables.
- It provide an *intermediate approach that is less constraining* than the global assumption of conditional independence made by the naive Bayes classifier.

Bayesian Belief Networks



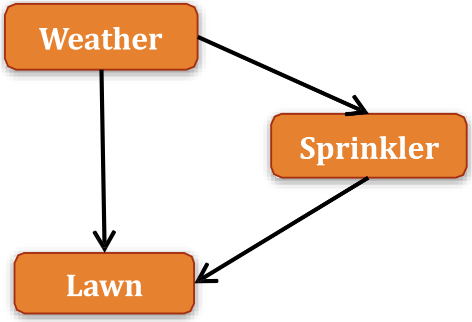
<u>Bayesian Belief Networks (BBN)</u>

- Graphical (Directed Acyclic Graph) Model
- **Nodes** are the features:
  - Each has a set of possible parameters/values/ states
  - Weather = {sunny, cloudy, rainy}; Sprinkler = {off, on}; Lawn = {dry, wet}
- **Edges / Links** represent relations between features
- BBN is a probabilistic graphical model (PGM)
- Each node/feature is a random variable

Bayesian Belief Networks

Bayesian Belief Networks

- We call these probabilities of occurring states - **Beliefs**
  - Example: our belief in the state {*coin*=‘*head*’} is 50%
- All beliefs of all possible states of a node are gathered in a single CPT - Conditional Probability Table
- *BBN is a 2-component model:*
  - Graph
  - CPTs



Bayesian Belief Networks

Joint Space

- set of random variables $Y_1 \ldots Y_n$,
- each variable Y_i can take on the set of possible values $V(Y_i)$.
- The ***joint space*** of the set of variables Y is the cross product $V(Y_1) \times V(Y_2) \times \ldots V(Y_n)$.
- The probability distribution over this joint' space is called the ***joint probability distribution***

Bayesian Belief Networks

- **A and B are independent**

$$P(A,B) = P(A)P(B)$$

- **A and B are <u>conditionally independent</u> given C**

$$P(A,B \mid C) = P(A \mid C)P(B \mid C)$$

$$P(A \mid C,B) = P(A \mid C)$$

Bayesian Belief Networks

- Bayesian Belief Network represents the **full joint distribution** over the set of variables more compactly with a **smaller number of parameters.**
- **Full joint distribution** is defined in terms of local conditional distributions (obtained via the chain rule).

- **How do we get the local parameterizations?**

Answer:

- **Graphical structure** encodes **conditional and marginal independences** among random variables

Bayesian Belief Networks

Network represents joint probability distribution over all variables

- In general,

$$\boxed{P(y_1, \ldots, y_n) = \prod_{i=1}^{n} P(y_i | Parents(Y_i))}$$

where $Parents(Y_i)$ denotes immediate predecessors of Y_i in graph

Bayesian Belief Networks

Network represents joint probability distribution over all variables

- In general,

$$\boxed{P(y_1, \ldots, y_n) = \prod_{i=1}^{n} P(y_i | Parents(Y_i))}$$

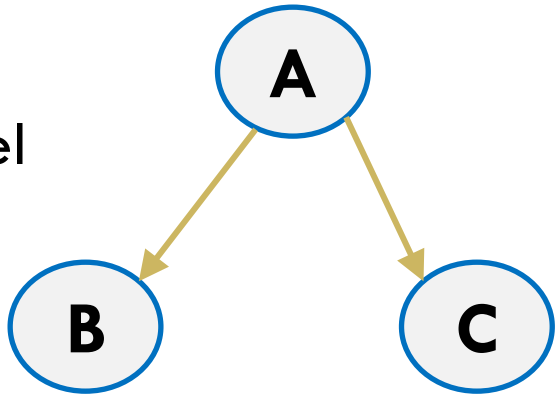
  where $Parents(Y_i)$ denotes immediate predecessors of Y_i in graph

- So joint distribution is fully defined by graph, plus the $P(y_i | Parents(Y_i))$

Bayesian Belief Networks … Example

Conditionally independent effects:

- $P(A,B,C) = P(B|A)P(C|A)P(A)$
- B and C are conditionally independent given A
  - For example, A is a disease, and we model B and C as conditionally independent symptoms given A
  - For example, A is Fire, B is Heat, C is Smoke.
    - "Where there's Smoke, there's Fire."
    - If we see Smoke, we can infer Fire.
    - If we see Smoke, observing Heat tells us very little additional information.



