CS-215 ASSIGNMENT-1

REPORT

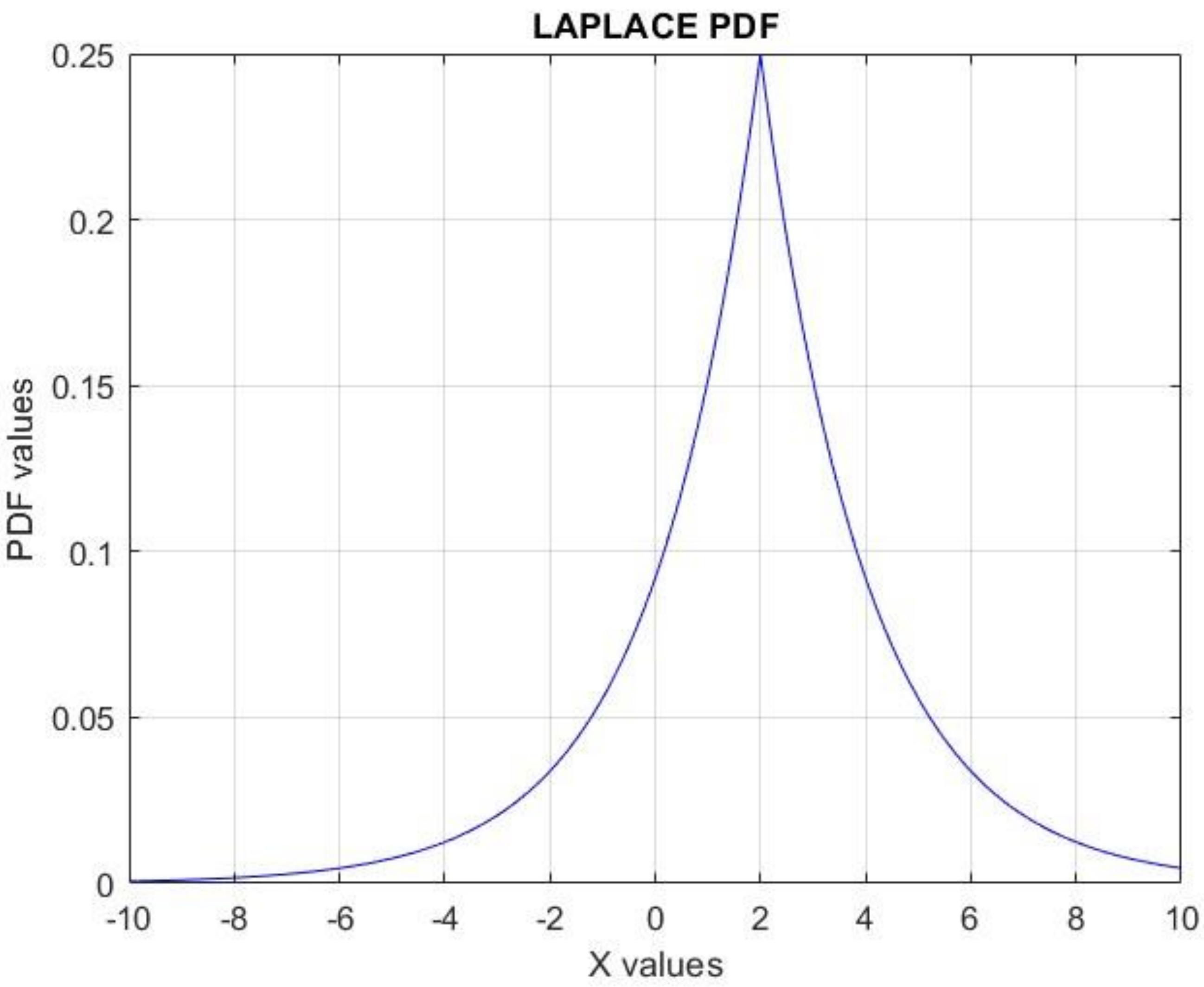
Shubham Hazra

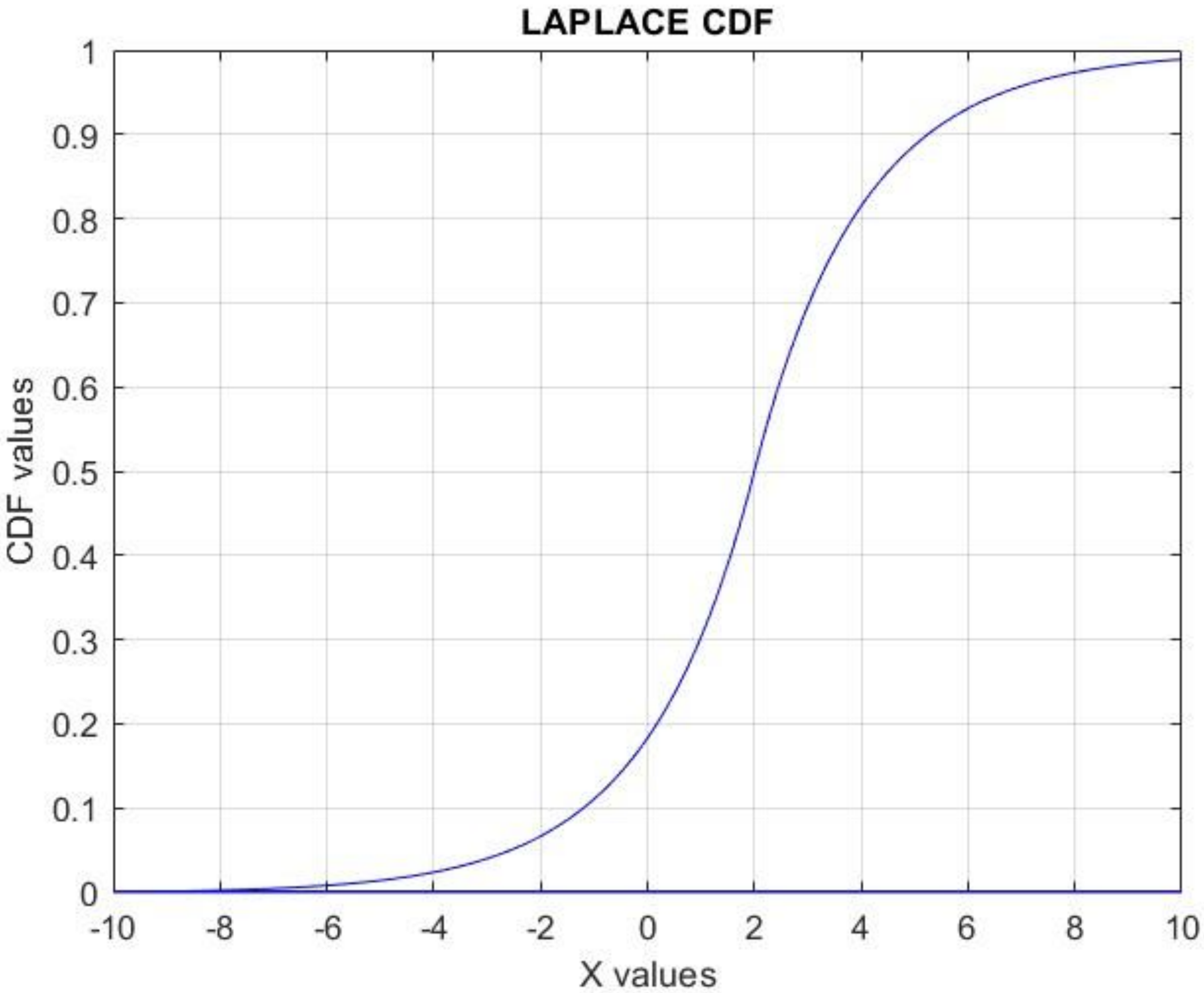
Om Godage

Q1. Plots of the PDF, CDF, and the variance of the following distributions are:

<u>LAPLACE DISTRIBUTION</u>

- Location Parameter = 2
- Scale Parameter = 2





- <u>VARIANCE</u>

The calculated variance using Reimann sums:

Calculated Variance = 7.9996

Whereas the theoretical variance:

Theoretical Variance = 8

Error:

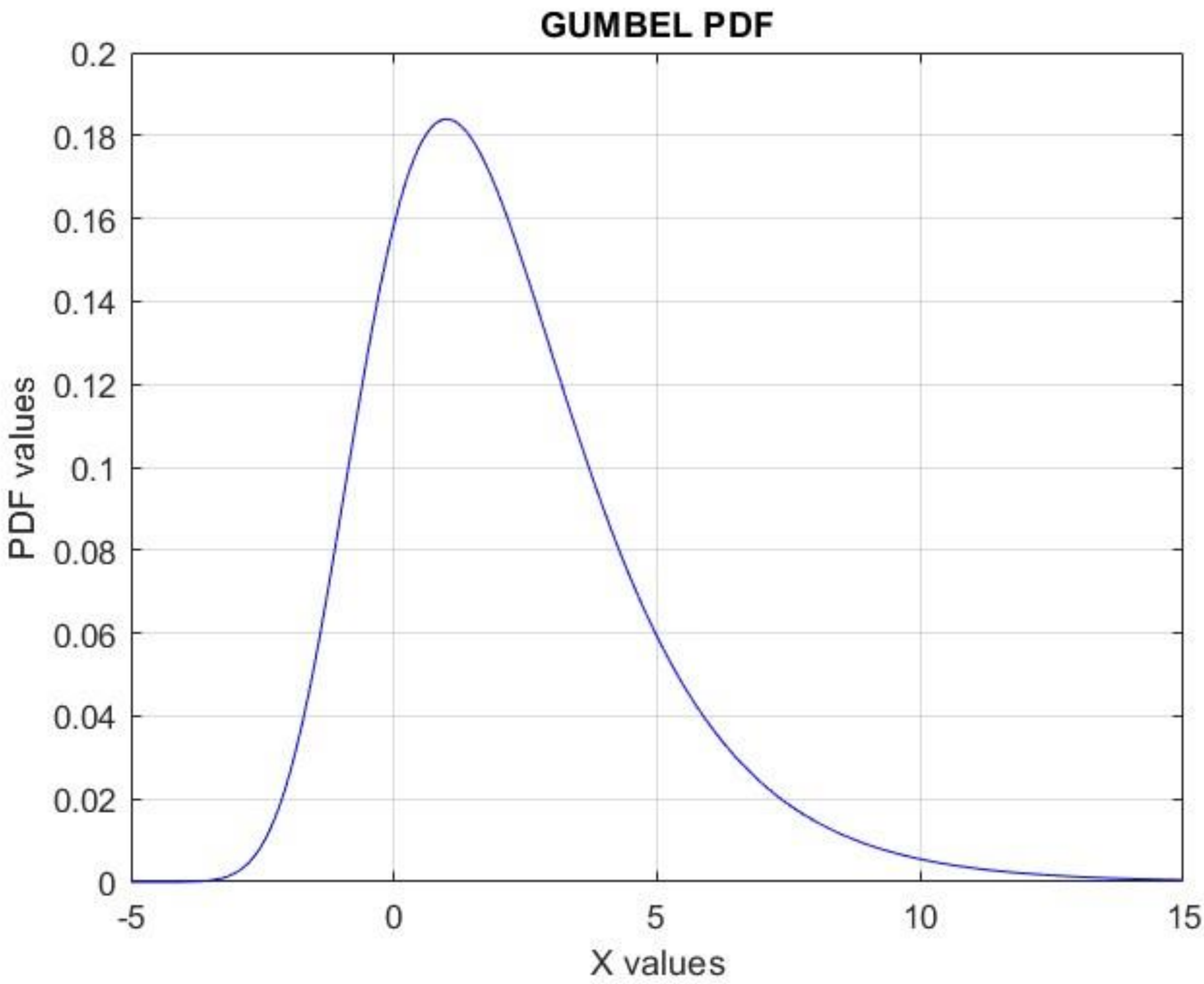
Error = Theoretical Variance – Calculated Variance

