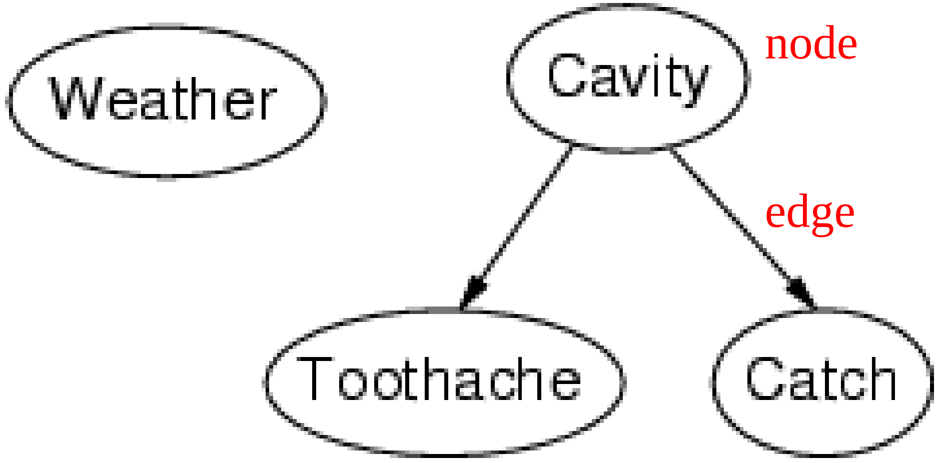
Bayesian Belief Network (BBN)

Dr. Kuppusamy .P
Associate Professor / SCOPE

Bayesian network

- **Bayesian network** is a probabilistic graphical model that represents a set of variables and their conditional dependencies using a directed acyclic graph.
- Also called a Bayesian belief network, belief network, decision network, or Bayesian model.
- It solves solve decision problems under uncertain knowledge using probabilistic events.
- Bayesian Network can be used for building models from data. It consists of two parts:
  - Directed Acyclic Graph
  - Conditional probabilities Table.
  - Topology + CPTs = compact representation of joint distribution
- Influence diagram represents Bayesian network that solves decision problems under uncertain knowledge.
- Each node represents the random variable that can be continuous or discrete.
- Directed Edge represent the relationship or conditional probabilities between random variables. (Directed edges => direct dependence)
- Edge represent that one node directly influence the other node, and if there is no directed edge that nodes are independent with each other. i.e., Absence of an edge => conditional independence



E.g.,

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

Bayesian Belief network

Bayesian network contains two components:

- Causal Component - describes the graph structure of the domain i.e., dependencies between variables
- Actual numbers - numerical probabilities (for each variable given its parents)

- Each node contains condition probability distribution $P(X_i|Parent(X_i))$, it determines the effect of the parent on that node.
- Bayesian network works based on Joint probability distribution and conditional probability.

Joint probability distribution:

- Let's consider variables $x_1, x_2 \ldots x_n$, then the probabilities of a different combination of $x_1, x_2 \ldots x_n$, are known as Joint probability distribution.
- $P(x_1, x_2 \ldots x_n) = P(x_1|\, x_2 \ldots x_n)\, \mathrm{P}(x_2 \ldots x_n)$

$$= P(x_1|\, x_2 \ldots x_n)\, \mathrm{P}(x_2|x_3 \ldots x_n) \ldots \mathrm{P}(x_{n-1}|x_n)\, P(x_n)$$

Note:

- BBN represent the full joint distribution over random variables more compactly using the product of local conditionals.
- Compute Local parameterizations from BBN that encodes conditional and marginal independencies among variables.
- When A & B are independent P(A, B) = P(A) * P(B)
- When A & B are conditionally independent given C.

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

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

Example - Bayesian Belief network

Harshith has fixed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary and minor earthquakes. Harshith has two neighbors David and Sophia, who have taken a responsibility to inform Harshith at work when they hear the alarm. David always calls Harshith when he hears the alarm, but sometimes he got confused with the phone ringing and calls at that time too. Sophia likes to listen to high music, so sometimes she misses to hear the alarm. compute the probability of Burglary Alarm.

