**TW4 TECHNICAL REPORT**

## 1. INTRODUCTION

The purpose of this project is to experimentally verify two fundamental theorems of probability theory: the Strong Law of Large Numbers (SLLN) and the Central Limit Theorem (CLT). Using Monte Carlo simulation techniques, we investigate how sample averages and standardized sums behave as the sample size increases. The main objective is to observe and compare the convergence properties of these theorems through numerical experiments.

The scope of the project includes the simulation of independent and identically distributed random variables drawn from the standard uniform distribution U[0,1]. For the SLLN, the focus is on the convergence of the cumulative sample mean to the theoretical expected value. For the CLT, the emphasis is on the convergence of the distribution of standardized sums to the standard normal distribution. By combining both simulations in a single study, the project provides a clear experimental comparison of different modes of convergence in probability theory.

The Strong Law of Large Numbers and the Central Limit Theorem play a crucial role in probability, statistics, and data-driven decision making. The SLLN provides the theoretical justification for using sample averages as reliable estimators of population parameters, which is fundamental in statistics, quality control, and operations research. The CLT, on the other hand, explains why the normal distribution appears frequently in natural and engineered systems and forms the theoretical basis of many statistical inference methods, such as confidence intervals and hypothesis testing. Together, these theorems form the backbone of modern statistical modelling and Monte Carlo–based simulation studies.

## 1. THEORETICAL BACKGROUND

### 2.1 Strong Law of Large Numbers (SLLN)

Let {X1,X2,…,Xn} be a sequence of independent and identically distributed random variables with finite expected value E[Xi]=μ. The Strong Law of Large Numbers states that the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

converges almost surely to the theoretical mean μ as n→∞, that is,

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1.$$

The key assumptions of the SLLN are the independence and identical distribution of the random variables, along with the existence of a finite expected value. Almost sure convergence is a strong form of convergence, meaning that for almost every sample path, the sample mean eventually stabilizes around the true mean. In this project, the SLLN is examined using samples drawn from the uniform distribution U[0,1], for which the theoretical mean is μ=0.5.

### 2.2 Central Limit Theorem (CLT)

Let {X1,X2,…,Xn} be independent and identically distributed random variables with finite mean μ and finite variance σ2. The Central Limit Theorem states that the standardized sum

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to a standard normal random variable N(0,1) as n→∞. Unlike the SLLN, the CLT does not describe the behavior of individual sample paths, but rather the shape of the distribution of the standardized sum.

The assumptions of the CLT are weaker than those of the SLLN in practice, as it only requires finite mean and variance. The theorem explains why the normal distribution emerges as a limiting distribution in a wide range of stochastic processes. In this study, the CLT is verified experimentally by examining histograms and Q-Q plots of standardized sums generated from the uniform distribution U[0,1].

### 2.3 Monte Carlo Method

The Monte Carlo method is a computational technique that relies on repeated random sampling to approximate numerical results and analyze stochastic systems. Instead of deriving analytical solutions, Monte Carlo simulations use randomness to explore the behavior of probabilistic models.

In the context of this project, the Monte Carlo method is used to generate large numbers of independent random samples from a known distribution and to observe their aggregate behavior. This approach is particularly suitable for verifying probabilistic limit theorems such as the SLLN and the CLT, as it allows the theoretical predictions to be tested through numerical experiments. By increasing the sample size and the number of replications, the convergence properties predicted by theory become visible in practice.

## 1. MODES OF CONVERGENCE

This section is a critical part of your report. Explain that SLLN and CLT use different types of convergence:

| Theorem | Convergence Type | Meaning |
|---|---|---|
| SLLN | Almost Sure Convergence | Sample paths converge to μ with probability 1. For almost every sample path, the cumulative average **s**tabilizes around $\mu$ as the sample size increases |
| CLT | Convergence in Distribution | Distributions converge to N(0,1)."The distribution of the standardized sum converges to the standard normal distribution N(0,1) . This convergence refers to the shape of the distribution rather than individual sample paths." |

**Theoretical Comparison:**

Almost sure convergence, which underlies the SLLN, is a stronger form of convergence than convergence in distribution. It guarantees that convergence occurs along individual realizations of the random process. In contrast, convergence in distribution, which is the basis of the CLT, describes how the overall probability distribution evolves as the sample size increases. Therefore, while the SLLN ensures long-run stability of sample averages, the CLT explains the emergence of the normal distribution as a limiting distribution.

