Bayes' rule

What is Bayes' Theorem? Or Bayes Rule or Bayes Law

- Used to determine the conditional probability of event A when event B has already happened.
- The general statement of Bayes' theorem is "The conditional probability of an event A, given the occurrence of another event B, is equal to the product of the event of B, given A and the probability of A divided by the probability of event B." i.e.

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

This is known as Bayes' rule. Or **Bayes' theorem**. This equation is basic of most modern AI systems for **probabilistic inference**.

Bayes' rule Cont.....

- It shows the simple **relationship between joint and conditional probabilities**
- 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($A \land B$)= Joint probability of A and 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.

Bayes' rule Cont….

Terms Related to Bayes Theorem

- **Conditional Probability:** The probability of an event A based on the occurrence of another event B is termed conditional Probability. It is denoted as **P(A|B)** and represents the probability of A when event B has already happened.
- **Joint Probability:** When the probability of two more events occurring together and at the same time is measured it is marked as Joint Probability. For two events A and B, it is denoted by joint probability is denoted as, **P(A ^ B).**
- **Random Variables:** Real-valued variables whose possible values are determined by random experiments are called random variables.

Bayes' rule Cont….

Bayes' Theorem for multiple events is defined as,

- Consider the case where we have events A, B, C, ..., and Z, and we want to find the posterior probability of hypothesis H given all these events.
- The extended version of Bayes' Theorem for multiple events would be:

$$P(H \mid A, B, C, \ldots\ldots Z) = \frac{P(H)P(A, B, C, \ldots\ldots, Z \mid H)}{P(A, B, C, \ldots\ldots\ldots, Z)}$$

  - Where P(H|A, B, C….., Z) is the posterior probability of hypothesis H given events A, B, C, ..., Z.
  - P(H) is the prior probability of hypothesis H.
  - P(A,B,C,...,Z|H) is the joint probability of events A, B, C, ..., Z given hypothesis H.
  - P(A,B,C,...,Z) is the probability of events A, B, C, ..., Z occurring together.
  - [redacted] Z|H), you would need to consider the conditional probabiliti[redacted]nt given hypothesis H. For example, *P*(*A*|*H*), P(B|H), P(C|H)and so on.

Applying Bayes' Rule

- Suppose we want to perceive the effect of some unknown cause, and want to compute that cause, then the Bayes' rule becomes:

$$P(cause \mid effect) = \frac{P(effect \mid cause)\,P(cause)}{P(effect)}$$

Bayes' rule Cont….

Question: what is the probability that a patient has diseases meningitis with a stiff neck? i.e compute P(meningitis| stiff-neck) = ?

- Suppose a doctor knows that a meningitis causes a stiff neck in 80% of cases
  - Known probability that a patient has a stiff neck is 2%.
  - Doctor also has to know the incidence of meningitis in the population (1/30,000)
  - Let ***s*** be the proposition that patient has stiff neck and ***m*** be the proposition that patient has meningitis. , so we can calculate the following as:
  - P(s|m) = 0.8 (Probability of having a stiff neck given meningitis),
  - P(m) = 1/30000 = 0.0000333333333333
  - P(s) = 0.02 (Probability of having a stiff neck)

Bayes' rule Cont….

Question: Diseases meningitis with a stiff neck Cont…..

Using Bayes' rule, to calculate:

- P(m|s) (Probability of having meningitis given a stiff neck)

$$P(cause \mid effect) = \frac{P(effect \mid cause)\,P(cause)}{P(effect)}$$

$$P(m \mid s) = \frac{P(s \mid m)\,P(m)}{P(s)} = \frac{0.8 \times (1/30000)}{0.02} = 0.0001333333 \quad = 1/7500$$

- So, the calculated probability that a patient has meningitis given a stiff neck is 1/7500, which means approximately 0.00013 or about 0.013%.
- Hence, we can assume that 1 patient out of 7500 patients has meningitis [redacted]

Bayes' rule Cont….

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(King|Face), which means the drawn face card is a king card.

$$P(king \mid face) = \frac{P(face \mid king)P(king)}{P(face)} \quad \ldots\ldots\ldots\ldots\ldots\ldots 1$$

- P(king): probability that the card is King= 4/52= 1/13
- P(face): probability that a card is a face card= 3/13
  - There are 3 face cards in a deck of cards – J, Q, and K.
  - Now, consider that there are four types of cards namely spades, clubs, hearts and diamonds. Each of these types contain king, queen and jack. [redacted]e cards $=3\times 4=12$

Bayes' rule Cont….