Problem Statement:

Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harshith**.**

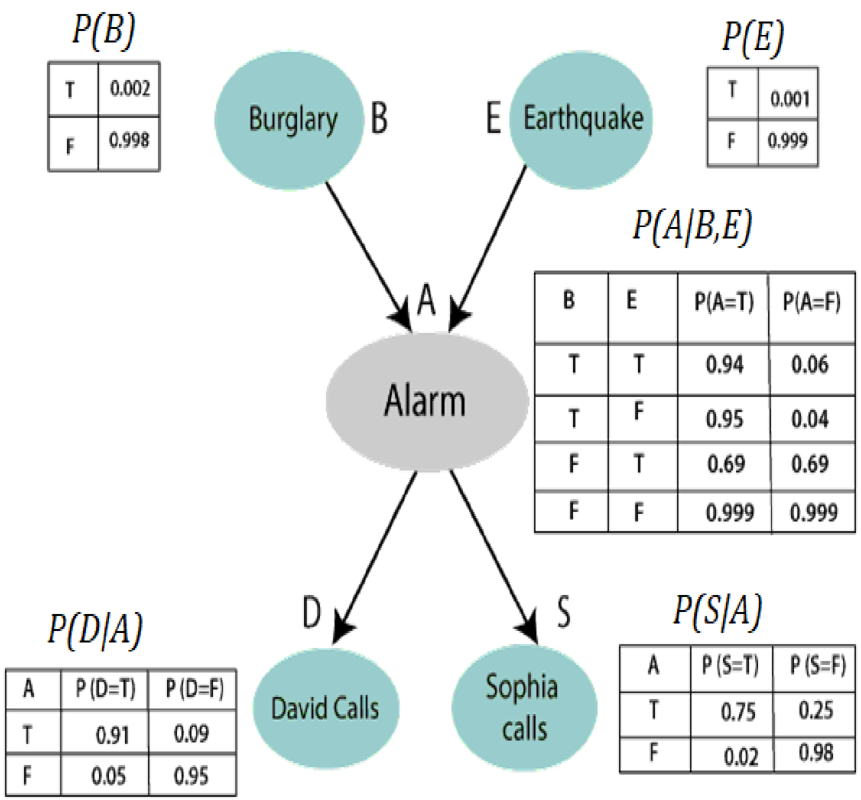
Network structure shows that burglary and earthquake is the parent node of the alarm and directly affecting the probability of alarm's going off, but David and Sophia's calls depend on alarm probability.

The network represents that our assumptions do not directly perceive the burglary, and also do not notice the minor earthquake, and they also not confer before calling.

The conditional distributions for each node are given as conditional probabilities table (CPT).

Each row in the CPT must be sum to 1 because all the entries in the table represent an exhaustive set of cases for the variable.

In CPT, a Boolean variable with k boolean parents contains 2^K probabilities. Hence, if there are two parents, then CPT will contain 4 probability values



$P(B)$

T	0.002
F	0.998

$P(E)$

T	0.001
F	0.999

$P(A|B,E)$

B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

$P(D|A)$

A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

$P(S|A)$

A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

Example - Bayesian Belief network

Variables (Events): Burglary (B), Earthquake (E), Alarm(A), DavidCalls(D), SophiaCalls(S)

Network topology reflects "causal" knowledge:

- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Sophia to call
- The alarm can cause David to call

Write equation based on Joint probability:

P[D, S, A, B, E] = P[D | S, A, B, E]. P[S, A, B, E]

=P[D | S, A, B, E]. P[S | A, B, E]. P[A, B, E]

= P [D| A]. P [S| A, B, E]. P[A, B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B |E]. P[E]

