

UNIT II

PROBABILISTIC REASONING

Acting under uncertainty – Bayesian inference – naïve bayes models. Probabilistic reasoning–Bayesian networks – exact inference in BN – approximate inference in BN – causal networks.

ACTING UNDER UNCERTAINTY

$A \rightarrow B$ means if A is true then B is true, if we are not sure about whether A is true or not then we cannot express this statement, this situation is called uncertainty.

To represent uncertain knowledge, uncertain reasoning or probabilistic reasoning is used.

Causes of uncertainty:

Following are some leading causes of uncertainty to occur in the real world.

1. Information occurred from unreliable sources.
2. Experimental Errors
3. Equipment fault
4. Temperature variation
5. Climate change

PROBABILISTIC REASONING:

Probabilistic reasoning is the way of Representing Knowledge in an Uncertain Domain. Probabilistic reasoning is a way of knowledge representation by applying the concept of probability to indicate the uncertainty in knowledge.

Probability: Probability can be defined as a chance that an uncertain event will occur. It is the numerical measure of the likelihood that an event will occur. Value of probability always remains between 0 and 1

- $0 \leq P(A) \leq 1$, where $P(A)$ is the probability of an event A.
- $P(A) = 0$, indicates total uncertainty in an event A.
- $P(A) = 1$, indicates total certainty in an event A.

Axioms Of Probability

- Given a set U (universe), a probability function is a function defined over the subsets of U that maps each subset to the real numbers and that satisfies the Axioms of Probability
- $P(U) = 1, P(A) [0,1]$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- if $A \cap B = \{\}$ then $P(A \cup B) = P(A) + P(B)$

Probability of an uncertain event can be found by using the below formula.

$$\text{Probability of occurrence} = \frac{\text{Number of desired outcomes}}{\text{Total number of outcomes}}$$

- $P(\neg A)$ = probability of a not happening event.
- $P(\neg A) + P(A) = 1$.

Event: Each possible outcome of a variable is called an event.

Sample space: The collection of all possible events is called sample space.

Random variables: Random variables are used to represent the events and objects in the real world.

Prior probability: Prior probability of an event is probability computed before observing new information.

Posterior Probability: Probability that is calculated after all evidence or information has taken into account. It is a combination of prior probability and new information.

Conditional probability:

Conditional probability is a probability of occurring an event when another event has already happened.

Let's suppose, we want to calculate the event A when event B has already occurred, "the probability of A under the conditions of B", it can be written as:

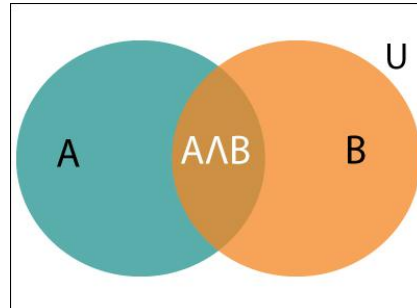
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Where $P(A \cap B)$ = Joint probability of a and B

$P(B)$ = Marginal probability of B.

If the probability of A is given and we need to find the probability of B, then it will be given as:

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



Example:

In a class, there are 70% of the students who like English and 40% of the students who likes English and mathematics, and then what is the percent of students those who like English also like mathematics?

Solution:

Let, A is an event that a student likes Mathematics

B is an event that a student likes English.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.4}{0.7} = 57\%$$

Hence, 57% are the students who like English also like Mathematics.

BAYES' THEOREM:

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**. **Bayesian inference** is an application of Bayes' theorem. It is a way to calculate the value of $P(B|A)$ with the knowledge of $P(A|B)$.

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 by product rule & conditional probability of event A with known event B:

As from product rule we can write:

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

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

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

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

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

The above equation (a) is called as **Bayes' rule** or **Bayes' theorem**. 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 Probability of hypothesis A when we have occurred an evidence B.

$P(B|A)$ is called **likelihood**, in which if 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.

In n general, we can write $P(B) = \sum_{i=1}^n P(A_i) * P(B|A_i)$, 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 events.

Applying Bayes' rule:

Bayes' rule allows us to compute single term $P(B|A)$ in terms of $P(A|B)$, $P(B)$, and $P(A)$. Suppose we want to perceive the effect of some unknown cause, and want to compute that cause, then Bayes' rule becomes:

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

Uses of Bayes' Theorem

In medical diagnosis, the goal is to determine identifications (diseases) given observations (symptoms). Bayes' Theorem provides such a relationship.

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

Suppose: A=Patient has measles, B =has a rash

$$\text{Then: } P(\text{measles}/\text{rash})=P(\text{rash}/\text{measles}) * P(\text{measles}) / P(\text{rash})$$

Example-1: A doctor is aware that disease meningitis causes a patient to have a stiff neck, and it occurs 80% of the time. 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%.

What is the probability that a patient has diseases meningitis with a stiff neck?

Given Data:

Let a be the proposition that patient has stiff neck and b be the proposition that patient has meningitis. , so

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

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

$$P(a) = .02$$

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} = \frac{0.8 * (\frac{1}{30000})}{0.02} = 0.0013333333.$$

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

Example-2: 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.

Solution:

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

P(king): probability that the card is King= 4/52= 1/13

P(face): probability that a card is a face card= 3/13

P(Face|King): probability of face card when we assume it is a king = 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 Artificial intelligence:

- 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.