Question: From a standard deck Cont….

There are 12 face cards in a pack.

- ∴ The number of favorable outcomes = 12
- Total number of outcomes = 52
- Hence, P (getting a face card) =Favorable outcomes/Total number of outcomes=12/52=3/13

- P(Face|King): probability of face card, when we assume it is a king = 1
- Putting all values in equation (1) we will get:

$$P(king \mid face) = \frac{1 * (\frac{1}{13})}{(\frac{3}{13})} = \frac{1}{3} = 0.3333$$

Bayes' rule Cont….

Example : There are three vases containing 3 white and 2 black balls; 2 white and 3 black balls; 1 black and 4 white balls respectively. There is an equal probability of each vase being chosen. One ball with equal probability is chosen at random. what is the probability that a white ball is drawn?

- **Solution:** **To Compute P(White-Ball)= ?**
- Let's denote the vases as follows:
  - Vase1: 3 white balls, 2 black balls
  - Vase2: 2 white balls, 3 black balls
  - Vase3: 1 black ball, 4 white balls
- Calculate the overall probability of drawing a white ball by considering the probabilities of each vase being chosen and then drawing a white ball from that vase. The probability of choosing each vase is equal, so .

[redacted] $P(Vase2) = P(Vase3) = 1/3$

Bayes' rule Cont….

Example 2: Three vases **Cont…**

- Now let's calculate the probability of drawing a white ball from each vase:
  - Vase 1: Probability of drawing a white ball

    $$P(WhiteBall/\ Vase1) = 3\ /\ (3+2)$$

  - Vase 2: Probability of drawing a white ball

    $$P(WhiteBall/Vase2) = 2\ /(2+3)$$

  - Vase 3: Probability of drawing a white ball

    $$P(WhiteBall/Vase3) = 4\ /(1+4)$$

- By theorem of total probability, we have

$$P(WhiteBall) = P(WhiteBall/\ Vase1)\ .\ P(Vase1) + P(WhiteBall/\ Vase2)\ .\ P(Vase2) + P(WhiteBall/\ Vase3)\ .\ P(Vase3)$$

$$= (3/5 \times 1/3) + (2/5 \times 1/3) + (4/5 \times 1/3)$$

[redacted]5 = 0.6

- So, the probab[redacted]ing a white ball is approximately 0.6 or 60%.

Bayes' rule Cont….

Application of Bayes' theorem in Artificial intelligence:

1. It is used to calculate the next step of the robot when the already executed step is given.
2. Bayes' theorem is helpful in weather forecasting.
3. It can solve the Monty Hall problem.

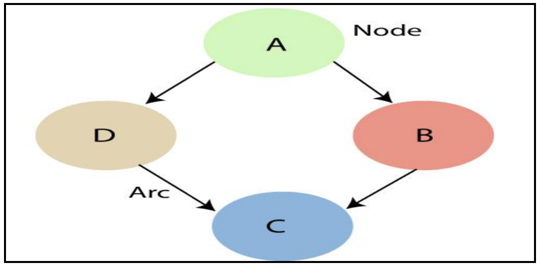
Bayesian Network

Introduction to Bayesian Network

- Bayesian Network, Bayesian Belief Network, **decision network** or a Probabilistic Graphical Model
- It is a graphical representation of probabilistic relationships among a set of variables.
- Powerful tool for modeling uncertainty and making predictions in situations where multiple variables interact with each other.
- Combines principles from probability theory and graph theory to model complex systems and capture dependencies between variables.
- Use probability theory for prediction and anomaly detection.
- It can also be used in various tasks including **prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction,** and **decision making under uncertainty**.

Key Concepts of Bayesian Networks

- **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 influence the other node, and if there is no directed link that means that nodes are independent with each other
  - An edge from Node A to Node B indicates that Node A has a direct influence on Node B. This can represent causal relationships, dependencies, or conditional probabilities.



Key Concepts of Bayesian Networks Cont….

