Report - Question 5

Atishay Jain & Megh Gohil

August 22, 2022

1 Uniform Distribution

A Uniform Distribution is a distribution that describes an experiment where there is an arbitary outcome which lies between certain bounds. That is, all the outcomes of a Uniform Distribution have the same probability. So, the true mean or the true expected value of such a distribution bounded in the interval [a,b] will be $\frac{a+b}{2}$.

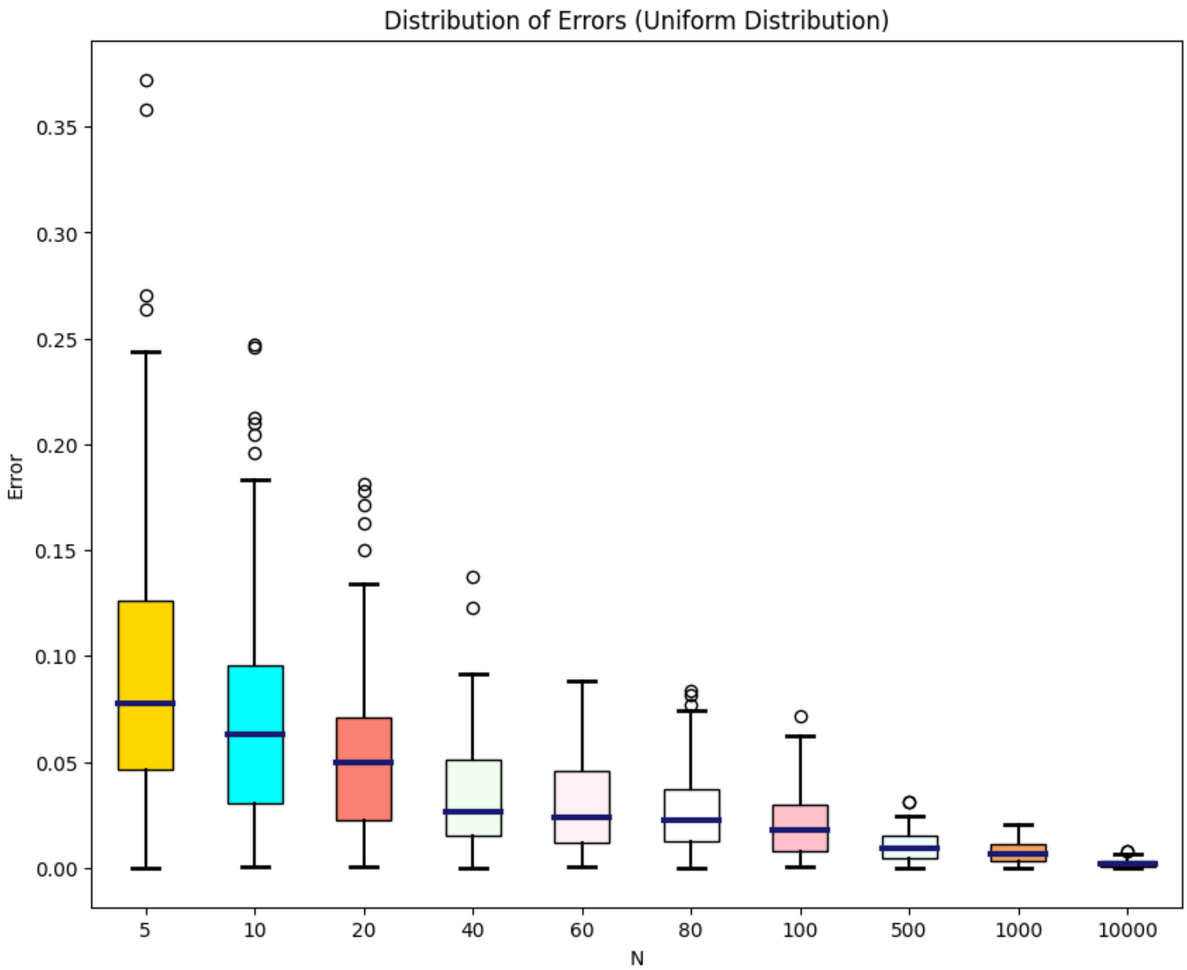

Figure 1: Distriubution of error in Mean (Uniform Distribution)

We generated a dataset comprising of N real numbers from the Uniform Distribution on $[0, 1]$. Considering various dataset sizes N = 5,10,20,40,60,80,100,500,1000,10000 for this experiment, we repeated the experiment of finding error between measured average and true mean for M := 100 times. In the code, we used uniform random number generator to generate Uniform random numbers for a given N. Then, computed their average $\hat{\mu}$. Since, the Uniform Distribution we are looking upon is on the bounded interval [0,1], its true mean will be $\mu = \frac{1}{2}$. Now, we measured the error between computed average $\hat{\mu}$ and true mean μ as $|\hat{\mu} - \mu_{true}|$. The above experiment was repeated 100 times for each of given N and we plotted the **Error** obtained for each **N** in the **Box Plot** above.

