

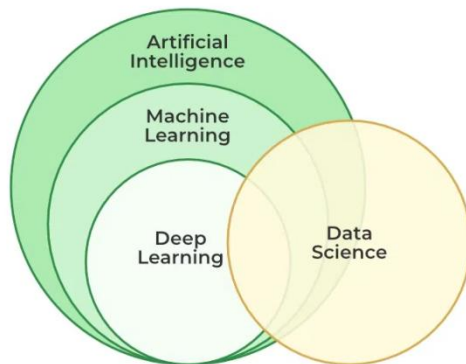
21CSC305P - Machine Learning

UNIT - I

Machine Learning

Machine learning is a branch of artificial intelligence that enables algorithms to uncover hidden patterns within datasets, allowing them to make predictions on new, similar data without explicit programming for each task. Traditional machine learning combines data with statistical tools to predict outputs, yielding actionable insights. This technology finds applications in diverse fields such as image and speech recognition, natural language processing, recommendation systems, fraud detection, portfolio optimization, and automating tasks.

For instance, recommender systems use historical data to personalize suggestions. Netflix, for example, employs collaborative and content-based filtering to recommend movies and TV shows based on user viewing history, ratings, and genre preferences. Reinforcement learning further enhances these systems by enabling agents to make decisions based on environmental feedback, continually refining recommendations.



Difference between Machine Learning and Traditional Programming

The Difference between Machine Learning and Traditional Programming is as follows:

Machine Learning	Traditional Programming	Artificial Intelligence
Machine Learning is a subset of artificial intelligence(AI) that focus on learning from data to develop an algorithm that can be used to make a prediction.	In traditional programming, rule-based code is written by the developers depending on the problem statements.	Artificial Intelligence involves making the machine as much capable, So that it can perform the tasks that typically require human intelligence.
Machine Learning uses a	Traditional programming	AI can involve many different

Machine Learning	Traditional Programming	Artificial Intelligence
data-driven approach, It is typically trained on historical data and then used to make predictions on new data.	is typically rule-based and deterministic. It hasn't self-learning features like Machine Learning and AI.	techniques, including Machine Learning and Deep Learning, as well as traditional rule-based programming.
ML can find patterns and insights in large datasets that might be difficult for humans to discover.	Traditional programming is totally dependent on the intelligence of developers. So, it has very limited capability.	Sometimes AI uses a combination of both Data and Pre-defined rules, which gives it a great edge in solving complex tasks with good accuracy which seem impossible to humans.
Machine Learning is the subset of AI. And Now it is used in various AI-based tasks like Chatbot Question answering, self-driven car., etc.	Traditional programming is often used to build applications and software systems that have specific functionality.	AI is a broad field that includes many different applications, including natural language processing, computer vision, and robotics.

Machine Learning works in the following manner.

A machine learning algorithm works by learning patterns and relationships from data to make predictions or decisions without being explicitly programmed for each task. Here's a simplified overview of how a typical machine learning algorithm works:

1. Data Collection:

First, relevant data is collected or curated. This data could include examples, features, or attributes that are important for the task at hand, such as images, text, numerical data, etc.

2. Data Preprocessing:

Before feeding the data into the algorithm, it often needs to be preprocessed. This step may involve cleaning the data (handling missing values, outliers), transforming the data (normalization, scaling), and splitting it into training and test sets.

3. Choosing a Model:

Depending on the task (e.g., classification, regression, clustering), a suitable machine learning model is chosen. Examples include decision trees, neural networks, support vector machines, and more advanced models like deep learning architectures.

4. Training the Model:

The selected model is trained using the training data. During training, the algorithm learns patterns and relationships in the data. This involves adjusting model parameters iteratively to minimize the difference between predicted outputs and actual outputs (labels or targets) in the training data.

5. Evaluating the Model:

Once trained, the model is evaluated using the test data to assess its performance. Metrics such as accuracy, precision, recall, or mean squared error are used to evaluate how well the model generalizes to new, unseen data.

6. Fine-tuning:

Models may be fine-tuned by adjusting hyperparameters (parameters that are not directly learned during training, like learning rate or number of hidden layers in a neural network) to improve performance.

7. Prediction or Inference:

Finally, the trained model is used to make predictions or decisions on new data. This process involves applying the learned patterns to new inputs to generate outputs, such as class labels in classification tasks or numerical values in regression tasks.

Machine Learning lifecycle:

The lifecycle of a machine learning project involves a series of steps that include:

1. Study the Problems:

The first step is to study the problem. This step involves understanding the business problem and defining the objectives of the model.

2. Data Collection:

When the problem is well-defined, we can collect the relevant data required for the model. The data could come from various sources such as databases, APIs, or web scraping.

3. Data Preparation:

When our problem-related data is collected, then it is a good idea to check the data properly and make it in the desired format so that it can be used by the model to find the hidden patterns. This can be done in the following steps:

- Data cleaning
- Data Transformation
- Explanatory Data Analysis and Feature Engineering
- Split the dataset for training and testing.

4. Model Selection:

The next step is to select the appropriate machine learning algorithm that is suitable for our problem. This step requires knowledge of the strengths and weaknesses of different algorithms. Sometimes we use multiple models and compare their results and select the best model as per our requirements.

5. Model building and Training:

- After selecting the algorithm, we have to build the model.
- In the case of traditional machine learning building mode is easy it is just a few hyperparameter tunings.
- In the case of deep learning, we have to define layer-wise architecture along with input and output size, number of nodes in each layer, loss function, gradient descent optimizer, etc.
- After that model is trained using the preprocessed dataset.

6. Model Evaluation:

Once the model is trained, it can be evaluated on the test dataset to determine its accuracy and performance using different techniques. like classification report, F1 score, precision, recall, ROC Curve, Mean Square error, absolute error, etc.

7. Model Tuning:

Based on the evaluation results, the model may need to be tuned or optimized to improve its performance. This involves tweaking the hyperparameters of the model.

8. Deployment:

Once the model is trained and tuned, it can be deployed in a production environment to make predictions on new data. This step requires integrating the model into an existing software system or creating a new system for the model.

9. Monitoring and Maintenance:

Finally, it is essential to monitor the model's performance in the production environment and perform maintenance tasks as required. This involves monitoring for data drift, retraining the model as needed, and updating the model as new data becomes available.