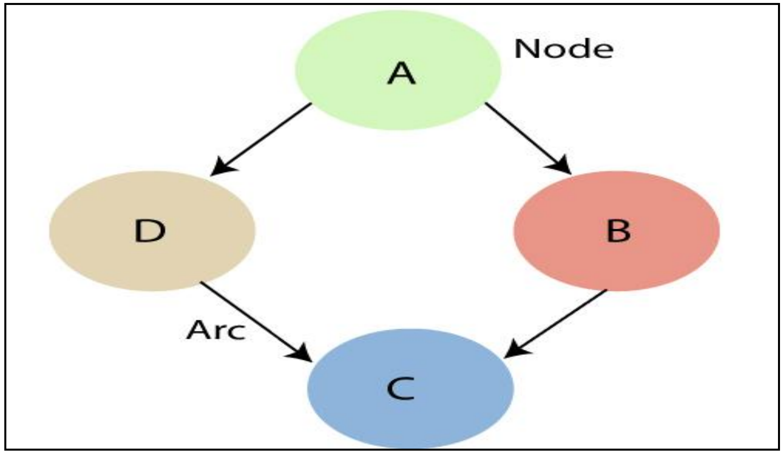
- **Conditional Probability Tables (CPTs)**: CPTs store the conditional probabilities associated with each node given its parent nodes. They quantify how the probability of a node's state changes based on the states of its parents.
  - The conditional distributions for each node are given as conditional 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 (True, True), (True, False), (False, True), and (False, [redacted]

Key Concepts of Bayesian Networks Cont….

- **Directed Acyclic Graph (DAG):** A Bayesian Network is represented as a DAG, which means the graph has no cycles.
- 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**.

Key Concepts of Bayesian Networks Cont…

- In the 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.



Joint Probability Distribution

- If we have variables x1, x2, x3,....., xn, then the probabilities of a different combination of x1, x2, x3.. xn, are known as Joint probability distribution.
- **$P[x_1, x_2, x_3, \ldots, x_n]$**, it can be written as the following way in terms of the joint probability distribution.

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

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

- In general for each variable Xi, we can write the equation as:
- $P(X_i | X_{i-1}, \ldots, X_1) = P(X_i \,|\, Parents(X_i))$

Steps to Build a Bayesian Network

- **Identify Variables:** Determine the variables that are relevant to the problem you're trying to model.
- **Define Relationships:** Specify the relationships between the variables, i.e., which variables influence each other.
- **Assign CPTs:** For each node, assign **Conditional Probability Tables** that represent the probabilities of the node's states given the states of its parent nodes.
- **Construct DAG:** Create a directed acyclic graph that represents the variables and their relationships.
- **Infer and Predict:** Use the network to perform probabilistic inference, make predictions, and analyze the behavior of the system.

Example on Bayesian Network

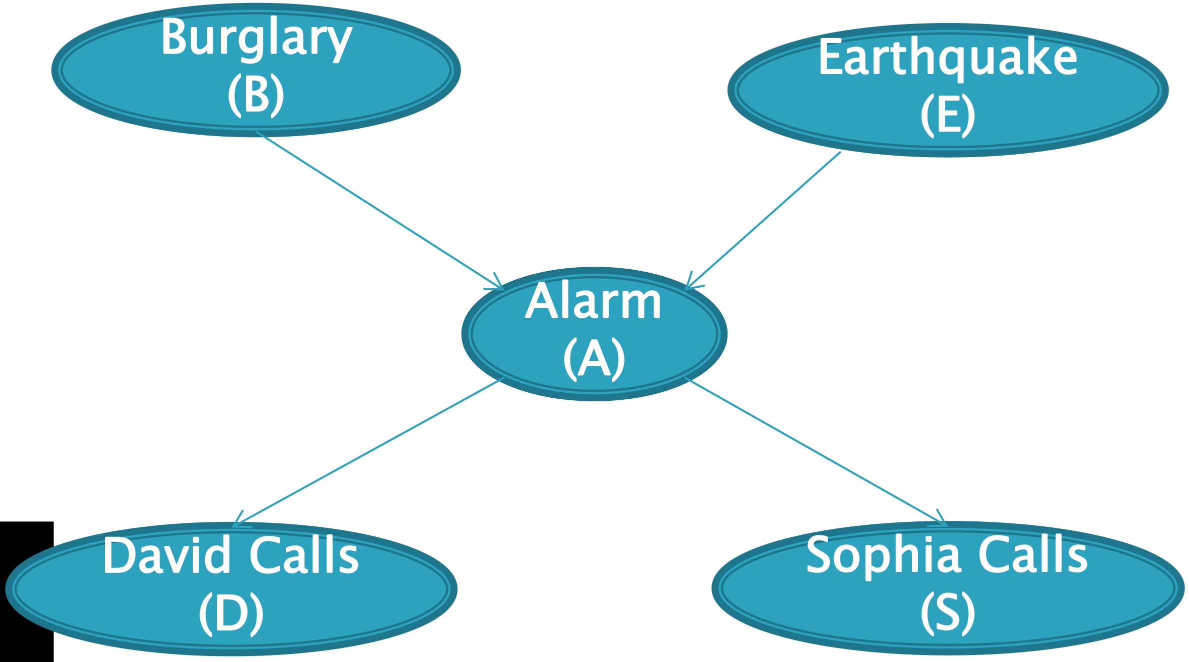
Example: Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary but also responds for minor earthquakes. Harry has two neighbors David and Sophia, who have taken a responsibility to inform Harry at work when they hear the alarm. David always 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 given Alarm i.e P(B|A).