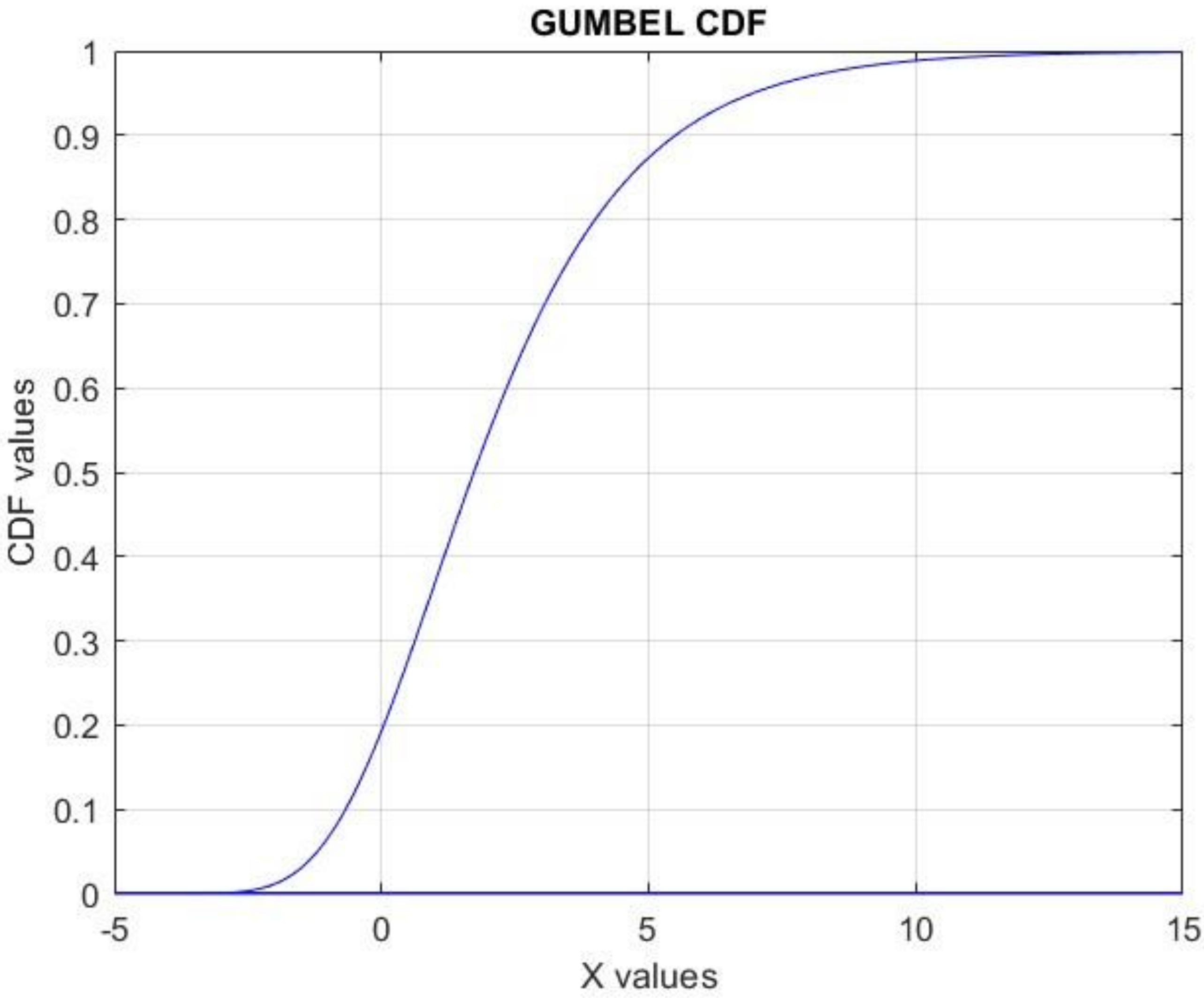
Error = 8 – 7.996

Error = 0.004

GUMBEL DISTRIBUTION

- Location Parameter = 1
- Scale Parameter = 2





- VARIANCE

The calculated variance using Reimann sums:

Calculated Variance = 6.5795

Whereas the theoretical variance:

Theoretical Variance = 6.5797

Error:

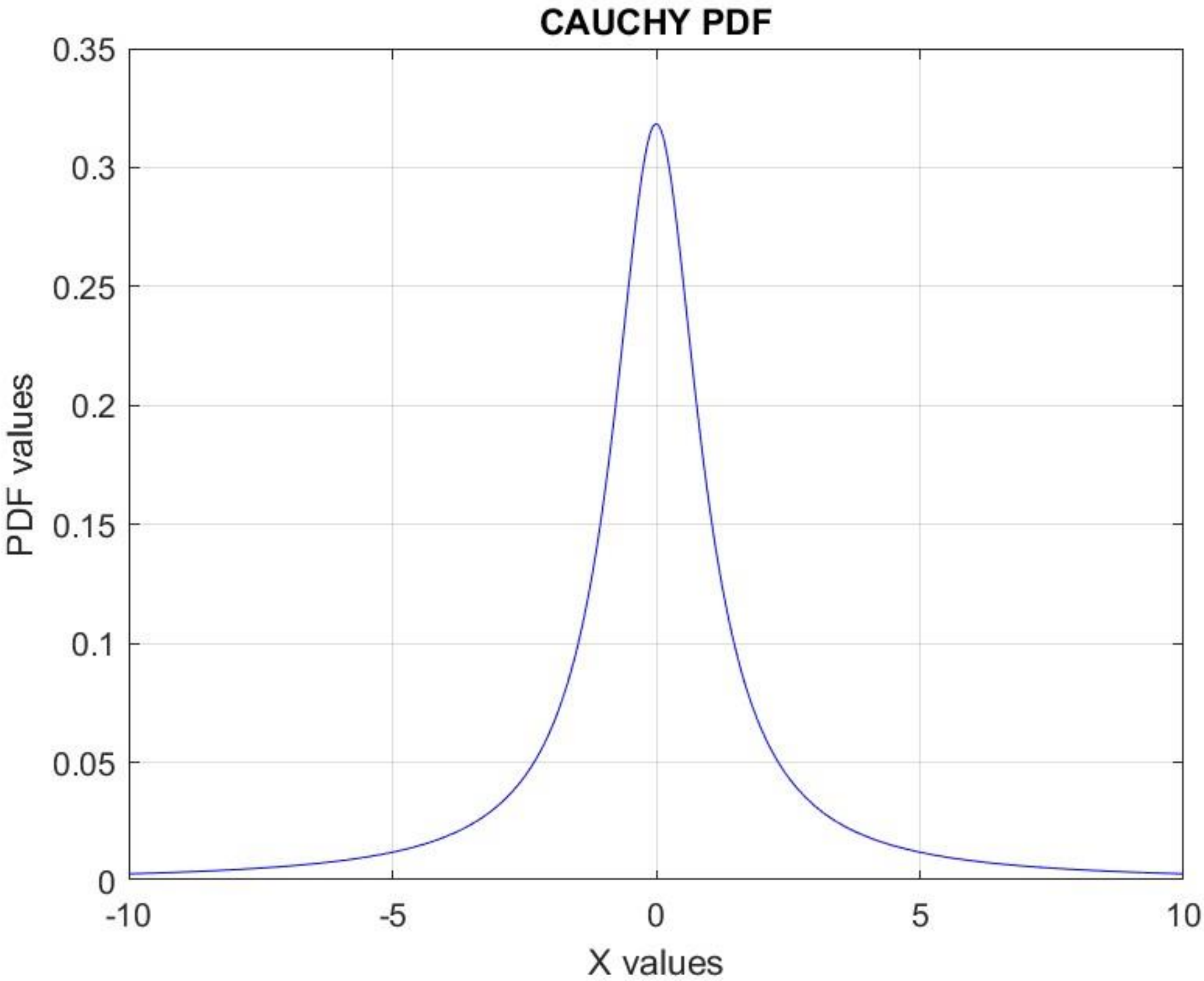
Error = Theoretical Variance – Calculated Variance