BAYESIAN BELIEF NETWORK

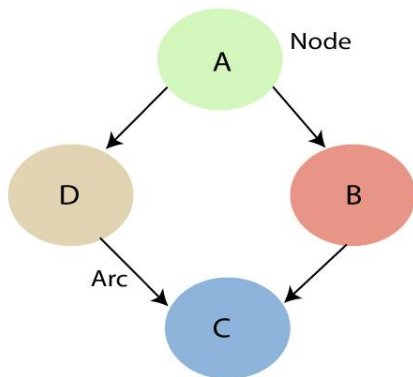
- A **Bayesian network** is a probabilistic graphical model(PGM) which represents a set of variables and their conditional dependencies using a directed acyclic graph(DAG).
- Each variable is associated with a conditional probability table which gives probability of the variables on which this node depends.
- Bayesian Belief Network is a Directed Acyclic Graph (or DAG),
 - Nodes represent random variables
 - Arcs represent direct influence
 - Nodes have conditional probability table that gives node probability
- The Bayesian network graph does not contain any cyclic graph. Hence, it is known as a directed acyclic graph or DAG.

Bayesian belief network deals with probabilistic events and to solve a problem which has uncertainty.

- 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.
- Bayesian Network consists of two parts:
 - **Directed Acyclic Graph**
 - **Table of conditional probabilities.**

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



A, B, C, and D are random variables represented by the nodes of the network graph.

- **Node B, which is connected with node A by a directed arrow, then node A is called parent of Node B.**
- **Node C is independent of node A.**

The Bayesian network has mainly two components:

- **Causal Component**
- **Actual numbers**

Example: Bayesian Network for burglar alarm

Problem: 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.

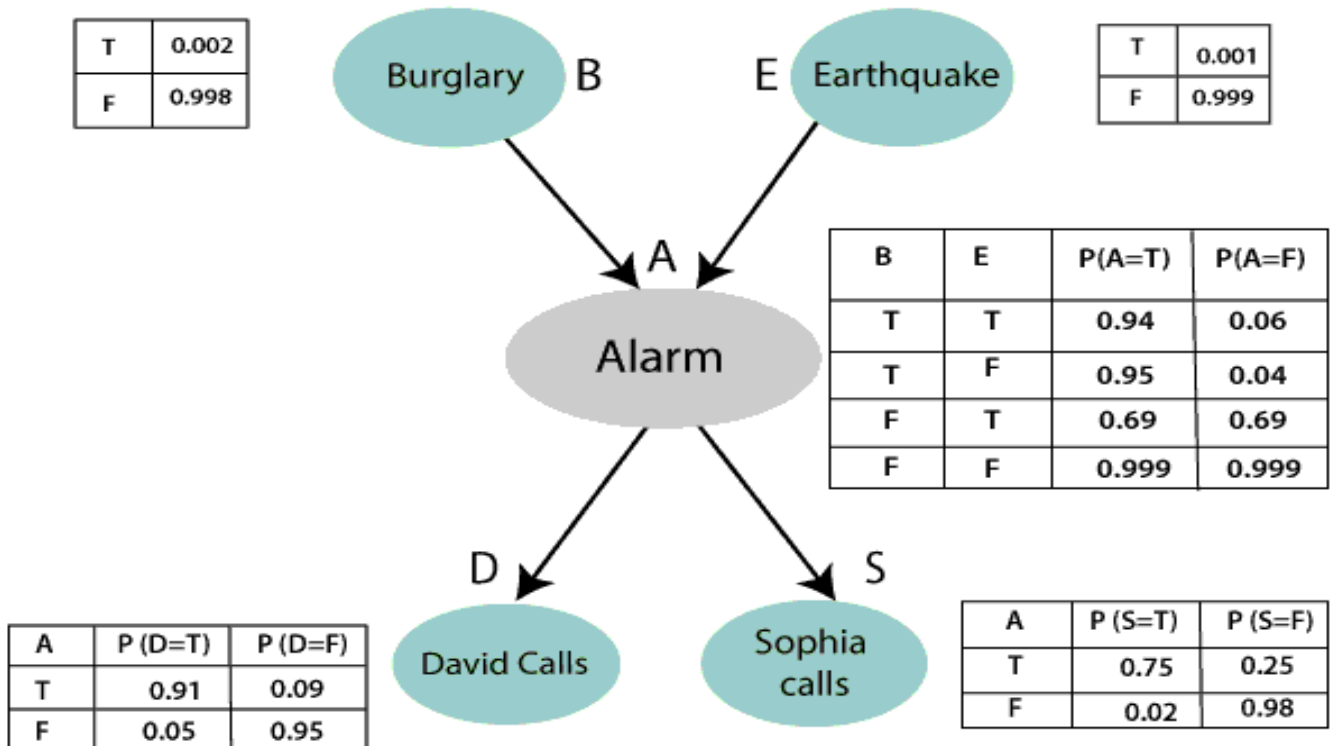
Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry.

Solution:

- Bayesian network is showing that burglary and earthquake is the parent node of alarm and directly affecting probability of alarm's going off, but David and Sophia's calls depend on alarm probability.
- The conditional distributions for each node are given as conditional probabilities table or CPT.

List of all events occurring in this network:

- **Burglary (B)**
- **Earthquake(E)**
- **Alarm(A)**
- **David Calls(D)**
- **Sophia calls(S)**



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.