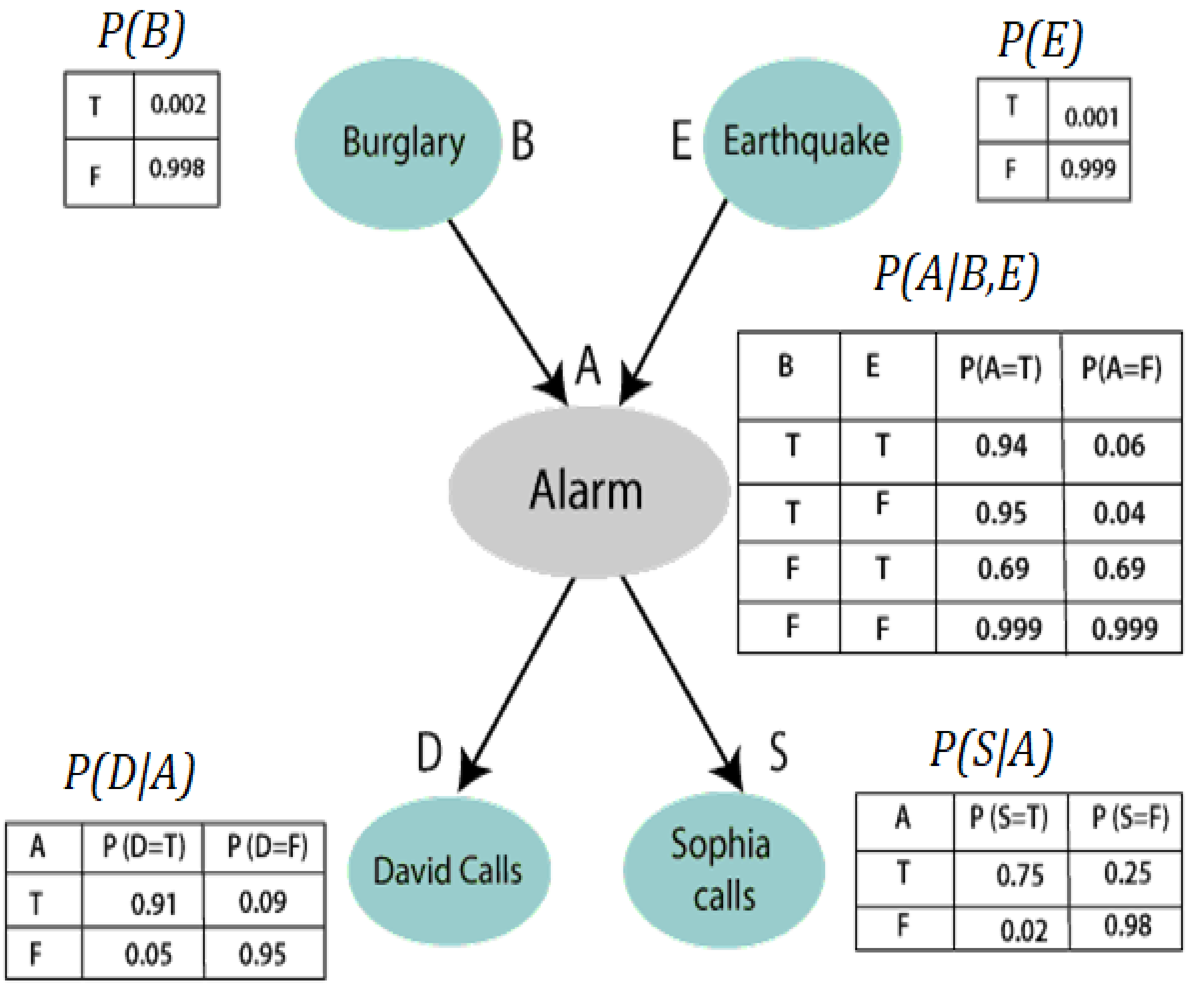
Observed probability for the burglary and earthquake:

P(B= True) = 0.002, the probability of burglary.

P(B= False)= 0.998, probability of no burglary.

P(E= True)= 0.001, probability of a minor earthquake

P(E= False)= 0.999, the probability that an earthquake not occurred.



P(B)

T	0.002
F	0.998

P(E)

T	0.001
F	0.999

P(A|B,E)

B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

P(D|A)

A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

P(S|A)

A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

Example - Bayesian Belief network

Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harshith.

