

Bayes' theorem

Bayes' theorem

- Bayes' theorem provides us with a formula to calculate probability of an event given probabilities of other events.
- This theorem determines the probability of an event with uncertain knowledge.
- 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.
- Bayes' theorem allows updating the probability prediction of an event by observing new information of the real world.

- **Example:** If rainfall corresponds to temperature of a region, then by using Bayes' theorem, we can determine the probability of getting rainfall more accurately with the help of temperature.
- (I.e) one factor is depending on other factor.

- Bayes' theorem can be derived using **product rule** and conditional probability of event A with known event B:
- As from product rule we can write:
- $P(A \wedge B) = P(A|B) P(B)$ or
- Similarly, the probability of event B with known event A:
- $P(A \wedge B) = P(B|A) P(A)$
- Equating right hand side of both the equations, we will get:

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

- This equation is called **Bayes' rule or Bayes' theorem**

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

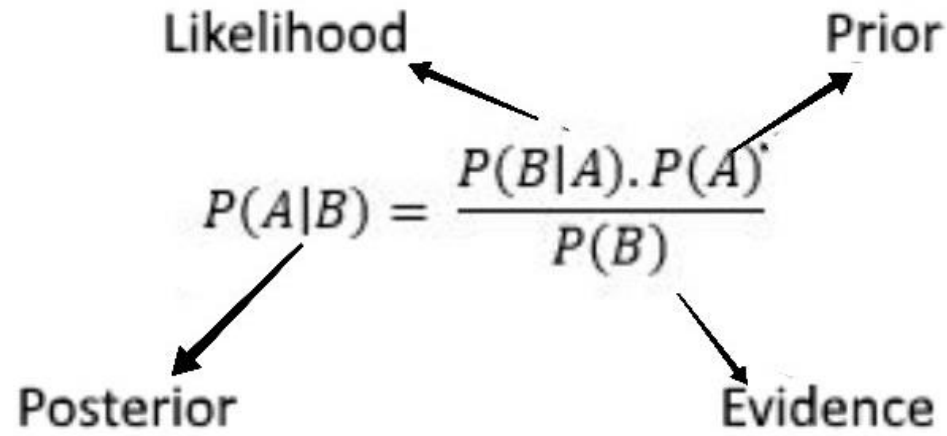
- Which tells us: how often A happens given that B happens, written $P(A|B)$,
- When we know: how often B happens given that A happens, written $P(B|A)$
and how likely A is on its own, written $P(A)$
and how likely B is on its own, written $P(B)$

- Let us say $P(\text{Fire})$ means how often there is fire, and $P(\text{Smoke})$ means how often we see smoke, then:
- $P(\text{Fire}|\text{Smoke})$ means how often there is fire when we can see smoke
- $P(\text{Smoke}|\text{Fire})$ means how often we can see smoke when there is fire

Example :

- dangerous fires are rare (1%)
- but smoke is fairly common (10%) due to barbecues,
- and 90% of dangerous fires make smoke
- **probability of dangerous Fire when there is Smoke:**

$$\begin{aligned}P(\text{Fire}|\text{Smoke}) &= P(\text{Fire}) P(\text{Smoke}|\text{Fire}) / P(\text{Smoke}) \\ &= (1\% \times 90\%) / 10\% \\ &= 9\%\end{aligned}$$



- $P(A|B)$ is known as **posterior**, which we need to calculate, and it will be read as Probability of hypothesis A when we have occurred an evidence B.
- $P(B|A)$ is called the **likelihood**, in which we consider that hypothesis is true, then we calculate the probability of evidence.
- $P(A)$ is called the **prior probability**, probability of hypothesis before considering the evidence
- $P(B)$ is called **marginal probability**, pure probability of an evidence.

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:
- A doctor is aware that 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%.

- Solution:

- Let a be the proposition that patient has stiff neck and b be the proposition that 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 600 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 posterior probability $P(\text{King}|\text{Face})$, which means the drawn face card is a king card.

- **Solution:**
$$P(\text{king}|\text{face}) = \frac{P(\text{Face}|\text{king}) \cdot P(\text{King})}{P(\text{Face})}$$

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

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

Homework

- You are planning a picnic today, but the morning is cloudy
- Oh no! 50% of all rainy days start off cloudy!
- But cloudy mornings are common (about 40% of days start cloudy)
- And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)
- **What is the chance of rain during the day?**

Probabilistic reasoning

- Probabilistic reasoning is a method of representation of knowledge where the concept of probability is applied to indicate the uncertainty in knowledge.