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

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

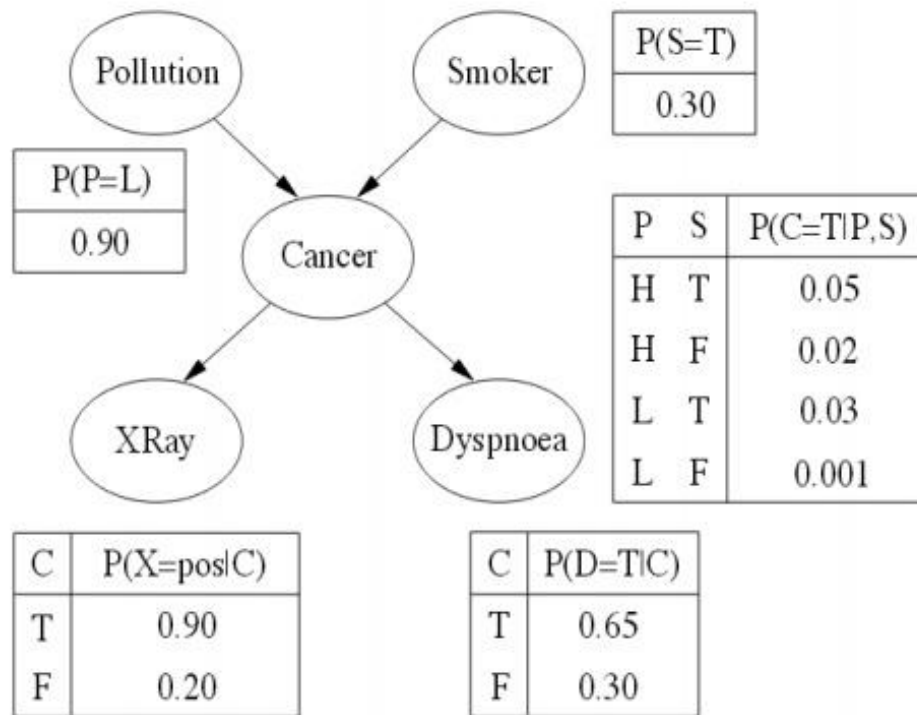
Conditional probability table for Sophia Calls:

The Conditional probability of Sophia that she calls is depending on its Parent Node "Alarm."

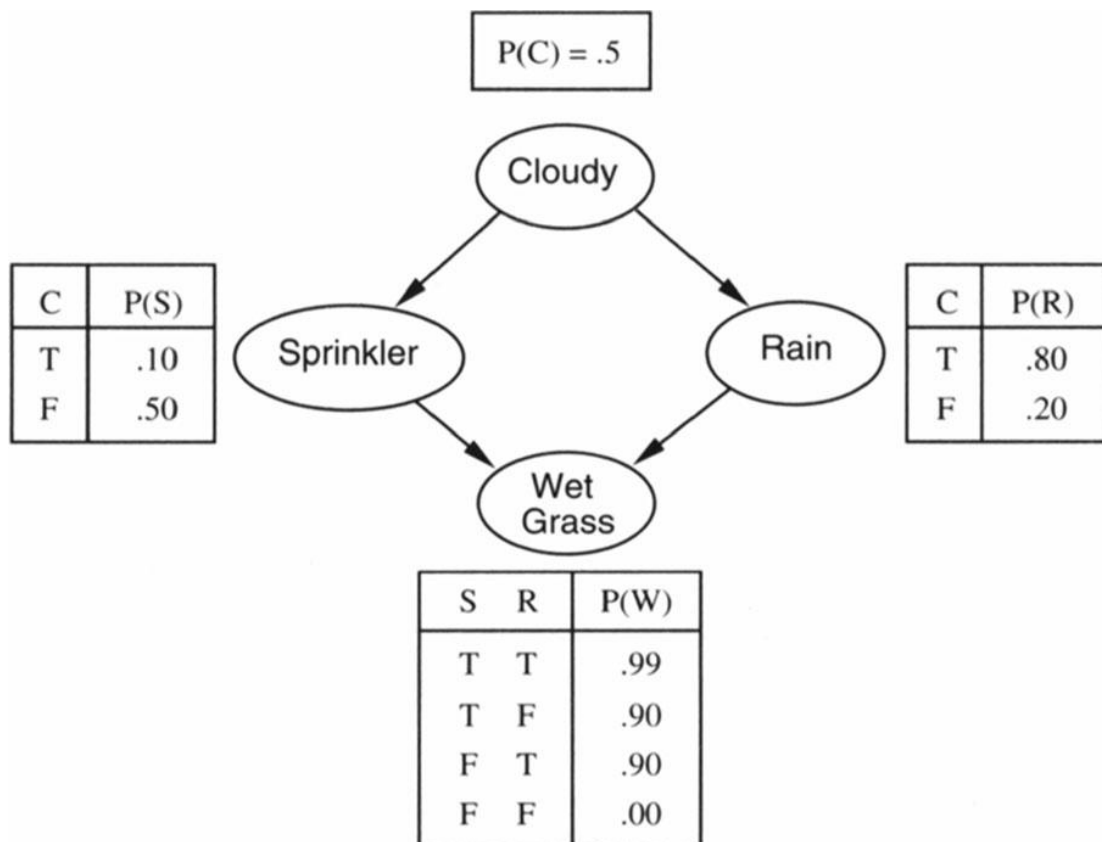
A	P(S= True)	P(S= False)
True	0.75	0.25
False	0.02	0.98

Example: BAYESIAN NETWORK FOR LUNG CANCER

Boolean nodes, which represent propositions, taking the binary values true (T) and false (F). In a medical diagnosis domain, the node Cancer would represent the proposition that a patient has cancer.