Supervised Machine Learning:

- Supervised learning is a type of machine learning in which the algorithm is trained on the labeled dataset. It learns to map input features to targets based on labeled training data. In supervised learning, the algorithm is provided with input features and corresponding output labels, and it learns to generalize from this data to make predictions on new, unseen data.
- There are two main types of supervised learning:
- Regression: Regression is a type of supervised learning where the algorithm learns to predict continuous values based on input features. The output labels in regression are continuous values, such as stock prices, and housing prices. The different regression algorithms in machine learning are: Linear Regression, Polynomial Regression, Ridge Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, etc
- Classification: Classification is a type of supervised learning where the algorithm learns to assign input data to a specific category or class based on input features. The output labels in classification are discrete values. Classification algorithms can be binary, where the output is one of two possible classes, or multiclass, where the output can be one of several classes. The different Classification algorithms in machine learning are: Logistic Regression, Naive Bayes, Decision Tree

Unsupervised Machine Learning:

- Unsupervised learning is a type of machine learning where the algorithm learns to recognize patterns in data without being explicitly trained using labeled examples. The goal of unsupervised learning is to discover the underlying structure or distribution in the data.
- There are two main types of unsupervised learning:
- Clustering: Clustering algorithms group similar data points together based on their characteristics. The goal is to identify groups, or clusters, of data points that are similar to each other, while being distinct from other groups. Some popular clustering algorithms include K-means, Hierarchical clustering, and DBSCAN.
- Dimensionality reduction: Dimensionality reduction algorithms reduce the number of input variables in a dataset while preserving as much of the original information as possible. This is

useful for reducing the complexity of a dataset and making it easier to visualize and analyze. Some popular dimensionality reduction algorithms include Principal Component Analysis (PCA), t-SNE, and Autoencoders

Reinforcement Machine Learning

- Reinforcement learning is a type of machine learning where an agent learns to interact with an environment by performing actions and receiving rewards or penalties based on its actions. The goal of reinforcement learning is to learn a policy, which is a mapping from states to actions, that maximizes the expected cumulative reward over time.
- There are two main types of reinforcement learning:
- Model-based reinforcement learning: In model-based reinforcement learning, the agent learns a model of the environment, including the transition probabilities between states and the rewards associated with each state-action pair. The agent then uses this model to plan its actions in order to maximize its expected reward. Some popular model-based reinforcement learning algorithms include Value Iteration and Policy Iteration.
- Model-free reinforcement learning: In model-free reinforcement learning, the agent learns a policy directly from experience without explicitly building a model of the environment. The agent interacts with the environment and updates its policy based on the rewards it receives. Some popular model-free reinforcement learning algorithms include Q-Learning, SARSA, and Deep Reinforcement Learning.

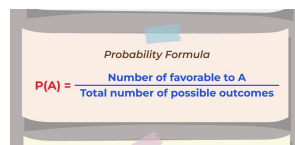
Applications of Machine Learning

- Now in this Machine learning tutorial, let's learn the applications of Machine Learning:
- Automation: Machine learning, which works entirely autonomously in any field without the need for any human intervention. For example, robots perform the essential process steps in manufacturing plants.
- Finance Industry: Machine learning is growing in popularity in the finance industry. Banks are mainly using ML to find patterns inside the data but also to prevent fraud.
- Government organization: The government makes use of ML to manage public safety and utilities. Take the example of China with its massive face recognition. The government uses Artificial intelligence to prevent jaywalking.
- Healthcare industry: Healthcare was one of the first industries to use machine learning with image detection.
- Marketing: Broad use of AI is done in marketing thanks to abundant access to data. Before the age of mass data, researchers develop advanced mathematical tools like Bayesian analysis to estimate the value of a customer. With the boom of data, the marketing department relies on AI to optimize customer relationships and marketing campaigns.
- Retail industry: Machine learning is used in the retail industry to analyze customer behavior, predict demand, and manage inventory. It also helps retailers to personalize the shopping experience for each customer by recommending products based on their past purchases and preferences.
- Transportation: Machine learning is used in the transportation industry to optimize routes, reduce fuel consumption, and improve the overall efficiency of transportation systems. It also plays a role in autonomous vehicles, where ML algorithms are used to make decisions about navigation and safety.

Probability Theory

Probability Theory: Probability is defined as the chance of happening or occurrences of an event. Generally, the possibility of analyzing the occurrence of any event concerning previous data is called probability. For example, if a fair coin is tossed, what is the chance that it lands on the head? These types of questions are answered under probability.

Probability measures the likelihood of an event's occurrence. In situations where the outcome of an event is uncertain, we discuss the probability of specific outcomes to understand their chances of happening. The study of events influenced by probability falls under the domain of statistics.


$$P(A) = \frac{\text{Number of favorable to A}}{\text{Total number of possible outcomes}}$$

Basics of Probability Theory

Various terms used in probability theory are discussed below,

Random Experiment

In probability theory, any event which can be repeated multiple times and its outcome is not hampered by its repetition is called a Random Experiment. Tossing a coin, rolling dice, etc. are random experiments.

Sample Space

The set of all possible outcomes for any random experiment is called sample space. For example, throwing dice results in six outcomes, which are 1, 2, 3, 4, 5, and 6. Thus, its sample space is (1, 2, 3, 4, 5, 6)

Event

The outcome of any experiment is called an event. Various types of events used in probability theory are,

- **Independent Events:** The events whose outcomes are not affected by the outcomes of other future and/or past events are called independent events. **For example**, the output of tossing a coin in repetition is not affected by its previous outcome.

- **Dependent Events:** The events whose outcomes are affected by the outcome of other events are called dependent events. **For example**, picking oranges from a bag that contains 100 oranges without replacement.
- **Mutually Exclusive Events:** The events that can not occur simultaneously are called mutually exclusive events. **For example**, obtaining a head or a tail in tossing a coin, because both (head and tail) can not be obtained together.
- **Equally likely Events:** The events that have an equal chance or probability of happening are known as equally likely events. **For example**, observing any face in rolling dice has an equal probability of $1/6$.

Random Variable

A variable that can assume the value of all possible outcomes of an experiment is called a random variable in Probability Theory. Random variables in probability theory are of two types which are discussed below,

Discrete Random Variable

Variables that can take countable values such as 0, 1, 2,... are called [discrete random variables](#).

Continuous Random Variable

Variables that can take an infinite number of values in a given range are called [continuous random variables](#).

Probability Theory Formulas

There are various formulas that are used in probability theory and some of them are discussed below,

- **Theoretical Probability Formula:** $(\text{Number of Favourable Outcomes}) / (\text{Number of Total Outcomes})$
- **Empirical Probability Formula:** $(\text{Number of times event A happened}) / (\text{Total number of trials})$
- **Addition Rule of Probability:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Complementary Rule of Probability:** $P(A') = 1 - P(A)$
- **Independent Events:** $P(A \cap B) = P(A) \cdot P(B)$
- **Conditional Probability:** $P(A | B) = P(A \cap B) / P(B)$
- **Bayes' Theorem:** $P(A | B) = P(B | A) \cdot P(A) / P(B)$

Bayes' Theorem

Bayes' Theorem is used to determine the conditional probability of an event. It was named after an English statistician, **Thomas Bayes** who discovered this formula in 1763. Bayes Theorem is a very important theorem in mathematics, that laid the foundation of a unique [statistical inference](#) approach called the **Bayes' inference**. It is used to find the probability of an event, based on prior knowledge of conditions that might be related to that event.

Bayes Theorem

Illustration: Guessing the Pet

Problem Statement: Is it a cat or a dog in the box?

