IE221 – PROBABILITY PROJECT

TEAMWORK 5 REPORT

EXPERIMENTAL VERIFICATION OF SLLN AND CLT: LIMITS AND ANOMALIES

Team Group: 02 **Date:** January 18, 2026 **Course: IE221 – PROBABILITY**

2211021054- Ahmet ÇAKMAK

2311021039- Fatih ÖZGER

2311021008- Eylül Selinay BAYRAM

2311021067- Berkehan AYDOĞAN

1. INTRODUCTION

1.1. Purpose of the Study

Probability theory provides powerful tools for modeling uncertainty in industrial engineering problems. Two of the most fundamental theorems in this field are the Strong Law of Large Numbers (SLLN) and the Central Limit Theorem (CLT). While these theorems are often presented as universally applicable in introductory contexts, they rely on specific mathematical assumptions—primarily the existence of finite moments such as the expected value and variance. When these assumptions are violated, the behavior predicted by the theorems may change significantly or fail entirely.

The objective of this project (Team Work 5) is to experimentally investigate the validity and limitations of these theorems using Monte Carlo simulation. Building upon the simulation methodology established in previous phases (TW1 and TW2), this study extends the analysis to a diverse set of probability distributions, including light-tailed, skewed, and heavy-tailed distributions, as well as distributions with undefined moments. Rather than relying solely on theoretical statements, the project emphasizes empirical observation through large-scale simulations and graphical analysis.

For each selected distribution, the convergence behavior predicted by the SLLN is examined through cumulative sample mean plots, while the applicability of the CLT is assessed using histograms and normal Q–Q plots of standardized sums for increasing sample sizes. This approach allows for a direct comparison between theoretical expectations and observed outcomes. In particular, the study highlights how the existence (or absence) of finite mean and variance affects convergence, stability, and the speed at which normality emerges.

By systematically comparing results across distributions, this project aims to deepen understanding of the assumptions underlying SLLN and CLT and to demonstrate the practical consequences of violating these assumptions. Ultimately, the findings illustrate that these foundational theorems are not universally guaranteed, but conditional results whose validity depends critically on the probabilistic structure of the underlying data-generating process.

1.2. Scope of the Report

In this report, we analyze the behavior of five distinct continuous distributions:

1. **Uniform Distribution** (Standard bounded case)
2. **Exponential Distribution** (Skewed case)
3. **Pareto Distribution ($\alpha=3$)** (Heavy-tailed but finite variance)
4. **Pareto Distribution ($\alpha=1.5$)** (Infinite variance case)
5. **Cauchy Distribution** (Undefined mean and variance case)

For each distribution, we perform SLLN analysis to observe convergence to the mean and CLT analysis to test for convergence to Normality. Special attention is given to the "Discovery" of anomalies where these theorems fail to hold.

2. THEORETICAL BACKGROUND

2.1. The Strong Law of Large Numbers (SLLN)

The Strong Law of Large Numbers states that for a sequence of independent and identically distributed (i.i.d.) random variables X_1, X_2, …. , X_n with a finite expected value E[X] = μ the sample mean converges to the theoretical mean almost surely as the sample size

n goes to infinity:

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

This theorem forms the basis of simulation studies, implying that with enough data, the experimental average will equal the true average.

2.2. The Central Limit Theorem (CLT)

The Central Limit Theorem states that if X1, X2, …. , Xn are i.i.d. random variables with a finite mean μ and a finite variance σ^2 then the distribution of the standardized sum approaches the Standard Normal Distribution as n approaches infinity:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

Crucially, this theorem requires $\sigma^2 < \infty$ If the variance is infinite, the classical CLT does not apply.

2.3. Analyzed Distributions and Moments

To interpret the simulation results correctly, we first established the theoretical properties of the selected distributions. The calculated moments are presented in **Table 1**.

