Lesson: Central Limit Theorem and Confidence Intervals

Section 1: Introduction to Central Limit Theorem (CLT)

Subsection 1.1: What is the Central Limit Theorem?

The Central Limit Theorem (CLT) is a fundamental result in probability theory and statistics. It states that under certain conditions, the sum of a large number of independent and identically distributed (i.i.d.) random variables (RVs) will converge to a Gaussian (normal) distribution, regardless of the original distribution of the individual RVs. This is a remarkable property of the CLT as it allows us to make inferences about the sum of RVs even when the individual RVs may not be normally distributed.

Subsection 1.2: Statement of the Central Limit Theorem

Let X1, X2, ..., Xn be a sequence of n i.i.d. random variables with mean μ and variance σ^2. Define the sample sum S\_n as:

\[ S\_n = X\_1 + X\_2 + \ldots + X\_n \]

Then, as n approaches infinity, the distribution of S\_n becomes approximately Gaussian:

\[ \lim\_{n \to \infty} P\left(\frac{S\_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int\_{-\infty}^{x} e^{-t^2/2} dt \]

Subsection 1.3: Implications of the Central Limit Theorem

The CLT has significant implications for many practical applications:

1. Approximating Non-Normal Distributions: It allows us to approximate the distribution of the sample sum S\_n with a Gaussian distribution, even if the individual RVs have non-normal distributions. This simplifies calculations and enables us to apply standard Gaussian-based statistical methods.

2. Estimation of Probabilities: The CLT enables us to estimate probabilities involving the sum of a large number of RVs by using Gaussian tables or software, making complex calculations more tractable.

3. Sample Mean Convergence: As a consequence of the CLT, the sample mean \(\bar{X} = \frac{S\_n}{n}\) also converges to a Gaussian distribution with mean μ and variance \(\frac{\sigma^2}{n}\) as n becomes large.

Section 2: Confidence Intervals

Subsection 2.1: Introduction to Confidence Intervals

Confidence intervals provide a powerful tool for estimating population parameters from sample data. Instead of varying the sample size n, confidence intervals focus on estimating the range of values that likely contains the true population parameter (e.g., mean or proportion) with a specified level of confidence.

Subsection 2.2: Constructing Confidence Intervals

The construction of a confidence interval involves three main components:

1. Point Estimation: We use the sample data to calculate a point estimate of the population parameter of interest. For example, the sample mean \(\bar{X}\) is a point estimate for the population mean μ.

2. Margin of Error: The margin of error is a measure of uncertainty around the point estimate. It depends on the level of confidence desired and the variability in the data.

3. Interval Construction: By combining the point estimate with the margin of error, we construct a range of values known as the confidence interval. It provides an interval estimate for the population parameter with a specified level of confidence.

Subsection 2.3: Confidence Interval for the Mean (Known Variance)

For the case where the population variance σ^2 is known, the confidence interval for the population mean μ at a confidence level (1 - α) is given by:

\[ \bar{X} \pm Z\_{\alpha/2} \frac{\sigma}{\sqrt{n}} \]

where Z\_{\alpha/2} is the critical value from the standard normal distribution corresponding to the desired confidence level (1 - α).

Subsection 2.4: Confidence Interval for the Mean (Unknown Variance)

In practice, the population variance σ^2 is often unknown. In this case, we use the sample variance \(S^2\) as an estimate for σ^2. The confidence interval for the population mean μ at a confidence level (1 - α) is then given by:

\[ \bar{X} \pm t\_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \]

where t\_{\alpha/2, n-1} is the critical value from the t-distribution with (n-1) degrees of freedom corresponding to the desired confidence level (1 - α).

Subsection 2.5: Interpreting Confidence Intervals

A confidence interval of (1 - α) level provides a range of values where we can be confident that the true population parameter lies with a probability of (1 - α). For example, a 95% confidence interval means that in repeated sampling, we expect 95% of such intervals to capture the true population parameter.

Conclusion:

The Central Limit Theorem (CLT) is a powerful tool that allows us to approximate the distribution of the sum of a large number of i.i.d. random variables to a Gaussian distribution. It has wide-ranging applications in statistics and data analysis. Confidence intervals provide a practical approach to estimate population parameters and account for uncertainty. Whether the variance is known or unknown, confidence intervals are valuable tools to make reliable inferences about population parameters based on sample data. Understanding the CLT and confidence intervals is essential for conducting rigorous statistical analysis in engineering and other disciplines.