TOPIC : 2.11 – Central Limit Theorem

The **Central Limit Theorem** states that the distribution of the sum (or average) of a large number of independent, identically distributed (i.i.d.) random variables will approach a normal (Gaussian) distribution, no matter the shape of the original distribution, as long as the variables have finite mean and variance.

In simpler terms, if you take a sufficiently large sample from any population (with any distribution), the sampling distribution of the sample mean will be approximately normally distributed, regardless of the original population's distribution.

Key Points of CLT:

1. **Independence:** The random variables in the sample must be independent of each other.
2. **Identically Distributed**: The random variables should have the same probability distribution.
3. **Sample Size**: As the sample size increases (typically n ≥ 30 is considered large enough), the distribution of the sample mean approaches a normal distribution.
4. **Mean and Variance**: The mean of the sample mean distribution will be equal to the population mean, and the variance of the sample mean distribution will be the population variance divided by the sample size.

Mathematical Statement:

Let $X_1, X_2, X_3, ..., X_n$ be a sample of independent random variables with the same distribution. If μ is the population mean and σ^2 is the population variance, then as n becomes large:

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

where $\bar{X}$ is the sample mean, and $\mathcal{N}(0, 1)$ denotes the standard normal distribution.

Applications of CLT:

- **Estimating Population Parameters**: In hypothesis testing, the CLT allows us to use normal distribution approximations for large sample sizes even when the underlying population is not normal.
- **Quality Control**: CLT is applied in process monitoring and defect detection in manufacturing processes.
- **Finance**: In stock returns, the CLT can be used to estimate the distribution of the sample mean of returns over time.

Problems Related to CLT:

1. Approximation of Sample Mean:

Suppose a population has a mean (μ) of 50 and a standard deviation (σ) of 10. A random sample of size $n = 40$ is drawn. What is the probability that the sample mean will be between 48 and 52?

Solution Outline:

- Use the CLT to approximate the sample mean distribution: $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.
- $\mu_{\bar{X}} = 50$ and $\sigma_{\bar{X}} = \frac{10}{\sqrt{40}}$.
- Standardize the values to use the normal distribution table or a calculator.

2. Non-Normal Population Distribution:

Imagine a population of incomes that follows an exponential distribution with a mean of 1000 and a standard deviation of 500. A random sample of 100 individuals is selected. What is the approximate distribution of the sample mean?

Solution Outline:

- Apply the Central Limit Theorem: For large sample sizes, the sample mean will follow a normal distribution with mean 1000 and standard deviation $\frac{500}{\sqrt{100}} = 50$.

3. Estimation in Real-World Scenarios:

In a factory, the amount of product in bottles is normally distributed with a mean of 300 ml and a standard deviation of 5 ml. If 40 bottles are sampled, what is the probability that the average amount of product in the bottles is between 298 and 302 ml?

Solution Outline:

- Use the CLT to find the distribution of the sample mean.
- Find the probability using the normal distribution formula.

Summary of CLT Problems:

- The CLT helps approximate the distribution of sample means, even for non-normal populations, given a large enough sample size.
- The sample size needed depends on the original distribution's skewness and kurtosis, but for most distributions, $n \geq 30$ is typically considered sufficient.
- In practical scenarios, CLT is widely used for making inferences about population parameters based on sample data.