Bayesian Networks

Dr Babu Rao K,
Professor in CSE,
SoET, CMR University

Agenda

- What ia a Bayesian Network?
- What is a Directed Acyclic Graph?
- Understanding Bayesian Network with an Example
- Bayesian Network Application

What is Bayesian Network

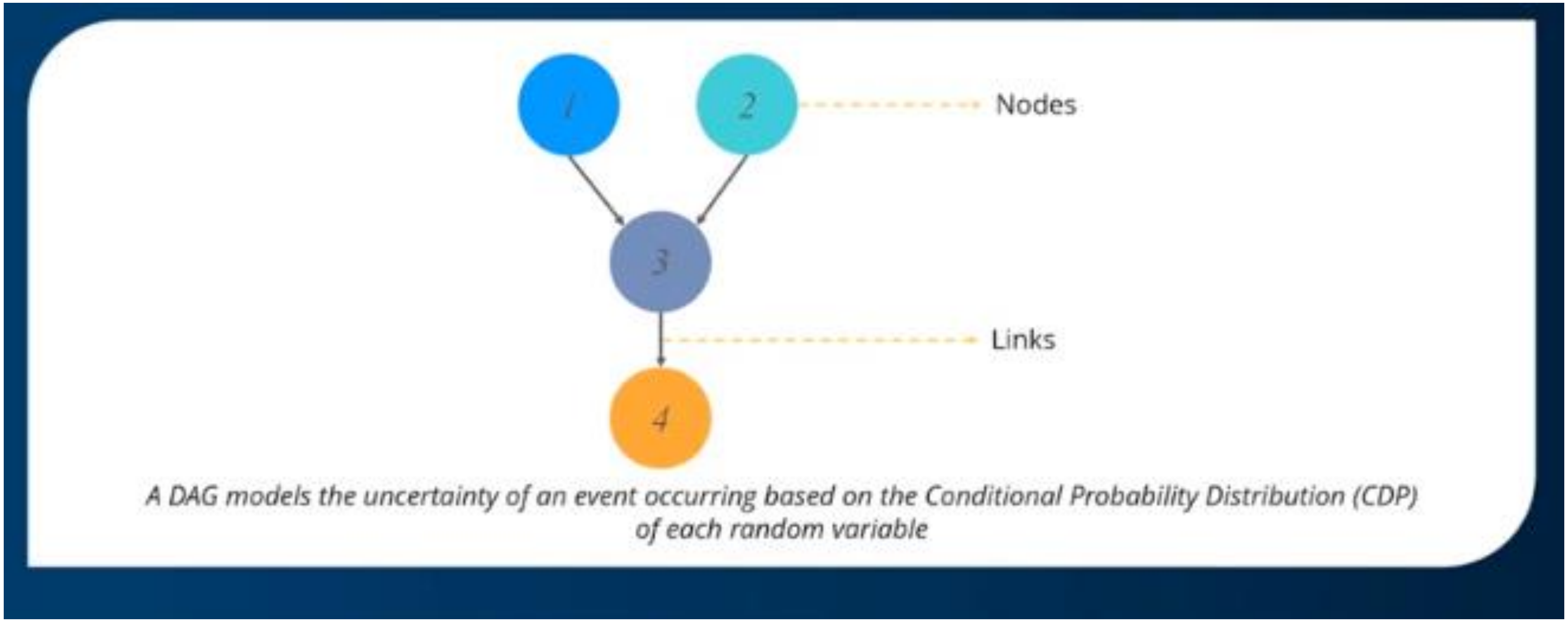
- A Bayesian Network fall under the category of **Probabilistic Graphical Modeling (PGM)** technique that is used to compute uncertainties by using the concept of probability.

Bayesian Networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions
- Syntax
  - **a set of nodes, one per variable**
  - **a directed, acyclic graph (link "directly influences")**
  - **a conditional distribution for each node given its parents:**

    P(Xi|Parents(Xi))
- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the
- distribution over Xi for each combination of parent values

What is Directed Acyclic Graph



- A Directed Acyclic Graph is used to represent a Bayesian Network and like any other statistical graph, DAG contains a set of nodes and links, where the links denote the relationship between the nodes.

Joint Probability & Conditional Probability

Joint Probability is a measure of two events happening at the same time, i.e., p(A and B). The probability of the intersection of A and B may be written $p(A \cap B)$.

Conditional Probability of an event B is the probability that the event will occur given that an event A has already occurred.