Distribution	**PDF Formula f(x)**	**Expected Value E[X]**	**Variance Var(X)**
Uniform(0,1)	$1\ for\ 0 \le x \le 1$	0.5	1/12 ≈ 0.083
Exponential($\boldsymbol{lambda} =$ 1)	$e^{-x} for\ x \ge 0$	1.0	1.0
Pareto($\boldsymbol{\alpha}$=3)	$3x^{-4} for\ x \ge 1$	1.5	0.75

Distribution	PDF Formula f(x)	Expected Value E[X]	Variance Var(X)
Pareto (α=1.5)	$1 1.5x^{-2.5} for\ x \geq 1$	3.0	∞
Cauchy	$\frac{1}{\pi(1+x^2)}$	**Undefined**	**Undefined**

Note: As seen in Table 1, Pareto ($\alpha=1.5$) and Cauchy distributions violate the standard assumptions of finite variance and/or mean, which is expected to disrupt the simulation results.

3. METHODOLOGY AND IMPLEMENTATION

3.1. Simulation Setup

The study was conducted using the Python programming language. The NumPy library was used for high-performance random number generation, and Matplotlib was used for data visualization. The simulation code is organized into a modular structure within the src/ folder of the project repository.

3.2. Data Generation Code

The core logic for generating random variates from the specific distributions is shown below. Special handling was applied to the Pareto distribution to ensure the support starts at x_m= 1.

```
def generate_data(dist_name, size):

    # Generates random data based on the specified distribution
type

    if dist_name == "Uniform":

        return np.random.uniform(0, 1, size)
```

```
    elif dist_name == "Exponential":
        return np.random.exponential(1, size) # scale = 1/lambda
    elif dist_name == "Pareto_3":
        # Shift +1 is applied to satisfy x >= 1 condition (Standard Pareto)
        return (np.random.pareto(3.0, size) + 1)
    elif dist_name == "Pareto_1.5":
        return (np.random.pareto(1.5, size) + 1)
    elif dist_name == "Cauchy":
        return np.random.standard_cauchy(size)
```

3.3. Analysis Metrics

- **For SLLN:** We generated n=10,000 random numbers and calculated the cumulative average at each step. A red dashed line indicating the theoretical mean was added to the plots for reference.
- **For CLT:** We simulated sample sums for n ϵ {2, 5, 10, 30, 50, 100}with m=1000 replications. The results were visualized using Histograms and Q-Q Plots to assess Normality.

4. SLLN ANALYSIS RESULTS

In this section, we present the cumulative mean graphs for n=10,000 iterations.