Initial Belief: 50/50 chances

→ $P(\text{Cat}) = 0.5$

→ $P(\text{Dog}) = 0.5$

We have a CLUE: Pet is quiet

→ $P(\text{Quiet} | \text{Cat}) = 80\% \text{ or } 0.8$

→ $P(\text{Quiet} | \text{Dog}) = 30\% \text{ or } 0.3$

From Bayes Theorem probability that pet is a cat given it is quiet is:

P(Cat | Quiet): $\frac{P(\text{Quiet Cat}) \times P(\text{Cat})}{P(\text{Quiet})}$

For example, if we want to find the probability that a white marble drawn at random came from the first bag, given that a white marble has already been drawn, and **there are three bags each containing some white and black marbles**, then we can use Bayes' Theorem.

This article explores the Bayes theorem including its statement, proof, derivation, and formula of the theorem, as well as its applications with various examples.

What is Bayes' Theorem?

Bayes theorem (also known as the **Bayes Rule** or **Bayes Law**) is used to determine the conditional probability of event A when event B has already occurred.

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.

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

where,

- **P(A)** and **P(B)** are the probabilities of events A and B
- **P(A|B)** is the probability of event A when event B happens
- **P(B|A)** is the probability of event B when A happens

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

where,

- **P(A)** and **P(B)** are the probabilities of events A and B also P(B) is never equal to zero.
- **P(A|B)** is the probability of event A when event B happens
- **P(B|A)** is the probability of event B when A happens

Discrete Random Variable

Discrete Random Variables are an essential concept in probability theory and statistics. Discrete Random Variables play a crucial role in modelling real-world phenomena, from the number of customers who visit a store each day to the number of defective items in a production line.

Understanding discrete random variables is essential for making informed decisions in various fields, such as finance, engineering, and healthcare.

In this article, we'll delve into the fundamentals of discrete random variables, including their definition, [probability mass function](#), expected value, and variance. By the end of this article, you'll have a solid understanding of discrete random variables and how to use them to make better decisions.

Discrete Random Variable Definition

In probability theory, a [discrete random variable](#) is a type of random variable that can take on a finite or countable number of distinct values. These values are often represented by integers or whole numbers, other than this they can also be represented by other discrete values.

For example, the number of heads obtained after flipping a coin three times is a discrete random variable. The possible values of this variable are 0, 1, 2, or 3.

Examples of a Discrete Random Variable

A very basic and fundamental example that comes to mind when talking about discrete random variables is the rolling of an unbiased standard die. An unbiased standard die is a die that has six faces and equal chances of any face coming on top. Considering we perform this experiment, it is pretty clear that there are only six outcomes for our experiment.

Thus, our random variable can take any of the following discrete values from 1 to 6. Mathematically the collection of values that a random variable takes is denoted as a set. In this case, let the random variable be X .

Thus, $X = \{1, 2, 3, 4, 5, 6\}$

Another popular example of a discrete random variable is the number of heads when tossing of two coins. In this case, the random variable X can take only one of the three choices i.e., 0, 1, and 2.

Other than these examples, there are various other examples of random discrete variables. Some of these are as follows:

- The number of cars that pass through a given intersection in an hour.
- The number of defective items in a shipment of goods.
- The number of people in a household.
- The number of accidents that occur at a given intersection in a week.
- The number of red balls drawn in a sample of 10 balls taken from a jar containing both red and blue balls.
- The number of goals scored in a soccer match.

Probability Distributions for Discrete Random Variables

The probability distribution of a discrete random variable is described by its **probability mass function (PMF)**, which assigns a probability to each possible value of the variable. The key properties of a PMF are:

- Each probability is non-negative.
- The sum of all probabilities is equal to 1.

Common examples of discrete probability distributions include the **binomial distribution**, **Poisson distribution**, and **geometric distribution**.

Continuous Random Variable

Consider a generalized experiment rather than taking some particular experiment. Suppose that in your experiment, the outcome of this experiment can take values in some interval (a, b) .

That means that each and every single point in the interval can be taken up as the outcome values when you do the experiment. Hence, you do not have discrete values in this set of possible values but rather an interval.

Thus, $X = \{x: x \text{ belongs to } (a, b)\}$

Example of a Continuous Random Variable

Some examples of Continuous Random Variable are:

- The height of an adult male or female.
- The weight of an object.
- The time is taken to complete a task.
- The temperature of a room.
- The speed of a vehicle on a highway.

Random Variable

Random variable is a fundamental concept in statistics that bridges the gap between theoretical probability and real-world data. **random variable** in statistics is a function that assigns a real value to an outcome in the sample space of a random experiment. **For example:** if you roll a die, you can assign a number to each possible outcome.

There are two basic types of random variables,

- Discrete Random Variables (which take on specific values)
- Continuous Random Variables (assume any value within a given range)

In this article, we will learn about **random variable statistics, their types, random variable examples, and others in detail.**

What is a Random Variable?

Random Variable Probability is a mathematical concept that assigns numerical values to outcomes of a [sample space](#). They can **describe the outcomes of objective randomness (like tossing a coin) or subjective randomness(results of a cricket game).**

There are two types of Random Variables- **Discrete and Continuous.**

A random variable is considered a discrete random variable when it takes specific, or distinct values within an interval. Conversely, if it takes a continuous range of values, then it is classified as a continuous random variable.

Random Variable Definition

Random variable in statistics is a variable whose **possible values are numerical outcomes of a random phenomenon**. It is a function that assigns a real number to each outcome in the sample space of a random experiment.

We define a random variable as a **function that maps from the sample space of an experiment to the real numbers**. Mathematically, Random Variable is expressed as,

$$X: S \rightarrow R$$

where,

- X is Random Variable (It is usually denoted using capital letter)
- S is Sample Space
- R is Set of [Real Numbers](#)

Random Variable Example

Example 1

If two unbiased coins are tossed then find the random variable associated with that event.

Solution:

Suppose Two (unbiased) coins are tossed

X = number of heads. [X is a random variable or function]

Here, the sample space $S = \{HH, HT, TH, TT\}$

Example 2

Suppose a random variable X takes m different values i.e. sample space

$X = \{x_1, x_2, x_3, \dots, x_m\}$ with probabilities

$$P(X = x_i) = p_i$$

where $1 \leq i \leq m$

The probabilities must satisfy the following conditions :

- $0 \leq p_i \leq 1$; where $1 \leq i \leq m$
- $p_1 + p_2 + p_3 + \dots + p_m = 1$ Or we can say $0 \leq p_i \leq 1$ and $\sum p_i = 1$

Hence possible values for random variable X are 0, 1, 2.