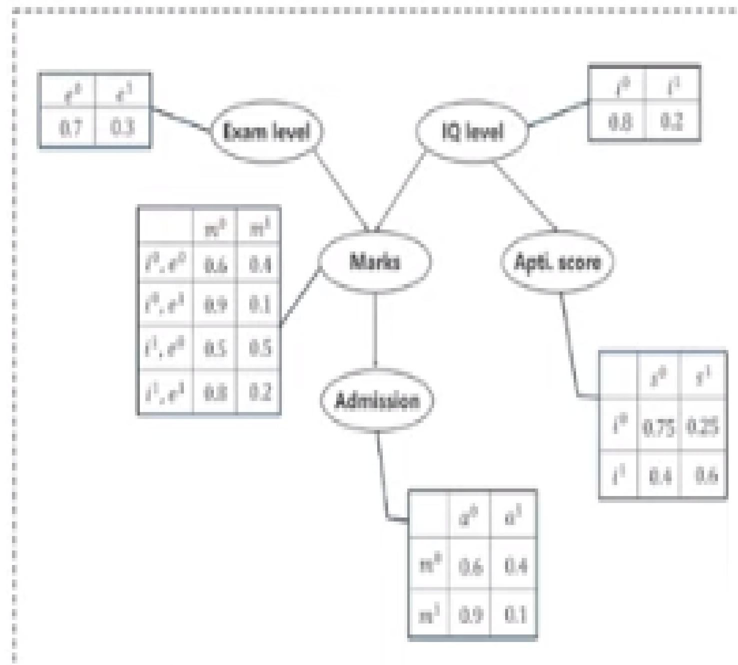
$p(B|A)$: probability of event B occurring, given that event A occurs.
If A and B are **dependent** events : $P(B|A) = P(A \text{ and } B) / P(A)$
If A and B are **independent** events: $P(B|A) = P(B)$

Bayesian Network Example

Create a Bayesian Network that will model the marks (m) of a student on his examination.

The marks will depend on:

- ***Exam level*** *(e): (difficult, easy)*
- ***IQ*** *of the student (i): (high, low)*
- *Marks ->* ***admitted (a)*** *to a university*
- *The IQ ->* ***aptitude score (s)*** *of the student*