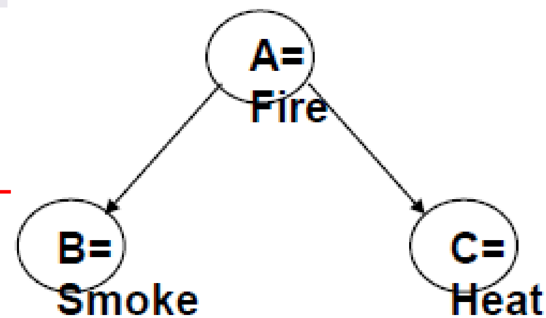
Bayesian Belief Networks … Example

- Smoke and Heat are conditionally independent given Fire.

$$P(A, B, C) = P(B|A)P(C|A)P(A)$$

What is the probability of
$P(Fire = t \,|Smoke = t, Heat = f)$**?**

P(Fire)
0. 1

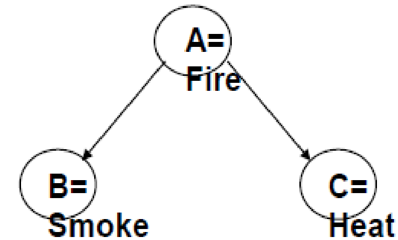


Fire	P(Smoke)
t	.90
f	.001

Fire	P(Heat)
t	.99
f	.0001

Bayesian Belief Networks … Example

P(Fire)
0. 1



Fire	P(Smoke)
t	.90
f	.001

Fire	P(Heat)
t	.99
f	.0001

What is $P(Fire = t \mid Smoke = t, Heat = f)$?

$$\boldsymbol{P(Fire = t \wedge Smoke = t, Heat = f) =}$$

$$= P(Fire = t \wedge Smoke = t \wedge Heat = f)$$

$$= P(Fire = t)P(Smoke = t|Fire = t)P(Heat = f|Fire = t)$$

$$= 0.1 \times .90 \times .01$$

$$= \ 0.0009$$

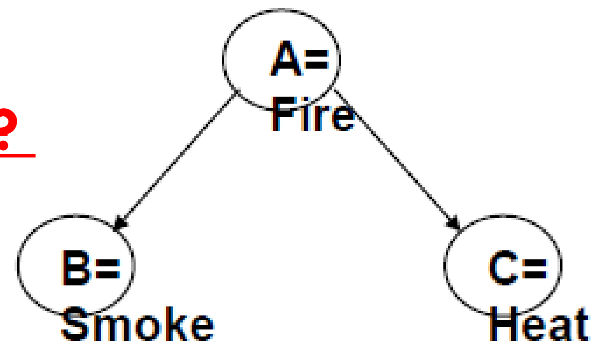
Bayesian Belief Networks … Example

- Smoke and Heat are conditionally independent given Fire.

$$P(A,B,C) \;=\; P(B|A)P(C|A)P(A)$$

What is $P(Fire = t \mid Smoke = t)$?

P(Fire)
0. 1

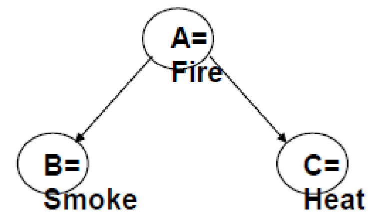


Fire	P(Smoke)
t	.90
f	.001

Fire	P(Heat)
t	.99
f	.0001

Bayesian Belief Networks … Example

P(Fire)
0. 1



Fire	P(Smoke)
t	.90
f	.001

Fire	P(Heat)
t	.99
f	.0001

What is $P(Fire = t \mid Smoke = t)$?