$X = \{0, 1, 2\}$ where $m = 3$

- $P(X = 0) = (\text{Probability that number of heads is 0}) = P(TT) = 1/2 \times 1/2 = 1/4$
- $P(X = 1) = (\text{Probability that number of heads is 1}) = P(HT \mid TH) = 1/2 \times 1/2 + 1/2 \times 1/2 = 1/2$

- $P(X = 2) = (\text{Probability that number of heads is 2}) = P(HH) = 1/2 \times 1/2 = 1/4$

Here, you can observe that, $(0 \leq p_1, p_2, p_3 \leq 1/2)$

$$p_1 + p_2 + p_3 = 1/4 + 2/4 + 1/4 = 1$$

For example,

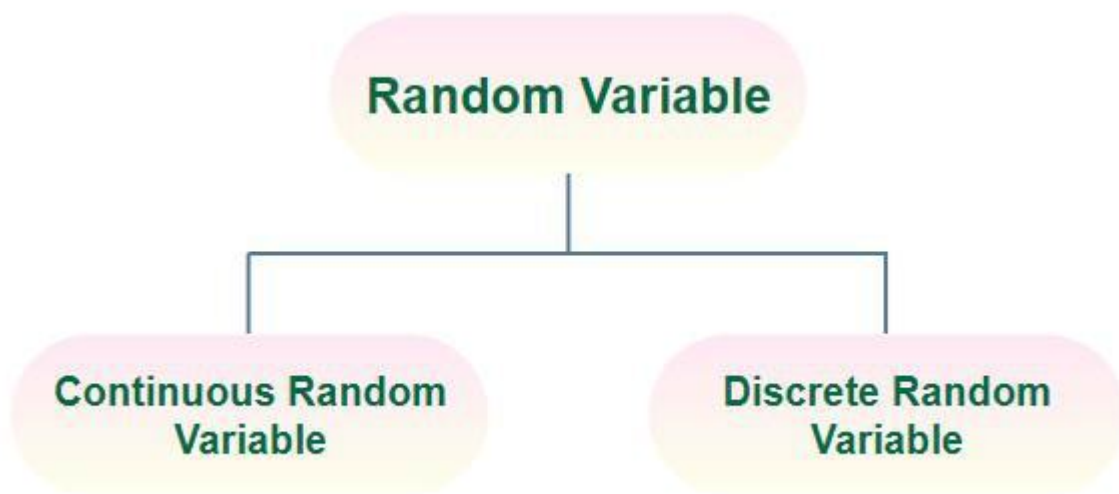
Suppose a dice is thrown ($X = \text{outcome of the dice}$). Here, the sample space $S = \{1, 2, 3, 4, 5, 6\}$. The output of the function will be:

- $P(X=1) = 1/6$
- $P(X=2) = 1/6$
- $P(X=3) = 1/6$
- $P(X=4) = 1/6$
- $P(X=5) = 1/6$
- $P(X=6) = 1/6$

Types of Random Variable

Random variables are of two types that are,

- Discrete Random Variable
- Continuous Random Variable



Discrete Random Variable

A [Discrete Random Variable](#) takes on a finite number of values. The probability function associated with it is said to be **PMF**.

PMF([Probability Mass Function](#))

If X is a discrete random variable and the PMF of X is $P(x_i)$, then

- $0 \leq p_i \leq 1$
- $\sum p_i = 1$ where the sum is taken over all possible values of x

Discrete Random Variables Example

Example: Let $S = \{0, 1, 2\}$

x_i	0	1	2
$P_i(X = x_i)$	0.3	0.5	

Find the value of $P(X = 0)$

Solution:

We know that the sum of all probabilities is equal to 1. And $P(X = 0)$ be P_1

$$P_1 + 0.3 + 0.5 = 1$$

$$P_1 = 0.2$$

Then, $P(X = 0)$ is 0.2

Continuous Random Variable

Continuous Random Variable takes on an infinite number of values. The probability function associated with it is said to be [PDF \(Probability Density Function\)](#).

PDF ([Probability Density Function](#))

If X is a continuous random variable. $P(x < X < x + dx) = f(x)dx$ then,

- $0 \leq f(x) \leq 1$; for all x
- $\int f(x) dx = 1$ over all values of x

Then $P(X)$ is said to be a PDF of the distribution.

Continuous Random Variables Example

Find the value of $P(1 < X < 2)$

Such that,

- $f(x) = kx^3; 0 \leq x \leq 3 = 0$

Otherwise $f(x)$ is a density function.

Solution:

If a function f is said to be a density function, then the sum of all probabilities is equal to 1.

Since it is a continuous random variable Integral value is 1 overall sample space s .

$$\int f(x) dx = 1$$

$$\int kx^3 dx = 1$$

$$K[x^4]/4 = 1$$

Given interval, $0 \leq x \leq 3 = 0$

$$K[3^4 - 0^4]/4 = 1$$

$$K(81/4) = 1$$

$$K = 4/81$$

Thus,

$$P(1 < X < 2) = k \times [X^4]/4$$

$$P = 4/81 \times [16-1]/4$$

$$P = 15/81$$

Random Variable Formulas

There are two main random variable formulas,

- Mean of Random Variable
- Variance of Random Variable

Let's learn about the same in detail,

Mean of Random Variable

For any random variable X where P is its respective probability we define its mean as,

$$\text{Mean}(\mu) = \sum X \cdot P$$

where,

- X is the random variable that consist of all possible values.
- P is the probability of respective variables

Variance of Random Variable

The variance of a random variable tells us how the random variable is spread about the mean value of the random variable. Variance of Random Variable is calculated using the formula,

$$\text{Var}(x) = \sigma^2 = E(X^2) - \{E(X)\}^2$$

where,

- $E(X^2) = \sum X^2 P$
- $E(X) = \sum X P$

Random Variable Functions

For any random variable X if it assume the values x_1, x_2, \dots, x_n where the probability corresponding to each random variable is $P(x_1), P(x_2), \dots, P(x_n)$, then the expected value of the variable is,

Expectation of X , $E(x) = \sum x \cdot P(x)$

Now for any new random variable Y in which the random variable X is its input, i.e. $Y = f(X)$, then the cumulative distribution function of Y is,

$$F_y(Y) = P(g(X) \leq y)$$

Random Variable Example with Solutions

Here are some of the solved examples on Random variable. Learn random variables by practicing these solved examples.

