

# **Introduction to AI**

**Lecture 19**

## **Bayes' Theorem**

**Dr. Tamal Ghosh**  
**Department of CSE**  
**Adamas University**

# Bayes' theorem in AI

- Bayes' theorem is also known as **Bayes' rule**, **Bayes' law**, or **Bayesian reasoning**, which determines the probability of an event with uncertain knowledge.
- In probability theory, it relates the conditional probability and marginal probabilities of two random events.
- Bayes' theorem was named after the British mathematician **Thomas Bayes**. The **Bayesian inference** is an application of Bayes' theorem, which is fundamental to Bayesian statistics.
- It is a way to calculate the value of  $P(B|A)$  with the knowledge of  $P(A|B)$ .

# Contd...

- Bayes' theorem allows updating the probability prediction of an event by observing new information in the real world.
- **Example:** If cancer corresponds to one's age then by using Bayes' theorem, we can determine the probability of cancer more accurately with the help of age.
- Bayes' theorem can be derived using the product rule and conditional probability of event A with known event B:
- As from the product rule we can write:

$$P(A \wedge B) = P(A | B) P(B) \text{ or}$$

Similarly, the probability of event B with known event A:

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

# Contd...

- Equating right-hand side of both equations, we will get:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)} \quad \dots(a)$$

- The above equation (a) is called as **Bayes' rule** or **Bayes' theorem**. This equation is basic of most modern AI systems for **probabilistic inference**.
- It shows the simple relationship between joint and conditional probabilities. Here,
- $P(A|B)$  is known as **posterior**, which we need to calculate, and it will be read as the Probability of hypothesis A when we have evidence B already occurred.
- $P(B|A)$  is called the likelihood, in which we consider that the hypothesis is true, and then we calculate the probability of evidence.
- $P(A)$  is called the **prior probability**, the probability of the hypothesis before considering the evidence
- $P(B)$  is called **marginal probability**, pure probability of evidence.

# Contd...

In the equation (a), in general, we can write  $P(B) = \sum P(A_i) * P(B | A_i)$ , hence the Bayes' rule can be written as:

$$P(A_i | B) = \frac{P(A_i) * P(B | A_i)}{\sum_{i=1}^k P(A_i) * P(B | A_i)}$$

Where  $A_1, A_2, A_3, \dots, A_n$  is a set of mutually exclusive and exhaustive events.

# Applying Bayes' rule

- Bayes' rule allows us to compute the single term  $P(B|A)$  in terms of  $P(A|B)$ ,  $P(B)$ , and  $P(A)$ . This is very useful in cases where we have a good probability of these three terms and want to determine the fourth one. Suppose we want to perceive the effect of some unknown cause, and want to compute that cause, then the Bayes' rule becomes:

$$P(\text{cause} | \text{effect}) = \frac{P(\text{effect} | \text{cause}) P(\text{cause})}{P(\text{effect})}$$

- Example-1:
- Question: what is the probability that a patient has diseases meningitis with a stiff neck?
- Given Data:

# Example-1

A doctor is aware that the disease meningitis causes a patient to have a stiff neck, and it occurs 80% of the time. He is also aware of some more facts, which are given as follows:

- The Known probability that a patient has meningitis disease is 1/30,000.
- The Known probability that a patient has a stiff neck is 2%.

Let **a** be the proposition that the patient has a stiff neck and **b** be the proposition that the patient has meningitis. , so we can calculate the following as:

$$P(a|b) = 0.8$$

$$P(b) = 1/30000$$

$$P(a) = .02$$

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} = \frac{0.8 * (\frac{1}{30000})}{0.02} = 0.001333333.$$

Hence, we can assume that 1 patient out of 750 patients has meningitis disease with a stiff neck.

## Example-2

- Question: From a standard deck of playing cards, a single card is drawn. The probability that the card is king is  $4/52$ , then calculate the posterior probability  $P(\text{King}|\text{Face})$ , which means the drawn face card is a king card.