$$\boldsymbol{P(Fire = t \wedge Smoke = t) =}$$

$$= \sum_{heat} P(Fire = t \wedge Smoke = t \wedge heat)$$

$$= \sum_{heat} P(Fire = t)P(Smoke = t \wedge heat|Fire = t)$$

$$= \Sigma_{heat} P(Fire = t)P(Smoke = t|Fire = t)P(heat|Fire = t)$$

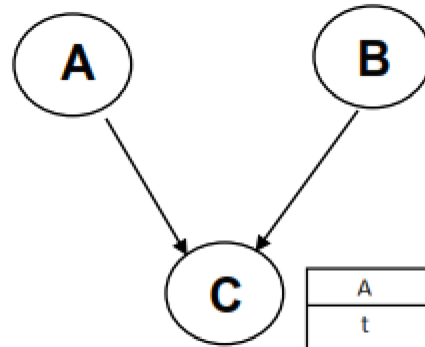
$$= P(Fire = t)P(Smoke = t|Fire = t)P(heat = t|Fire = t)$$
$$+ P(Fire = t)P(Smoke = t|Fire = t)P(heat = f|Fire = t)$$

$$= (0.1 \times 0.90 \times 0.99) + (0.1 \times 0.90 \times .01)$$
$$= 0.09$$

Bayesian Belief Networks

- $P(A,B,C) = P(C|A,B)P(A)P(B)$

P(A)
0.33



P(B)
0.67

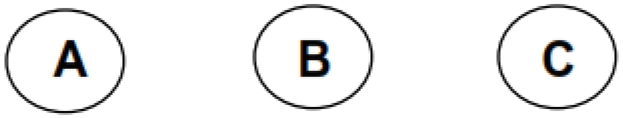
A	B	P(C)
t	t	0.2
t	f	0.4
f	t	0.3
f	f	0.3

- Dependence/independence represented via a directed graph:
  - Node = random variable
  - Directed Edge = conditional dependence
  - Absence of Edge = conditional independence

- Allows concise view of joint distribution relationships:
  - Graph nodes and edges show conditional relationships between variables.
  - Tables provide probability data.

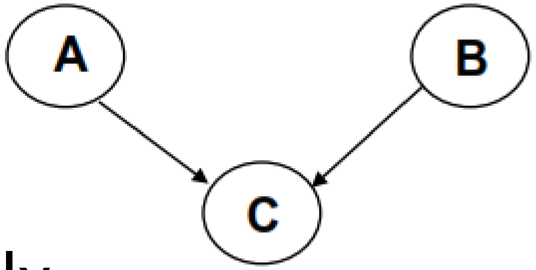
Bayesian Belief Networks

Marginal Independence:



- $P(A, B, C) = P(A)\,P(B)\,P(C)$

Independent Causes:

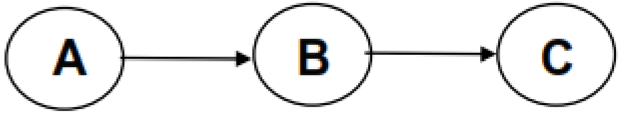


- $P(A, B, C) = P(C|A, B)P(A)P(B)$

"Explaining away" effect:

- Given C, observing A makes B less likely
- A and B are (marginally) independent but become dependent once C is known