1. Find the mean value for the continuous random variable, $f(x) = x^2$, $1 \leq x \leq 3$

Solution:

Given,

$$f(x) = x^2$$

$$1 \leq x \leq 3$$

$$E(x) = \int_1^3 x \cdot f(x) dx$$

$$E(x) = \int_1^3 x \cdot x^2 \cdot dx$$

$$E(x) = \int_1^3 x^3 \cdot dx$$

$$E(x) = [x^4/4]_1^3$$

$$E(x) = 1/4 \{3^4 - 1^4\} = 1/4 \{81 - 1\}$$

$$E(x) = 1/4 \{80\} = 20$$

2. Find the mean value for the continuous random variable, $f(x) = e^x$, $1 \leq x \leq 3$

Solution:

Given,

$$f(x) = e^x$$

$$1 \leq x \leq 3$$

$$E(x) = \int_1^3 x \cdot f(x) dx$$

$$E(x) = \int_1^3 x \cdot e^x \cdot dx$$

$$E(x) = [x \cdot e^x - e^x]_1^3$$

$$E(x) = [e^x(x - 1)]_1^3$$

$$E(x) = e^3(2) - e(0)$$

3. Find the mean value for the continuous random variable, $f(x) = x^2$, $1 \leq x \leq 3$

Solution :

$$f(x) = x^2, 1 \leq x \leq 3$$

$$\text{Mean} = \int (x^3) dx / \int (x^2) dx \text{ from } 1 \text{ to } 3$$

$$= [(1/4)x^4]_1^3 / [(1/3)x^3]_1^3$$

$$= (81/4 - 1/4) / (27/3 - 1/3)$$

$$= 20 / 26/3$$

$$= 20 * 3/26 = 60/26 \approx 2.31$$

4. Find the mean value for the continuous random variable, $f(x) = 2x + 1$, $0 \leq x \leq 4$

Solution :

$$f(x) = 2x + 1, 0 \leq x \leq 4$$

$$\text{Mean} = \int (x(2x+1)) dx / \int (2x+1) dx \text{ from } 0 \text{ to } 4$$

$$= [(2/3)x^3 + (1/2)x^2]_0^4 / [x^2 + x]_0^4$$

$$= (170/3) / 20$$

$$= 17/6 \approx 2.83$$

5. Find the mean value for the continuous random variable, $f(x) = x^3$, $-1 \leq x \leq 2$

Solution :

$$f(x) = x^3, -1 \leq x \leq 2$$

$$\text{Mean} = \int (x^4) dx / \int (x^3) dx \text{ from } -1 \text{ to } 2$$

$$= [(1/5)x^5]_{-1}^2 / [(1/4)x^4]_{-1}^2$$

$$= (32/5 + 1/5) / (16/4 + 1/4)$$

$$= 33/5 / 17/4$$

$$= 334 / (517) \approx 1.55$$

6. Find the mean value for the continuous random variable, $f(x) = \sqrt{x}$, $1 \leq x \leq 9$

Solution :

$$f(x) = \sqrt{x}, 1 \leq x \leq 9$$

$$\text{Mean} = \int (x \sqrt{x}) dx / \int (\sqrt{x}) dx \text{ from } 1 \text{ to } 9$$

$$= [(2/5)x^{(5/2)}]_1^9 / [(2/3)x^{(3/2)}]_1^9$$

$$= (486 - 2/5) / (18 - 2/3)$$

$$\approx 5.2$$

7. Find the mean value for the continuous random variable, $f(x) = 3x^2 - 2x$, $0 \leq x \leq 3$

Solution :

$$f(x) = 3x^2 - 2x, 0 \leq x \leq 3$$

$$\text{Mean} = \int (x(3x^2 - 2x)) dx / \int (3x^2 - 2x) dx \text{ from } 0 \text{ to } 3$$

$$= [(3/4)x^4 - (2/3)x^3]_0^3 / [x^3 - x^2]_0^3$$

$$= (81/4 - 18) / (27 - 9)$$

$$= 27/4 / 18 = 27/72 = 3/8 \approx 0.375$$

8. Find the mean value for the continuous random variable, $f(x) = \sin(x)$, $0 \leq x \leq \pi$

Solution :

$$f(x) = \sin(x), 0 \leq x \leq \pi$$

$$\text{Mean} = \int (x \sin(x)) dx / \int (\sin(x)) dx \text{ from } 0 \text{ to } \pi$$

$$= [-x \cos(x) + \sin(x)]_0^\pi / [-\cos(x)]_0^\pi$$

$$= (\pi + 1) / 2$$

$$\approx 2.07$$

9. Find the mean value for the continuous random variable, $f(x) = e^x$, $0 \leq x \leq 2$

Solution :

$$f(x) = e^x, 0 \leq x \leq 2$$

$$\text{Mean} = \int_0^2 (x e^x) dx / \int_0^2 (e^x) dx \text{ from } 0 \text{ to } 2$$

$$= [x e^x - e^x]_0^2 / [e^x]_0^2$$

$$= (2e^2 - e^2 + 1) / (e^2 - 1)$$

$$\approx 1.54$$

10. Find the mean value for the continuous random variable, $f(x) = \ln(x)$, $1 \leq x \leq e$

Solution :

$$f(x) = \ln(x), 1 \leq x \leq e$$

$$\text{Mean} = \int_1^e (x \ln(x)) dx / \int_1^e (\ln(x)) dx \text{ from } 1 \text{ to } e$$

$$= [(1/4)x^2(2\ln(x)-1)]_1^e / [x \ln(x) - x]_1^e$$

$$= (e^2/2 - 1/4) / (e - 1)$$

$$\approx 1.95$$

Difference Between Discrete Random Variable And Continuous Random Variable

The key differences between discrete and continuous random variables are as follows:

	Discrete Random Variable	Continuous Random Variable
--	---------------------------------	-----------------------------------

Definition	Takes on a finite or countably infinite set of possible values.	Takes on any value within a range or interval i.e., can be uncountably infinite as well.
Probability Distribution	Described by a probability mass function (PMF), which gives the probability of each possible value.	Described by a probability density function (PDF), which gives the probability density at each possible value.
Example	Number of heads in three coin tosses.	Height of a person selected at random.
Probability of a single value	Non-zero probability at each possible value.	Zero probability at each possible value.
Cumulative Distribution Function	Describes the probability of getting a value less than or equal to a particular value.	Describes the probability of getting a value less than or equal to a particular value.

Mean and Variance	Mean and variance can be calculated directly from the PMF.	Mean and variance can be calculated using the PDF and integration.
Probability of an Interval	The probability of an interval is the sum of the probabilities of each value in the interval.	The probability of an interval is the area under the PDF over the interval.

Independence and conditional independence

The conditional probability of A given B is represented by $P(A|B)$. The variables A and B are said to be independent if $P(A) = P(A|B)$ (or alternatively if $P(A,B) = P(A) P(B)$ because of the formula for conditional probability).

Example1 Suppose Norman and Martin each toss separate coins. Let A represent the variable "Norman's toss outcome", and B represent the variable "Martin's toss outcome". Both A and B have two possible values (Heads and Tails). It would be uncontroversial to assume that A and B are independent. Evidence about B will not change our belief in A.

Example2 Now suppose both Martin and Norman toss the same coin. Again let A represent the variable "Norman's toss outcome", and B represent the variable "Martin's toss outcome". Assume also that there is a possibility that the coin is biased towards heads but we do not know this for certain. In this case A and B are not independent. For example, observing that B is Heads causes us to increase our belief in

A being Heads (in other words $P(a|b) > P(b)$ in the case when $a=\text{Heads}$ and $b=\text{Heads}$).

