# Central Limit Theorem

Shashank Pyati

July 11, 2023

# Introduction

- The Central Limit Theorem (CLT) is a fundamental concept in probability theory and statistics.
- It states that the sum (or average) of a large number of independent and identically distributed random variables tends to follow a normal distribution.

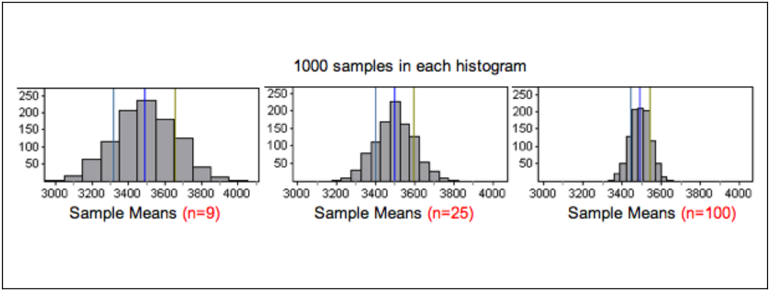
# Statement of the Central Limit Theorem

**Central Limit Theorem**

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with a common mean $\mu$ and standard deviation $\sigma$. As $n$ approaches infinity, the distribution of the sample mean $\bar{X}$ approaches a normal distribution with mean $\mu$ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

# Visualizing the Central Limit Theorem

- To understand the Central Limit Theorem visually, consider the following example:

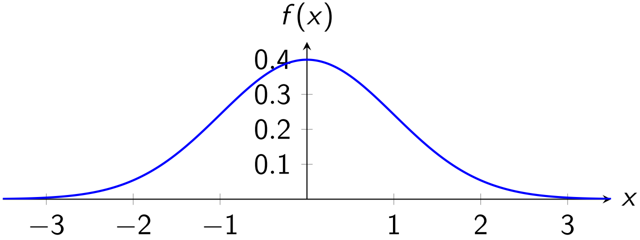



(a) Hypothesis Testing

- Each histogram represents the distribution of sample means obtained by repeatedly sampling from a population.

# Normal Distribution

- The normal distribution, also known as the Gaussian distribution, is a continuous probability distribution that is symmetric and bell-shaped.




- The graph shows a smooth curve that is symmetric around the mean. The mean and standard deviation determine the exact shape and location of the distribution.
- Many natural phenomena and statistical processes follow a normal distribution.

# Effect of Sample Size

- The Central Limit Theorem highlights the impact of sample size on the shape of the distribution of sample means.

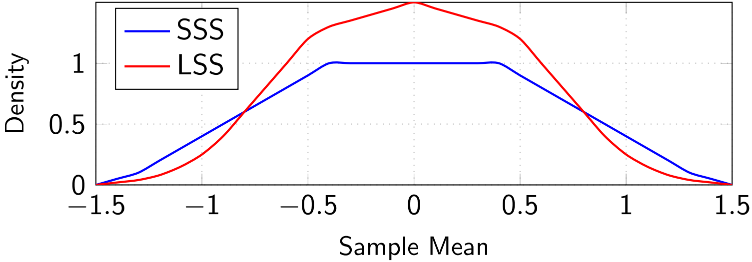



Figure: Comparison of Distribution of sample means for small and large sample sizes.

*LSS-Large Sample Size, SSS-Small Sample Size

- The distribution becomes more bell-shaped and closer to a normal distribution as the sample size increases.

# Proof of the Central Limit Theorem

- To prove the Central Limit Theorem, we can use the characteristic function approach.
- The characteristic function of the sample mean $\bar{X}$ is given by:

$$\phi_{\bar{X}}(t) = \left[\phi\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

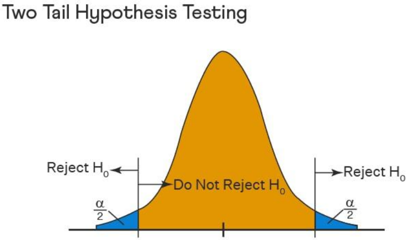
  where $\phi(t)$ is the characteristic function of the individual random variable $X$.
- Using Taylor series expansion, we can approximate the characteristic function of $\bar{X}$ as:

$$\phi_{\bar{X}}(t) \approx 1 + it\mu - \frac{t^2\sigma^2}{2n}$$

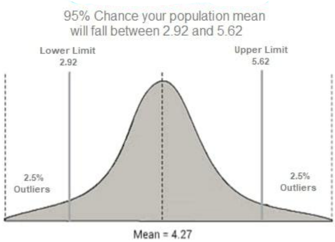
- This approximation tends to a normal distribution as $n$ approaches infinity, proving the Central Limit Theorem.

# Applications of the Central Limit Theorem

- The Central Limit Theorem has extensive applications in various fields:




(a) Hypothesis Testing




(b) Confidence Intervals

- It provides the foundation for statistical inference and allows us to make inferences about population parameters based on sample statistics.

# Z-test

- The Z-test is a statistical test used to compare sample means with a known population mean when the population standard deviation is also known.
- It is based on the standard normal distribution.
- The test statistic for the Z-test is calculated as:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

  where $\bar{X}$ is the sample mean, $\mu$ is the population mean, $\sigma$ is the population standard deviation, and $n$ is the sample size.
- If the absolute value of the test statistic is greater than a critical value (usually based on a significance level), we reject or fail to reject the null hypothesis.

# Sampling Distribution of the Sample Mean

- The Central Limit Theorem focuses on the sampling distribution of the sample mean.
- As the sample size increases, the sampling distribution becomes more concentrated around the population mean.
- The sampling distribution of the sample mean is important because it allows us to make inferences about the population mean based on the sample mean. It helps us understand the variability of sample means and provides a basis for estimating the population mean and constructing confidence intervals. The smaller the standard error, the more precise our estimate of the population mean is likely to be.

# Assumptions of the Central Limit Theorem

- The Central Limit Theorem relies on a few assumptions:
- The random variables should be independent and identically distributed for the theorem to hold.
- The theorem assumes a finite population variance or, in the case of infinite variance, a slower rate of convergence.

# Limitations of the Central Limit Theorem

- The Central Limit Theorem has some limitations:
- It requires the assumption of independence and identical distribution of random variables.
- The theorem holds only asymptotically as the sample size approaches infinity.
- It may not apply to distributions with heavy tails or infinite variance.
- The convergence rate to the normal distribution depends on the underlying distribution.

# Conclusion

- The Central Limit Theorem is a fundamental concept in statistics.
- It provides insights into the behavior of sample means and allows us to make statistical inferences.
- Understanding the Central Limit Theorem is crucial in various statistical applications.