- **Solution:**

$$P(\text{king}|\text{face}) = \frac{P(\text{Face}|\text{king}) \cdot P(\text{King})}{P(\text{Face})} \dots\dots(i)$$

- $P(\text{king})$ : probability that the card is King =  $4/52 = 1/13$
- $P(\text{face})$ : the probability that a card is a face card =  $3/13$
- $P(\text{Face}|\text{King})$ : probability of king card when we assume it is a face = 1
- Putting all values in equation (i) we will get:

$$P(\text{king}|\text{face}) = \frac{1 * (\frac{1}{13})}{(\frac{3}{13})} = 1/3, \text{ it is a probability that a face card is a king card.}$$

# Application of Bayes' theorem in AI

- Following are some applications of Bayes' theorem:
  - It is used to calculate the next step of the robot when the already executed step is given.
  - Bayes' theorem is helpful in weather forecasting.
  - It can solve the Monty Hall problem.

# Bayesian Belief Network in AI

- Bayesian belief network is key computer technology for dealing with probabilistic events and solving a problem that has uncertainty. We can define a Bayesian network as:
- "A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph."
- It is also called a **Bayes network**, **belief network**, **decision network**, or **Bayesian model**.
- Bayesian networks are probabilistic, because these networks are built from a **probability distribution**, and also use probability theory for prediction and anomaly detection.
- Real-world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network.

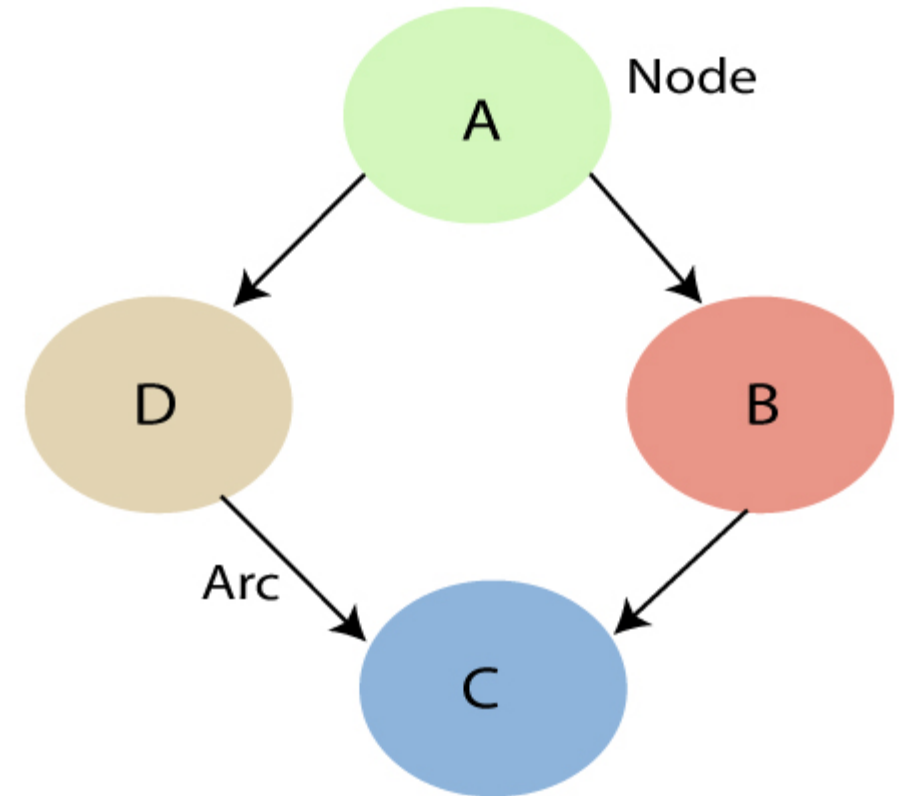
# Contd...

- It can also be used in various tasks including **prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction, and decision making under uncertainty.**
- Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:
  - **Directed Acyclic Graph**
  - **Table of conditional probabilities**
- The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an **Influence diagram.**

# Directed Acyclic Graph

A Bayesian network graph is made up of nodes and Arcs (directed links), where:

- Each **node** corresponds to the random variables, and a variable can be **continuous** or **discrete**.
- **Arc or directed arrows** represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph.
- These links represent that one node directly influences the other node, and if no directed link means that nodes are independent of each other.