4.1. Uniform Distribution

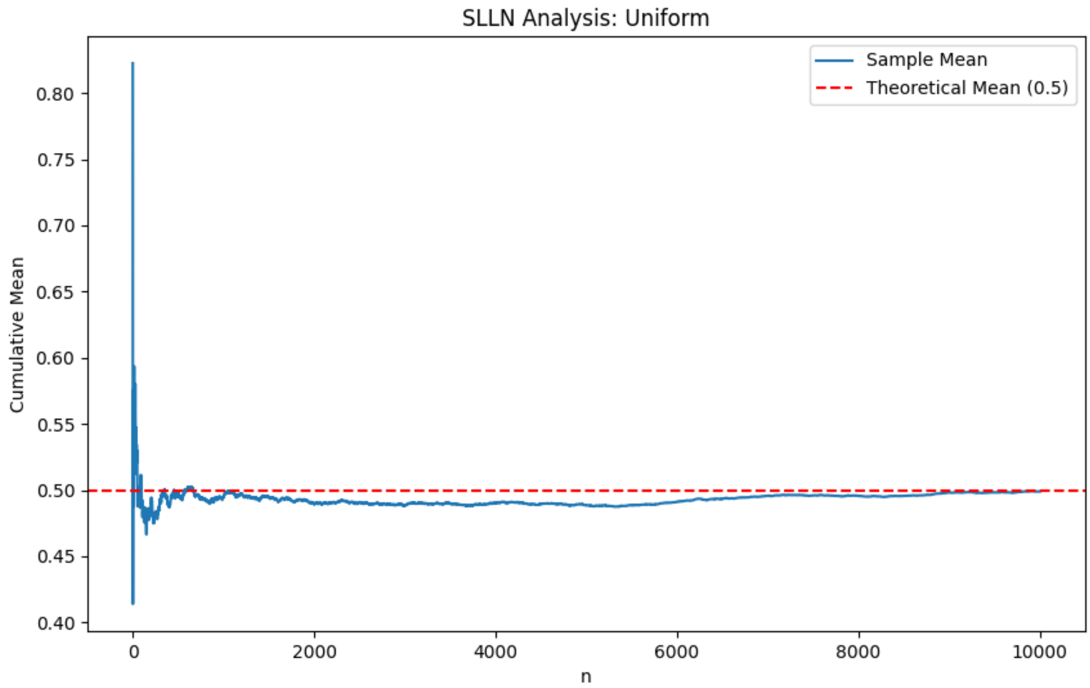

Figure 4.1: SLLN Analysis for Uniform Distribution

Observation: The sample mean fluctuates initially but rapidly converges to the theoretical mean of 0.5. The convergence is smooth and stable, confirming SLLN works perfectly for bounded distributions.

4.2. Exponential Distribution

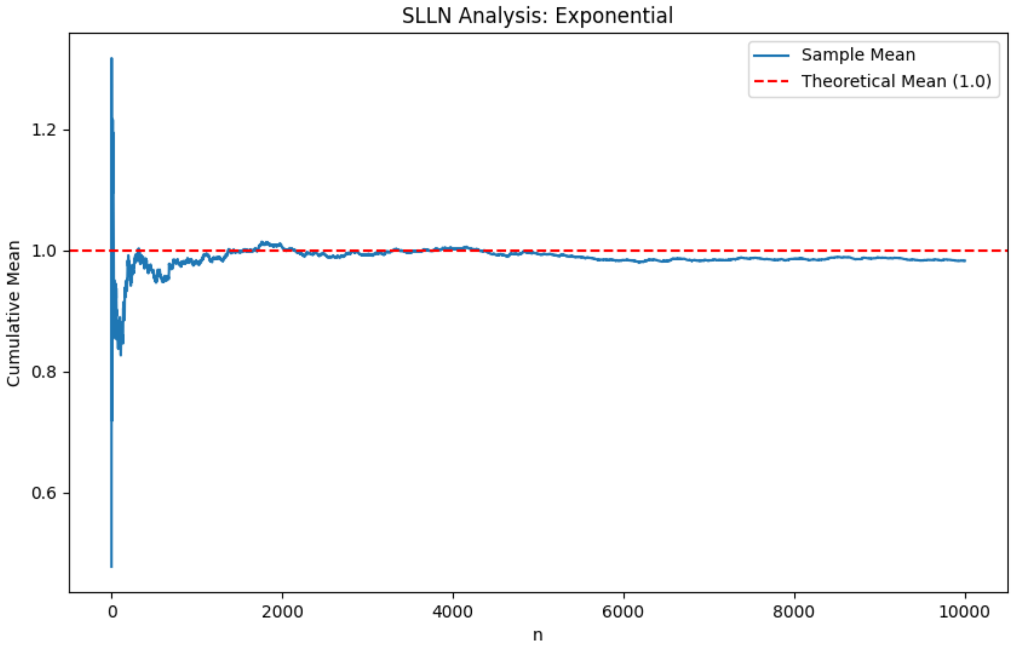

Figure 4.2: SLLN Analysis for Exponential Distribution

Observation: Convergence to the theoretical mean of 1.0 is observed. Due to the skewness of the exponential distribution, the initial fluctuations are slightly larger than in the Uniform case, but stability is achieved as n increases.