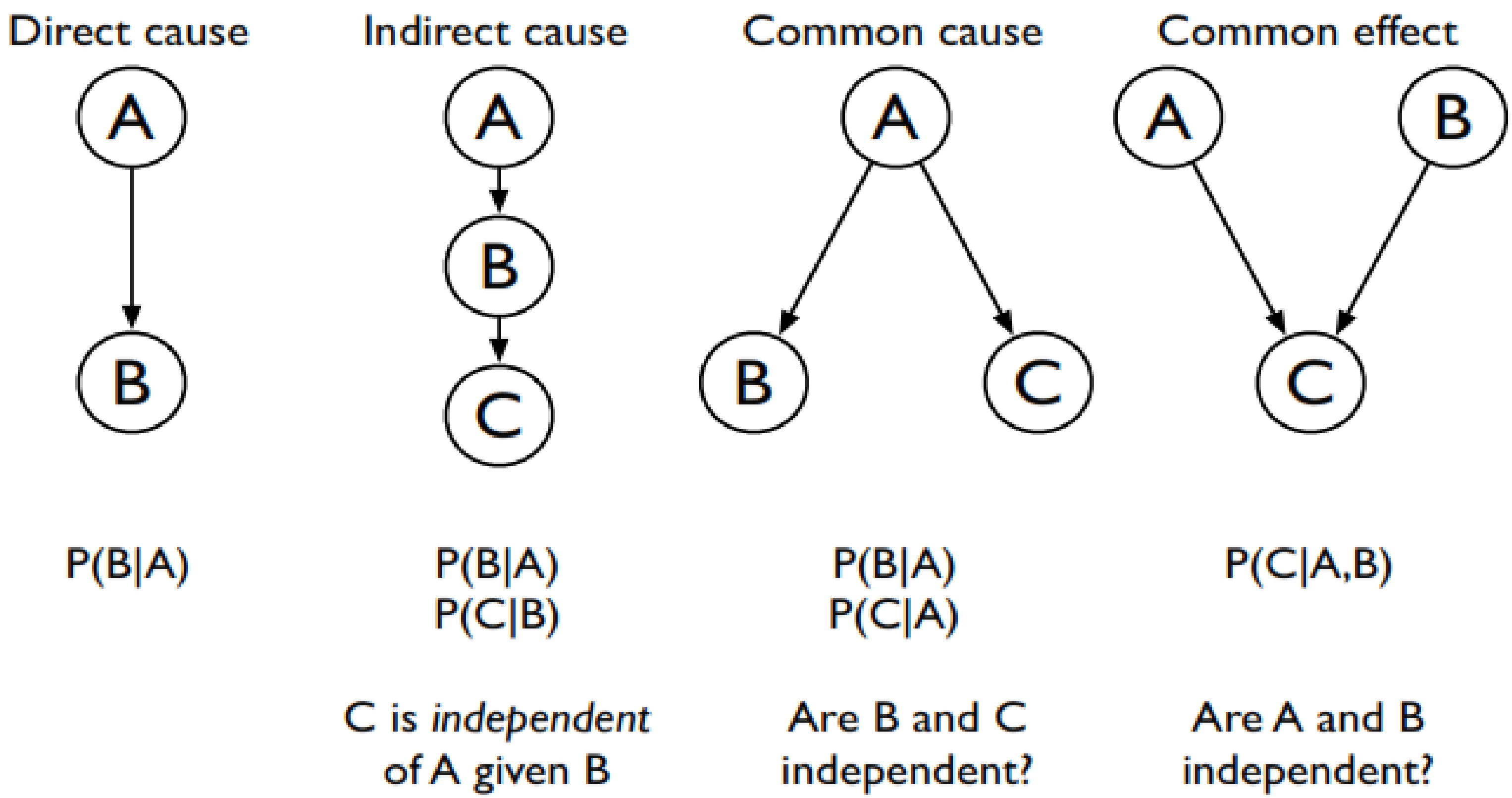
Markov dependence:

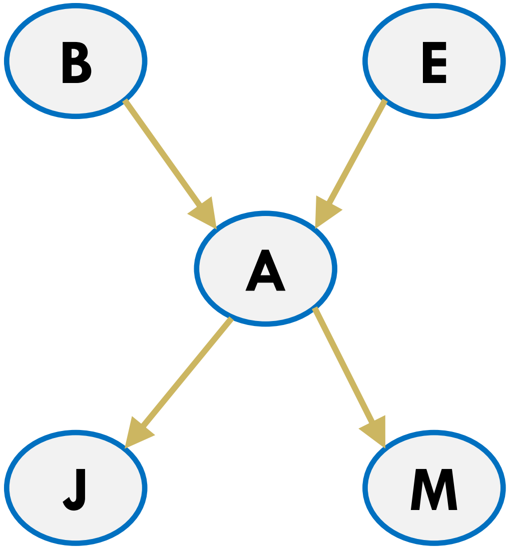


- $P(A, B, C) = P(C|B)\,P(B|A)P(A)$

Bayesian Belief Networks

- **Types of probabilistic relationships**





BAYESIAN BELIEF NETWORKS EXAMPLE

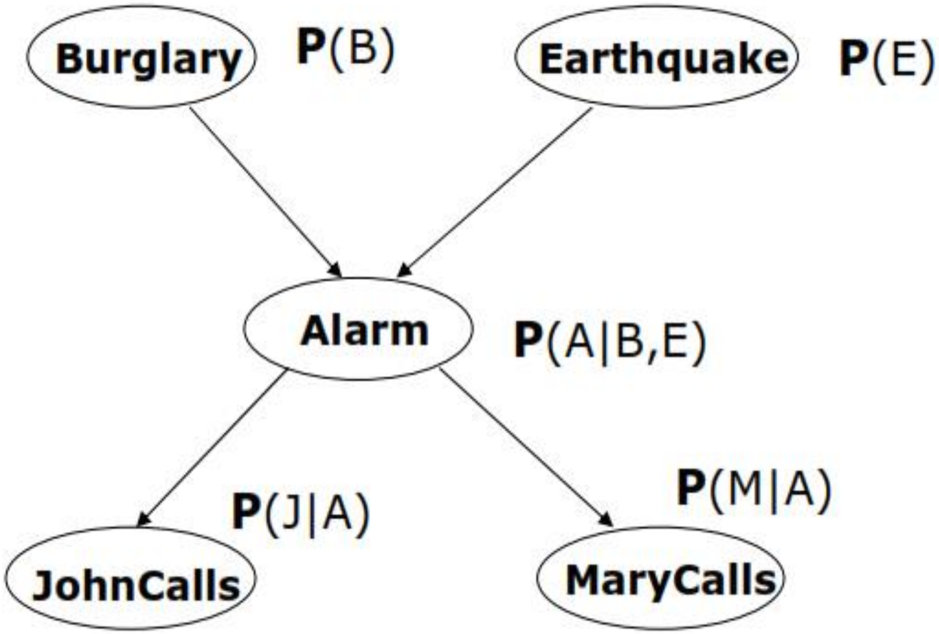
Example

- Assume your house has an **alarm system** against **burglary**. You live in the seismically active area and the alarm system can get occasionally set off by an **earthquake.**
- You have two neighbors, **Mary** and **John**, who do not know each other. *If they hear the alarm they call you, but this is not guaranteed*.
- We want to represent the probability distribution of events:
  - Burglary, Earthquake, Alarm, Mary calls and John calls

Example

Directed acyclic graph

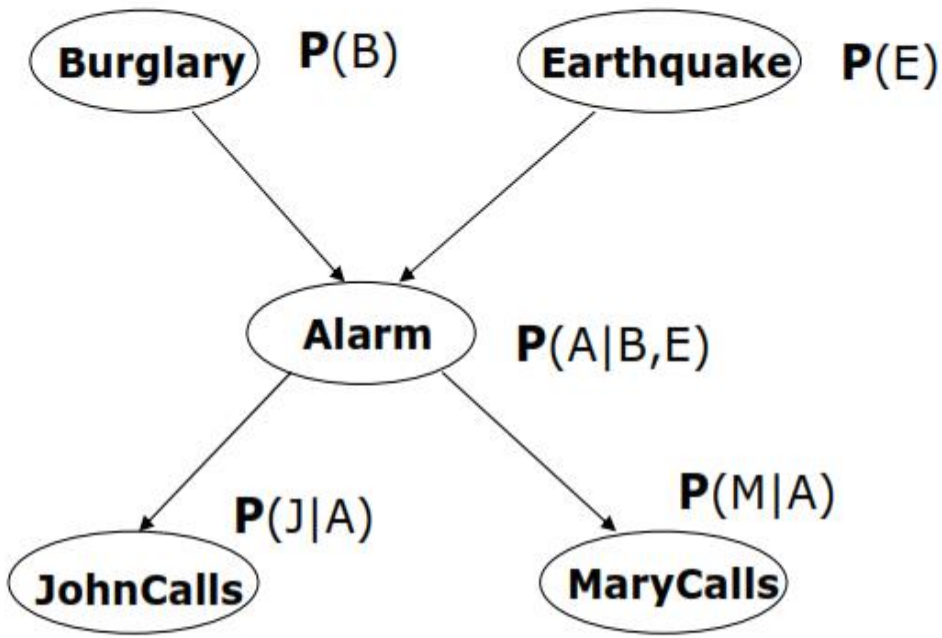
- **Nodes** = random variables (Burglary, Earthquake, Alarm, Mary calls and John calls)



