Assignment 1

Harshvardhan Agarwal
200050050

Parth Dwivedi
200050100

August 2021

1 Laplace, Gumbel and Cauchy Distribution

1.1 Laplace Distribution

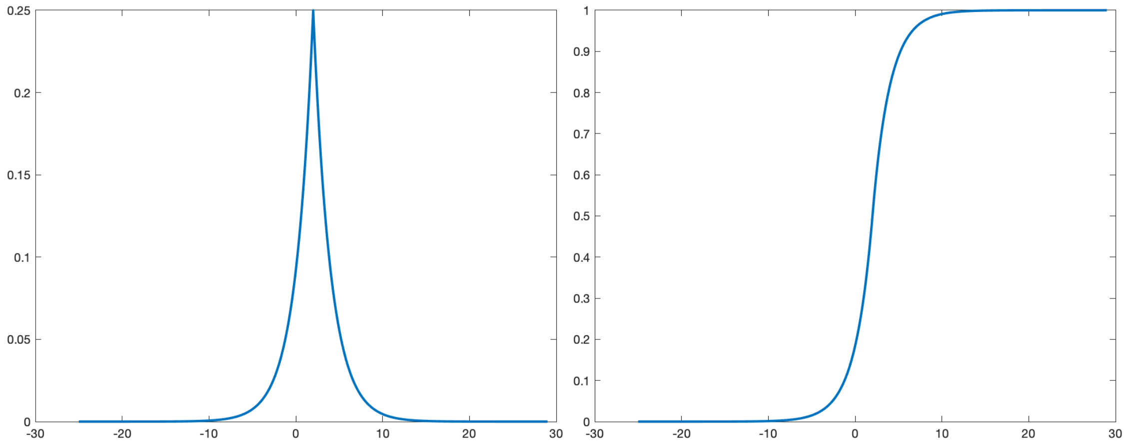


Figure 1: Laplace PDF and CDF

True Variance $= 2b^2 = 8$
Calculated Variance $= 7.998847$
Error $= -1.153 \times 10^{-3}$

1.2 Gumbel Distribution

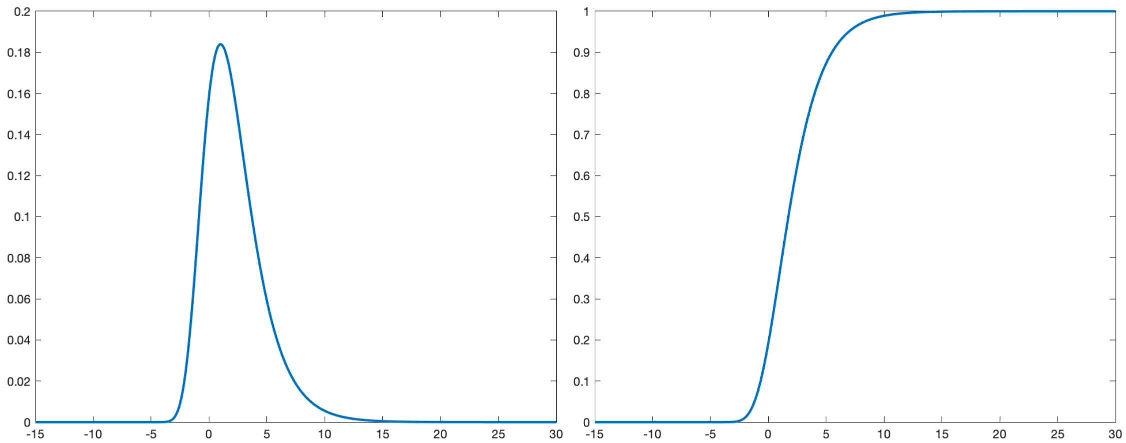


Figure 2: Gumbel PDF and CDF

True Variance $= \frac{\pi^2}{6}\beta^2 = 6.579736$
Calculated Variance $= 6.579287$
Error $= -4.493 \times 10^{-4}$