Example on Burglar Alarm Cont.....

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.

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



Example on Burglar Alarm Cont.....

- Let's take the observed probability for the Burglary and earthquake component:
  - P(B= True) = 0.002, which is the probability of burglary.
  - P(B= False)= 0.998, which is the probability of no burglary.
  - P(E= True)= 0.001, which is the probability of a minor earthquake
  - P(E= 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	0.05
False	True	0.31	0.69
False	False	0.001	0.999

Example on Burglar Alarm Cont.....

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

Computing Different Probabilities

Suppose to Compute the probability of Burglary given the David Calls i.e

$$P(B \mid D) = \frac{P(D \wedge B)}{P(D)} = \frac{P(DB)}{P(D)} \text{.........} \quad .1$$

To Compute **1. P(D^ B) = ?** and **2. P(D)= ?**

P(D ^ B) $= P(D\ B\ A) \quad + \quad P(D\ B\ A') \quad \text{...............2}$

$= P(D|A\ B) \times P(A\ B) + P(D|\ A'\ B) \times P(A'\ B)$

$= P(D|A) \times P(A\ B) + P(D|\ A') \times P(A'\ B)$

$= 0.91 \times$ **P (A B)** $+ 0.05 \times$ **P(A' B)**

Now To Compute **P (A B) and P(A' B)**

P (A B) $= P(A\ B\ E) + P(A\ B\ E') = P(A|B\ E) \times P(B\ E) + P(A|\ B\ E') \times P(B\ E') \text{ ....3}$

= [redacted] $\times P(E) + P(A|\ B\ E') \times P(B) \times P(E')$

= [redacted] $.001 + 0.95 \times 0.002 \times 0.999 =$ **0.00189998**

Similarly compute P(A' B)

$\mathbf{P\ (A'\ B)} = P\ (A'\ B\ E\ ) + P(A'\ B\ E') = P\ (A'\ |B\ E\ ) \times P\ (B\ E\ ) + P\ (A'|\ B\ E') \times P(B\ E')$ ….4

$= P\ (A'\ |B\ E\ ) \times P\ (B\ ) \times P\ (E\ ) + P\ (A'|\ B\ E') \times P\ (B\ ) \times P\ (E'\ )$

$= 0.06 \times 0.002 \times 0.001 + 0.05 \times 0.002 \times 0.999 = \mathbf{0.00010002}$

On Substituting values of Eq: 3 and 4 in 2 We get

$\mathbf{P(\ D \wedge B)} = 0.91 \times \mathbf{P\ (A\ B)} + 0.05 \times \mathbf{P(A'\ B)}$ ………5

$= 0.91 \times \mathbf{0.00189998} + 0.05 \times \mathbf{0.00010002} = \mathbf{0.0017339828}$

2. Now To Compute **P(D)= ?**

$\mathbf{P\ (D)} = P\ (\ D\ A) + P\ (D\ A') = P\ (\ D|\ A)\ P(A) + P\ (D|\ A') \times P(A')$ …………..6

[redacted]$91 \times \mathbf{P(A)} + 0.05 \times \mathbf{P(A')}$

Example on Burglar Alarm Cont…..