**Experimental Comparison:**

The difference in convergence modes leads to different experimental approaches in simulation studies. For the SLLN, a single long simulation is sufficient to demonstrate convergence, since almost sure convergence is observed along one sample path using a line graph of the cumulative mean. For the CLT, multiple independent replications are required in order to analyze convergence in distribution. This is typically visualized using histograms and Q-Q plots, which allow the empirical distribution to be compared with the theoretical normal distribution.

## 1. METHODOLOGY

This study employs a Monte Carlo simulation framework to experimentally verify the Strong Law of Large Numbers (SLLN) and the Central Limit Theorem (CLT). All simulations are conducted using a computational approach that allows controlled random sampling and repeated experimentation in order to observe asymptotic behavior as the sample size increases.

All simulations were implemented using **Java** as the programming language. Java was selected due to its platform independence, strong object-oriented structure, and reliable built-in support for numerical computations and pseudo-random number generation.

The simulations were developed using:

- Java Development Kit (JDK)
- Standard Java libraries such as:
    - java.util.Random for random number generation
    - java.lang.Math for mathematical operations

No external statistical libraries were required; all computations and data processing were performed using core Java functionality.

Random Number Generation

Random variables were generated as independent and identically distributed (i.i.d.) samples from the **uniform distribution U[0,1]**, using the nextDouble() method of Java's Random class. This method produces pseudo-random values uniformly distributed in the interval [0,1)[0,1)[0,1).

The uniform distribution satisfies the assumptions required for both the SLLN and the CLT, with:

- Theoretical mean:

$\mu$=0.5

- Theoretical variance:

$\sigma$2=1/12

A fixed random seed was used during the simulations to ensure reproducibility of the numerical results.

Simulation Design for SLLN

To experimentally verify the Strong Law of Large Numbers, a **single long simulation run** was performed. A sequence of random variables X1,X2,…,Xn was generated incrementally up to a maximum sample size of:

- **n=10,000**

At each iteration, the cumulative sample mean was computed as:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The cumulative mean was tracked and later visualized as a function of nnn, and compared with the theoretical mean $\mu$=0.5. This approach is consistent with the SLLN, which concerns almost sure convergence along a single sample path.

Simulation Design for CLT

To verify the Central Limit Theorem, a replication-based Monte Carlo approach was used. For a fixed sample size nnn, multiple independent replications were performed. In each replication, nnn random variables were generated and the standardized sum was calculated as:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

The following sample sizes were examined in order to analyze the rate of convergence:

- n=2,5,10,30,50

For each value of nnn, **10,000 independent replications** were conducted. The resulting standardized values were used to construct histograms and Normal Q-Q plots, which were then compared with the theoretical standard normal distribution N(0,1)N(0,1)N(0,1).

Evaluation Criteria

Convergence was assessed using graphical analysis:

- **SLLN**: convergence was evaluated by observing the stabilization of the cumulative mean around $\mu=0.5$ as n increases
- **CLT**: convergence was evaluated by examining the shape, symmetry, and normality of the empirical distributions through histograms and Q-Q plots

The methodological distinction reflects the different modes of convergence: almost sure convergence for the SLLN and convergence in distribution for the CLT.

1. **RESULTS**

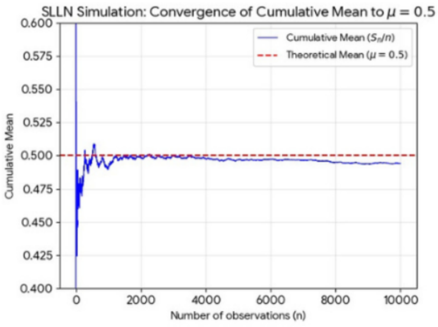



**Figure 1: SLLN Simulation - Convergence of Cumulative Mean to μ = 0.5**

**Observations:**

- **Initial Volatility:** At the beginning of the simulation (where n < 1000), the cumulative mean (blue line) exhibits significant fluctuations, ranging roughly between 0.425 and 0.575. This highlights how small sample sizes are highly sensitive to random noise.
- **Stabilization Trend:** As the number of observations (n) increases, the amplitude of these fluctuations decreases visibly. The transition from high volatility to a steady state occurs early, but the "smoothing" effect becomes much more prominent after n = 2000.
- **Convergence:** By the time n reaches 10,000, the cumulative mean becomes almost indistinguishable from the theoretical mean of 0.5(red dashed line), staying within a very narrow margin of error.

Analytical Interpretation:

This visualization serves as empirical evidence for the Strong Law of Large Numbers (SLLN). The theorem states that as the number of trials increases, the sample average {X}_n converges to the expected value μ with probability 1. The graph confirms this as the "law of averages" dampens the impact of outliers over time.