1.3 Cauchy Distribution

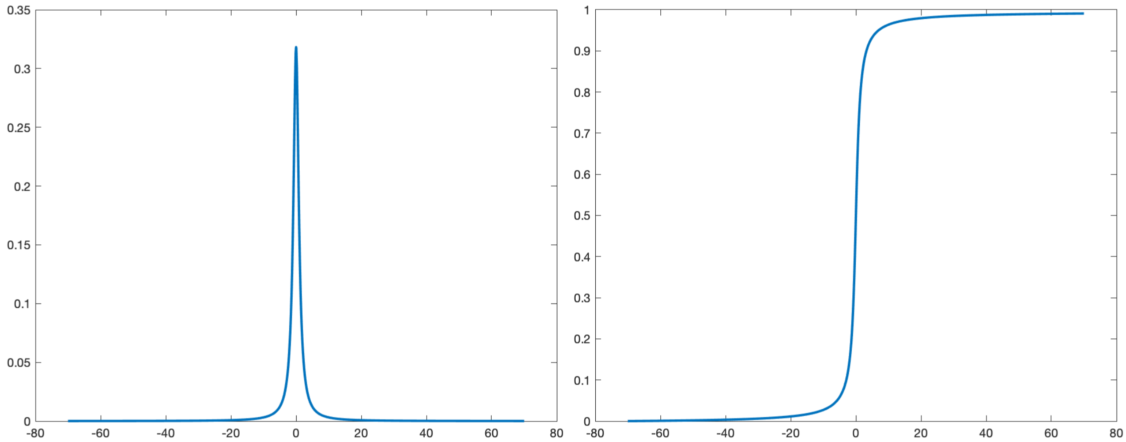


Figure 3: Cauchy PDF and CDF

True Variance $=$ Undefined

2 Poisson Distribution

2.1 Sum of Independent Poisson Random Variables

2.1.1 Analytical Distribution

As $Z = X + Y$, and X and Y are Poisson Distributions with parameters λ_1 and λ_2, we can write -

$$\begin{aligned} P(Z=k) &= \sum_{i=0}^{k} P(X=i)\, P(Y=k-i) \\ &= \sum_{i=0}^{k} \frac{e^{-\lambda_1}\lambda_1^i}{i!} \cdot \frac{e^{-\lambda_2}\lambda_2^{k-i}}{(k-i)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda_1^i \lambda_2^{k-i} \\ &= \frac{e^{-(\lambda_1+\lambda_2)} \cdot (\lambda_1+\lambda_2)^k}{k!} \end{aligned}$$

In this case, it was given that $\lambda_1 = 3$ and $\lambda_2 = 4$,

$$P(Z=k) = \frac{e^{-7}(7)^k}{k!} \tag{1}$$

2.1.2 Comparing Values

Finding expected values by multiplying probability with number of trials (=10^6) -

k	$P(Z=k)$	$\hat{P}(Z=k)$	k	$P(Z=k)$	$\hat{P}(Z=k)$
0	911	924	13	14188	14300
1	6383	6432	14	7094	7171
2	22341	22408	15	3310	3315
3	52129	52426	16	1448	1488
4	91226	91317	17	596	586
5	127716	127438	18	232	232
6	149002	148774	19	85	94
7	149002	148920	20	30	14
8	130377	130387	21	10	8
9	101404	101294	22	3	6
10	70983	71001	23	1	0
11	45171	45074	24	0	0
12	26350	26391	25	0	0

Table 1: Theoretical vs Empirical Values

2.2 Poisson Thinning

2.2.1 Analytical Distribution

As Z is obtained by Poisson thinning of Random Variable Y, with probability parameter p, we can write -

$$\begin{aligned}
P(Z=k) &= \sum_{i=k}^{\infty} P(Y=i, Z=k) \\
&= \sum_{i=k}^{\infty} P(Y=i) \cdot P(Z=k \mid Y=i) \\
&= \sum_{i=k}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{i!} \cdot \binom{i}{k} p^{k}(1-p)^{i-k} \\
&= \frac{e^{-(\lambda)}(\lambda p)^{k}}{k!} \sum_{i=k}^{\infty} \frac{(\lambda(1-p))^{j-k}}{(j-k)!} \\
&= \frac{e^{-(\lambda p)}(\lambda p)^{k}}{k!}
\end{aligned}$$