# Contd...

- In the above diagram, A, B, C, and D are random variables represented by the nodes of the network graph.
- If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.
- Node C is independent of node A.
- The Bayesian network has mainly two components:
- **Causal Component**
- **Actual numbers**
- Each node in the Bayesian network has condition probability distribution  $P(X_i | \text{Parent}(X_i))$ , which determines the effect of the parent on that node.
- Bayesian network is based on Joint probability distribution and conditional probability. So let's first understand the joint probability distribution:

# Joint Probability Distribution

If we have variables  $x_1, x_2, x_3, \dots, x_n$ , then the probabilities of a different combination of  $x_1, x_2, x_3, \dots, x_n$ , are known as Joint probability distribution.

$P[x_1, x_2, x_3, \dots, x_n]$ , it can be written in the following way in terms of the joint probability distribution.

$$= P[x_1 | x_2, x_3, \dots, x_n] P[x_2, x_3, \dots, x_n]$$

$$= P[x_1 | x_2, x_3, \dots, x_n] P[x_2 | x_3, \dots, x_n] \dots P[x_{n-1} | x_n] P[x_n].$$

In general, for each variable  $X_i$ , we can write the equation as:

$$P(X_i | X_{i-1}, \dots, X_1) = P(X_i | \text{Parents}(X_i))$$

# Explanation of Bayesian network

- **Example:** Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds to detecting a burglary but also responds to minor earthquakes. Harry has two neighbors David and Sophia, who have taken a responsibility to inform Harry at work when they hear the alarm. David always calls Harry when he hears the alarm, but sometimes he gets confused with the phone ringing and calls at that time too. On the other hand, Sophia likes to listen to high music, so sometimes she misses to hear the alarm. Here we would like to compute the probability of the Burglary Alarm.