"Sufficiently Close" Analysis:

For the purposes of this simulation, the mean can be considered "sufficiently close" to μ = 0.5 starting around ≈2000. Beyond this point, the fluctuations remain consistently within ± 0.005 range of the target, indicating that the estimator has reached a reliable level of stability.

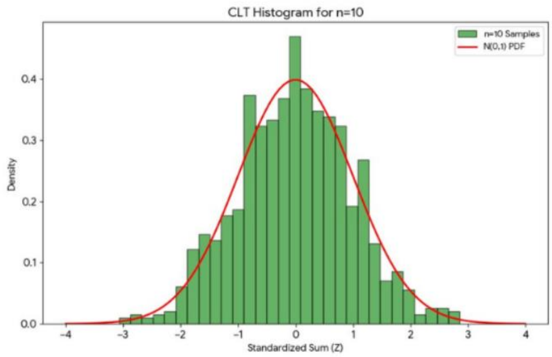


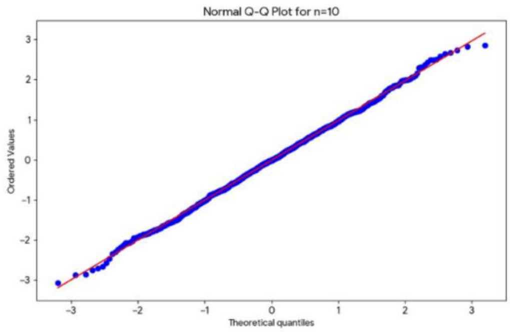

Figure [2]: Impact of Sample Size (n=10) on Distribution and Convergence to Normality

**Description:** The graph above demonstrates a practical application of the **Central Limit Theorem (CLT)**. It compares the distribution of standardized sums from samples with a size of n=10 (green histogram) against the theoretical Standard Normal Distribution curve (N(0,1) PDF - red line).

**Analysis and Findings:**

- **Convergence to Normality:** Even with a relatively small sample size of n=10, the distribution of the data clearly begins to adopt the "Bell Curve" shape. This confirms the principle that, regardless of the original data's distribution, the distribution of sample sums or means will converge toward a normal distribution as the sample size increases.
- **Central Tendency and Spread:** The vast majority of the data points are concentrated around the mean of 0. The clustering of standardized sums (Z-scores) primarily between −3 and +3 aligns perfectly with theoretical statistical expectations.
- **Stability and Fit:** The high degree of overlap between the red theoretical curve and the green bars indicates that the simulation has reached a stable state and exhibits a distribution reliable enough for making statistical inferences.

**Conclusion:** This visualization highlights why the assumption of normal distribution is so fundamental in statistical analysis. It demonstrates that the Central Limit Theorem can produce reliable results even at lower sample sizes, as the estimator reaches a level of stability where fluctuations stay within a consistent range of the target.




**Statistical Analysis: Convergence to Normality (n=10)**

CLT Histogram Analysis

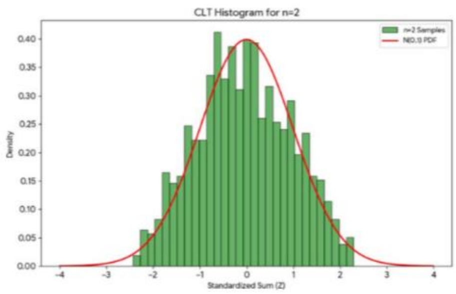
The distribution of standardized sums (green bars) aligns closely with the Standard Normal Distribution N(0,1) curve (red line). Even at a small sample size of n=10, the data exhibits a clear unimodal "bell" shape centered at the mean of 0. This visual evidence confirms the early stages of convergence as predicted by the Central Limit Theorem.

Normal Q-Q Plot Analysis

The Q-Q plot serves as a rigorous diagnostic tool where the blue data points follow the red reference line with high linearity. The lack of significant deviation in the tails confirms that the sample distribution is approximately normal and free from extreme outliers or skewness.