Example

Directed acyclic graph

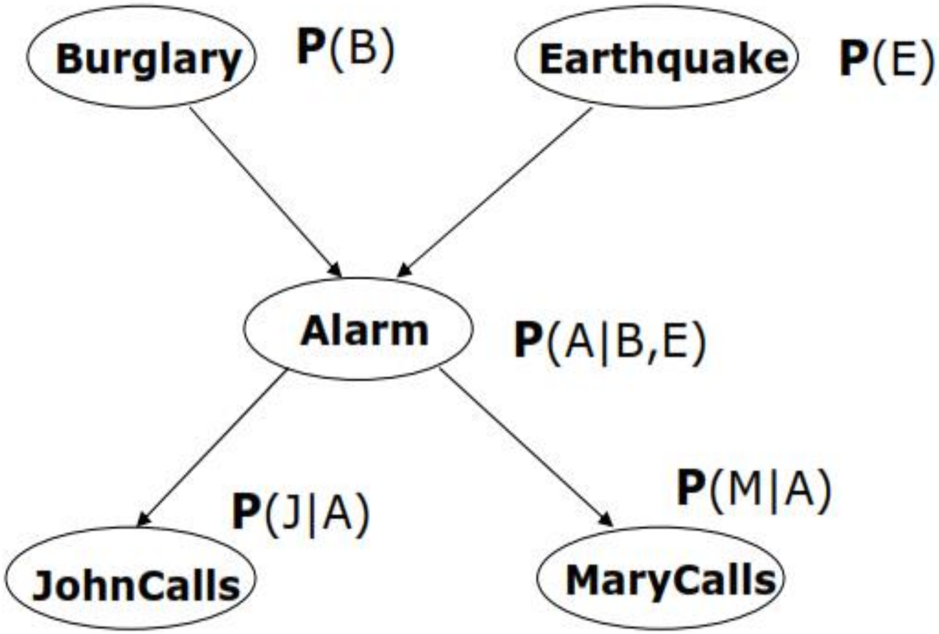
- **Links** = direct (causal) dependencies between variables.



Example

Directed acyclic graph

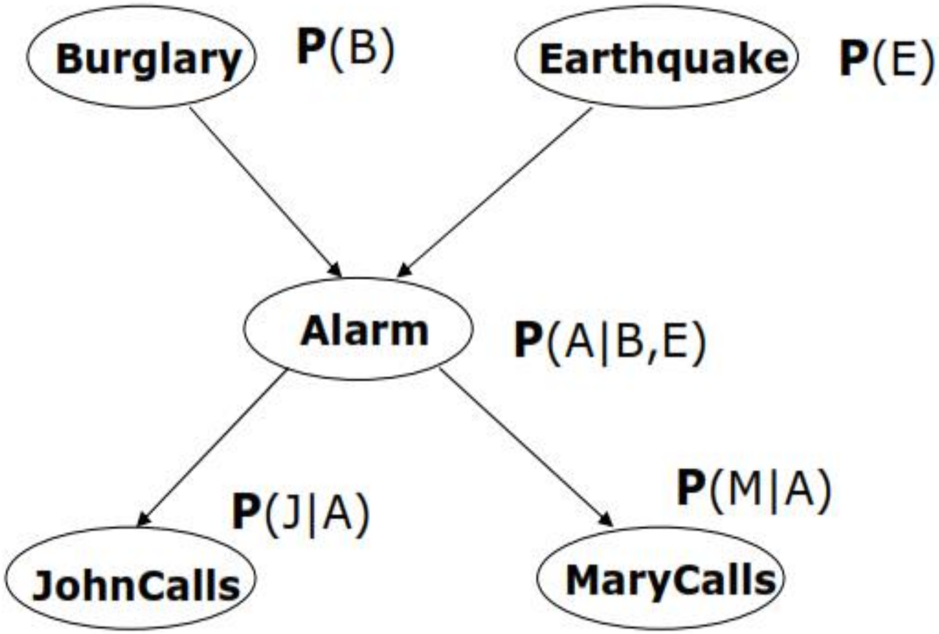
- The chance of Alarm is influenced by Earthquake, The chance of John calling is affected by the Alarm.