Example: BAYESIAN NETWORK FOR SPRINKLER RAIN PROBLEM:



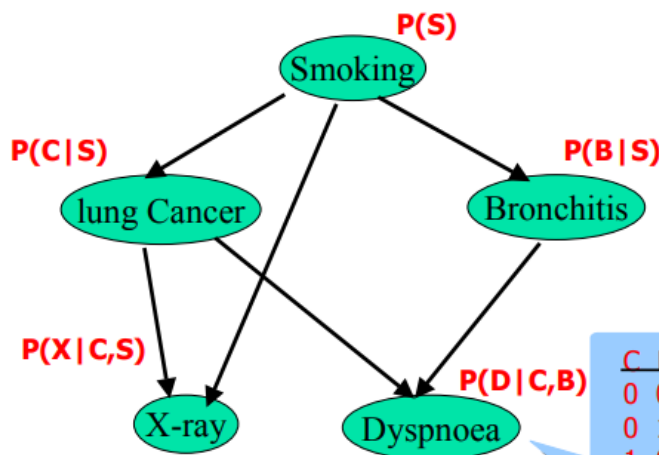
Another Example:

- Using this model, it is possible to perform inference and learning.

e.g. $P(\text{lung cancer}=\text{yes} \mid \text{smoking}=\text{no}, \text{positive X-ray}=\text{yes}) = ?$

For **example**, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. Efficient algorithms can perform inference and learning in Bayesian networks.

Bayesian Network: $BN = (G, \Theta)$



G - directed acyclic graph (DAG)
 nodes – random variables
 edges – direct dependencies

Θ - set of parameters in all conditional probability distributions (CPDs)

CPD:

C	B	D=0	D=1
0	0	0.1	0.9
0	1	0.7	0.3
1	0	0.8	0.2
1	1	0.9	0.1

CPD of node X:
 $P(X \mid \text{parents}(X))$

Compact representation of joint distribution in a **product form** (chain rule):

$$P(S, C, B, X, D) = P(S) P(C|S) P(B|S) P(X|C,S) P(D|C,B)$$

Applications:

- Real world applications are probabilistic in nature, and Bayesian network is used to represent the relationship between multiple events. It can also be used in various tasks including
 - Prediction
 - Anomaly detection
 - Diagnostics
 - Automated insight
 - Reasoning
 - Time series prediction
 - Decision making under uncertainty.

NAIVE BAYESIAN MODEL

Naive Bayesian classifier is based on Bayes' theorem with the independence assumptions between predictors. A Naive Bayesian model is easy to build, it particularly useful for very large datasets.

Why is it called Naïve Bayes?

Naïve Bayes algorithm is comprised of two words Naïve and Bayes,

- **Naïve:** It is called Naïve because occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify an apple without depending on each other.
- **Bayes:** It is called Bayes because it depends on the principle of [Bayes' Theorem](#).

Bayes' Theorem:

- Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:

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

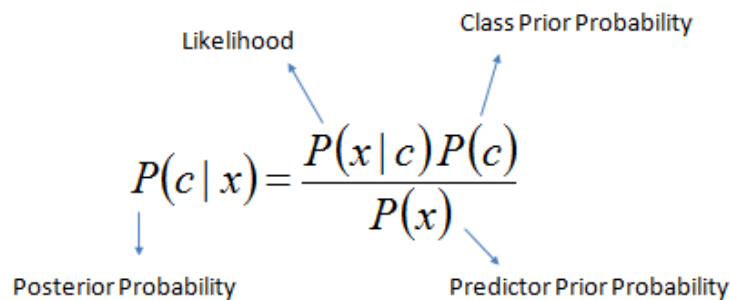
Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$


$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

- $P(c|x)$ is the posterior probability of *class (target)* given *predictor (attribute)*.
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

- Using Bayesian Theorem, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Likelihood of the Evidence given that the Hypothesis is True
 Prior Probability of the Hypothesis
 Posterior Probability of the Hypothesis given that the Evidence is True
 Prior Probability that the evidence is True

STEPS in NAÏVE BAYES MODEL:

Naive Bayes classifier calculates the probability of an event by following steps:

- Step 1: Calculate the prior probability for given class labels :
 - ✓ By Converting the given dataset into frequency tables.
- Step 2: Find Likelihood probability with each attribute for each class
 - ✓ By Generating Likelihood table by finding the probabilities of given features.
- Step 3: Use Bayes theorem to calculate the posterior probability.
- Step 4: See which class has a higher probability, then given input belongs to higher probability class.

Working of Naïve Bayes' Classifier:

Example: Dataset of **weather conditions** and corresponding target variable "**Play**". Using this dataset, Decide that whether to play or not on a particular day according to the weather conditions.

Problem: If the weather is sunny, then the Player should play or not?

Solution:

Consider the below **weather conditions** dataset:

	Outlook	Play
0	Rainy	Yes
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes

6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	No
10	Sunny	Yes
11	Rainy	No
12	Overcast	Yes
13	Overcast	Yes

Solution:

Given instance	Weather=Sunny	Play=?
-----------------------	---------------	--------

Frequency table for the Weather Conditions:

Weather	Yes	No
Overcast	5	0
Rainy	2	2
Sunny	3	2
Total	10	5

Likelihood table weather condition:

Weather	No	Yes	
Overcast	0	5	$5/14=0.35$
Rainy	2	2	$4/14=0.29$
Sunny	2	3	$5/14=0.35$
All	$4/14=0.29$	$10/14=0.71$	

Applying Bayes'theorem:

$$P(\text{Yes}|\text{Sunny})= P(\text{Sunny}|\text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$$

$$P(\text{Sunny}|\text{Yes})= 3/10= 0.3$$

$$P(\text{Sunny})= 0.35$$

$$P(\text{Yes})=0.71$$

$$\text{So } P(\text{Yes}|\text{Sunny}) = 0.3 * 0.71 / 0.35 = \mathbf{0.60}$$

$$P(\text{No}|\text{Sunny})= P(\text{Sunny}|\text{No}) * P(\text{No}) / P(\text{Sunny})$$

$$P(\text{Sunny}|\text{NO})= 2/4=0.5$$

$$P(\text{No})= 0.29$$

$$P(\text{Sunny})= 0.35$$

$$\text{So } P(\text{No}|\text{Sunny})= 0.5 * 0.29 / 0.35 = \mathbf{0.41}$$

So as we can see from the above calculation that $P(\text{Yes}|\text{Sunny}) > P(\text{No}|\text{Sunny})$

Hence on a Sunny day, Player can play the game.

