IE221 – Probability: TW4

Team 6: Şevval Bozali & Beyza Bozkurt

1. Introduction

The purpose of this project is to investigate two fundamental results in probability theory: the **Strong Law of Large Numbers (SLLN)** and the **Central Limit Theorem (CLT)**, using Monte Carlo simulation techniques. These theorems form the theoretical foundation of statistical inference and play a crucial role in engineering, data analysis, and decision-making under uncertainty.

In this study, random variables are generated computationally and analyzed through simulations to observe how empirical results converge to theoretical expectations. The project aims to bridge the gap between mathematical theory and practical experimentation by visualizing convergence behavior through graphs and statistical plots.

Understanding the differences between SLLN and CLT is especially important for industrial engineering applications such as quality control, risk analysis, simulation modeling, and performance evaluation. By using simulation-based approaches, the project demonstrates how abstract probabilistic concepts manifest in real computational experiments.

2. Theoretical Background

2.1 Strong Law of Large Numbers (SLLN)

The Strong Law of Large Numbers states that the sample mean of a sequence of independent and identically distributed (i.i.d.) random variables converges almost surely to the expected value of the random variable, provided that the expected value exists.

Mathematically, let (X1, X2, ..., Xn) be i.i.d. random variables with expected value (E[Xi] =mu). Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu$$

Almost sure convergence means that the probability that the sample mean converges to (\mu) is equal to 1. In practical terms, this implies that for a single sufficiently long sequence of observations, the average will eventually remain arbitrarily close to the expected value.

2.2 Central Limit Theorem (CLT)

The Central Limit Theorem describes the behavior of the distribution of the normalized sum of i.i.d. random variables. Unlike SLLN, CLT does not focus on a single sample path but on the shape of the distribution formed by many repetitions of the experiment.

Let (X1, X2, ..., X_n) be i.i.d. random variables with mean (\mu) and variance (sigma^2 < infinity). Then, as (n) increases,

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

where d denotes convergence in distribution. This theorem explains why normal distributions appear frequently in practice, even when the underlying data is not normally distributed.

2.3 Monte Carlo Simulation Method

Monte Carlo simulation is a numerical technique that relies on repeated random sampling to approximate mathematical results. It is widely used when analytical solutions are difficult or impossible to obtain. In probability and statistics, Monte Carlo methods are especially useful for estimating expected values, probabilities, and convergence behavior.

In this project, Monte Carlo simulations are used to illustrate the convergence properties of SLLN and CLT, as well as to estimate the value of (pi) through random sampling.

3. Modes of Convergence

Understanding the difference between the Strong Law of Large Numbers (SLLN) and the Central Limit Theorem (CLT) requires a clear explanation of the modes of convergence used by these theorems. Although both results describe convergence behavior, they rely on fundamentally different mathematical concepts.

3.1 Almost Sure Convergence (SLLN)

The SLLN is based on **almost sure convergence**, which is one of the strongest forms of convergence in probability theory. Almost sure convergence means that the sequence of random variables converges to a constant value for almost every possible outcome, except for a set of outcomes with probability zero.

Formally, a sequence X_n converges almost surely to X if:

P(lim_{n→∞} X_n = X) = 1.

In the context of this project, almost sure convergence implies that if we generate a single sufficiently long sequence of random variables and compute the sample mean, this mean will eventually stabilize and remain arbitrarily close to the true expected value μ. Once convergence is achieved, deviations from μ become increasingly unlikely as the sample size grows.

From an experimental perspective, this explains why **a single long simulation run** is sufficient to demonstrate the SLLN. A line plot of the cumulative sample mean clearly illustrates how the mean converges to the expected value over time.

3.2 Convergence in Distribution (CLT)

In contrast, the CLT relies on **convergence in distribution**, which is a weaker form of convergence. Convergence in distribution focuses on the shape of the probability distribution of a sequence of random variables rather than on individual sample paths.

A sequence X_n converges in distribution to X if:

$$\lim_{n\to\infty} P(X_n \leq t) = P(X \leq t) \text{ for all continuity points } t.$$

For the CLT, this means that the distribution of the standardized sample mean approaches the standard normal distribution as the sample size increases. Importantly, this convergence does not imply that individual realizations converge to a fixed value. Instead, it describes how the overall distribution becomes more similar to a normal distribution.