Compute P(A) and P(A')

$P(A) = P(A\,B\,E) + P(A\,B\,E') + P(A\,B'\,E) + P(A\,B'\,E')$ ………7

$= P(A|B\,E) \times P(B\,E) + P(A|\,B\,E') \times P(B\,E') + P(A|\,B'\,E) \times P(B'\,E)$

$+ P(A|B'\,E') \times P(B'\,E')$

$= P(A|B\,E) \times P(B) \times P(E) + P(A|\,B\,E') \times P(B) \times P(E')$

$+ P(A|\,B'\,E) \times P(B') \times P(E) + P(A|B'\,E') \times P(B') \times P(E')$

$= [0.94 \times 0.002 \times 0.001] + [0.95 \times 0.002 \times 0.999]$

$+ [0.31 \times 0.998 \times 0.001] + [0.001 \times 0.998 \times 0.999]$

$=$ **0.003206362**

$P(A') = P(A'\,B\,E) + P(A'\,B\,E') + P(A'\,B'\,E) + P(A'\,B'\,E')$ …………………8

$= P(A'|B\,E) \times P(B\,E) + P(A'|\,B\,E') \times P(B\,E') + P(A'|\,B'\,E) \times P(B'\,E)$

$+ P(A'|B'\,E') \times P(B'\,E')$

$= P(A'|B\,E)\,P(B) \times P(E) + P(A'|\,B\,E') \times P(B) \times P(E')$

$+ P(A'|\,B'\,E) \times P(B') \times P(E) + P(A'|B'\,E') \times P(B') \times P(E')$

[redacted]01] + [0.05× 0.002 × 0.999]

+ [redacted]69× 0.998 × 0.001] + [0.999× 0.998 × 0.999]

= **0.996793638**

Example on Burglar Alarm Cont…..

On Substituting 7^{th} and 8^{th} Eq Value in Eq. 6, we get

$$P(D) = P(D\ A) + P(D\ A') = P(D|\ A)\ P(A) + P(D|\ A') \times P(A') \quad \ldots\ldots\ldots\ldots 9$$

$$= 0.91 \times P(A) + 0.05 \times P(A')$$

$$= 0.91 \times 0.003206362 + 0.05 \times 0.996793638$$

$$= 0.062914583058$$

On Substituting 5^{th} and 9^{th} Eq Value in Eq. 2, we get

$$P(D|B) = P(D \wedge B)\ /P(D) = 0.0017339828 / 0.062914583058$$

$$= 0.027560904256513431124898674044 39$$

$$= 0.0275609 = 0.027561$$

Example on Burglar Alarm Cont.....

Compute Another probability:

- **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: From the formula of joint distribution, we can write the problem statement in the form of probability distribution:

$$P(S, D, A, \neg B, \neg E) = P(S|A) \times P(D|A) \times P(A|\neg B \wedge \neg E) \times P(\neg B) \times P(\neg E)$$

$$= 0.75 \times 0.91 \times 0.001 \times 0.998 \times 0.999$$

$$= \mathbf{0.00068045.}$$

Advantages of Bayesian Networks

- **Modeling Uncertainty**: Bayesian Networks are excellent for handling situations with incomplete or uncertain information. They allow you to model and quantify uncertainty through probabilities.
- **Causal Inference**: Bayesian Networks can help identify causal relationships between variables. By analyzing the network's structure, you can gain insights into how changes in one variable might affect others.
- **Prediction and Inference**: Given some observed variables, Bayesian Networks allow you to perform probabilistic inference to predict the values of other variables.
- **Decision-Making**: Bayesian Networks can be extended to include decision nodes that represent actions or decisions to be made. This allows you to model decision-making scenarios under uncertainty.
- **[redacted]sis**: Bayesian Networks provide a visual representat[redacted]ex relationships, aiding in understanding and interpretation of the un[redacted]ing system.

Importance Bayesian Network

- Bayesian Networks find applications in various fields such as healthcare (diagnosis), finance (risk assessment), natural language processing (language models), and more.
- They offer a structured and intuitive way to represent and reason about uncertain information in complex systems.