Alexandria University
Faculty of Engineering

CS 112: Statistics
Spring 2025
Lab 1
Assigned: 28/2/2025
Due: 6/3/2025

Understanding Random Variables and Statistical Distributions in Python

Objective

This lab introduces students to key statistical concepts using Python, including the Law of Large Numbers, the Central Limit Theorem, and variance estimation. Through simulations and visualizations, students will explore how sample statistics behave and converge to theoretical expectations.

Requirements

- Write and run an interactive Python notebook (ipynb)
- You can use Colab or your local machine
- Write a clean and commented code and do the explanations in markdown cells.
- Use matplotlib, numpy and stats libraries

Submission

- You have to submit the ipynb
- Work on this lab individually

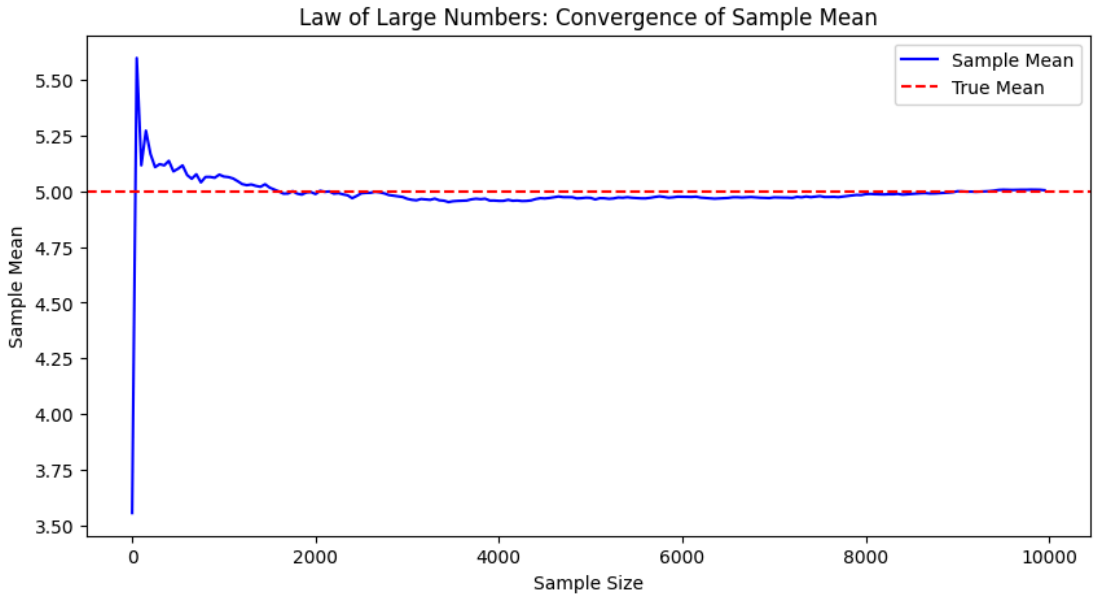
Task 1: Simulating the Law of Large Numbers

Task

- Generate a large number of random samples from a normal distribution **N(0,1)**
- Compute the sample mean as the sample size increases (e.g., from 10 to 10,000).
- Plot the sample mean vs. sample size to illustrate how it converges to the true mean.

Outcome

- A plot showing the sample mean stabilizing around the true mean as the sample size increases.



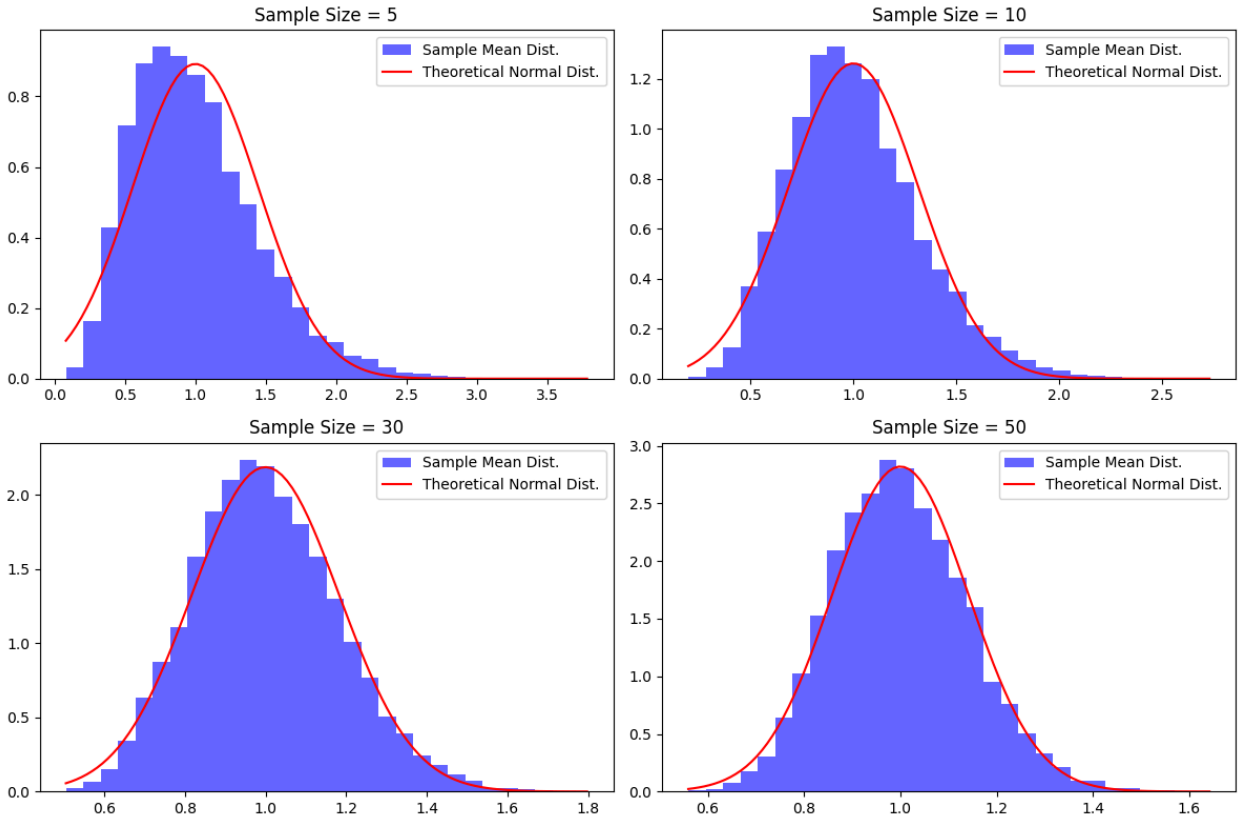
Task 2: Central Limit Theorem in Action

Task

- Generate 10,000 samples of size **n** from an **exponential distribution** (which is not normal).
- Compute the mean of each sample and plot the histogram of these means.
- Show that as **n** increases (e.g., from 5 to 50), the sample means form a normal distribution.

Outcome

- The histogram of sample means should resemble a normal distribution as **n** increases.



Task 3: Simulating Variance of Small Samples

Task

- Generate 10,000 samples of small size **n** (eg, 5, 10, 20, 50) from a **normal distribution**.
- Compute sample variance using both biased (dividing by n) and unbiased (dividing by n−1) estimators.
- Compare the average of sample variances to the true variance.

Outcome

- Unbiased estimators should better approximate the population variance.