In this case, it was given that $p = 0.8$, therefore -

$$P(Z=k) = \frac{e^{-3.2}(3.2)^{k}}{k!} \tag{2}$$

2.2.2 Comparing Values

Finding expected values by multiplying probability with number of trials (=100000), found by combining Poisson with Binomial Distribution in MATLAB Code -

k	$P(Z=k)$	$\hat{P}(Z=k)$	k	$P(Z=k)$	$\hat{P}(Z=k)$
0	4076	4092	13	2	0
1	13044	13070	14	1	0
2	20870	20862	15	0	1
3	22262	22238	16	0	0
4	17809	17962	17	0	0
5	11398	11372	18	0	0
6	6079	6097	19	0	0
7	2779	2744	20	0	0
8	1112	1039	21	0	0
9	395	359	22	0	0
10	126	124	23	0	0
11	37	30	24	0	0
12	10	10	25	0	0

Table 2: Theoretical vs Empirical Values

3 Independant Random Walkers

3.1 Final Locations

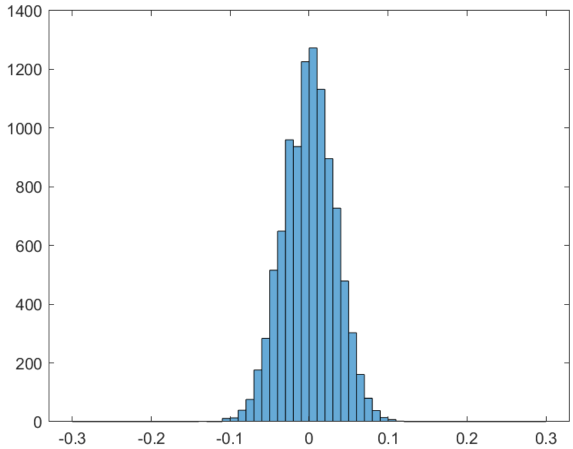


Figure 4: Histogram of final locations of 10^4 random walkers

3.2 Space-Time Curves

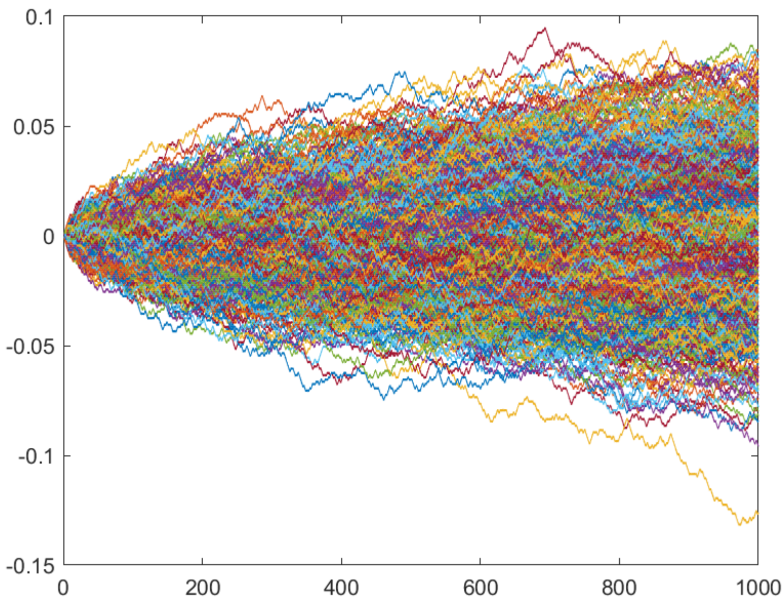


Figure 5: Space-Time Curves of 10^3 random walkers

3.3 Proofs

Let us consider the variance of $\hat{M}$:

$$\begin{aligned}
Var(\hat{M}) &= Var\left(\frac{(X_1 + \cdots + X_N)}{N}\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} Var\left(X_i\right) \\
&= \frac{1}{N^2} N \cdot Var(X) \\
&= \frac{Var(X)}{N}
\end{aligned}$$

Clearly, we can see that in the limit $N \longrightarrow \infty$, $Var(\hat{M}) \longrightarrow 0$. Therefore, $\hat{M}$ converges to $E[\hat{M}]$.

$$\begin{aligned}
E[\hat{M}] &= E\left[\frac{X_1 + \cdots + X_N}{N}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E[X_i] \\
&= \frac{1}{N} N \cdot E[X] \\
&= E[X]
\end{aligned}$$