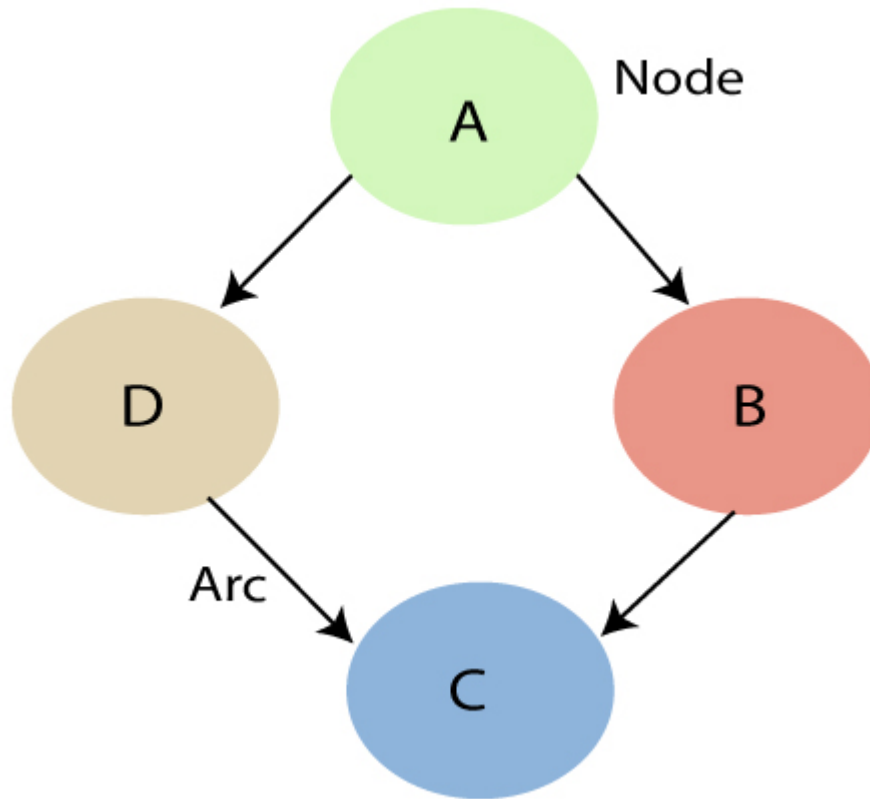
Probabilistic reasoning is used in AI:

- When we are unsure of the predicates
- When the possibilities of predicates become too large to list down
- When it is known that an error occurs during an experiment

Bayesian Belief Network in artificial intelligence

- Bayesian belief network is key computer technology for dealing with probabilistic events and to solve a problem which 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**.

- Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network. 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.**



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.

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

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 influence the other node, and if there is no directed link that means that nodes are independent with each other

- 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 as the following way in terms of the joint probability distribution.
- $= P[x_1 \mid x_2, x_3, \dots, x_n]P[x_2, x_3, \dots, x_n]$
- $= P[x_1 \mid x_2, x_3, \dots, x_n]P[x_2 \mid x_3, \dots, x_n] \dots P[x_{n-1} \mid x_n]P[x_n]$.
- In general for each variable X_i , we can write the equation as:
- $P(X_i \mid X_{i-1}, \dots, X_1) = P(X_i \mid \text{Parents}(X_i))$

Explanation of Bayesian network by example

- **Example:** Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary but also responds for 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 got 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 Burglary Alarm.
- **Problem:**
- **Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry.**

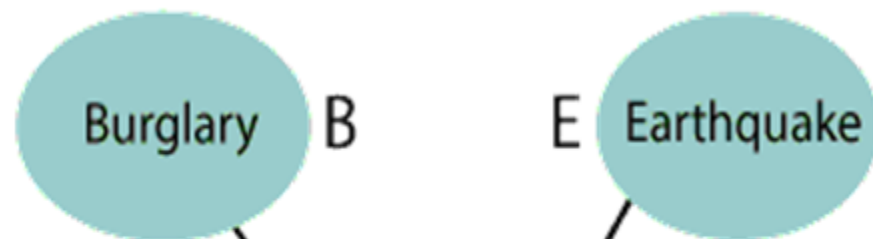
- **Solution:**

- The Bayesian network for the above problem is given below. The network structure is showing 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 is representing 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 or 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

- **List of all events occurring in this network:**
- **Burglary (B)**
- **Earthquake(E)**
- **Alarm(A)**
- **David Calls(D)**
- **Sophia calls(S)**

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

T	0.002
F	0.998



T	0.001
F	0.999

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

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



A	P(S=T)	P(S=F)
T	0.75	0.25
F	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.