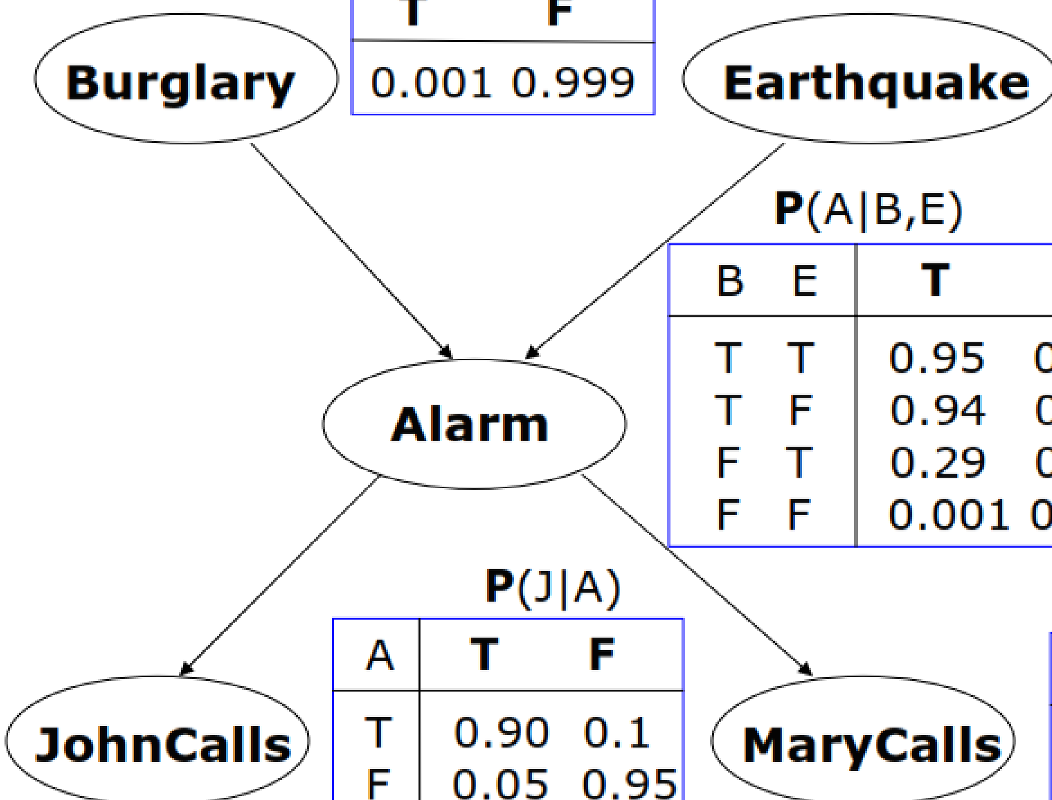
Example

Local conditional distributions

- relate variables and their parents.



