MA4240 Report

Dishank, Chirag, Mulugu, Datta

April 2022

1 Introduction

Central Limit Theorem states that normalised sum of independent and identically distributed random variables tends towards a normal distribution, irrespective of the distribution of random variables.

$$Z = \lim_{n \to \infty} \left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{n}} \right) \tag{1}$$

In this project, we propose to verify the correctness of Central Limit Theorem by running simulations beginning with a variety of distributions covered in the course.

2 Central Limit Theorem and imperial approximation

While equation (1) suggests that n should be a very large number. In practice, we tend to use the theorem for $n > 30$.

2.1 Proof of CLT

THE PROOF GOES HERE

3 Hypothesis

We formulate our hypothesis in the following manner

$$H_A : \text{CLT doesn't hold}$$
$$H_0 : \text{CLT holds}$$

Here, type 2 error is when CLT is actually false but we fail to reject it. It is more dangerous because several models are built on the assumption that CLT is indeed correct.

4 Procedure

We are generating a batch of 100 samples from the distribution, we find the sample mean of this batch, call it $\bar{X}$. From Central Limit Theorem, we know

$$\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{100}) \tag{2}$$

To verify the claim, we repeat this experiment 5000 times, then perform normality tests, which can be classified into two parts

1. Graphical Methods
   - Q-Q plot
   - Histogram
2. Frequentist tests
   - Shapiro-wilk test

4.1 Shapiro-Wilk test

lorum ipsum dolor sit amet

4.2 Distributions used

1. **Standard Normal**: The pdf of standard normal distribution is given by:

$$f_X(x) = \frac{1}{\sqrt{(2\pi)}} exp(-\frac{x^2}{2})$$

The mean is 0 and standard deviation is 1. Figure **??** shows the PDF of standard normal distribution.

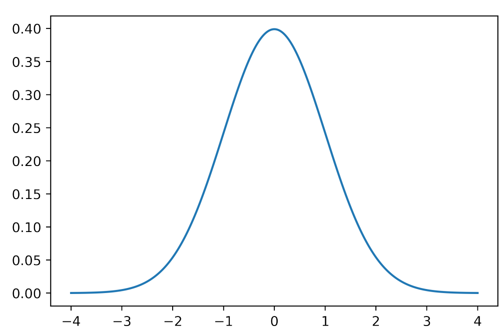

Figure 1: PDF of standard normal distribution

2. **Continuous uniform distribution**: Here, we have used U(0, 1). The PMF is given by

$$f_X(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & otherwise \end{cases}$$

The mean is 0.5 and standard deviation is 0.289. Figure **??** shows the PDF of the uniform distribution.

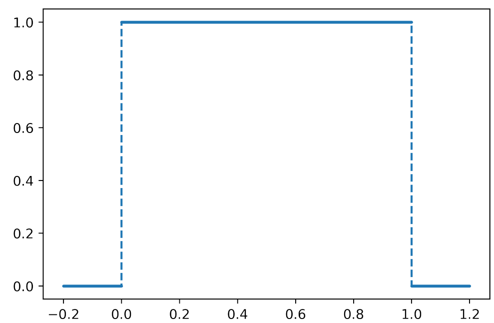

Figure 2: PDF of uniform distribution

3. **Geometric Distribution**: Unlike the other distributions that we used, this distribution is for a discrete random variable. The PDF is given by

$$f_X(k) = (1-p)^{k-1}p, \; k = 1, 2, 3, ...$$

In our experiments, we arbitrarily chose to use $p = 0.35$. The mean is $\dfrac{1}{p}$ and the standard deviation is $\dfrac{\sqrt{(1-p)}}{p}$. The PMF is given in figure **??**.

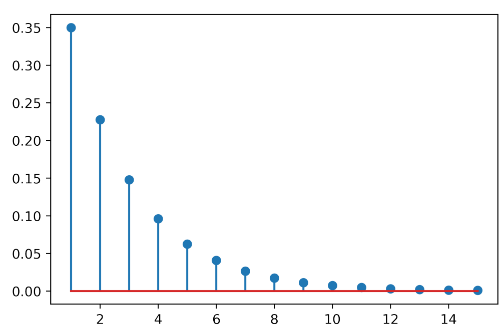

Figure 3: PMF of geometric distribution

4. **Standard cauchy distribution**: The PDF is given by

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

Neither the mean nor the standard deviation are finite. Thus CLT should not apply on this distribution. Figure **??** shows the PDF of standard cauchy distribution.

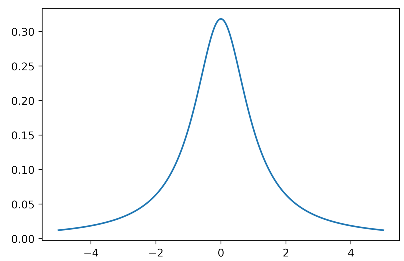

Figure 4: PDF of Cauchy Distribution

5 Shapiro-Wilk Test

6 Results

7 Conclusion