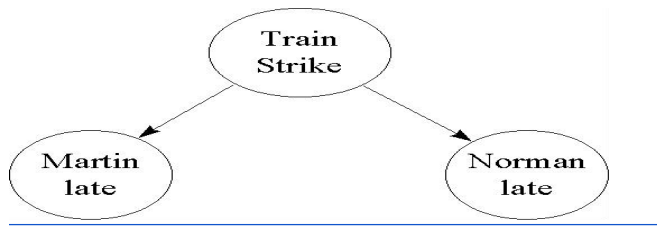
In Example 2 the variables A and B are both dependent on a separate variable C, "the coin is biased towards Heads" (which has the values True or False). Although A and B are not independent, it turns out that once we know for certain the value of C then any evidence about B cannot change our belief about A. Specifically:

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

In such case we say that A and B are conditionally independent given C.

In many real life situations variables which are believed to be independent are actually only independent conditional on some other variable.

Example 3 Suppose that Norman and Martin live on opposite sides of the City and come to work by completely different means, say Norman comes by train while Martin drives. Let A represent the variable "Norman late" (which has values true or false) and similarly let B represent the variable "Martin late". It would be tempting in these circumstances to assume that A and B must be independent. However, even if Norman and Martin lived and worked in different countries there may be factors (such as an international fuel shortage) which could mean that A and B are not independent. In practice any model of uncertainty should take account of all reasonable factors. Thus while, say, a meteorite hitting the Earth might be reasonably excluded it does not seem reasonable to exclude the fact that both A and B may be affected by a Train strike (C). Clearly $P(A)$ will increase if C is true; but $P(B)$ will also increase because of extra traffic on the roads. Thus the situation is represented in the following animation (which is actually a BBN)



Quantiles in Machine Learning

Quantiles offers valuable insights into data distribution and helping in various aspects of analysis. This article describes quantiles, looks at **how to calculate** them, and talks about **how important they are for machine learning applications**. We also discuss the problems with quantiles and how box plots may be used to represent them. For anybody dealing with data in the field of machine learning, having a firm understanding of quantiles is crucial.

What are Quantiles?

Quantiles divide the dataset into equal parts based on rank or percentile. They represent the values at certain points in a dataset sorted in increasing order. General quantiles include the [median](#) (50th percentile), quartiles (25th, 50th, and 75th percentiles), and percentiles (values ranging from 0 to 100).

In machine learning and data science, quantiles play an important role in understanding the data, detecting outliers and evaluating model performance.

Types of Quantiles

- **Quartiles:** [Quartiles](#) divide a dataset into four equal parts, representing the 25th, 50th (median), and 75th percentiles.
- **Quintiles:** Quintiles divide a dataset into five equal parts, each representing 20% of the data.
- **Deciles:** Deciles divide a dataset into ten equal parts, with each decile representing 10% of the data.

- **Percentiles:** Percentiles divide a dataset into 100 equal parts, with each percentile representing 1% of the data.

Steps to Calculate Quantiles

The steps for calculating quantiles involve:

1. **Sorting the Data:** Arrange the dataset in increasing order.
2. **Determine the Position:** Calculate the position of the desired quantile based on the given formula: “**Position=(quantile×(n+1))/100**”, where n is the total number of observations.
3. **Interpolation (if needed):** Interpolate between two adjacent values to find the quantile if the position is not an integer.

Example with Mathematical Imputation:

Let's consider a dataset: [5, 10, 15, 20, 25, 30, 35, 40, 45, 50].

1. **Median (Q2):** There are 10 observations, so the median position is $(2 \times (10+1))/2 = 5.5$. Since, 5.5 is not an integer, we interpolate between the 5th and 6th observations: $\text{Median} = (25+30)/2 = 27.5$.
2. **First Quartile (Q1):** $(25 \times (10+1))/4 = 13.75$. Interpolating between the 13th and 14th observations: $Q1 = (15+20)/2 = 17.5$.
3. **Third Quartile (Q3):** $(75 \times (10+1))/4 = 41.25$. Interpolating between the 41st and 42nd observations: $Q3 = (40+45)/2 = 42.5$.

Uses of Quantiles in Machine Learning

Quantiles play a crucial role in various aspects of machine learning and [data analysis](#). Here are some key uses:

1. **Descriptive Statistics:** Quantiles help summarize the distribution of a dataset, providing insights into its spread and central tendency.
2. **Outlier Detection:** Observations that fall far from certain quantiles may be considered outliers, aiding in anomaly detection.

3. **Probability Distributions:** Quantiles are used to describe the distribution of random variables, facilitating the analysis of probability distributions in machine learning models.
4. **Comparative Analysis:** By comparing quantiles across different datasets, analysts can make informed decisions about the relative standing and characteristics of the datasets.
5. **Risk Assessment:** In finance and other fields, quantiles are used to assess the risk of investments by determining the potential for loss or gain based on the distribution of data.

Understanding these uses is essential for effectively utilizing quantiles in machine learning and data analysis tasks.

Challenges and Limitations of Quantiles

1. **Influence of Outliers:** Quantiles can be sensitive to outliers, especially when calculating quartiles. Outliers can significantly affect the position of quantiles, potentially leading to a misrepresentation of the data's central tendency and spread.
2. **Skewed Distributions:** Quantiles may not fully capture the characteristics of skewed distributions. For highly skewed datasets, the quantiles may not provide a complete picture of the data distribution, especially in the tails.
3. **Variability in Calculations:** Different methods and software packages may use different algorithms for calculating quantiles, leading to variability in results. This can be a challenge when comparing quantiles across different datasets or when using quantiles for decision-making.

Probability Density Estimation & Maximum Likelihood Estimation

Probability Density: Assume a random variable x that has a probability distribution $p(x)$. The relationship between the outcomes of a random variable and its probability is referred to as the probability density.

The problem is that we don't always know the full probability distribution for a random variable. This is because we only use a small subset of observations to derive the outcome. This problem is referred to as **Probability Density Estimation** as we use only a random sample of observations to find the general density of the whole sample space.

Probability Density Function (PDF)

A PDF is a function that tells the probability of the random variable from a sub-sample space falling within a particular range of values and not just one value. It tells the likelihood of the range of values in the random variable sub-space being the same as that of the whole sample.

By definition, if X is any continuous random variable, then the function $f(x)$ is called a [probability density function](#) if:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

```
where,  
a -> lower limit  
b -> upper limit  
X -> continuous random variable  
f(x) -> probability density function
```

Steps Involved:

```
Step 1 - Create a histogram for the random set of observations to understand the
density of the random sample.

Step 2 - Create the probability density function and fit it on the random sample.
Observe how it fits the histogram plot.

Step 3 - Now iterate steps 1 and 2 in the following manner:
3.1 - Calculate the distribution parameters.
3.2 - Calculate the PDF for the random sample distribution.
3.3 - Observe the resulting PDF against the data.
3.4 - Transform the data to until it best fits the distribution.
```

Most of the histogram of the different random sample after fitting should match the histogram plot of the whole population.