Example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

P(A|B,E)

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

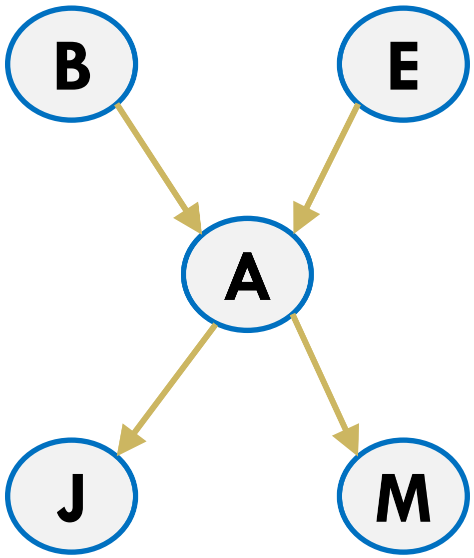
P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

Example

- **Full joint distribution:** local conditional distributions

$$\mathbf{P}(X_1, X_2, \ldots, X_n) = \prod_{i=1,\ldots n} \mathbf{P}(X_i \mid pa(X_i))$$



- Assume the assignment of values to random variables

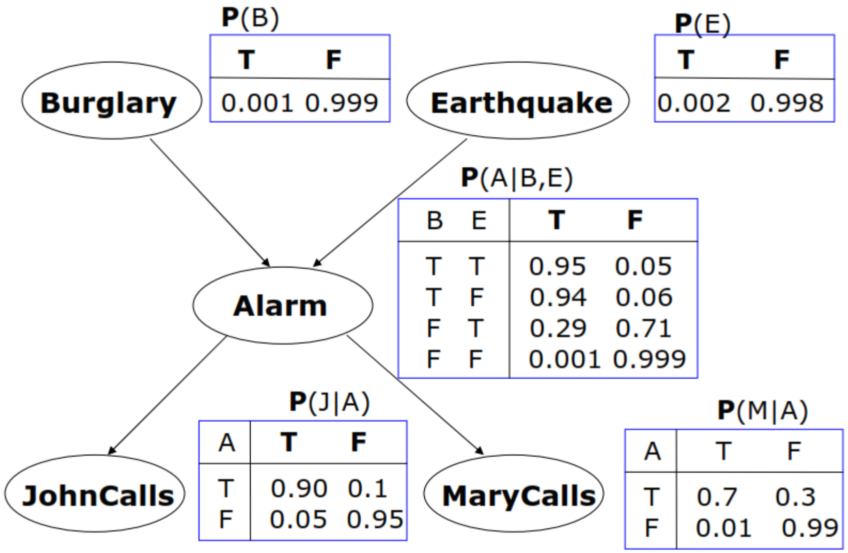
Example 1:

- $\underline{\boldsymbol{B = t, E = t, A = t, J = t, M = f}}$
- Then its probability is:

$= P(B = t, E = t, A = t, J = t, M = f)$
$= P(B = t)P(E = t)P(A = t \mid B = t, E = t)P(J = t \mid A = t)P(M = f \mid A = t)$

Example

Example 1:



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

P(A|B,E)

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

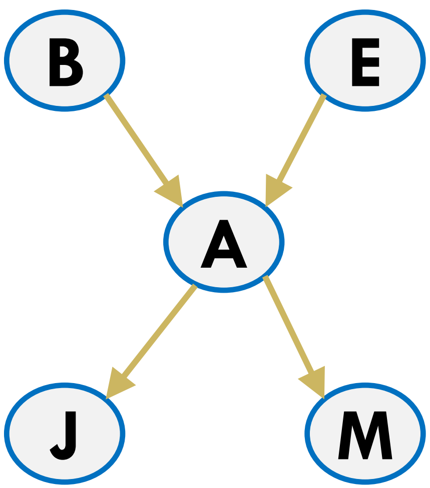
- $\boldsymbol{B = t, E = t, A = t, J = t, M = f}$

$= P(B = t, E = t, A = t, J = t, M = f)$

$= P(B = t)P(E = t)P(A = t \mid B = t, E = t)P(J = t \mid A = t)P(M = f \mid A = t)$

$= 0.001 \times 0.002 \times 0.95 \times 0.90 \times 0.3$

$= 5.13e - 07$



Acknowledgement

Tom Mitchel, Russel & Norvig, Wolfram Burgard, Maren Bennewitz, Marco Ragni