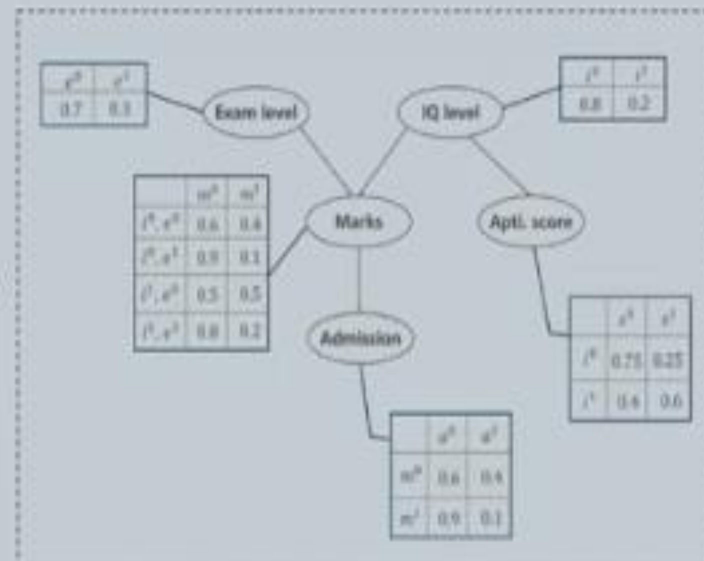


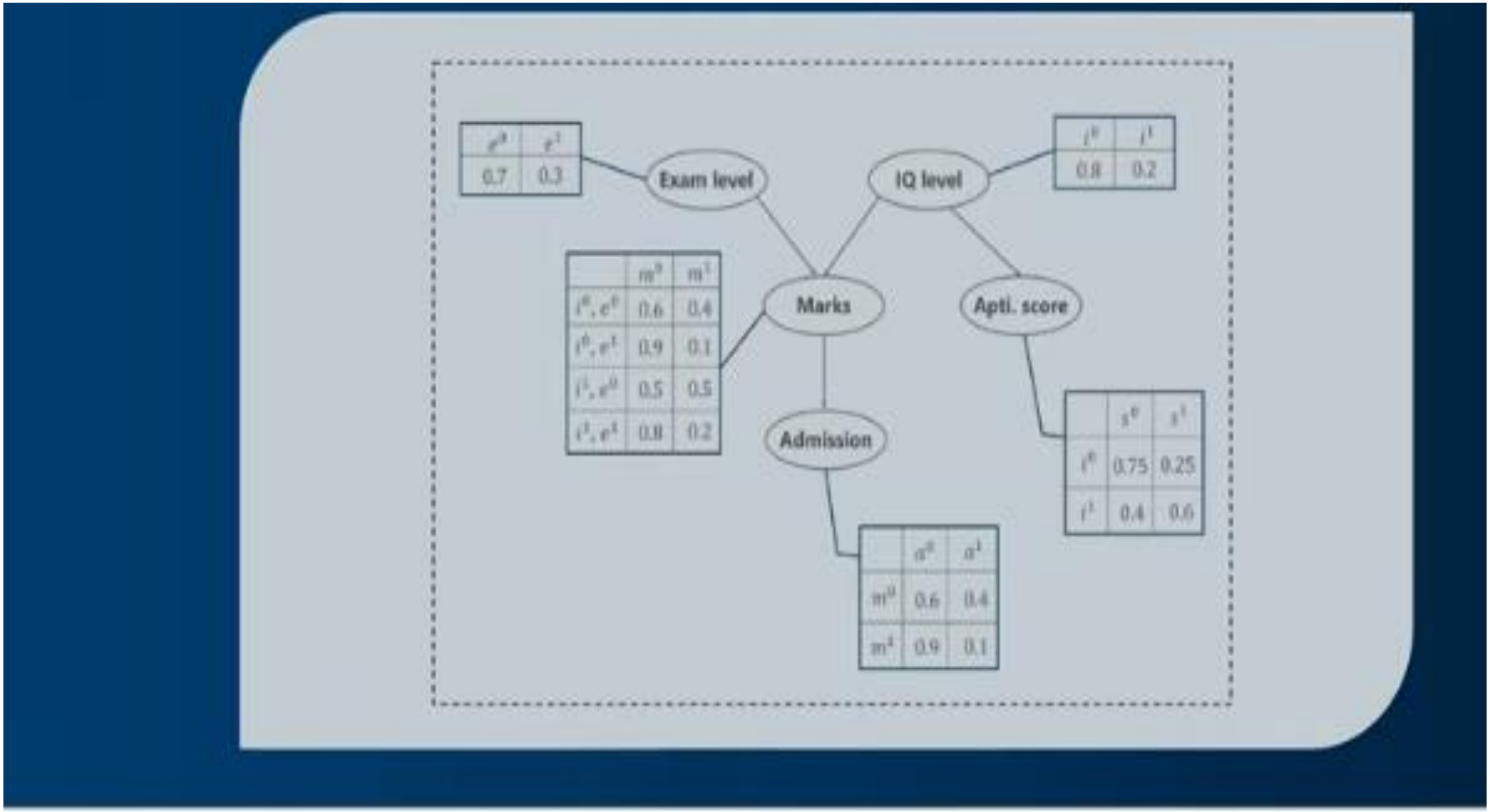
Bayesian Network Example – Create a Bayesian Network that will model the marks (m) a student on his examination

Factorizing Joint Probability Distribution:

p(a, m, i, e, s) = p(a | m) p(m | i, e) p(i) p(e) p(s | i)

- p(a | m): CP of student admit -> marks
- p(m | i, d): CP of the student's marks -> (IQ & exam level)
- p(i): Probability -> IQ level
- p(d): Probability -> exam level
- p(s | i) CP of aptitude scores ->IQ level



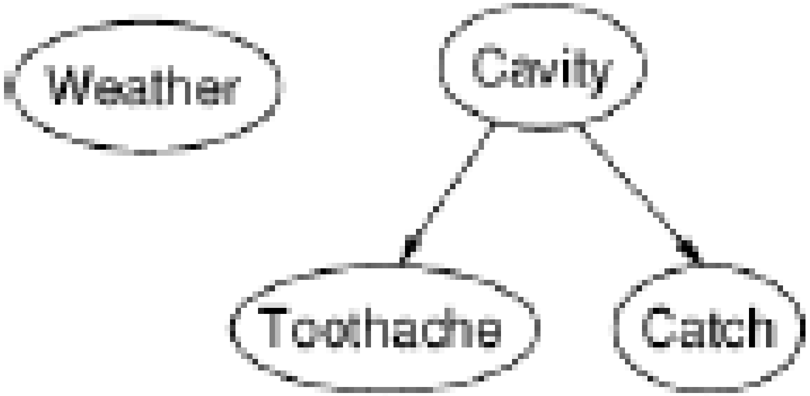


The probability of a random variable depends on his parents. Therefore, we can formulate Bayesian Networks as:

$$P(X_1, \dots, X_n) = \Pi_{i=1}^{n} p(X_i \mid Parents(X_i))$$

Example

- Topology of network encodes conditional independence assertions:



- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

Example

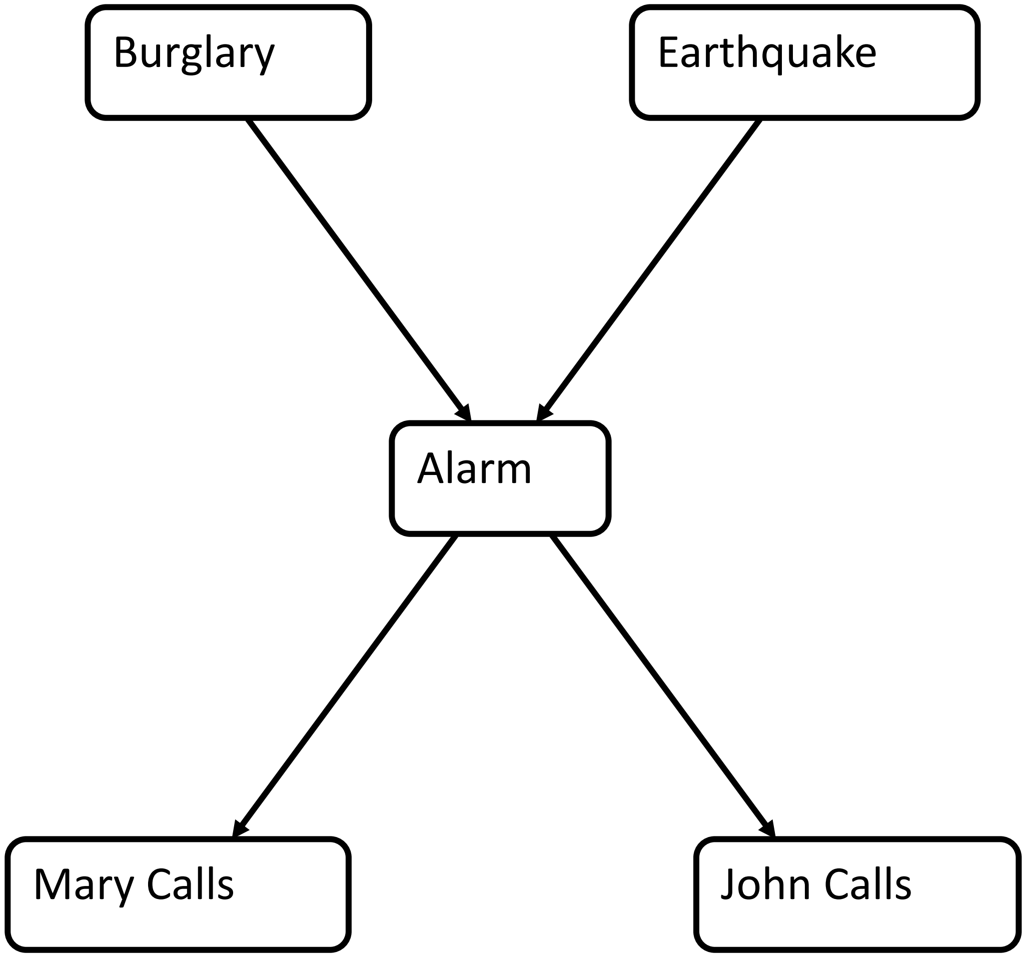
- *I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes.*

  Is there a burglar?

- Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$

- Network topology reflects "causal" knowledge
  - 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

- 3 Earthquake or Burglar



Tour through Probability

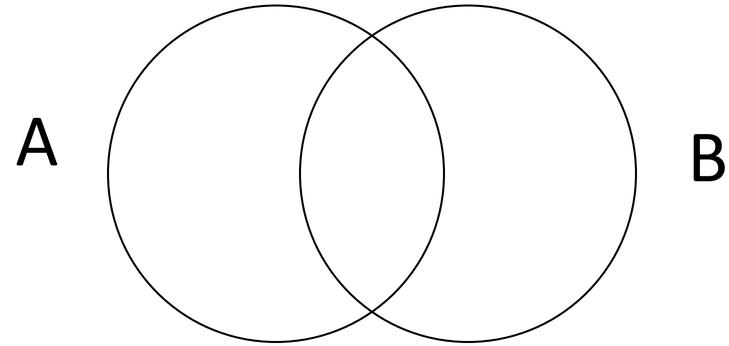
- All probabilities are between 0 and 1

- Necessarily true propositions have probability=1 and necessarily false propositions have probability=0

