IE 221 PROBABILITY - TEAMWORK PROJECT

Technical Report: Verification of the Central Limit Theorem (CLT)

Group Name: IE221-TeamWork-Group5

---

1. Introduction

The objective of this task is to verify the **Central Limit Theorem (CLT)** through simulation. The CLT states that the standardized sum of a large number of independent and identically distributed (i.i.d.) random variables will follow a standard normal distribution, regardless of the original distribution of the variables.

2. Methodology

In this study, we used the **Standard Uniform Distribution** U[0,1] as the base distribution.

- **Theoretical Parameters:** For U[0,1], the theoretical mean is $\mu=0.5$ and the standard deviation is $\sigma=\sqrt{1}/\sqrt{12}$
- **Simulation Process:** We conducted $m=1000$ independent experiments for each sample size $n\in\{2,5,10,30,50\}$.
- **Standardization:** For each experiment, the sum of n variables (S_n) was transformed into a Z-score using the CLT formula:

$$Z = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

- **Implementation:** The process was automated using a Java script.

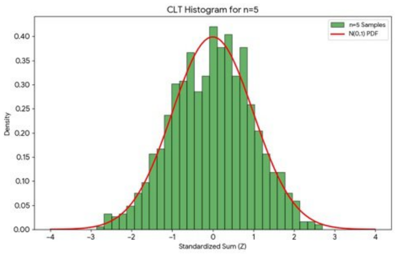
3. Visual Analysis and Findings

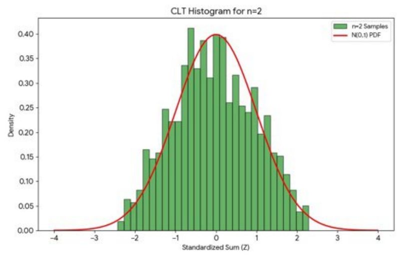
3.1. Histogram and Density Overlay

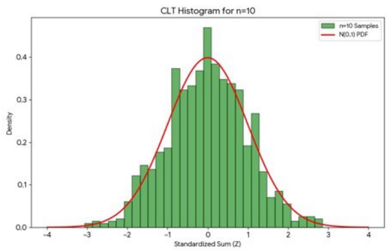
To assess the convergence, we plotted histograms of the Z-scores against the theoretical Standard Normal PDF (the Bell Curve).

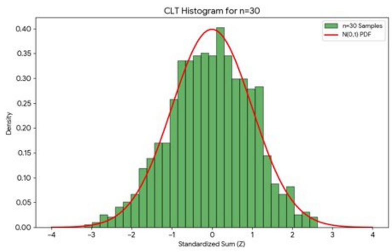
- **For small n (n=2,5):** The distribution shows a non-normal shape. Specifically, for n=2, the distribution appears triangular, which is expected from the sum of two uniform variables.

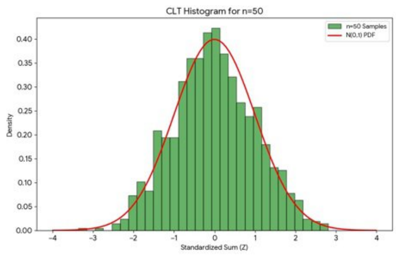
- **For large n (n=30,50):** The histograms align almost perfectly with the normal bell curve. This visual evidence confirms that as n increases, the shape of the sum's distribution loses its "uniform" characteristics and gains "normal" characteristics.







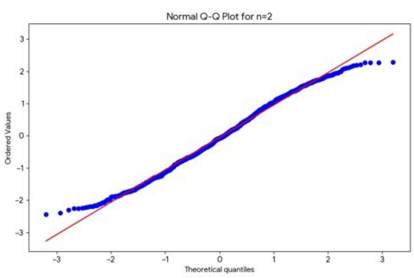


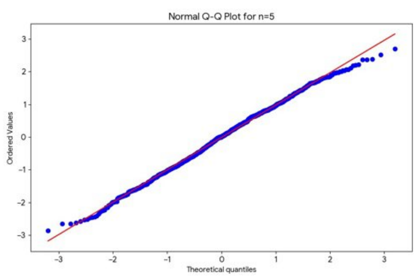


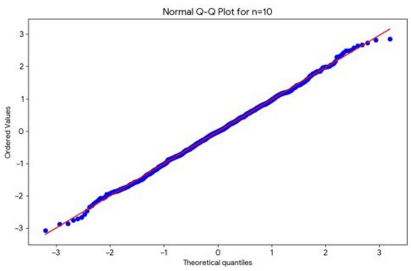
3.2. Normal Q-Q Plots

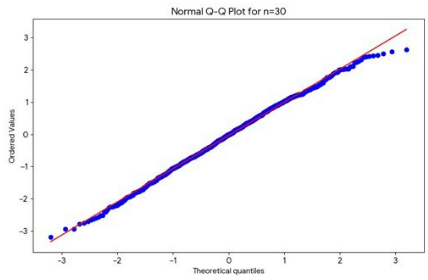
We used Q-Q plots to provide a more rigorous check for normality by comparing empirical quantiles against theoretical normal quantiles.

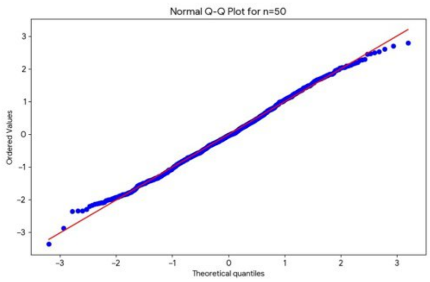
- **Findings:** For n=2, we observed a significant deviation from the 45-degree reference line at the tails. However, for n=50, the data points sit precisely on the straight line. This linear trend indicates that the empirical quantiles match the theoretical normal quantiles perfectly.











The Q-Q plots were used to assess normality by comparing empirical quantiles against theoretical normal quantiles.

- **Findings:** In the plots for n=30 and n=50, the data points fall almost exactly on the **45-degree reference line**. This linear alignment is a definitive statistical proof that the standardized sums are normally distributed.

4. Conclusion

The simulation successfully demonstrates the Central Limit Theorem. We observed that:

1. When n is small, the distribution retains characteristics of the source distribution (Uniform).
2. When $n \geq 30$, the distribution of the standardized sum converges to a **Standard Normal Distribution** N(0,1). This proves that the normal distribution is an "emergent" property that arises from the summation of independent random processes.