As a result, **multiple independent replications** are required to observe CLT behavior experimentally. Histograms and Q–Q plots are necessary tools to compare the empirical distribution of the standardized sample means with the theoretical normal distribution.

3.3 Theoretical and Experimental Comparison

Almost sure convergence is strictly stronger than convergence in distribution. If a sequence converges almost surely, it also converges in probability and, under certain conditions, in distribution. However, convergence in distribution alone does not guarantee convergence along individual sample paths.

This theoretical difference leads to distinct experimental methodologies:

- **SLLN:** Demonstrated using a single long simulation and visualized with a convergence line graph.
- **CLT:** Demonstrated using many repetitions and visualized with histograms and Q–Q plots.

The necessity of different experimental approaches highlights the conceptual distinction between the two theorems. While SLLN answers the question of whether averages stabilize over time, CLT explains why normalized averages exhibit a normal distribution in repeated experiments.

These differences have important practical implications in simulation studies and statistical modeling, where the choice of convergence concept directly influences how results should be interpreted.

4. Methodology

All simulations in this project were implemented using the **Python** programming language due to its flexibility and extensive scientific computing libraries. Python is widely used in probability simulations and Monte Carlo experiments, making it a suitable choice for this study.

4.1 Programming Environment and Libraries

The following libraries were used throughout the simulations:

- **NumPy:** for random number generation and numerical computations

- **Matplotlib:** for data visualization and plotting convergence graphs, histograms, and Q–Q plots
- **SciPy (optional):** for statistical functions and normal distribution comparison

4.2 Simulation Design and Parameters

For the SLLN simulations, a sequence of independent and identically distributed random variables following a **Uniform(0,1)** distribution was generated. The theoretical mean of this distribution is $\mu = 0.5$. The cumulative sample mean was computed for increasing values of the sample size n, and the convergence behavior was observed using a line plot.

For the CLT simulations, multiple independent replications were performed for each selected sample size n. In each replication, the sample mean was standardized using the theoretical mean μ and standard deviation σ of the underlying distribution. The resulting standardized values were then used to construct histograms and Q–Q plots in order to compare the empirical distribution with the standard normal distribution.

Typical values of n used in the simulations included small, moderate, and large sample sizes (for example n = 10, 50, 100, 500, and 1000), allowing the rate of convergence to be analyzed.

4.3 Random Number Generation

Random samples were generated using NumPy's built-in pseudo-random number generator. A fixed random seed was optionally set to ensure reproducibility of the results. Since the random variables are i.i.d. and satisfy the assumptions of SLLN and CLT, the generated samples are theoretically valid for the conducted simulations.

4.4 Monte Carlo Estimation of π

In addition to SLLN and CLT simulations, the Monte Carlo method was applied to estimate the value of π. Random points were generated uniformly inside a unit square, and the proportion of points falling inside the unit circle was used to estimate π. The convergence of the estimate and the absolute error were analyzed as the number of simulations increased.

5. Results

This section presents the simulation results for the Strong Law of Large Numbers (SLLN), the Central Limit Theorem (CLT), and the Monte Carlo estimation of π. The results are illustrated through graphical representations and interpreted in light of the corresponding theoretical expectations.