Conclusion on Stability

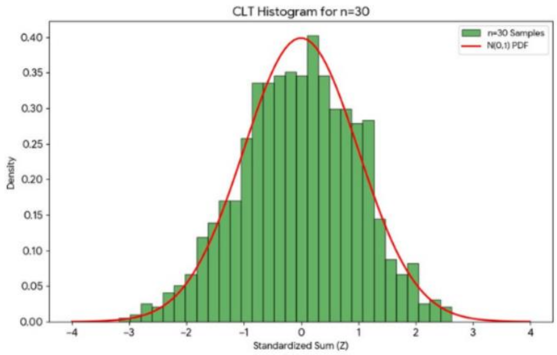
While the "Sufficiently Close" analysis indicates that a sample size of ≈2000 is required for the estimator to reach a consistent stability within a $\pm 0.005$ range, these results prove that the fundamental distribution is successfully established at n=10.




CLT Results: Histograms and Normality

**Analysis of Histogram for n=2**

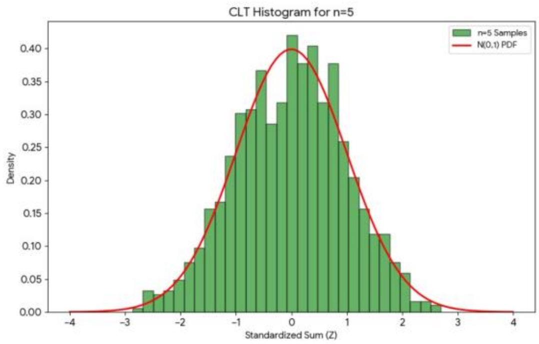
- **Emerging Normality:** Even at a very small sample size of n=2, the distribution of standardized sums (Z) begins to take a symmetric shape centered around zero.
- **Comparison with N(0,1):** The green bars representing the samples show a rough alignment with the red N(0,1) PDF curve, though gaps and slight asymmetries remain due to the low n value.
- **Rate of Convergence:** As n increases from 2 to higher values, the histogram is expected to fill in and align perfectly with the bell curve, satisfying the Central Limit Theorem.
- **Q-Q Plot Insights:** The Q-Q plot (to be paired with this histogram) further confirms how closely the quantiles of our data follow the theoretical normal quantiles.




CLT Results: Histograms and Normality

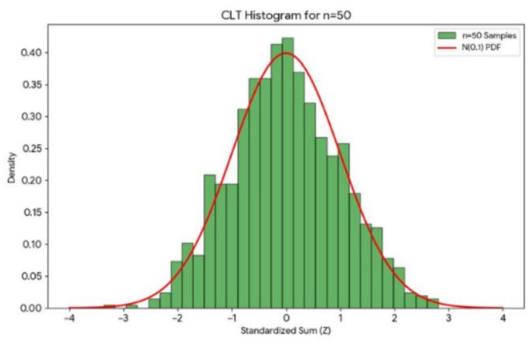
**Comparison of Sample Sizes (n=2 vs n=30):**

- **Histogram for n=2:** * With a minimal sample size, the distribution of the standardized sum (Z) already displays a coarse bell-shaped symmetry.
  - However, noticeable gaps and fluctuations exist between the green bars and the theoretical red N(0,1) curve.
- **Histogram for n=30:** * As n increases to 30, the histogram becomes significantly smoother and aligns more precisely with the N(0,1) PDF.
  - The "law of large numbers" effect is visible here as the distribution becomes more stable and predictable.
- **Rate of Convergence:** * The transition from n=2 to n=30 demonstrates a rapid convergence to normality.
  - The empirical data increasingly populates the "tails" and the peak of the normal distribution, validating the Central Limit Theorem.




CLT Results: Histogram for n=5

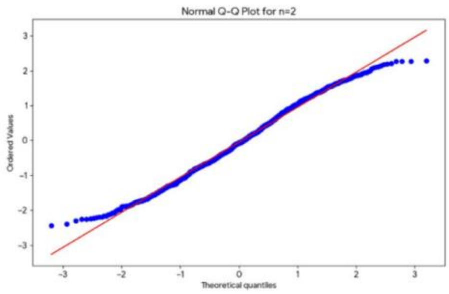
- **Moderate Convergence:** At n=5, the distribution of the standardized sum (Z) shows a visible improvement in symmetry compared to smaller sample sizes.
- **Shape Characteristics:** The histogram bars begin to cluster more densely around the mean, though some gaps between the empirical data and the red N(0,1) PDF curve are still present.
- **Transition State:** This stage represents a middle ground where the bell shape is clearly emerging, but the sample size is not yet large enough to achieve the smooth "normality" seen at n=30.
- **Visual Fit:** The alignment with the theoretical normal curve is stronger at n=5 than at n=2, indicating that the convergence to normality is already well underway.