Conjunctions and Disjunctions

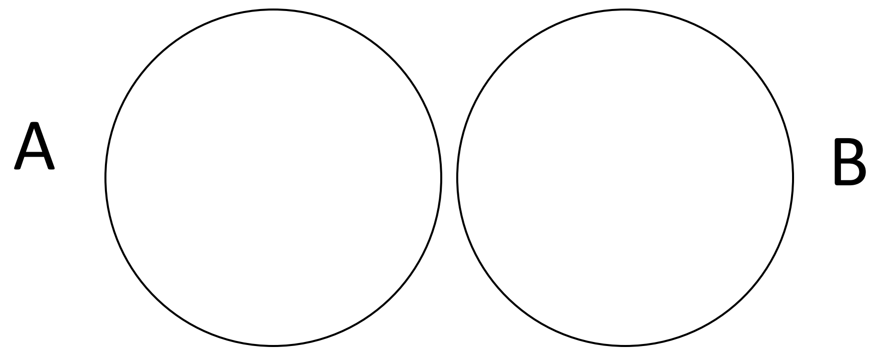
Venn Diagrams

- $P(A \& B) = P(A) \times P(B)$



- $P(A \vee B) = P(A) + P(B)$

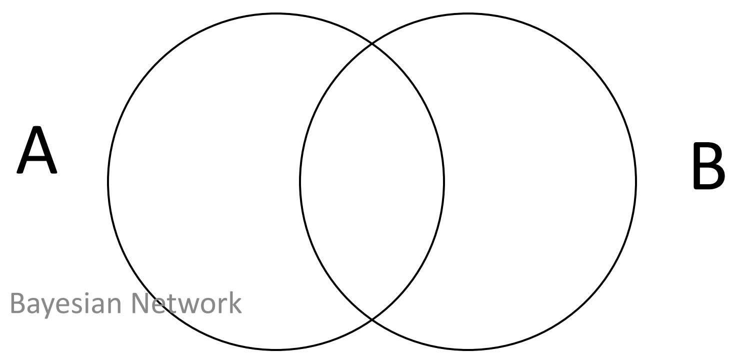
  (mutually exclusive)



- $P(A \vee B)$

  $= P(A) + P(B) - P(A \& B)$

  (not mutually exclusive)



Conditional probability & independence

- Probability of B “given” A:

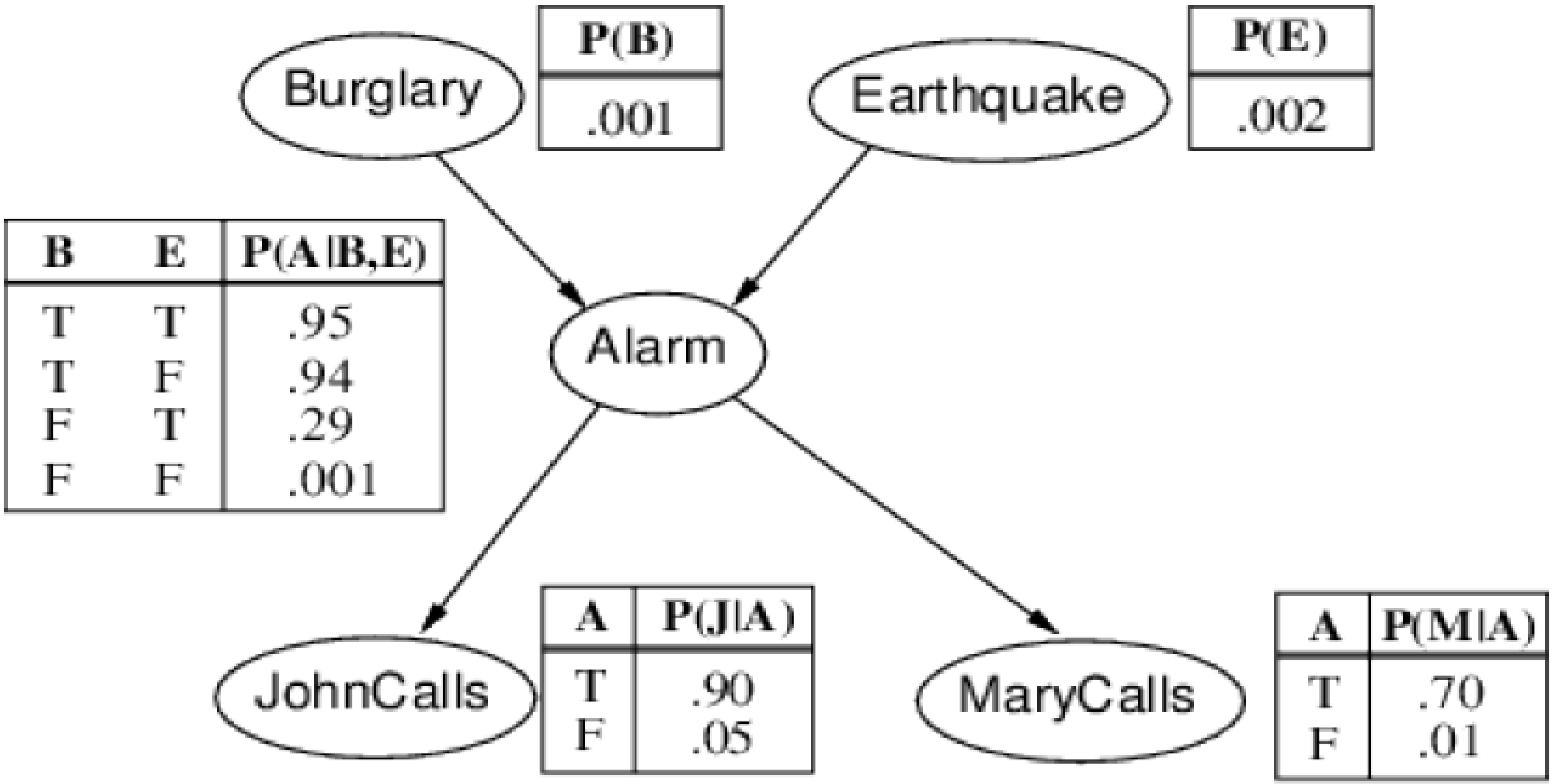
E.g. P(Hearts|Heart last time)

$$P(B|A) = \frac{P(A\&B)}{P(A)}$$

- Independence:

E.g. P(Heads|Even) = P(Heads)

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



P(B)
.001

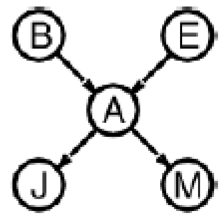
P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Compactness



- A conditional probability table for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values

- Each row requires one number p for $X_i = true$
(the number for $X_i = false$ is just $1 - p$)

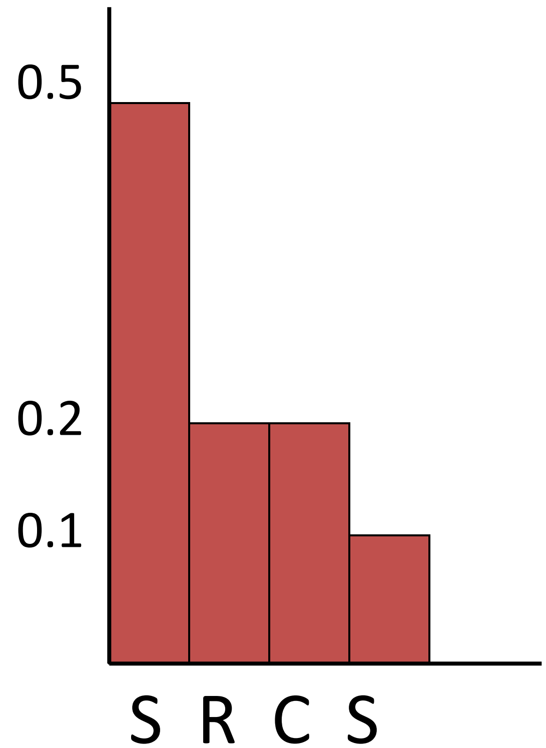
- If each variable has no more than k parents,
the complete network requires $O(n \cdot 2^k)$ numbers

- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution

- For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5 - 1 = 31$)

Probability Distributions

- Probability Distribution:
  - p(Weather=Sunny) = 0.5
  - p(Weather=Rain)= 0.2
  - p(Weather=Cloud)= 0.2
  - p(Weather=Snow)= 0.1



- NB Distribution sums to 1.