Hence, in the limit $N \longrightarrow \infty$, $\hat{M} \longrightarrow E[X]$. **QED**

For the second part,

$$\begin{aligned}
E[\hat{V}] &= E\left[\sum_{i=1}^{N} \frac{(X_i - \hat{M})^2]}{N}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[X_i^2 - 2\hat{M}X_i + \hat{M}^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E[X_i^2] - \frac{2}{N}\sum_{i=1}^{N} E[\hat{M}X_i] + \frac{1}{N}\sum_{i=1}^{N} E[\hat{M}^2]
\end{aligned}$$

Analysing each of the terms individually,

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} E[X_i^2] &= \frac{1}{N} N \cdot E[X^2] \\
&= E[X^2] \\
&= Var(X) + E[X]^2 \\
&= Var(X) + M^2
\end{aligned}$$

$$
\begin{aligned}
E[\hat{M}X_i] &= E\left[\frac{X_1X_i + \cdots + X_i^2 + \cdots + X_NX_i}{N}\right] \\
&= \frac{1}{N}\left(E[X_i^2] + \sum_{j=1, j\neq i}^{N} E[X_jX_i]\right) \\
&= \frac{1}{N}\left(Var(X) + M^2 + (N-1)E[X]^2\right) \\
&= \frac{1}{N}\left(Var(X) + M^2 + (N-1)M^2\right) \\
&= M^2 + \frac{Var(X)}{N}
\end{aligned}
$$

$$
\begin{aligned}
E[\hat{M}^2] &= Var(\hat{M}) + E[\hat{M}]^2 \\
&= \frac{Var(X)}{N} + M^2 \qquad \text{(Proved Previously)}
\end{aligned}
$$

Putting these values back -

$$
\begin{aligned}
E[\hat{V}] &= Var(X) + M^2 - \frac{2}{N}N \cdot \left(\frac{Var(X)}{N} + M^2\right) + \frac{1}{N}N \cdot \left(\frac{Var(X)}{N} + M^2\right) \\
&= Var(X) + M^2 - 2\frac{Var(X)}{N} - 2M^2 + \frac{Var(X)}{N} + M^2 \\
&= \frac{N-1}{N}\,Var(X)
\end{aligned}
$$

Therefore as $N \longrightarrow \infty$, $E[\hat{V}] \longrightarrow Var(X)$. **QED**

3.4 Empirical Values and Errors

Empirically Computed Mean $(\hat{M}) = -1.386000 \times 10^{-4}$
Empirically Computed Variance $(\hat{V}) = 1.004869 \times 10^{-3}$

Since each step taken is independent of each other, and probability of moving left or right is equal, the number of steps taken to the right can be modeled as a binomial random variable with $n = 1000$ and $p = 0.5$ (Let this be represented by Y)

We know that if Y is a binomial random variable with probability = p and number of trials = n,

$$
\begin{aligned}
E[Y] &= np \\
Var(Y) &= npq
\end{aligned}
$$

Also, we can see that -

$$
\begin{aligned}
X &= 0.001(Y) + (-0.001)(1000 - Y) \\
&= \frac{Y}{500} - 1
\end{aligned}
$$

Therefore,

$$\begin{aligned} E[X] &= \frac{E[Y]}{500} - 1 \\ &= \frac{np}{500} - 1 \\ &= \frac{1000 \cdot 0.5}{500} - 1 = 0 \end{aligned}$$

$$\begin{aligned} Var(X) &= Var\left(\frac{Y}{500} - 1\right) \\ &= Var\left(\frac{Y}{500}\right) \\ &= \frac{Var(Y)}{500^2} = \frac{npq}{500^2} \\ &= \frac{1000 \cdot 0.5 \cdot 0.5}{500^2} = 0.001 \end{aligned}$$

Error in Mean $= \hat{M} - M = -1.386 \times 10^{-4}$
Error in Variance $= \hat{V} - V = 4.869 \times 10^{-6}$

4 M-shaped Distribution Function

4.1 Inverse Transform Sampling

It is a method to generate random values that fit a certain probability distribution.