Solution:

Given instance	Weather=Sunny	Play=Yes
-----------------------	---------------	----------

Advantages of Naïve Bayes Classifier:

- Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- It can be used for Binary as well as Multi-class Classifications.
- It performs well in Multi-class predictions as compared to the other Algorithms.
- It is the most popular choice for **text classification problems**.

Disadvantages of Naïve Bayes Classifier:

- Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.

Applications of Naïve Bayes Classifier:

- It is used for **Credit Scoring**.
- It is used in **medical data classification**.
- It can be used in **real-time predictions** because Naïve Bayes Classifier is an eager learner.
- It is used in Text classification such as **Spam filtering** and **Sentiment analysis**.

Types of Naïve Bayes Model:

There are three types of Naive Bayes Model, which are given below:

- **Gaussian:** The Gaussian model assumes that features follow a normal distribution. This means if predictors take continuous values instead of discrete, then the model assumes that these values are sampled from the Gaussian distribution.
- **Multinomial:** The Multinomial Naïve Bayes classifier is used when the data is multinomial distributed. It is primarily used for document classification problems, it means a particular document belongs to which category such as Sports, Politics, education, etc. The classifier uses the frequency of words for the predictors.
- **Bernoulli:** The Bernoulli classifier works similar to the Multinomial classifier, but the predictor variables are the independent Booleans variables. Such as if a particular word is present or not in a document. This model is also famous for document classification tasks.

Example 2: Weather dataset

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

The posterior probability can be calculated by first, constructing a frequency table for each attribute against the target. Then, transforming the frequency tables to likelihood tables and finally use the Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

$$P(x | c) = P(\text{Sunny} | \text{Yes}) = 3 / 9 = 0.33$$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$$P(x) = P(\text{Sunny}) = 5 / 14 = 0.36$$

$$P(c) = P(\text{Yes}) = 9 / 14 = 0.64$$

Posterior Probability: $P(c | x) = P(\text{Yes} | \text{Sunny}) = 0.33 \times 0.64 \div 0.36 = 0.60$



$$P(x | c) = P(\text{Sunny} | \text{No}) = 2 / 5 = 0.4$$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
		9	5	14

$$P(x) = P(\text{Sunny}) = 5 / 14 = 0.36$$

$$P(c) = P(\text{No}) = 5 / 14 = 0.36$$

Posterior Probability: $P(c | x) = P(\text{No} | \text{Sunny}) = 0.40 \times 0.36 \div 0.36 = 0.40$



The likelihood tables for all four predictors.

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3



		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1



		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1



		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3



		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

Outlook	Temp	Humidity	Windy	Play
Rainy	Cool	High	True	?



$$P(\text{Yes} | X) = P(\text{Rainy} | \text{Yes}) \times P(\text{Cool} | \text{Yes}) \times P(\text{High} | \text{Yes}) \times P(\text{True} | \text{Yes}) \times P(\text{Yes})$$

$$P(\text{Yes} | X) = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.00529 \rightarrow 0.2 = \frac{0.00529}{0.02057 + 0.00529}$$