Joint Probability

- Completely specifies all beliefs in a problem domain.
- Joint prob Distribution is an n-dimensional table with a probability in each cell of that state occurring.
- Written as $P(X_1, X_2, X_3 \ldots, X_n)$
- When instantiated as $P(x_1, x_2 \ldots, x_n)$

Joint Distribution Example

- Domain with 2 variables each of which can take on 2 states.

P(Toothache, Cavity)

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

Bayes' Theorem

Simple:

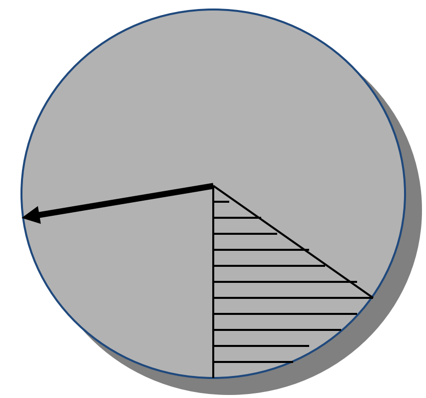
$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

General:

$$P(Y|X,E) = \frac{P(X|Y,E)P(Y|E)}{P(X|E)}$$

Bayesian Probability

- No need for repeated Trials
- Appear to follow rules of Classical Probability
- How well do we assign probabilities?