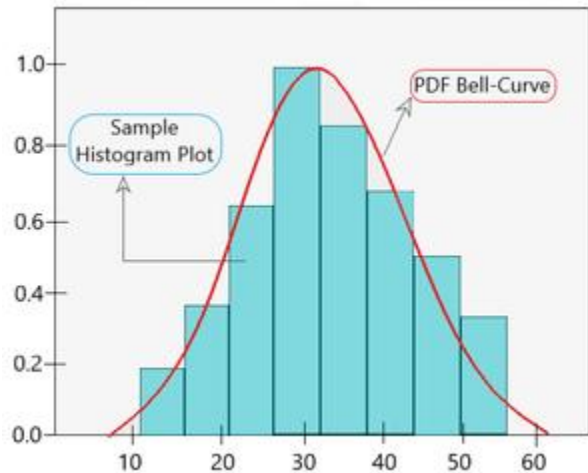
Density Estimation: It is the process of finding out the density of the whole population by examining a random sample of data from that population. One of the best ways to achieve a density estimate is by using a histogram plot.

Parametric Density Estimation

A normal distribution has two given parameters, mean and standard deviation. We calculate the sample mean and standard deviation of the random sample taken from this population to estimate the density of the random sample. The reason it is termed as '*parametric*' is due to the fact that the relation between the observations and its probability can be different based on the values of the two parameters.

Now, it is important to understand that the mean and standard deviation of this random sample is not going to be the same as that of the whole population due to its small size. A sample plot for parametric density estimation is shown below.

below.



PDF fitted over histogram plot with one peak value

Nonparametric Density Estimation

In some cases, the PDF may not fit the random sample as it doesn't follow a normal distribution (i.e. instead of one peak there are multiple peaks in the graph). Here, instead of using distribution parameters like mean and standard deviation, a particular algorithm is used to estimate the probability distribution. Thus, it is known as a '*nonparametric density estimation*'.

One of the most common nonparametric approach is known as **Kernel Density Estimation**. In this, the objective is to calculate the unknown density $f_h(x)$ using the equation given below:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)$$

where,

K -> kernel (non-negative function)

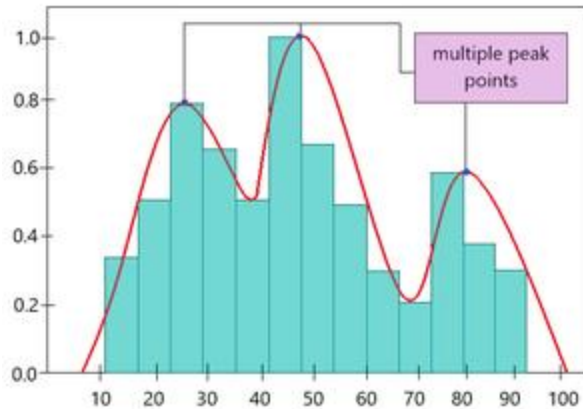
h -> bandwidth (smoothing parameter, $h > 0$)

K_h -> scaled kernel

$f_{h(x)}$ -> density (to calculate)

n -> no. of samples in random sample.

A sample plot for nonparametric density estimation is given below.



PDF plot over sample histogram plot based on KDE

Problems with Probability Distribution Estimation

Probability Distribution Estimation relies on finding the best PDF and determining its parameters accurately. But the random data sample that we consider, is very small. Hence, it becomes very difficult to determine what parameters and what probability distribution function to use. To tackle this problem, Maximum Likelihood Estimation is used.

Maximum Likelihood Estimation

It is a method of determining the parameters (mean, standard deviation, etc) of normally distributed random sample data or a method of finding the best fitting PDF over the random sample data. This is done by maximizing the likelihood function so that the PDF fitted over the random sample. Another way to look at it is that MLE function gives the mean, the standard deviation of the random sample is most similar to that of the whole sample.

NOTE: MLE assumes that all PDFs are a likely candidate to being the best fitting curve. Hence, it is computationally expensive method.

Intuition:

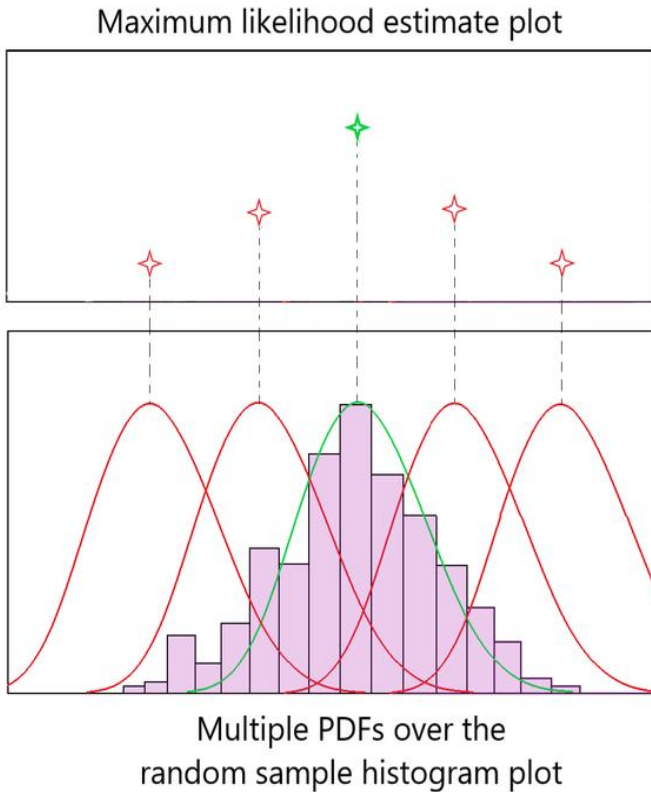


Fig 1 : MLE Intuition

Fig 1 shows multiple attempts at fitting the PDF bell curve over the random sample data. Red bell curves indicate poorly fitted PDF and the green bell curve shows the best fitting PDF over the data. We obtained the optimum bell curve by checking the values in Maximum Likelihood Estimate plot corresponding to each PDF.

As observed in Fig 1, the red plots poorly fit the normal distribution, hence their '*likelihood estimate*' is also lower. The green PDF curve has the maximum likelihood estimate as it fits the data perfectly. This is how the maximum likelihood estimate method works.

Mathematics Involved

In the intuition, we discussed the role that Likelihood value plays in determining the optimum PDF curve. Let us understand the math involved in MLE method.

We calculate Likelihood based on conditional probabilities. See the

$$L = F([X_1 = x_1], [X_2 = x_2], \dots, [X_n = x_n] | P) = \prod_{i=1}^n P^{x_i} (1 - P)^{1-x_i}$$

```
where,  
L -> Likelihood value  
F -> Probability distribution function  
P -> Probability  
X1, x2, ... xn -> random sample of size n taken from the whole population.  
x1, x2, ... xn -> values that these random sample (Xi) takes when determining the PDF.  
Π -> product from 1 to n.
```

In the above-given equation, we are trying to determine the likelihood value by calculating the joint probability of each X_i taking a specific value x_i involved in a particular PDF. Now, since we are looking for the maximum likelihood value, we differentiate the likelihood function w.r.t P and set it to 0 as given below.

$$\frac{\partial L}{\partial P} = 0$$

This way, we can obtain the PDF curve that has the maximum likelihood of fit over the random sample data.