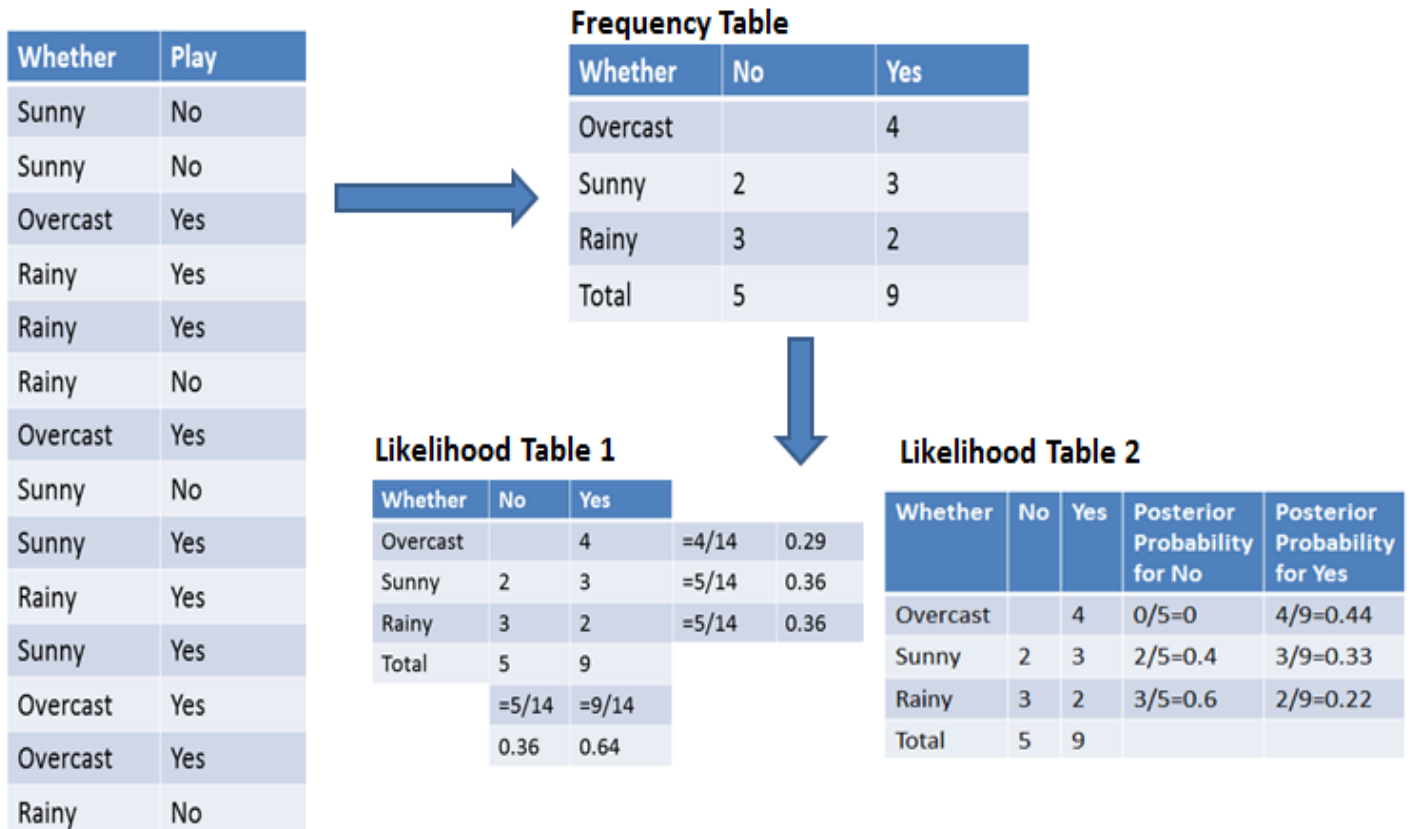
$$P(\text{No} | X) = P(\text{Rainy} | \text{No}) \times P(\text{Cool} | \text{No}) \times P(\text{High} | \text{No}) \times P(\text{True} | \text{No}) \times P(\text{No})$$

$$P(\text{No} | X) = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.02057 \rightarrow 0.8 = \frac{0.02057}{0.02057 + 0.00529}$$

RESULT: Play Golf=no

EXAMPLE-3

Two tables frequency and likelihood tables is used to calculate prior and posterior probability. Frequency table contains the occurrence of labels for all features. There are two likelihood tables. Likelihood Table 1 is showing prior probabilities of labels and Likelihood Table 2 is showing the posterior probability.



Calculate the probability of playing when the weather is overcast.

Solution:Given new Instance :If **Weather =overcast** then **Play = ?**

Probability of playing:

$$P(\text{Yes} | \text{Overcast}) = P(\text{Overcast} | \text{Yes}) P(\text{Yes}) / P(\text{Overcast}) \dots\dots\dots(1)$$

Calculate Prior Probabilities:

$$P(\text{Overcast}) = 4/14 = 0.29$$

$$P(\text{Yes}) = 9/14 = 0.64$$

Calculate Posterior Probabilities:

$$P(\text{Overcast} | \text{Yes}) = 4/9 = 0.44$$

Put Prior and Posterior probabilities in equation (1)

$$P(\text{Yes} | \text{Overcast}) = 0.44 * 0.64 / 0.29 = 0.98(\text{Higher})$$

Probability of not playing:

$$P(\text{No} | \text{Overcast}) = P(\text{Overcast} | \text{No}) P(\text{No}) / P(\text{Overcast}) \dots\dots\dots(2)$$

Calculate Prior Probabilities:

$$P(\text{Overcast}) = 4/14 = 0.29$$

$$P(\text{No}) = 5/14 = 0.36$$

Calculate Posterior Probabilities:

$$P(\text{Overcast} | \text{No}) = 0/9 = 0$$

Put Prior and Posterior probabilities in equation (2)

$$P(\text{No} | \text{Overcast}) = 0 * 0.36 / 0.29 = 0$$

The probability of a 'Yes' class is higher. So if the weather is overcast then players will play the sport.

Solution: For Given Instance: If Weather = overcast then Play = Yes.

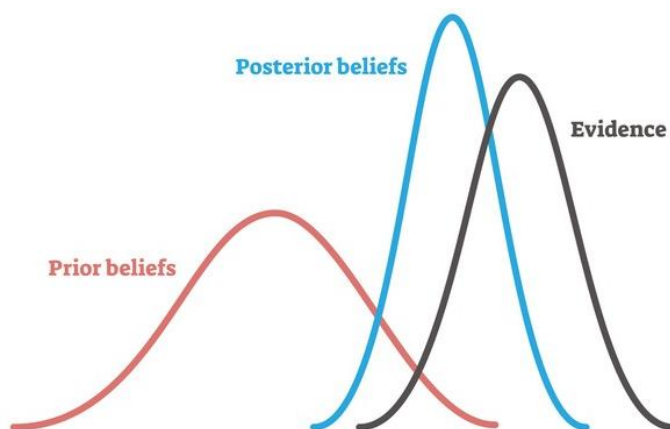
BAYESIAN INFERENCE

Bayesian inference is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available.