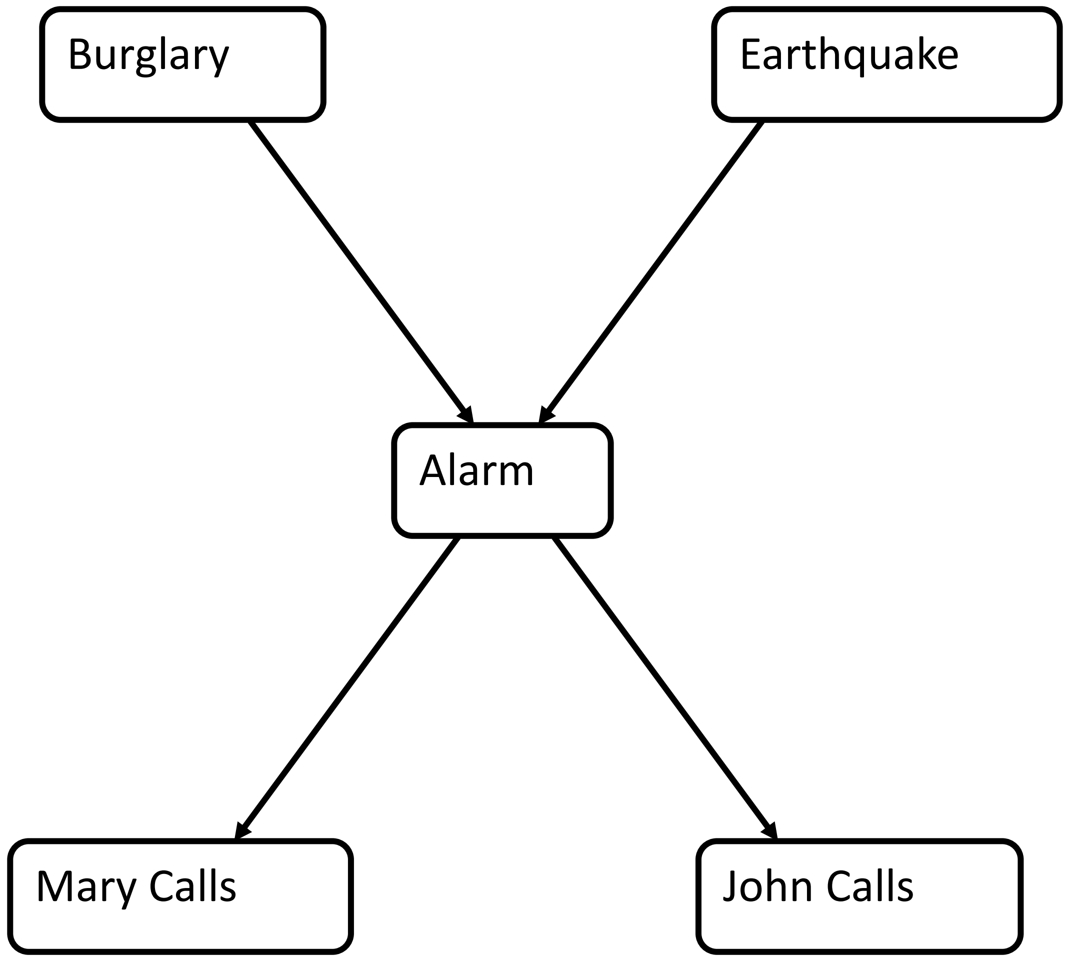


The Probability Wheel:
A Tool for Assessing Probabilities

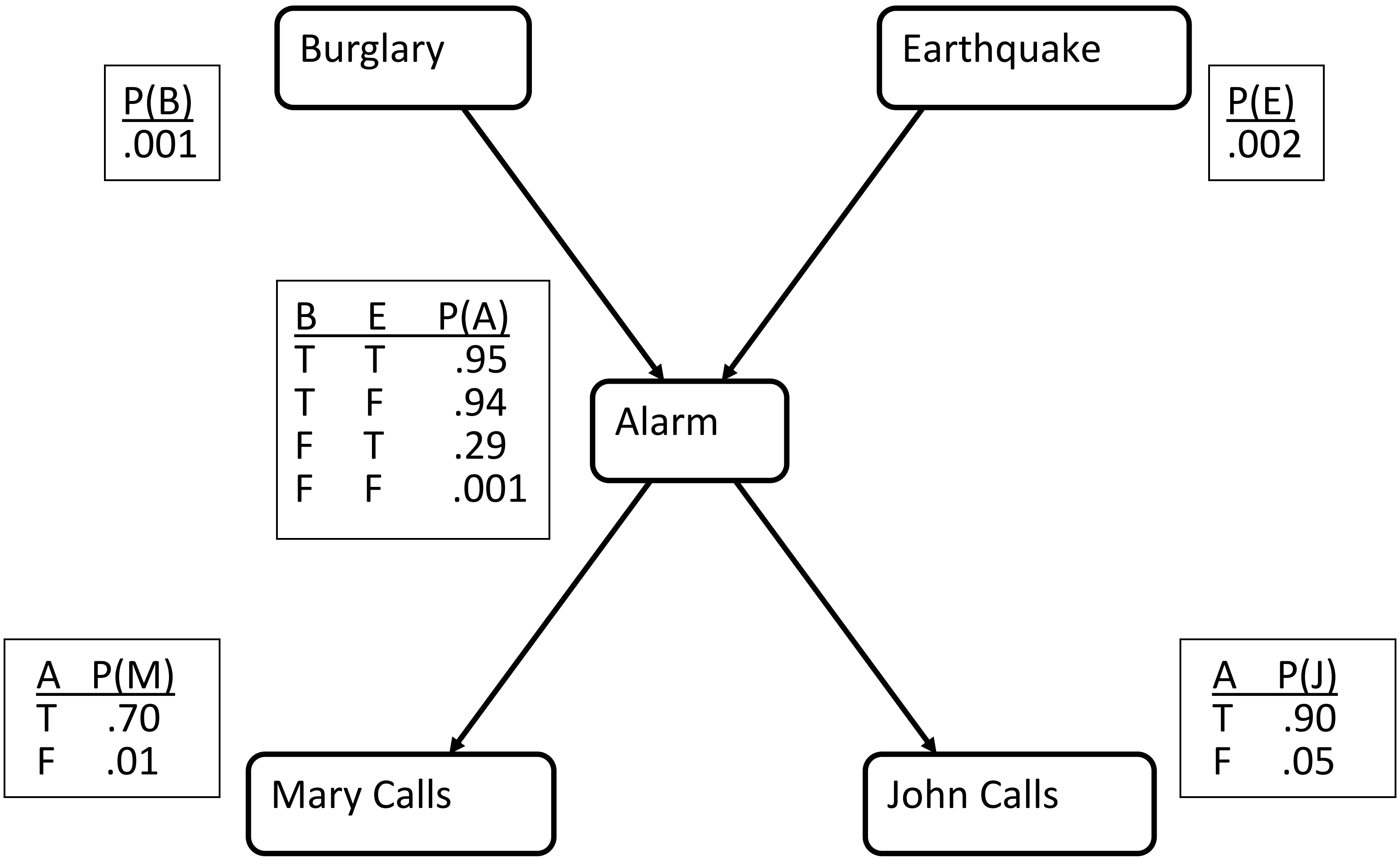
Bayesian Network - Definition

- Causal Structure
- Interconnected Nodes
- Directed Acyclic Links
- Joint Distribution formed from conditional distributions at each node.

Earthquake or Burglar



Bayesian Network for Alarm Domain



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(M)
T	.70
F	.01

A	P(J)
T	.90
F	.05