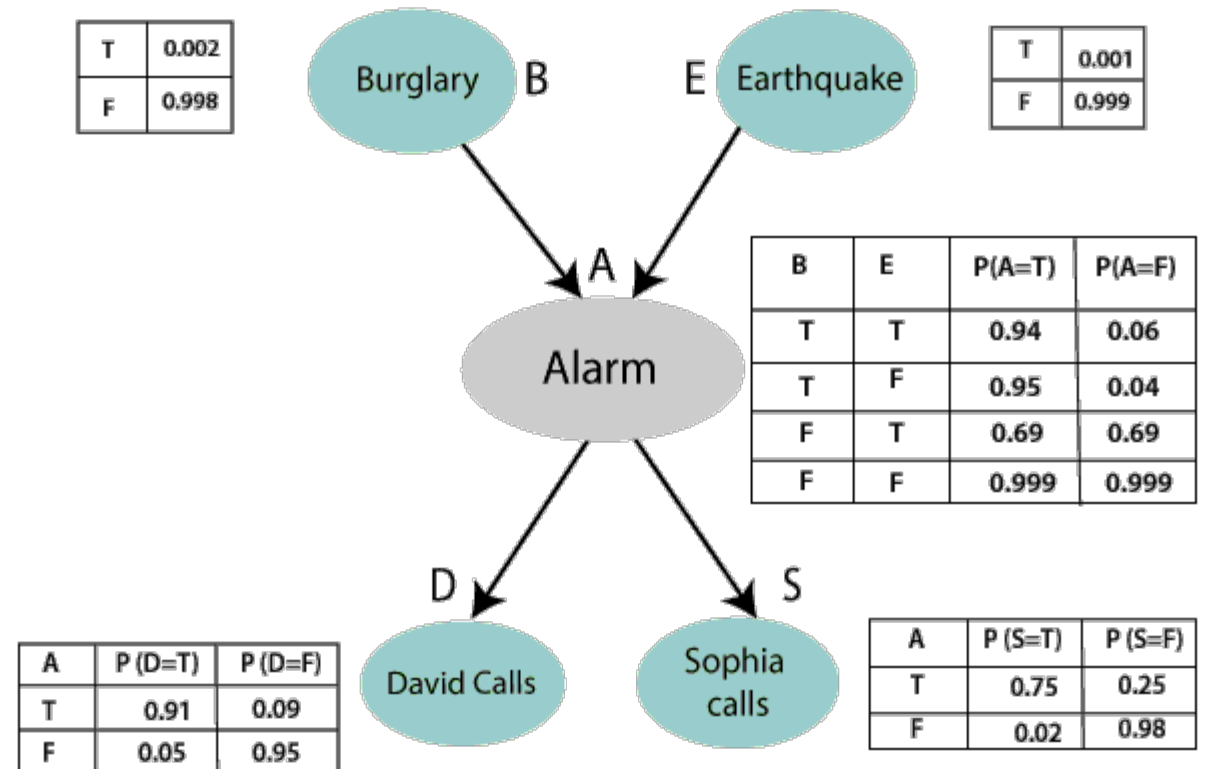
## **Problem:**

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

- **Solution:**
- The Bayesian network for the above problem is given below. The network structure shows that burglary and earthquake are the parent nodes of the alarm and directly affect the probability of the alarm 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 do not confer before calling.
- The conditional distributions for each node are given as conditional probabilities table or CPT.
- Each row in the CPT must be summed 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.
- **List of all events occurring in this network:**
  - **Burglary (B)**
  - **Earthquake(E)**
  - **Alarm(A)**
  - **David Calls(D)**
  - **Sophia calls(S)**

# Contd...

- We can write the events of the problem statement in the form of probability:  $P[D, S, A, B, E]$ , can rewrite the above probability statement using joint probability distribution:
- $P[D, S, A, B, E] = P[D | S, A, B, E] \cdot P[S, A, B, E]$
- $= P[D | S, A, B, E] \cdot P[S | A, B, E] \cdot P[A, B, E]$
- $= P[D | A] \cdot P[S | A, B, E] \cdot P[A, B, E]$
- $= P[D | A] \cdot P[S | A] \cdot P[A | B, E] \cdot P[B, E]$
- $= P[D | A] \cdot P[S | A] \cdot P[A | B, E] \cdot P[B | E] \cdot P[E]$



# Contd...

- Let's take the observed probability for the Burglary and earthquake component:
- $P(B= \text{True}) = 0.002$ , which is the probability of burglary.
- $P(B= \text{False})= 0.998$ , which is the probability of no burglary.
- $P(E= \text{True})= 0.001$ , which is the probability of a minor earthquake
- $P(E= \text{False})= 0.999$ , Which is the probability that an earthquake not occurred.
- We can provide the conditional probabilities as per the below tables:
- **Conditional probability table for Alarm A:**
- The Conditional probability of Alarm A depends on Burglar and earthquake:

B	E	P(A= True)	P(A= False)
True	True	0.94	0.06
True	False	0.95	0.04
False	True	0.31	0.69
False	False	0.001	0.999

# Contd...

- **Conditional probability table for David Calls:**
- The Conditional probability of David that he will call depends on the probability of Alarm.

A	P(D= True)	P(D= False)
True	0.91	0.09
False	0.05	0.95

- **Conditional probability table for Sophia Calls:**
- The Conditional probability of Sophia that she calls is depending on its Parent Node "Alarm."

A	P(S= True)	P(S= False)
True	0.75	0.25
False	0.02	0.98

- From the formula of joint distribution, we can write the problem statement in the form of probability distribution:
- $P(S, D, A, \neg B, \neg E) = P(S|A) * P(D|A) * P(A|\neg B \wedge \neg E) * P(\neg B) * P(\neg E)$ .
- $= 0.75 * 0.91 * 0.001 * 0.998 * 0.999$
- $= 0.00068045$ .
- Hence, a Bayesian network can answer any query about the domain by using Joint distribution.
- **The semantics of the Bayesian Network:**
- There are two ways to understand the semantics of the Bayesian network, which is given below:
  - 1. To understand the network as the representation of the Joint probability distribution.
  - It is helpful to understand how to construct the network.
  - 2. To understand the network as an encoding of a collection of conditional independence statements.
  - It helps design inference procedures.