Bayesian inference derives the posterior probability as a consequence of two antecedents: a prior probability and a "likelihood function" derived from a statistical model for the observed data. Bayesian inference computes the posterior probability according to Bayes' theorem:

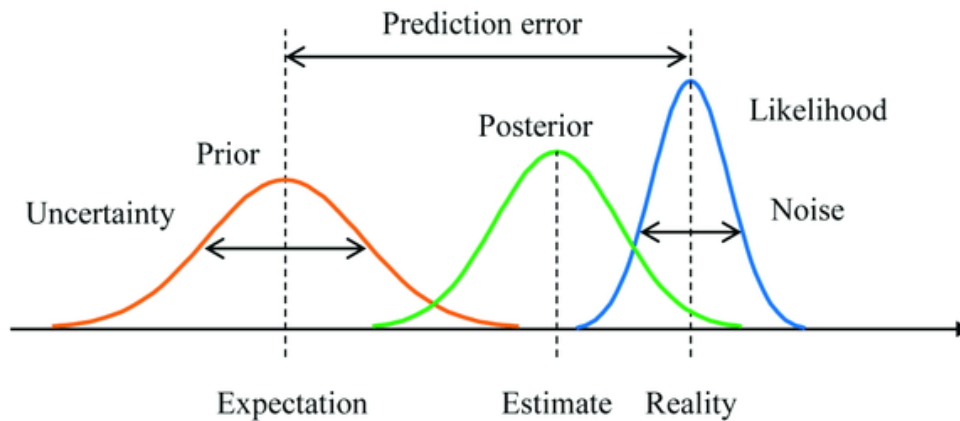
where

- $$P(H | E) = \frac{P(E | H) \cdot P(H)}{P(E)}$$
- H stands for any *hypothesis*
- P(H) is the *prior probability*, is the estimate of the probability of the hypothesis
- E is the *evidence*, corresponds to new data that were not used in computing the prior probability.
- P(H/E) is the *posterior probability*
- P(E/H) is called the *likelihood*.



Example:

Bayesian inference with a prior distribution, a posterior distribution, and a likelihood function. The prediction error is the difference between the prior expectation and the peak of the likelihood function (i.e., reality). Uncertainty is the variance of the prior. Noise is the variance of the likelihood function



Types of Bayesian Inference

- Exact inference in BN – Guessing is Exact
Example:
 - Junction Tree Algorithm (JTA)
 - Belief Propagation Method
 - Sum product method
- Approximate inference in BN – Guessing is Approximate
 - Stochastic and
 - Deterministic

Example:

- Markov Chain Monte Carlo Algorithm
- Expectation Propagation Method

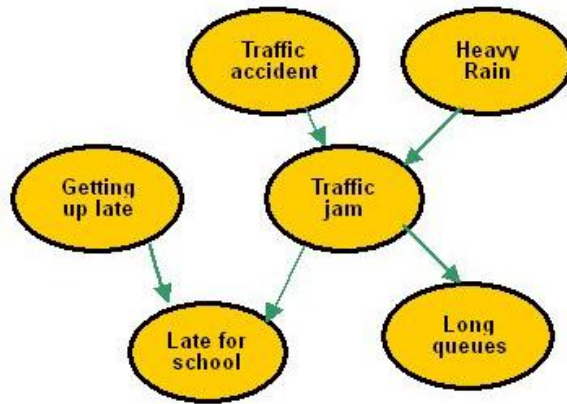
Applications:

Bayesian inference is an important technique in statistics, and especially in mathematical statistics. Bayesian updating is particularly important in the dynamic analysis of a sequence of data.

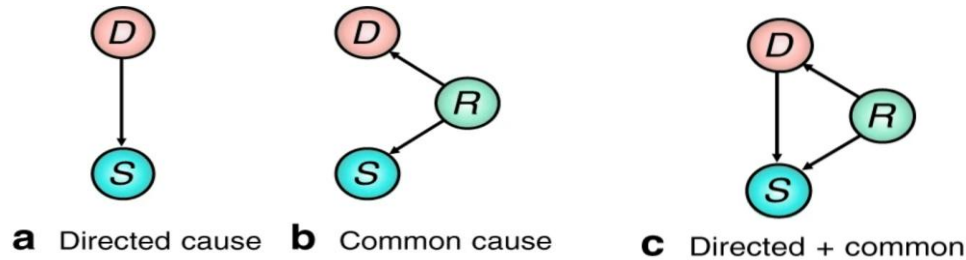
Bayesian inference has found application in a wide range of activities, including science, engineering, philosophy, medicine, sport, and law.

CAUSAL NETWORKS

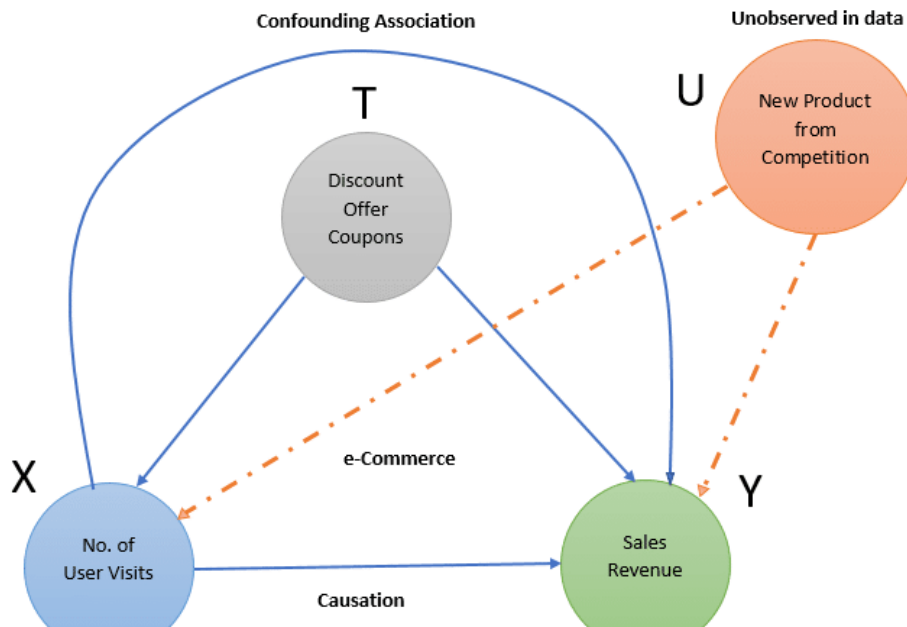
Causal Networks is a Acyclic Digraph, based on Cause and Effect relationships rather than Correlational Relationships. Causal networks are diagrams that indicate causal connections using arrows, where an arrow from A to B indicates that A is the cause of B.