Let $P(X)$ be the probability distribution and $\mathbf{F}(x) = P(X \le x)$ be the corresponding cumulative distributive function. Let us call U to be the uniformly distributed random variable from 0 to 1.

So, $P(U \le x) = x \quad \forall\, x\, \epsilon\, [0, 1]$

Now,

$$\begin{aligned} P(X \le x) &= \mathbf{F}(x) \\ &= P(U \le \mathbf{F}(x)) \\ &= P(\mathbf{F}^{-1}(U) \le x) \qquad (\text{As } \mathbf{F}(x)\, \epsilon\, [0,1]\ ) \\ \implies X &= \mathbf{F}^{-1}(U) \end{aligned}$$

Therefore, we can find random values satisfying any probability function, using random values from a uniform distribution only.

4.2 Application

To solve the question, we first find the Cumulative Distribution Function of the given function. This can be done by integrating the probability distribution function, which gives the CDF -

$$\mathbf{F}(x) = \begin{cases} 0 & x < -1 \\ \frac{1-x^2}{2} & -1 \le x < 0 \\ \frac{1+x^2}{2} & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

Now, we have to find $\mathbf{G} = \mathbf{F}^{-1}(x)$, which is done by taking the piece wise inverse of $\mathbf{F}(x)$ -

$$\mathbf{G}(x) = \begin{cases} -\sqrt{1-2x} & 0 \le x < \frac{1}{2} \\ \sqrt{2x-1} & \frac{1}{2} \le x \le 1 \end{cases}$$

This function is used in the code, along with a random number generator, to give numbers corresponding to the given probability distribution.

4.3 Histogram and Cumulative Distribution Function

The constructed function is used to make 10^5 draws to acquire this Histogram and CDF plot.

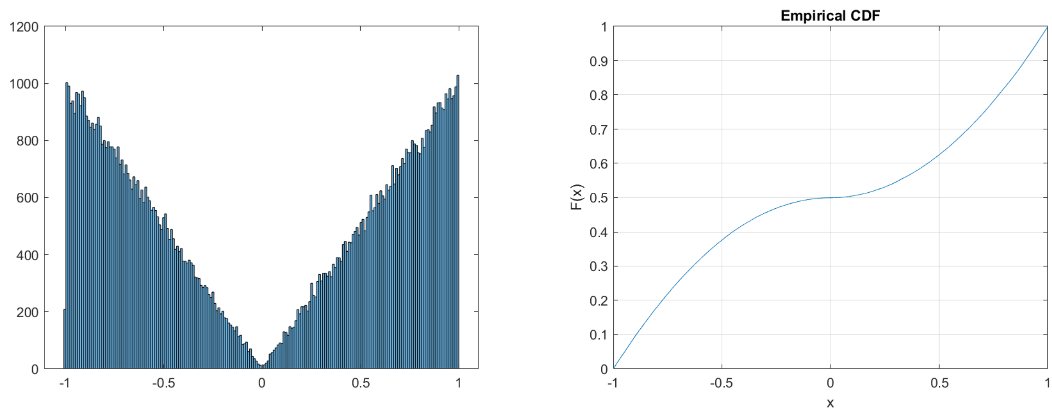


Figure 6: Histogram(200 bins) and CDF

4.4 Sample Mean as a Random Variable

In this section, we discuss distribution of sample mean. A function is defined which generate random draws of average of N draws of M-shaped Distribution. 10^4 draws were made to generate the histogram and CDF plot for various N.