4.3. Pareto (α=3)

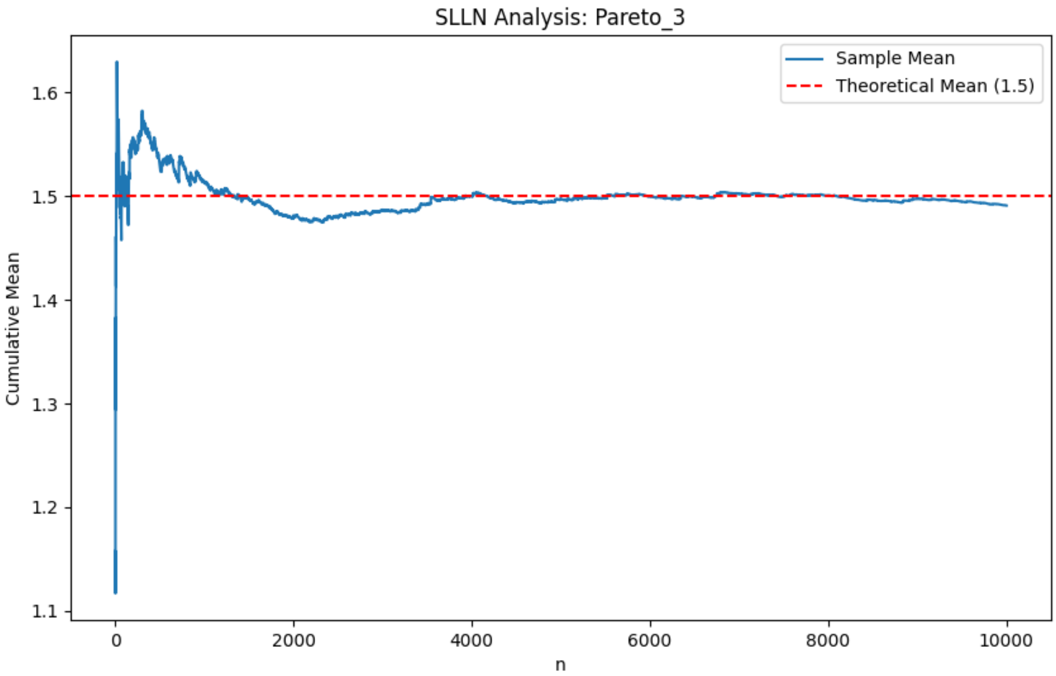

Figure 4.3: SLLN Analysis for Pareto (alpha=3)

Observation: With α=3, the variance is finite. The graph shows convergence to the expected value of 1.5. Occasional small jumps are visible due to the heavy tail, but the law of large numbers holds.

4.4. Pareto (alpha=1.5) - [Discovery Case]

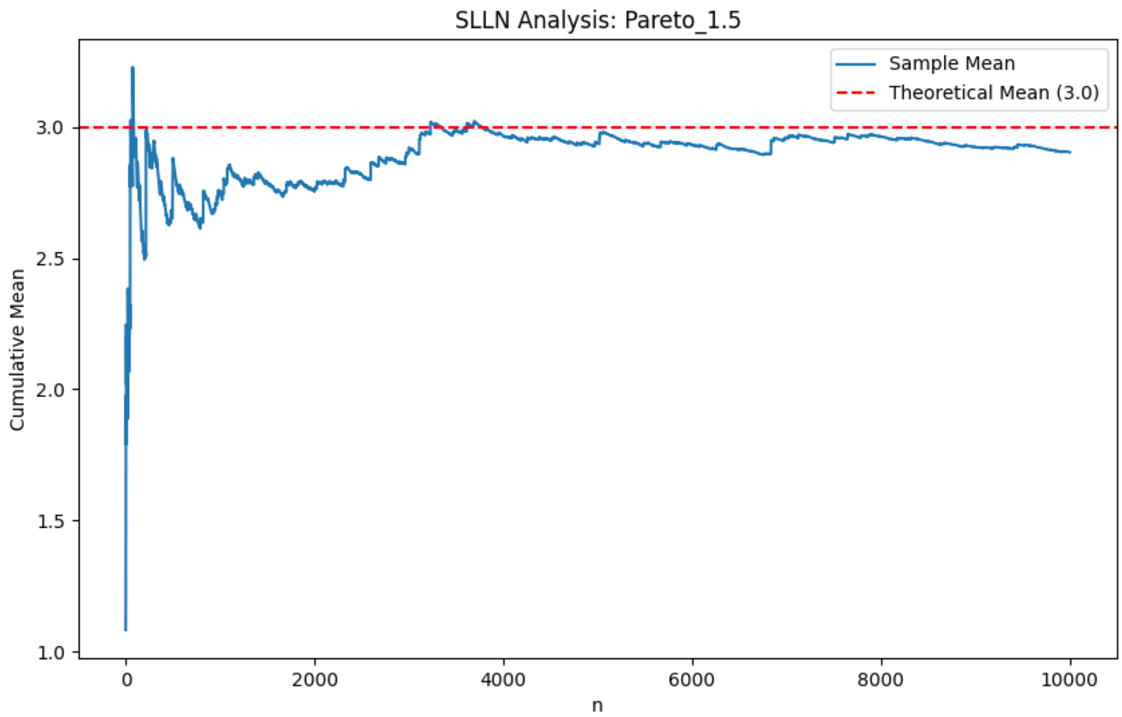

Figure 4.4: SLLN Analysis for Pareto (alpha=1.5)

