✅ 1. Discrete and Continuous Random Variables

A **random variable** is a numerical value assigned to the outcome of a random experiment. Random variables are mainly classified into **discrete** and **continuous** types.

A **discrete random variable** takes a **finite or countable number of distinct values**. These variables often come from counting outcomes, such as the number of heads when tossing a coin three times, or the number of students present in a classroom. Each value has a certain probability, given by the **Probability Mass Function (PMF)**.
For example:

If X is the number of heads in 2 tosses, then $P(X=0)=\frac{1}{4}, P(X=1)=\frac{1}{2}, P(X=2)=\frac{1}{4}$

A **continuous random variable** takes **uncountably infinite values**, usually over a continuous range. These variables come from measurements like height, weight, or time. They are described using the **Probability Density Function (PDF)**.

The probability for a continuous variable over an interval is given by:

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

This means we don't find the probability of exact values, but rather the chance of values falling within a range.

The main difference is that **discrete variables are used when outcomes are counted**, and **continuous variables are used when outcomes are measured**.

✅ 2. Independence

In probability, two events are called **independent** when the occurrence of one **does not affect** the occurrence of the other. It means both events are unrelated.

For two events A and B, they are independent if:

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

This means the chance of both happening together is the product of their individual probabilities.

A simple example is tossing two coins. The outcome of the first toss (e.g., heads or tails) does not change the outcome of the second toss. Therefore, these events are independent.

In terms of random variables, X and Y are independent if the value of one does not give any information about the value of the other. This idea is very useful in machine learning. For example, in **Naive Bayes classification**, we assume features are independent to simplify computations.

Independence helps simplify complex probability models and is a key concept in both theory and practical applications.

✅ 3. Covariance

Covariance is a measure that shows how two random variables **vary together**. It tells us the **direction** of their relationship.

- If the values of both variables increase together, the covariance is **positive**.
- If one increases while the other decreases, the covariance is **negative**.
- If there is no relationship, the covariance is close to **zero**.

The formula for covariance is:

The formula for covariance is:

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Or, an easier form is:

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

Covariance is useful in determining whether variables move together. However, it does not indicate the **strength** of the relationship — just the direction. For strength and scale-free comparison, we use **correlation**.

Covariance is important in fields like finance (portfolio risk) and data science (feature relationships).

✅ 4. Central Limit Theorem (CLT)

The **Central Limit Theorem** is a key principle in statistics. It states that if we take a large number of samples from any population, the **distribution of the sample means** will be **approximately normal**, even if the population itself is not normally distributed.

Mathematically, if we take n independent samples $X_1, X_2, \ldots, X_n$ from a population with mean μ and standard deviation σ, the sample mean $\bar{X}$ is:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

As $n \to \infty$, the distribution of $\bar{X}$ becomes:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

This theorem is very useful in real-world applications like polling, quality control, and experiment analysis, where we work with averages. It also justifies the use of the **normal distribution** in hypothesis testing and confidence intervals.

✅ 5. Chebyshev's Inequality

Chebyshev's Inequality is a general rule in probability that applies to **any distribution**, regardless of shape. It gives a **guaranteed minimum proportion** of values that lie within a certain distance (in terms of standard deviations) from the mean.

The inequality is written as:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Which means, at least $1 - \frac{1}{k^2}$ of the data lies within k standard deviations from the mean.

For example:

- When $k = 2$, at least 75% of values lie within 2 standard deviations.
- When $k = 3$, at least 88.89% of values lie within 3 standard deviations.

This rule is very useful when we do not know the exact shape of the distribution. It is commonly used in **risk management**, **outlier detection**, and **data validation**, as it helps understand how spread out the data is, even without knowing its full details.

✅ 6. Diverse Continuous and Discrete Distributions

In probability, there are many **standard distributions** used to model various real-world situations. These distributions are divided into two types: **discrete** and **continuous**.

🔹 **Common Discrete Distributions:**

- **Bernoulli Distribution**: Used for a single trial with two outcomes — success (1) or failure (0).
  Example: Tossing a coin once.
- **Binomial Distribution**: Used for a fixed number of independent trials, each with the same probability of success.
  Example: Number of correct answers in a multiple-choice test.
- **Poisson Distribution**: Used to model the number of events occurring in a fixed interval of time or space.
  Example: Number of customer arrivals per hour.

🔹 **Common Continuous Distributions:**

- **Uniform Distribution**: All values in a range are equally likely.
  Example: Picking a random number between 0 and 1.
- **Normal Distribution**: Bell-shaped curve, most values close to the mean.
  Example: Heights or weights of people.
- **Exponential Distribution**: Time between events in a random process.
  Example: Time between two incoming calls.

These distributions are important because they help in solving real-life problems using mathematical models. Choosing the correct distribution allows for better predictions and analysis in statistics, data science, and artificial intelligence.

✅ 1. Descriptive Statistics

Descriptive statistics are used to **summarize and describe** the main features of a dataset. They help us understand the general pattern, central values, and variability of the data.

🔹 **Key Measures:**

- **Measures of Central Tendency:**
  - **Mean:** Average value → $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
  - **Median:** Middle value when data is sorted.
  - **Mode:** Most frequently occurring value.
- **Measures of Dispersion:**
  - **Range:** Difference between the maximum and minimum values.
  - **Variance:** Spread of data → $\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$
  - **Standard Deviation:** Square root of variance.
  - **Interquartile Range (IQR):** Range between the 1st and 3rd quartiles.

🔹 **Importance:**

- Descriptive statistics give a **quick overview** of the data.
- They are the **first step** before applying advanced analysis or building models.
- Useful in identifying trends, errors, or outliers in the dataset.