$$\mathbf{Y_N} := \frac{1}{N}\sum_{i=1}^{N}\mathbf{X_i}$$

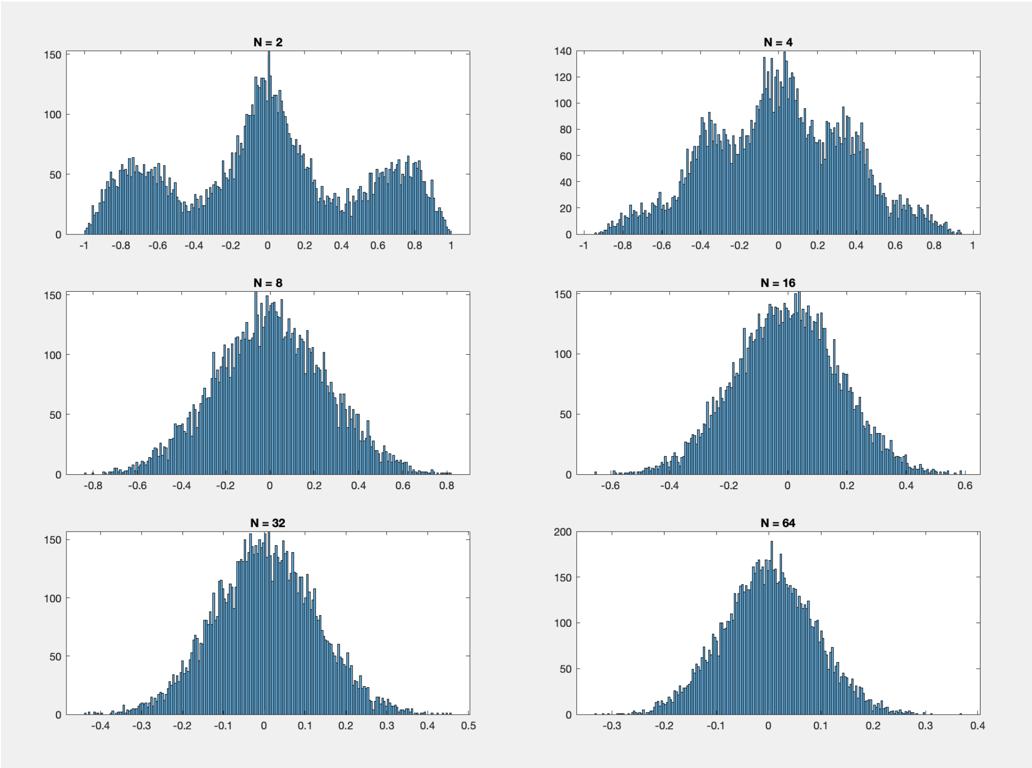


Figure 7: Histograms(200 bins) for various N

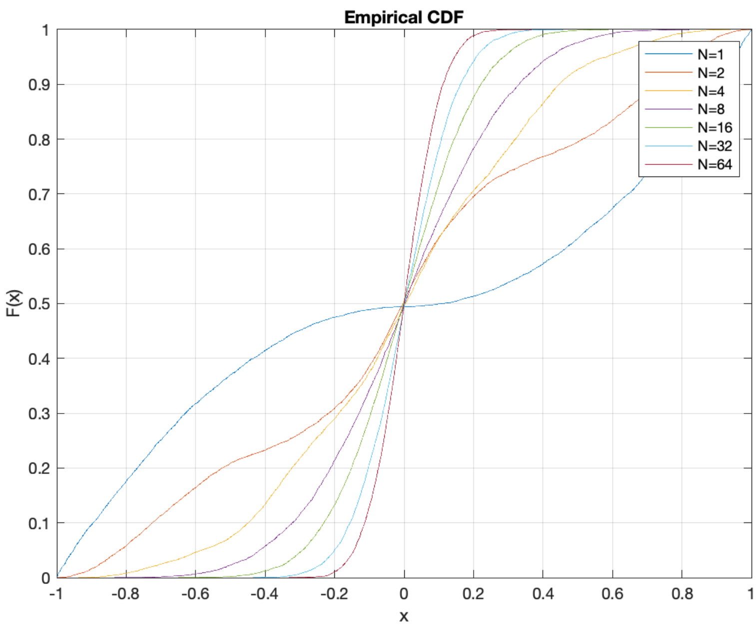


Figure 8: CDFs for various N

5 Convergence of Mean

In the box and whisker plot, we see that as the value of N increases, average of the absolute value of the error becomes closer and closer to 0.

In addition to the empirical mean approaching the true mean, we also see that the variance or "the spread" of the trials also decreases.