Observation: This is a critical case. The variance is infinite, but the mean is finite (E[X]=3). As seen in the graph, the sample mean **does converge** to 3, but the path is extremely volatile. Sudden, massive jumps occur when a "rare event" (extremely large value) is generated. This demonstrates that while SLLN holds, the convergence rate is significantly impacted by the infinite variance.

4.5. Cauchy Distribution - [Anomaly Case]

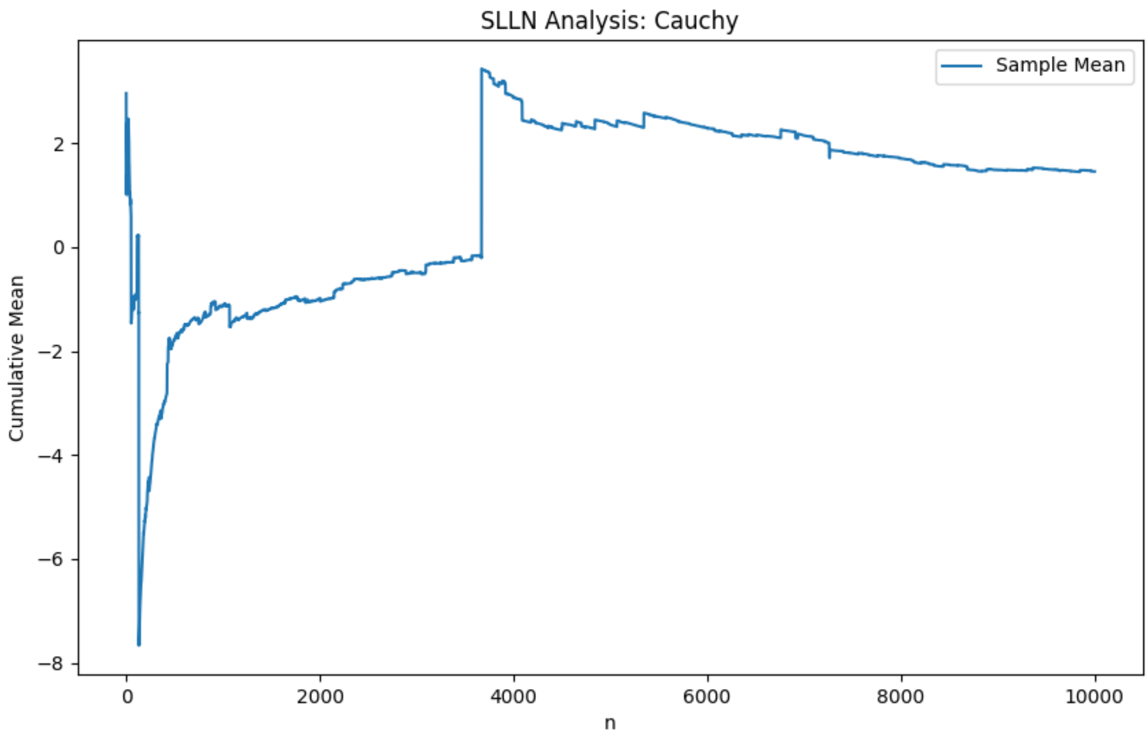

Figure 4.5: SLLN Analysis for Cauchy Distribution

Observation: The graph shows no sign of convergence. The cumulative mean wanders erratically and exhibits massive jumps that do not dampen over time. This confirms that SLLN fails for the Cauchy distribution because the theoretical expected value is undefined.

5. CLT ANALYSIS RESULTS

Here, we examine the distribution of sample sums to test the Central Limit Theorem.