But, if you observe carefully, differentiating L w.r.t P is not an easy task as all the probabilities in the likelihood function is a product. Hence, the calculation becomes computationally expensive. To solve this, we take the log of the Likelihood function L .

Log Likelihood

$$\log(L) = \log(\prod_{i=1}^n P^{x_i} (1 - P)^{1-x_i})$$

Taking the log of likelihood function gives the same result as before due to the increasing nature of Log function. But now, it becomes less computational due to the property of logarithm:

$$\log(a) + \log(b)$$

Thus, the equation becomes:

$$\begin{aligned}\log(L) &= \log[\prod_{i=1}^n P^{x_i} (1 - P)^{1-x_i}] \\ &= \sum_{i=1}^n \log[P^{x_i} (1 - P)^{1-x_i}]\end{aligned}$$

Now, we can easily differentiate log L wrt P and obtain the desired result.

Covariance and Correlation

Covariance and correlation are the two **key concepts in Statistics** that help us **analyze the relationship between two variables**. Covariance measures how two variables change together, **indicating whether they move in the same or opposite directions**.

However, its magnitude can be difficult to interpret because it's not standardized. Correlation, refines this measure by normalizing covariance, Correlation explains the proportion in which the second variable change. **Correlation varies between -1 to +1**. If the correlation value is 0 then it means there is no Linear Relationship between variables however other functional relationship may exist. This allows for a clearer understanding of both the strength and direction of the relationship between variables.

What is Covariance?

Covariance is a statistical measure that indicates the direction of the linear relationship between two variables. It assesses how much two variables change together from their mean values.

Types of Covariance:

- **Positive Covariance:** When one variable increases, the other variable tends to increase as well, and vice versa.
- **Negative Covariance:** When one variable increases, the other variable tends to decrease.
- **Zero Covariance:** There is no linear relationship between the two variables; they move independently of each other.

Covariance is calculated by taking the average of the product of the deviations of each variable from their respective means. It is useful for understanding the direction of the relationship but not its strength, as its magnitude depends on the units of the variables.

It is an essential tool for understanding how variables change together and are widely used in various fields, including finance, economics, and science.

Covariance:

1. It is the relationship between a pair of random variables where change in one variable causes change in another variable.
2. It can take any value between $-\infty$ to $+\infty$, where the negative value represents the negative relationship whereas a positive value represents the positive relationship.
3. It is used for the linear relationship between variables.
4. It gives the direction of relationship between variables.

Covariance Formula

For Population:

$$\text{Covari}(x, y) = \frac{\sum_{i=1}^n (x_i - x') (y_i - y')}{n}$$

For Sample:

$$\text{Covari}(x, y) = \frac{\sum_{i=1}^n (x_i - x') (y_i - y')}{n - 1}$$

Here, x' and y' = mean of given sample set n = total no of sample x_i and y_i = individual sample of set

Example –



What is Correlation?

Correlation is a standardized measure of the strength and direction of the linear relationship between two variables. It is derived from covariance and ranges between -1 and 1. Unlike covariance, which only indicates the direction of the relationship, correlation provides a standardized measure.

- **Positive Correlation (close to +1):** As one variable increases, the other variable also tends to increase.
- **Negative Correlation (close to -1):** As one variable increases, the other variable tends to decrease.

- Zero Correlation: There is no linear relationship between the variables.

The [correlation coefficient](#) ρ (rho) for variables X and Y is defined as:

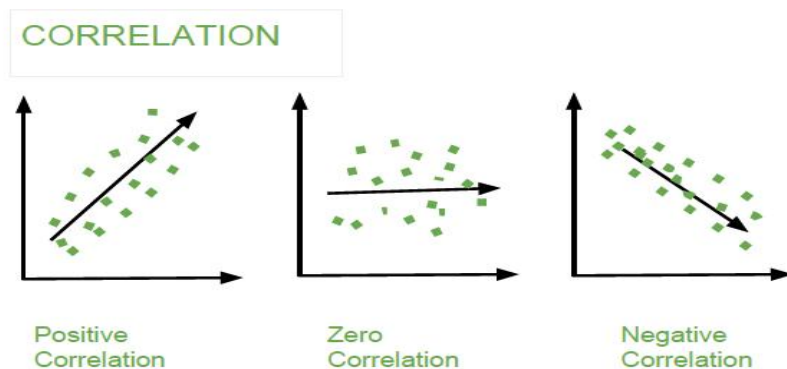
1. It show whether and how strongly pairs of variables are related to each other.
2. Correlation takes values between -1 to +1, wherein values close to +1 represents strong positive correlation and values close to -1 represents strong negative correlation.
3. In this variable are indirectly related to each other.
4. It gives the direction and strength of relationship between variables.

Correlation Formula

$$\text{Corr}(x, y) = \frac{\sum_{i=1}^n (x_i - x') (y_i - y')}{\sqrt{\sum_{i=1}^n (x_i - x')^2 \sum_{i=1}^n (y_i - y')^2}}$$

Here, x' and y' = mean of given sample set n = total no of sample x_i and y_i = individual sample of set

Example –



Difference between Covariance and Correlation

This table shows the [difference between Covariance and Covariance](#):

Covariance

Covariance is a measure of how much two random variables vary together

Involves the relationship between two variables or data sets

Lie between $-\infty$ and $+\infty$

Measure of correlation

Provides direction of relationship

Dependent on scale of variable

Have dimensions

Correlation

Correlation is a statistical measure that indicates how strongly two variables are related.

Involves the relationship between multiple variables as well

Lie between -1 and +1

Scaled version of covariance

Provides direction and strength of relationship

Independent on scale of variable

Dimensionless

Applications of Covariance and Correlation

Applications of Covariance

- **Portfolio Management in Finance:** Covariance is used to measure how different stocks or financial assets move together, aiding in portfolio diversification to minimize risk.
- **Genetics:** In genetics, covariance can help understand the relationship between different genetic traits and how they vary together.

- **Econometrics:** Covariance is employed to study the relationship between different economic indicators, such as the relationship between GDP growth and inflation rates.
- **Signal Processing:** Covariance is used to analyze and filter signals in various forms, including audio and image signals.
- **Environmental Science:** Covariance is applied to study relationships between environmental variables, such as temperature and humidity changes over time.

Applications of Correlation

- **Market Research:** Correlation is used to identify relationships between consumer behavior and sales trends, helping businesses make informed marketing decisions.
- **Medical Research:** Correlation helps in understanding the relationship between different health indicators, such as the correlation between blood pressure and cholesterol levels.
- **Weather Forecasting:** Correlation is used to analyze the relationship between various meteorological variables, such as temperature and humidity, to improve weather predictions.
- **Machine Learning:** Correlation analysis is used in feature selection to identify which variables have strong relationships with the target variable, improving model accuracy.