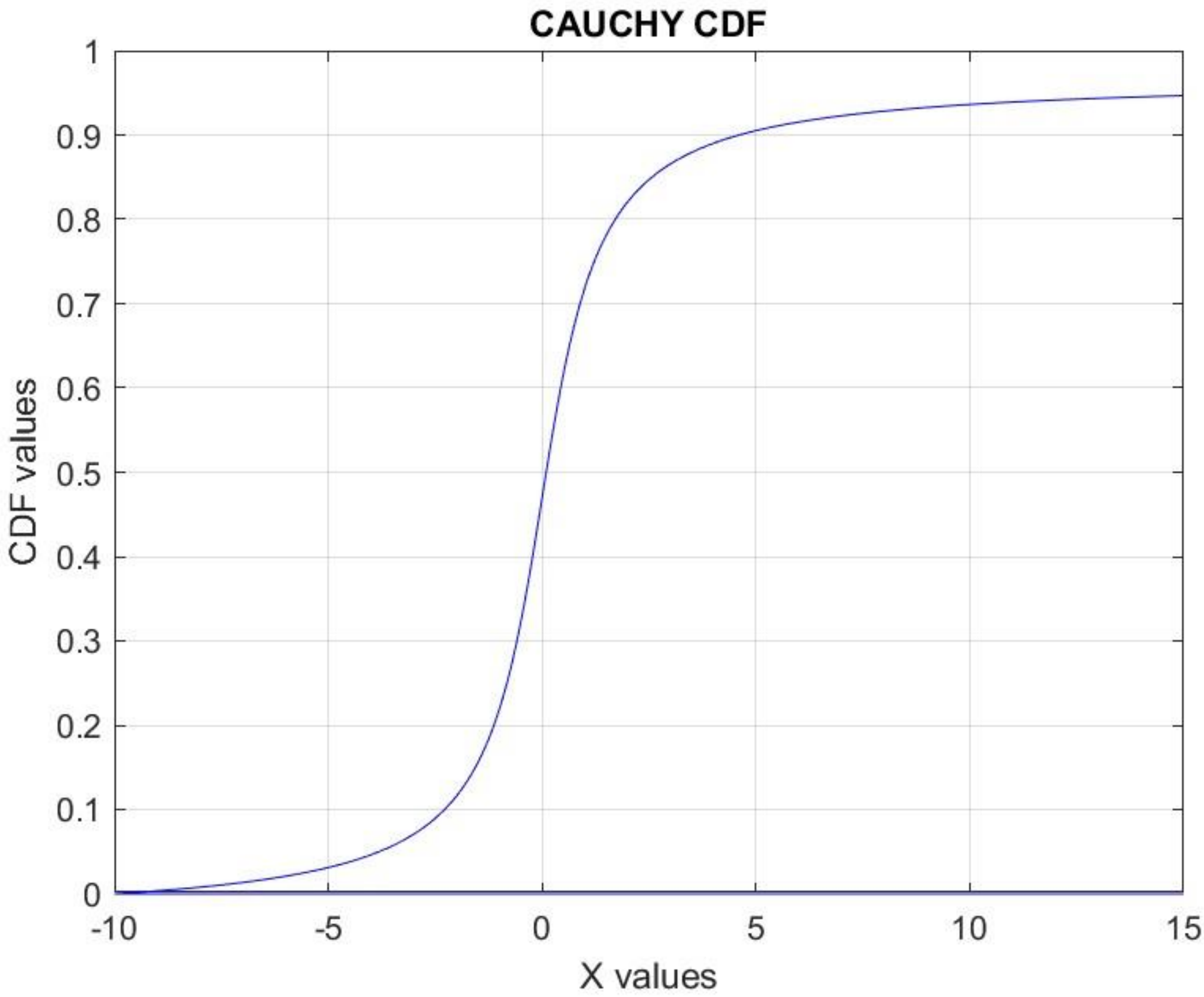
Error = 6.5797 – 6.5795

Error = 0.0002

CAUCHY DISTRIBUTION

- Location Parameter = 0
- Scale Parameter = 1





- <u>VARIANCE</u>

The calculated variance using Reimann sums:

Calculated Variance for different intervals:

Between -10 and 20 = 9.4493

Between -100 and 200 = 94.4492

Between -500 and 1000 = 471.2728

Between -1000 and 2000 = 940.1129

Clearly the variance is diverging and keeps increasing on increasing the intervals and hence the variance is undefined for this distribution.