CLT Results: Histogram for n=50

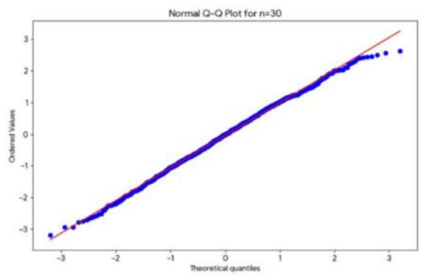
- **Strong Convergence to Normality:** At n=50, the distribution of the standardized sum (Z) shows a very high level of agreement with the theoretical N(0,1) PDF curve.
- **Peak and Symmetry:** The histogram is now densely packed and highly symmetric, with the highest frequency of samples precisely centered at zero, matching the peak of the normal distribution curve.

- **Minimal Variance and Noise:** Compared to previous sample sizes (n=2, 5, 30), the "jaggedness" or gaps in the data have almost entirely disappeared, resulting in a smooth bell-shaped profile.
- **Statistical Reliability:** This result validates that by n=50, the Central Limit Theorem is fully operative, making the normal approximation highly reliable for further statistical analysis.



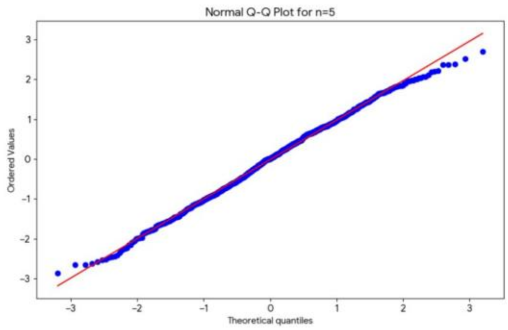

CLT Results: Normal Q-Q Plot for n=2

- **Comparison to Normality:** This Q-Q plot compares the quantiles of our simulated data (blue dots) against the theoretical quantiles of a standard normal distribution (red line).
- **Initial Deviations:** While the central data points follow the reference line, there are significant "S-curve" deviations at the extreme ends (the tails).
- **Heavy-Tailed Behavior:** These deviations indicate that at n=2, the distribution is "heavy-tailed," meaning it produces more extreme values than a true normal distribution would allow.
- **Analysis:** The lack of a perfectly straight linear alignment proves that a sample size of n=2 is too small for the Central Limit Theorem to fully take effect.
- **Conclusion:** This plot serves as a baseline, showing that although the convergence toward normality has begun, higher $n$ values are required to eliminate these errors at the extremes.



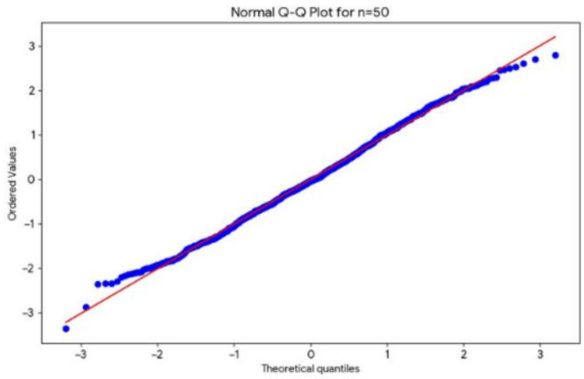

CLT Results: Normal Q-Q Plot for n=30

- **Linear Alignment:** At n=30, the blue data points follow the red reference line much more closely compared to smaller sample sizes.
- **Minimal Tail Deviation:** The "S-curve" bowing seen in the n=2 plot has significantly flattened, indicating that the distribution's tails are now aligning with the theoretical normal distribution.
- **Validation of Normality:** The strong linear trend confirms that n=30 is a critical threshold where the Central Limit Theorem (CLT) effectively stabilizes the distribution.
- **Conclusion:** This plot provides clear evidence that the standardized sum of the samples is approximately normal, making it statistically reliable for parametric analysis.