Let us try to formalize these statements -

Assuming A to be the random variable whose trials we are taking (Uniform Distribution in Case 1, and Gaussian Distribution in Case 2)

The Random Variable we are plotting is the absolute error of the empirical average and the true average over N trials, i.e. if we define -

$$Y = \frac{A_1 + \cdots + A_N}{N}$$

where $A, A_1, \ldots, A_N$ are independent and identically distributed random variables.
Then the random variable under consideration is -

$$X = |Y - E[Y]|$$

We see that -

$$\begin{aligned}
E[Y] &= E\left[\frac{A_1 + \cdots + A_N}{N}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E[A_i] \\
&= E[A]
\end{aligned}$$

$$\begin{aligned}
Var(Y) &= Var\left[\frac{A_1 + \cdots + A_N}{N}\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} Var(A_i) \\
&= \frac{Var(A)}{N}
\end{aligned}$$

Jensen's Inequality for concave functions states that -

$$E[f(Z)] \leq f(E[Z])$$

Taking $Z = (Y - E[Y])^2$ and $f(x) = \sqrt{x}$ (concave function), we get -

$$\begin{aligned} E[\sqrt{(Y - E[Y])^2}\,] &\leq \sqrt{E[(Y - E[Y])^2]} \\ \implies E[X] &\leq \sqrt{Var(Y)} \\ \implies E[X] &\leq \sqrt{\frac{Var(A)}{N}} \end{aligned}$$

As $E[X] \geq 0$,

$$0 \leq E[X] \leq \sqrt{\frac{Var(A)}{N}}$$

Also,

$$\begin{aligned} Var(X) &= E[|Y - E[Y]|^2] - E[X]^2 \\ &= Var(Y) - E[X]^2 \end{aligned}$$

Using bounds on $E[X]$ derived earlier,

$$\begin{aligned} Var(Y) - \left(\sqrt{\frac{Var(A)}{N}}\right)^2 &\leq Var(X) \leq Var(Y) - (0)^2 \\ \implies Var(Y) - \frac{Var(A)}{N} &\leq Var(X) \leq Var(Y) \\ \implies \qquad & 0 \leq Var(X) \leq Var(Y) \\ \implies \qquad & 0 \leq Var(X) \leq \frac{Var(A)}{N} \end{aligned}$$

Since $Var(A)$ is finite for both Uniform and Gaussian Distribution,

$$\begin{aligned} \lim_{N\to\infty} Var(Y) &= \lim_{N\to\infty} \frac{Var(A)}{N} \\ &= 0 \end{aligned}$$

So, by Sandwich Theorem,

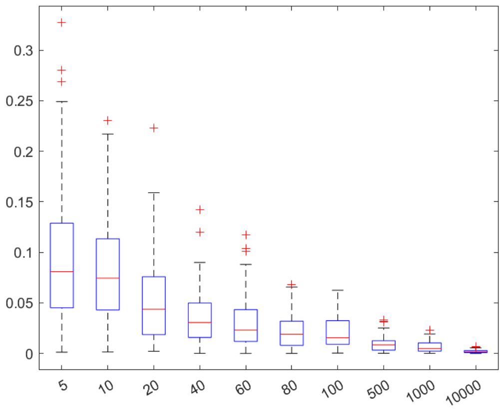
$$\begin{aligned} \lim_{N\to\infty} E[X] &= 0 \\ \lim_{N\to\infty} Var(X) &= 0 \end{aligned}$$

And in the box plot we see the same result, that as the value of N increases the empirical mean of the absolute error as well as the variance of the absolute error, both tend to 0.

Also,

Variance of the Uniform Distribution $= \int_0^1 (x - 0.5)^2 dx = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}$
Variance of Gaussian Distribution $= 1$

Since, the variance of the Uniform Distribution is less than the Gaussian, we see that the deviation from the expected value should be more in the case of Gaussian Distribution than in the case of Uniform Distribution, as is the case.



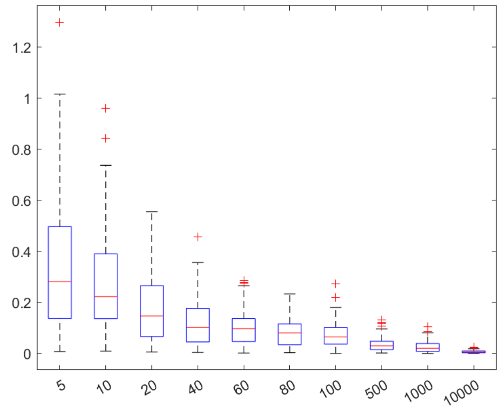


Figure 9: Box Plot for Uniform and Gaussian Distributions