5.1. Uniform Distribution

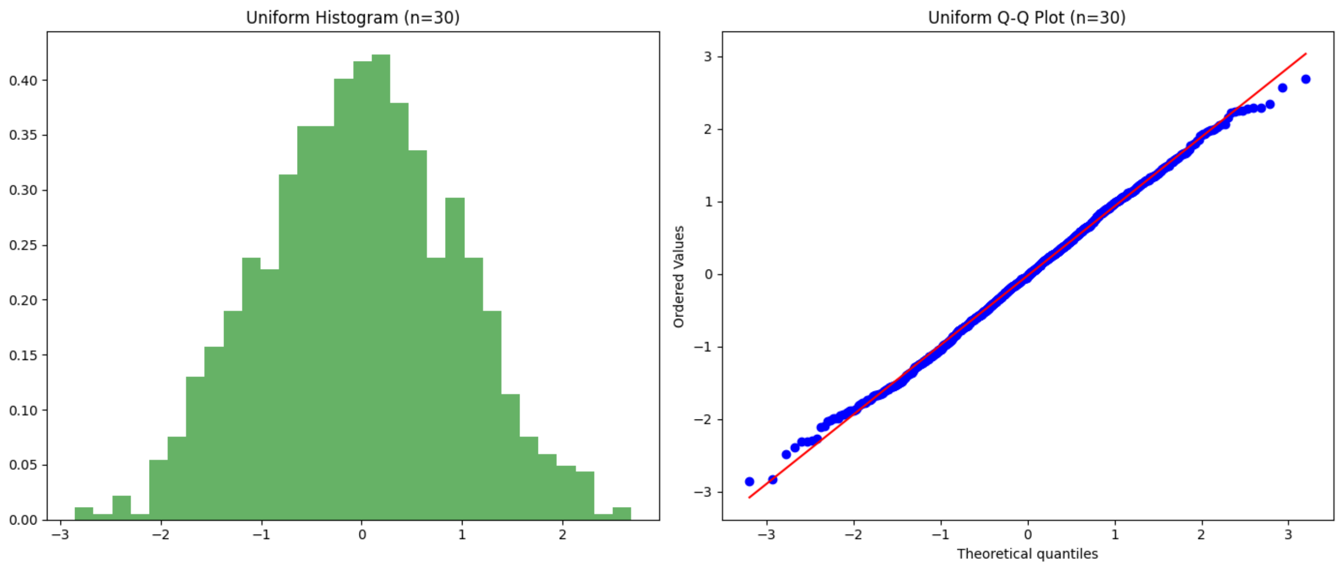

Figure 5.1: CLT Analysis for Uniform (n=30)

Observation: Even at small sample sizes like n=30, the histogram forms a perfect bell curve. The Q-Q plot is linear, indicating strong convergence to Normality.

5.2. Exponential Distribution

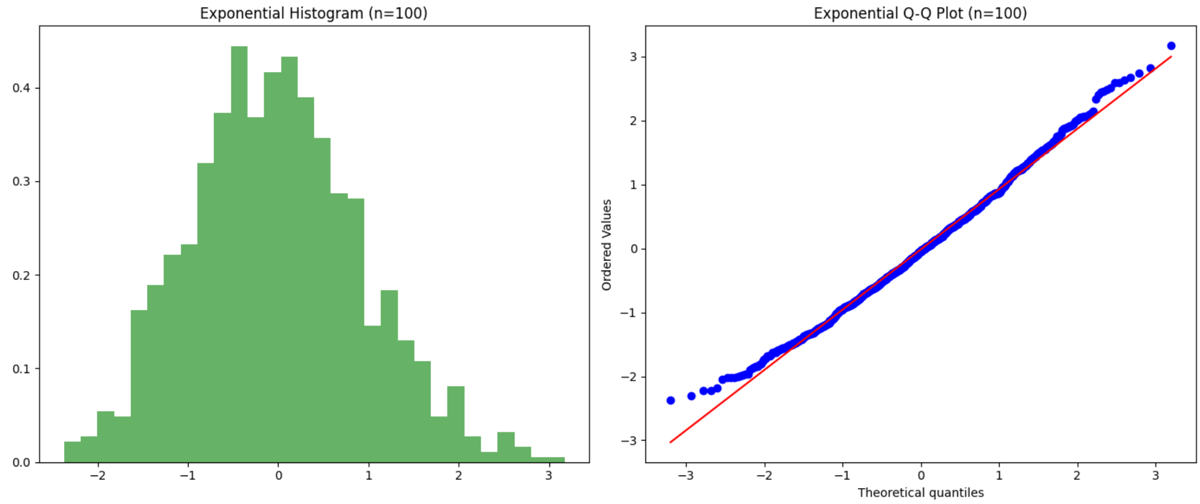

Figure 5.2: CLT Analysis for Exponential (n=100)