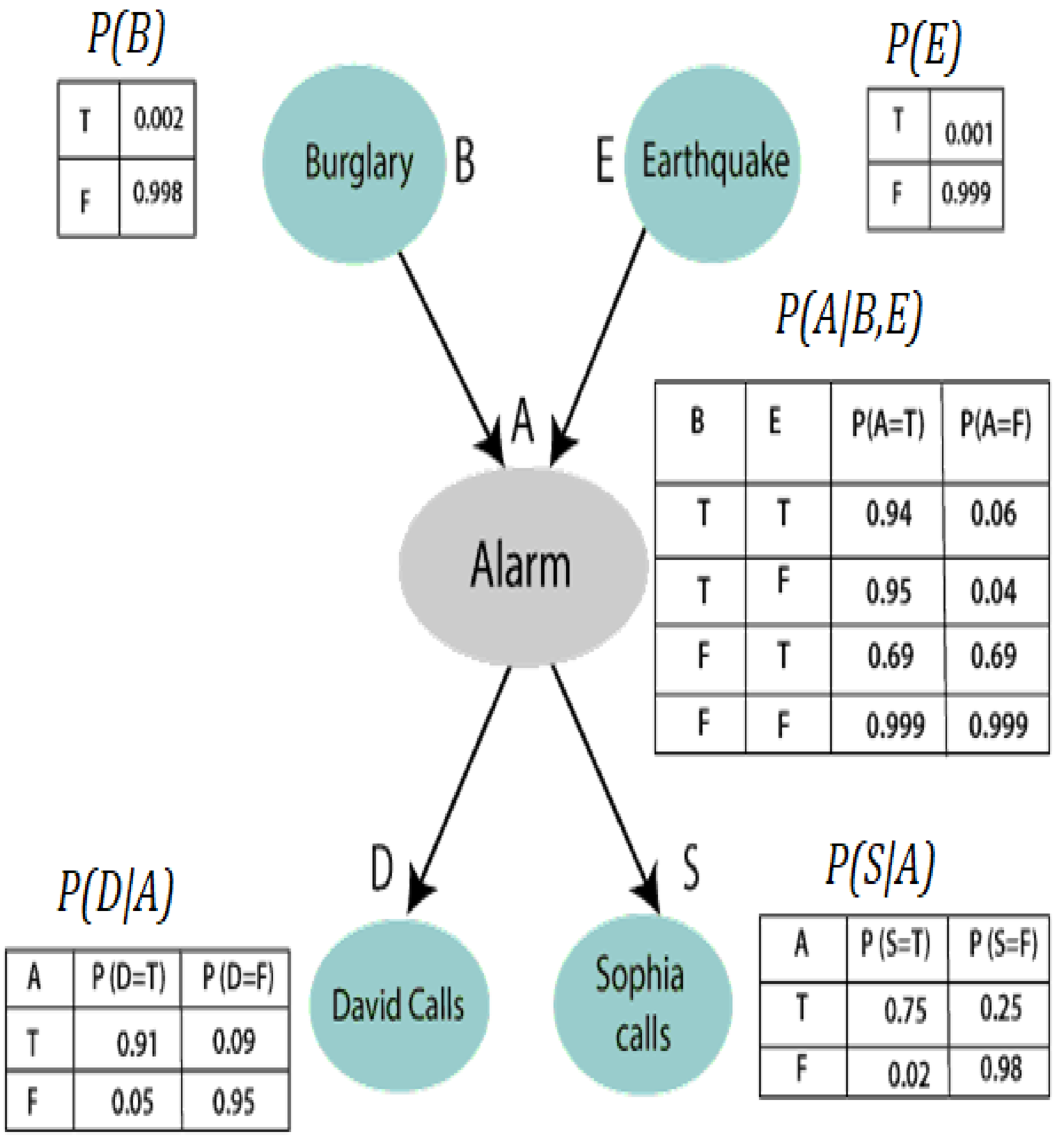
- Burglary, not occurred P(¬B)
- Earthquake not occurred P(¬E)

Write the joint probability distribution for the above problem:

P(S, D, A, ¬B, ¬E) = P (S|A) *P (D|A)*P (A|¬B ^ ¬E) *P (¬B) *P (¬E).

= 0.75* 0.91* 0.001* 0.998*0.999

= 0.00068045.



Example 2- Bayesian Belief network

Calculate the probability Only David called the Harshith.

Solution:

We are unknown about other events.

Assumptions

- David will call the Harshith when alarm goes off. P (D|A)
- David will call the Harshith without alarm goes off. P (D| ¬ A)

Write the joint probability distribution for the above problem:

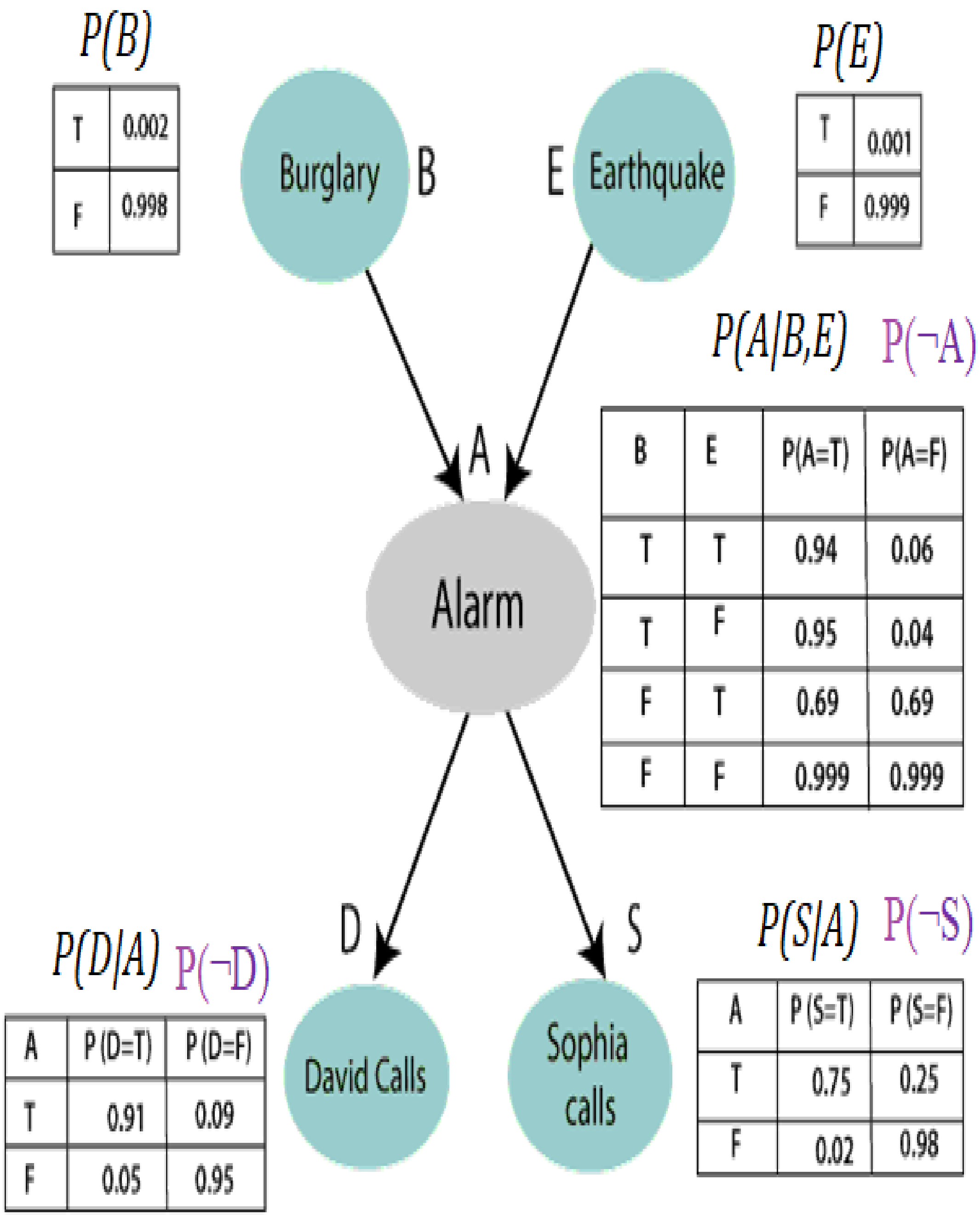
P(D) = P (D|A)*P (A) + P (D| ¬A) *P (¬A).

- Here P(A) depends on P(B) and P(E). But we don't know about these events i.e., Burgalary and earthquack occured or not. So, Consider all possibilities P(B,E), P(¬B,E), P(B, ¬E), P(¬B, ¬E).
- Also unknown about alarm is ringing or not. So, consider P(A) and P(¬A).

= P (D|A)* {P(A|B,E)*P(B,E) + P(A|¬B, E) * P(¬B, E) + P(A|B, ¬E) * P(B, ¬E) + P(A|¬B, ¬E)*P(¬B, ¬E)} + P (D| ¬ A)*{P(¬ A|B,E)*P(B,E) + P(¬ A|¬B, E) * P(¬B, E) + P(¬ A|B, ¬E) * P(B, ¬E) + P(¬ A|¬B, ¬E)*P(¬B, ¬E)}

= ?

Note: B,E independent of other events. So P(B,E)=P(B)*P(E) =0.002 * 0.001



Bayesian network Applications

- Medical diagnosis systems
- Manufacturing system diagnosis
- Computer systems diagnosis
- Network systems diagnosis
- Helpdesk troubleshooting
- Information retrieval

References

1. Tom M. Mitchell, Machine Learning, McGraw Hill , 2017.

2. EthemAlpaydin, Introduction to Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 2017.

3. Wikipedia