2 Gaussian Distribution

A Gaussian Distribution is a very important distribution in Statistics. For a Gaussian Distribution (with given μ and σ) the true mean is given by μ. For this question, we have $\mu = 0$ and $\sigma^2 = 1$. So, the true mean is 0. We are repeating the above same experiment on Gaussian distribution by using Random Normal distribution generator in the code with the given μ and σ. The plot below shows the error in true mean and computed average obtained for each N.

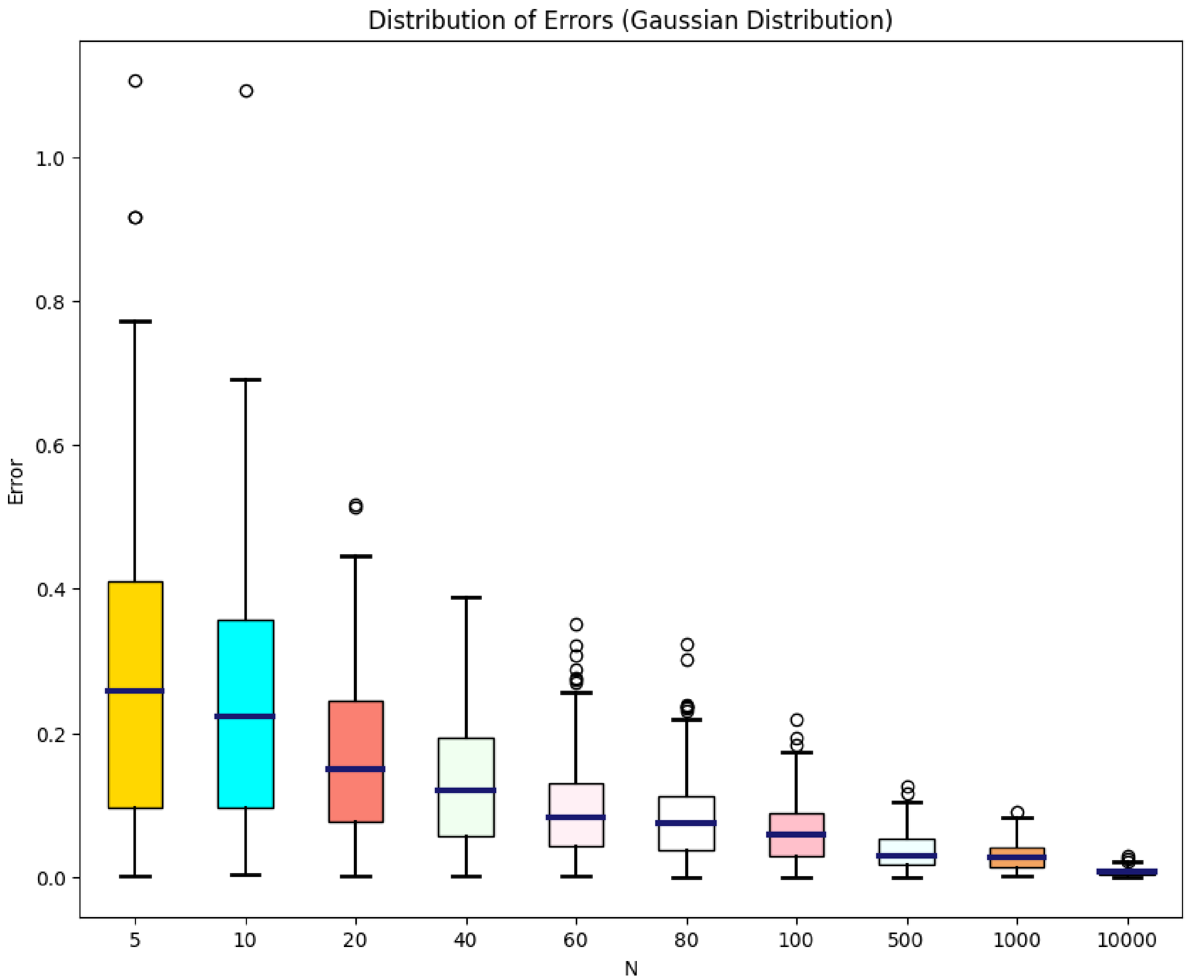

Figure 2: Distribution of error in Mean (Gaussian Distribution)

3 Interpretation of Graphs obtained

- The above graph shows the distribution of deviation of mean of randomly generated numbers from true mean
- The median of this hundreds of deviation is shown by a black line inside each box, so from the diagram above the median is tending to zero as the size of data increases
- The lower small side of rectangle in each boxplot tells about 25^{th} percentile which means the median of dataset starting from lowest number to the median while the higher small side of the rectangle tells about the 75^{th} percentile which means the median of dataset starting from median number to highest number in the dataset. It can be interpreted from the graph that the gap difference

between the 25^{th} and 75^{th} percentile is decreasing as the size of dataset is increasing, and this can be explained with the help of law of large numbers. Similarlly the gap difference between highest value and lowest value is also decreasing with large dataset.

- The empirical side of law of large numbers is clearly depicted by above graphs i.e. $P(|x - \mu| \geq \epsilon) \leq \frac{v}{n\epsilon^2}$. As the size of data increases the deviation of errors from true mean decreases.