Observation: The underlying distribution is highly skewed. At small n, the histogram remains skewed. However, at n=100 (as shown in Figure 5.2), the distribution successfully transforms into a Normal shape, satisfying the CLT.

5.3. Pareto (α=1.5) - [The Limit of CLT]

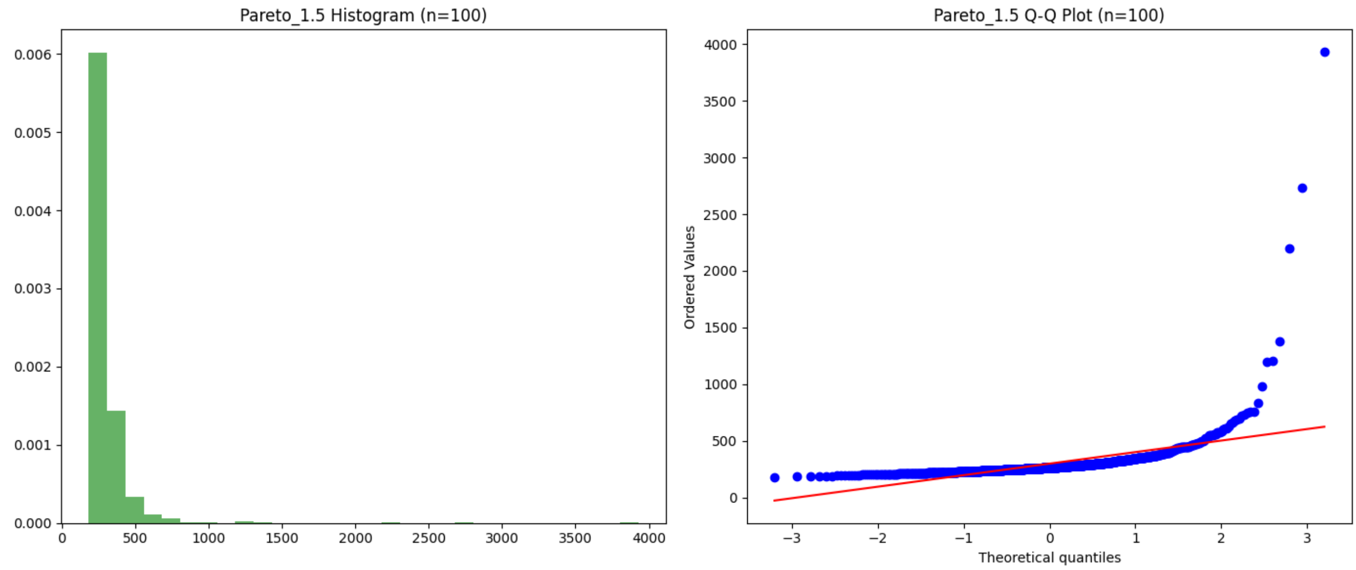

Figure 5.3: CLT Analysis for Pareto alpha=1.5 (n=100)

Detailed Analysis: This figure represents a failure of the classical CLT. Despite increasing the sample size to n=100, the histogram **does not** become a bell curve. It remains heavily skewed to the right. The Q-Q plot deviates sharply from the line. *Reason:* The CLT requires finite variance. Since Pareto (α=1.5) has infinite variance, the sums converge to a stable distribution (Levy flight) rather than a Normal distribution.