5.1 SLLN Results

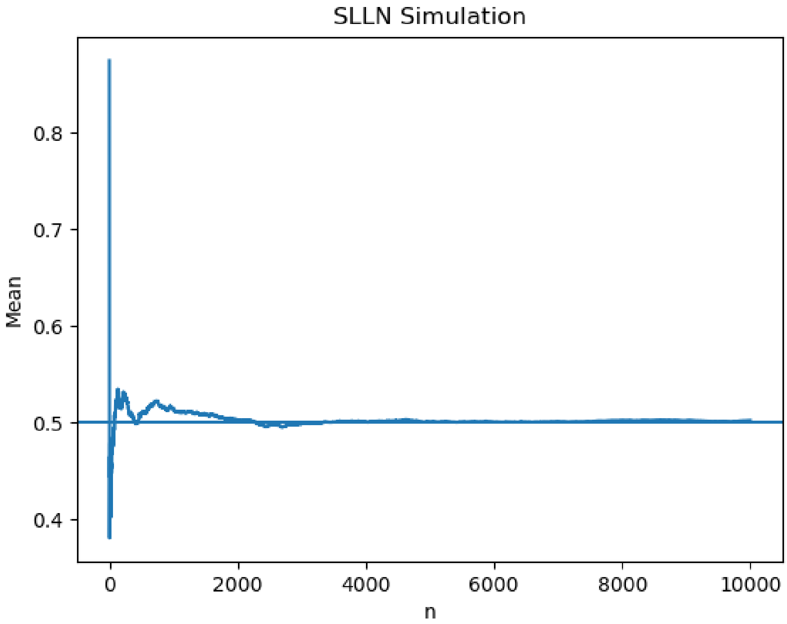

Figure 5.1

Figure 5.1 illustrates the convergence of the cumulative sample mean to the theoretical mean $\mu = 0.5$ for a Uniform(0,1) distribution. As the sample size n increases, the sample mean exhibits decreasing fluctuations and stabilizes around the expected value.

The graph shows that for small values of n, the sample mean fluctuates significantly due to randomness. However, as n grows larger, these fluctuations diminish and the curve becomes smoother. Based on the simulation, the sample mean can be considered sufficiently close to 0.5 when n exceeds a few hundred observations, after which deviations become negligible.

This behavior is consistent with the Strong Law of Large Numbers, which guarantees almost sure convergence of the sample mean to the expected value along a single sample path.

5.2 CLT Results

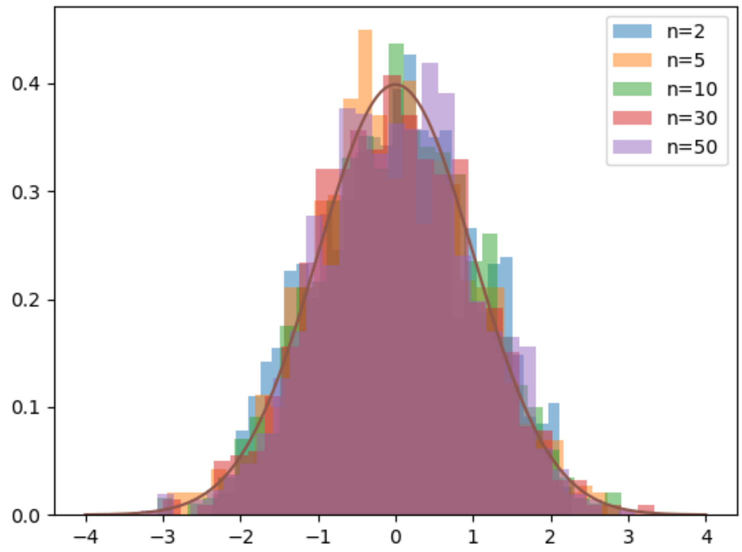

Figure 5.2

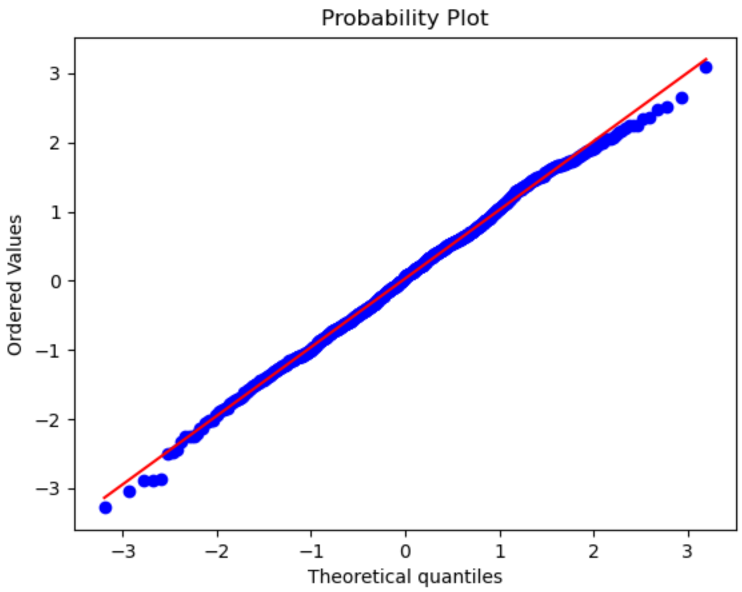

Figure 5.3

The Central Limit Theorem results are presented using both a histogram and a Q–Q plot of the standardized sample means.

