Course Title

The Central Limit Theorem: Theory, Simulation, and Applications

Course Description

This course develops a deep understanding of the Central Limit Theorem from intuition to proof to real-world application. Students learn why normal distributions appear everywhere, how sampling distributions behave, and how the CLT justifies statistical inference and machine learning assumptions.

Learning Objectives

By the end of the course students will be able to:

- Explain why sample means become normally distributed
- Derive the mean and variance of sampling distributions
- Apply the CLT to approximate probabilities
- Use the CLT to justify confidence intervals and hypothesis tests
- Simulate convergence to normality in code
- Understand when the CLT fails or converges slowly

Module Breakdown

Module 1: Motivation and Intuition (Week 1)

Topics

- Why normal distributions appear everywhere
- Law of Large Numbers vs Central Limit Theorem
- Sampling distributions
- Visual intuition using simulations

Key Ideas

- Averages stabilize
- Variability shrinks as n increases
- Shape becomes normal even if the population is not

Lab

- Simulate rolling dice and plot distribution of sample means
- Compare small n vs large n

Module 2: Formal Statement of the CLT (Week 2)

Topics

- Independent and identically distributed random variables
- Standardization
- The formal CLT statement
- Meaning of convergence in distribution

Core Formula

If
$X_1, X_2, \ldots, X_n$ are i.i.d. with mean μ and variance σ^2,

Then

$\sqrt{n} (\bar{X} - \mu) / \sigma \to \text{Normal}(0,1)$

Discussion

- Why standardization is required
- What changes and what stays fixed

Module 3: Sampling Distributions (Week 3)

Topics

- Mean of $\bar{X}$
- Variance of $\bar{X}$
- Standard error
- Relationship between population and sampling distribution

Key Results

$E[\bar{X}] = \mu$
$Var(\bar{X}) = \sigma^2 / n$

Applications

- Margin of error
- Estimating population parameters

Module 4: Applications in Statistics (Week 4)

Topics

- Confidence intervals
- Hypothesis testing
- Z-scores
- Approximation of binomial with normal

Real Examples

- Polling
- A/B testing
- Quality control
- Machine learning performance estimates

Module 5: Rate of Convergence and Limitations (Week 5)

Topics

- When $n \geq 30$ rule works and when it does not
- Heavy-tailed distributions
- Skewed distributions
- Berry-Esseen bound (conceptually)

Key Questions

- How fast does convergence happen?
- When should you not trust CLT?

Module 6: Computational Perspective (Week 6)

Topics

- Monte Carlo simulations
- Empirical verification of CLT
- Visualizing convergence
- Bootstrap connection

Lab

- Simulate exponential distribution means
- Compare histograms as n increases
- Implement bootstrap and compare to CLT approximation

Assessment Structure

- Weekly problem sets (theory + simulation)
- One simulation project
- Final applied project

Capstone Project Options

Students choose one:

- Empirically test CLT on different distributions
- Compare CLT vs bootstrap confidence intervals
- Analyze convergence speed under skewed data
- Apply CLT to large dataset and evaluate approximation error

---

Advanced Section (Optional)

- Multivariate CLT
- Lindeberg condition
- CLT for dependent variables
- Connection to stochastic processes

What Students Should Walk Away Understanding

- The CLT explains why normal distributions dominate statistics
- It justifies almost all classical inference
- It connects probability theory to real-world data
- It is powerful but not magic
- Sample size and distribution shape matter