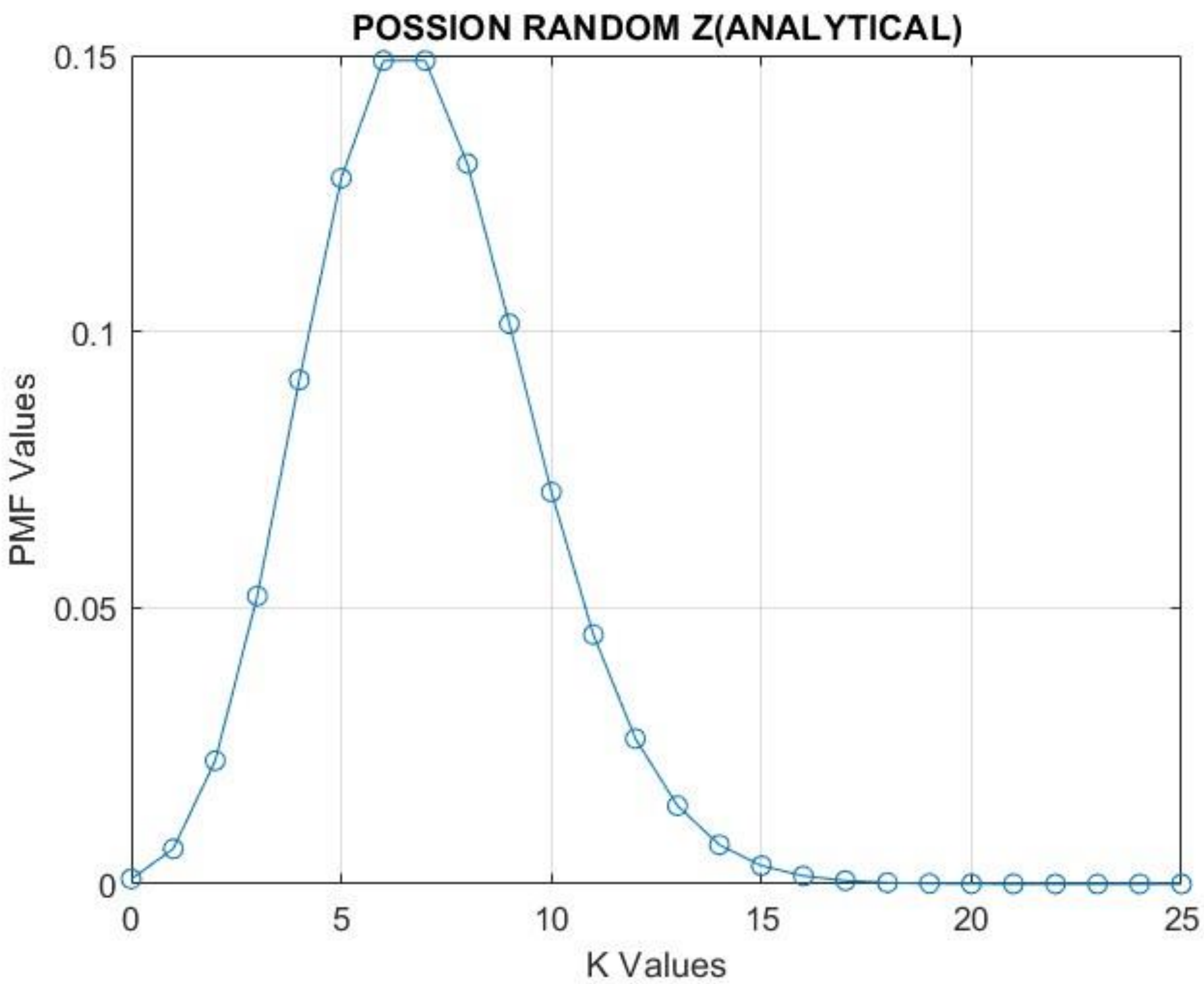
Q2. Two Poisson random variables – X and Y were defined