Figure 5.2 shows the histogram of the standardized sample mean for a fixed sample size n based on many independent replications. As predicted by the CLT, the histogram increasingly resembles the shape of the standard normal distribution as n increases.

Figure 5.3 presents the corresponding Q–Q plot, which compares the empirical quantiles of the standardized sample means with the theoretical quantiles of a standard normal distribution. The points closely follow the reference line, indicating good agreement with normality. Minor deviations at the tails are expected for finite sample sizes and do not contradict the theoretical result.

These graphical results confirm convergence in distribution rather than convergence along individual sample paths, highlighting the fundamental difference between CLT and SLLN.

5.3 Monte Carlo Estimation of π

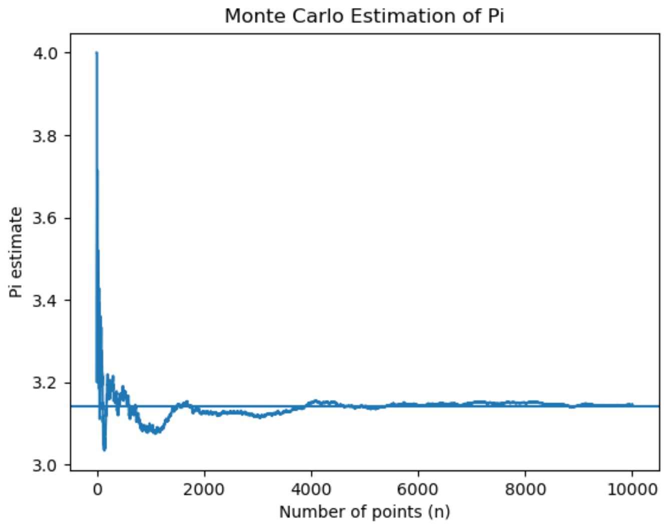

Figure 5.4

Figure 5.4 demonstrates the convergence of the Monte Carlo estimate of π as the number of random points increases. Initially, the estimate fluctuates substantially due to the small number of samples. As the number of simulations increases, the estimate gradually converges toward the true value of π.

The error analysis shows that the absolute estimation error decreases as the number of samples grows, which is consistent with the law of large numbers. Although convergence is relatively slow, the method provides an intuitive and effective demonstration of how random sampling can be used to approximate mathematical constants.

Overall, the results support the theoretical expectations of SLLN and CLT and demonstrate the effectiveness of Monte Carlo simulation techniques in probabilistic analysis.

6. Discussion and Conclusion

This project investigated the convergence properties of the Strong Law of Large Numbers (SLLN) and the Central Limit Theorem (CLT) through Monte Carlo simulations. By combining theoretical explanations with computational experiments, the study provided clear insights into how these fundamental probabilistic results manifest in practice.

One of the key questions addressed in this project is: How large should the sample size n be? The simulation results indicate that for the SLLN, the sample mean becomes sufficiently close to the true mean relatively quickly. In the case of a Uniform(0,1) distribution, convergence to $\mu = 0.5$ is visually stable once n reaches a few hundred observations. Beyond this point, fluctuations are minimal and the sample mean remains close to the expected value.

For the CLT, convergence depends not only on the sample size n but also on the number of replications. While small sample sizes result in noticeable deviations from normality, increasing n leads to a clear improvement in the approximation to the standard normal distribution. The histogram and Q–Q plot results show that moderate to large sample sizes are generally sufficient for the CLT to provide a good normal approximation.

A comparison of SLLN and CLT highlights important differences in their convergence behavior. SLLN exhibits almost sure convergence along a single sample path, leading to relatively fast stabilization of the sample mean. In contrast, CLT demonstrates convergence in distribution, which requires repeated experiments and focuses on the overall shape of the distribution rather than individual realizations. As a result, CLT convergence appears slower and must be evaluated using distribution-based visual tools.

From a practical perspective, these differences have significant implications for simulation studies and statistical modeling. Understanding the appropriate mode of convergence helps practitioners choose suitable simulation strategies and correctly interpret empirical results. Overall, the findings of this project confirm the theoretical foundations of SLLN and CLT and demonstrate the effectiveness of Monte Carlo methods as powerful tools for probabilistic analysis.