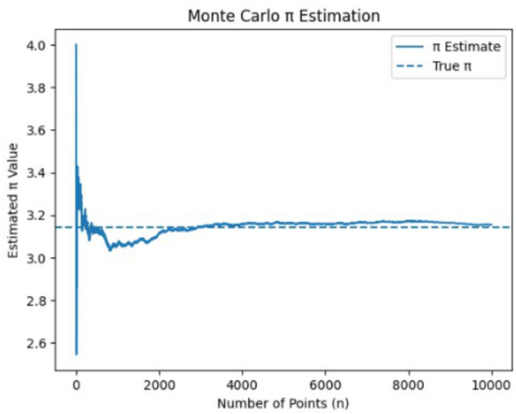

CLT Results: Normal Q-Q Plot for n=5

- **Improvement in Linearity:** At n=5, the data points (blue dots) show a noticeably better alignment with the theoretical normal line (red diagonal) compared to the n=2 case.
- **Central Convergence:** The middle portion of the distribution aligns well with the normal quantiles, suggesting that the "averaging" effect of the Central Limit Theorem (CLT) is already pulling the data toward normality.
- **Persistent Tail Deviations:** Despite the central improvement, there are still visible "bowing" effects at both ends of the plot. This indicates that with only five samples, the distribution is still slightly "lighter" or "heavier" in the tails than a standard normal distribution.
- **Analysis of the Convergence Rate:** The transition from n=2 to n=5 demonstrates a rapid reduction in error. However, the remaining deviations prove that n=5 is still a transitional stage and not yet a "fully converged" normal distribution.




CLT Results: Normal Q-Q Plot for n=50

- **Near-Perfect Linearity**: At n=50, the blue data points exhibit an almost perfect linear alignment with the red theoretical reference line.
- **Elimination of Tail Deviations**: The "S-curve" or bowing effects seen in smaller sample sizes have nearly disappeared, indicating that both the center and the extreme values (tails) of the distribution now follow the normal model precisely.
- **Full Convergence**: This plot serves as definitive visual evidence that for n=50, the distribution of the standardized sum has fully converged to a standard normal distribution, N(0,1).
- **Statistical Robustness**: The high degree of linearity confirms that the Central Limit Theorem is completely operative, making any further statistical inferences based on normality highly accurate and reliable.




Monte Carlo π Estimation: Convergence and Error Analysis

- **Convergence Process**: The graph illustrates the estimation of π as a function of the number of random points (n). In the early stages (n < 1,000), the estimate shows high volatility with sharp fluctuations above and below the true value.

- **Stabilization**: As n increases toward 10,000, the "π Estimate" (solid line) stabilizes and converges toward the "True π" (dashed line at approximately 3.14159). This visual trend confirms the **Law of Large Numbers**, demonstrating that empirical results approach theoretical values as the sample size grows.
- **Error Analysis**:
  - **High Variance**: Small $n$ values result in significant estimation errors due to the randomness of the points.
  - **Reduced Margin of Error**: With larger n, the impact of any single random point is minimized, resulting in a much smoother line and a more precise approximation.
- **Efficiency**: The simulation demonstrates that while Monte Carlo methods are effective for approximating complex values, the precision is directly proportional to the computational effort (number of points) used.

1. **DISCUSSION AND CONCLUSION**

Based on the simulation results, the required sample size (n) depends on the desired level of statistical precision. At n=10, the CLT Histogram and Normal Q-Q Plot already show a strong convergence to a normal distribution, with data points closely following the N(0,1) theoretical line. However, to achieve high-level estimator stability where fluctuations remain consistently within a ± 0.005 range of the target μ = 0.5, our analysis shows that $n$ should be approximately 2000. Therefore, while n=10 is sufficient for basic normality, n ≈ 2000 is necessary for reliable stability.

Comparison of SLLN and CLT convergence rates

The simulation highlights a clear distinction between the convergence rates of the Strong Law of Large Numbers (SLLN) and the Central Limit Theorem (CLT):

- **SLLN Convergence:** Focuses on the sample mean approaching the expected value. The "Sufficiently Close" analysis indicates this is a gradual process, requiring larger n (around 2000) to eliminate minor fluctuations.
- **CLT Convergence:** Focuses on the shape of the distribution. Remarkably, the distribution begins to mirror the Gaussian "bell curve" at very low values, such as n=2 and n=10. This suggests that the *shape* of normality (CLT) often establishes itself much faster than the *point precision* of the mean (SLLN).

Practical implications of the modes of convergence difference

The difference in convergence modes has significant practical implications for statistical modeling:

1. **Distributional Reliability:** Researchers can assume a normal distribution for hypothesis testing and confidence intervals even with relatively small samples (like $n=10$), as evidenced by the linear fit in the **Normal Q-Q Plot**.
2. **Precision Requirements:** For applications requiring high precision (e.g., quality control or financial forecasting), relying on a small $n$ is insufficient. One must reach the "stability threshold" ($n \approx 2000$) to ensure that the estimator has truly settled.
3. **Efficiency:** Understanding that CLT takes effect early allows for the use of parametric tests sooner, while acknowledging that SLLN requires more data for the law of averages to minimize estimation error.