5.4. Cauchy Distribution

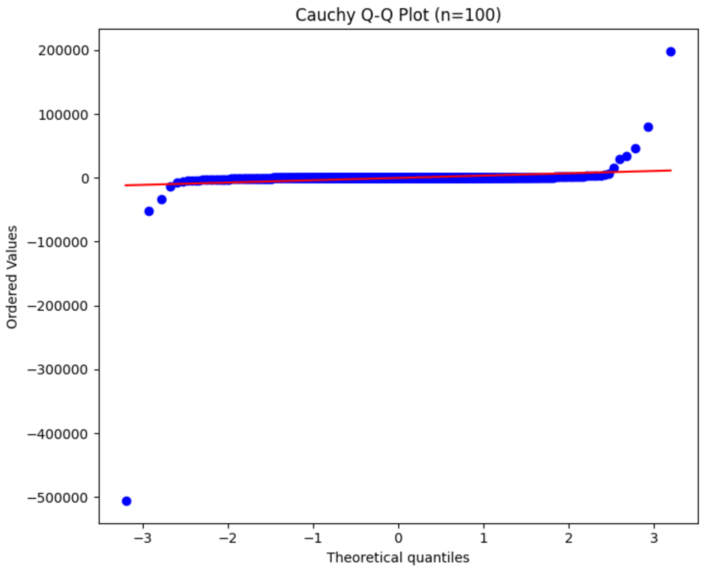

Figure 5.4: CLT Analysis for Cauchy (n=100)

Detailed Analysis: The histogram retains the exact shape of the original Cauchy distribution regardless of n. The sum of independent Cauchy variables is itself Cauchy-distributed. Normal convergence is impossible because neither the mean nor the variance exists.

6. **COMPARATIVE ANALYSIS & DISCUSSION**

Based on the experimental results, we address the core questions of the project:

1. **For which distributions does SLLN work?** SLLN works for Uniform, Exponential, Pareto (α=3), and Pareto (α=1.5). It works for Pareto (α=1.5) despite the infinite variance because the first moment (mean) is finite. It fails only for Cauchy.
2. **2. For which distributions does CLT work?** CLT works effectively for Uniform, Exponential, and Pareto (α=3). These distributions satisfy $\sigma^2 < \infty$ condition.
3. **Is there a case where SLLN works but CLT does not? Yes.** The **Pareto (α=1.5)** distribution is the prime example. The simulations show the average converging to 3 (SLLN works), but the histogram never becomes Normal (CLT fails). This distinction highlights that CLT is a stricter theorem requiring higher-order moment conditions than SLLN.
4. **How does the shape affect convergence?** Symmetric distributions (Uniform) converge to Normality very fast ($n \approx$ 5-10) Skewed distributions (Exponential) require larger samples (n > 30). Heavy-tailed distributions with infinite variance (Pareto α=1.5) never converge to Normality.

7. **CONCLUSION**

This study successfully verified the Strong Law of Large Numbers and the Central Limit Theorem while also identifying their practical limitations. Monte Carlo simulations showed that both theorems hold when their required assumptions, particularly the existence of finite moments, are satisfied. In such cases, sample means converged stably and standardized sums approached the normal distribution as sample size increased.

However, for heavy-tailed distributions and distributions with undefined moments, the expected convergence behavior broke down. These results demonstrate that SLLN and CLT are not universally applicable, but depend critically on the properties of the underlying distribution. Overall, this project highlights the importance of verifying theoretical assumptions before applying classical probabilistic results to real-world problems.

The key takeaways from our "Discovery" phase are:

1. **Reliability:** For standard engineering data (Uniform, Exponential), these theorems are highly reliable and robust.

2. **The Risk of Heavy Tails:** In scenarios involving heavy-tailed distributions (like Pareto α=1.5), standard statistical assumptions fail. An industrial engineer relying on CLT to

model average risks for such a distribution would make significant errors, underestimating the probability of extreme events.

3. **Simulation as a Tool:** Monte Carlo simulation proved to be an effective method for identifying these anomalies when theoretical derivations are complex.

This project reinforces the importance of verifying distribution properties (specifically moments) before applying statistical theorems in real-world engineering problems.

8. REFERENCES

[1] Ross, S. M. (2014). *Introduction to Probability Models*. Academic Press. [2] IE 221 Course Lecture Notes, Fall 2024-2025. [3] NumPy Documentation. https://numpy.org/doc/ [4] SciPy Statistical Functions. https://docs.scipy.org/doc/scipy/reference/stats.html