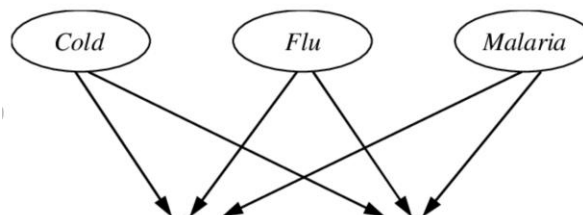
Types of Causes:



Example-1 Casual Network For E-Commerce Application



Example:2 Casual Network For Disease and Symptoms



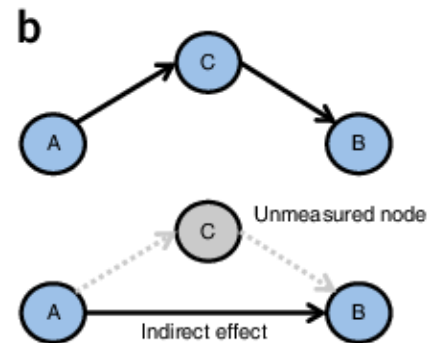
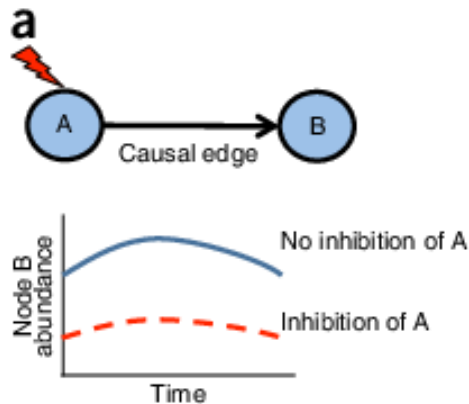
Example of a causal Bayesian network with causes (diseases) Cold, Flu, and Malaria and effects (symptoms) Nausea and Headache.

Advantages:

- More accurate insights and Decision Making Capability

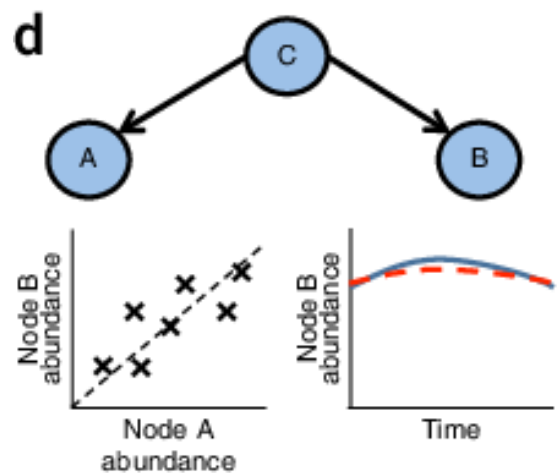
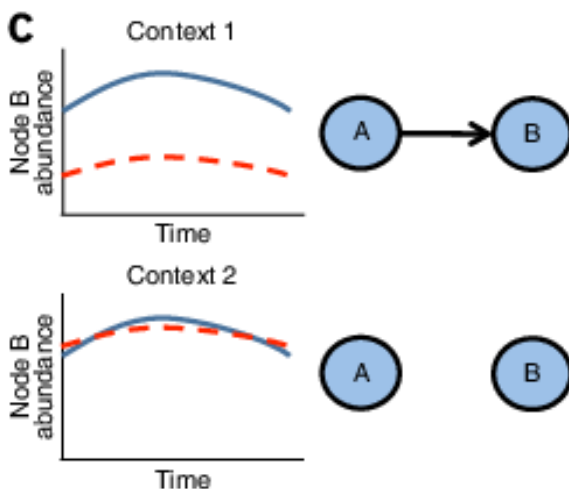
Types of Causal Networks:

(a) **Directed Casual edge may represent direct effects** denotes that inhibition of parent node A can change the abundance of the child node B.



(b) **Indirect Causal edges may represent indirect effects** that occur via unmeasured intermediate nodes. If node A causally influences node B via measured node C, the causal network should contain edges from A to C and from C to B. However, if node C is not measured (and is not part of the network), the causal network should contain an edge from A to B (bottom).

(c) **Causal edges depend on biological context**; Causal edge from A to B appears in context1, not in context2



(d) **Correlation and causation.** Nodes A and B are correlated owing to regulation by the same node (C), but in this example no sequence of mechanistic events links A to B, and thus inhibition of A does not change the abundance of B (lines in bottom right graph are as defined in a). There is no causal edge from A to B.