- Lambda X = 3
- Lambda Y = 4

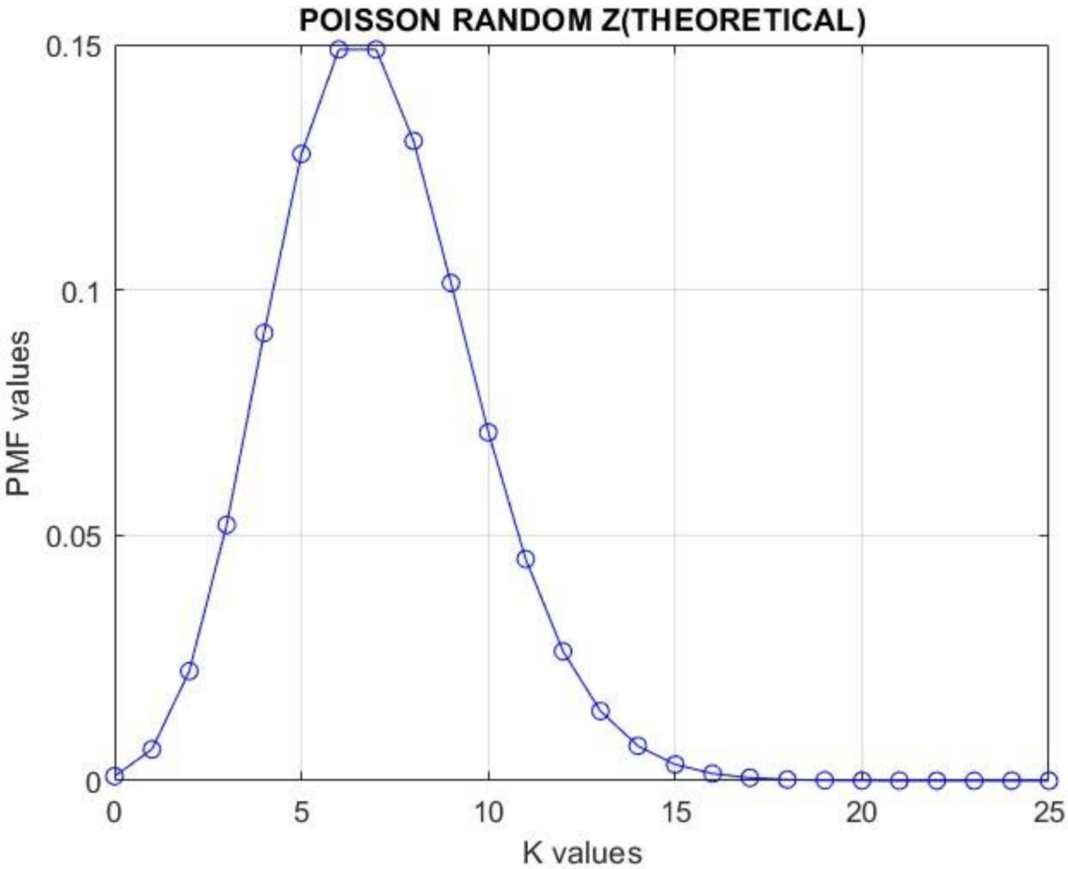
Another Poisson random variable Z was defined as :

Z = X + Y

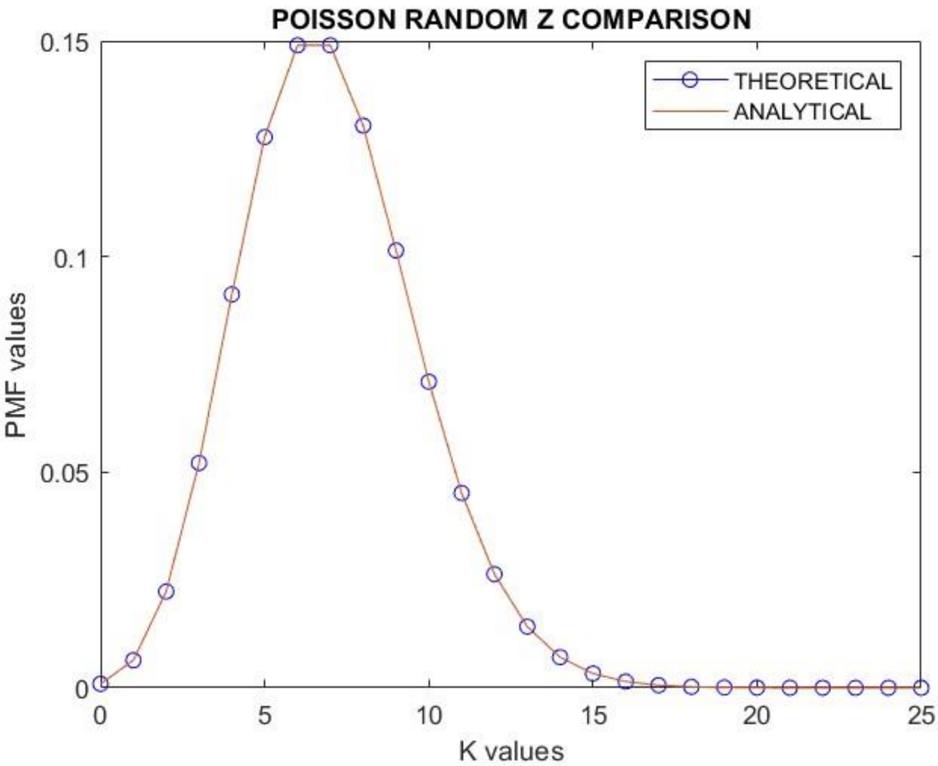
Plot of PMF of Z analytically:



Plot of PMF of Z theoretically (With lambda z = lambda x + lambda y):

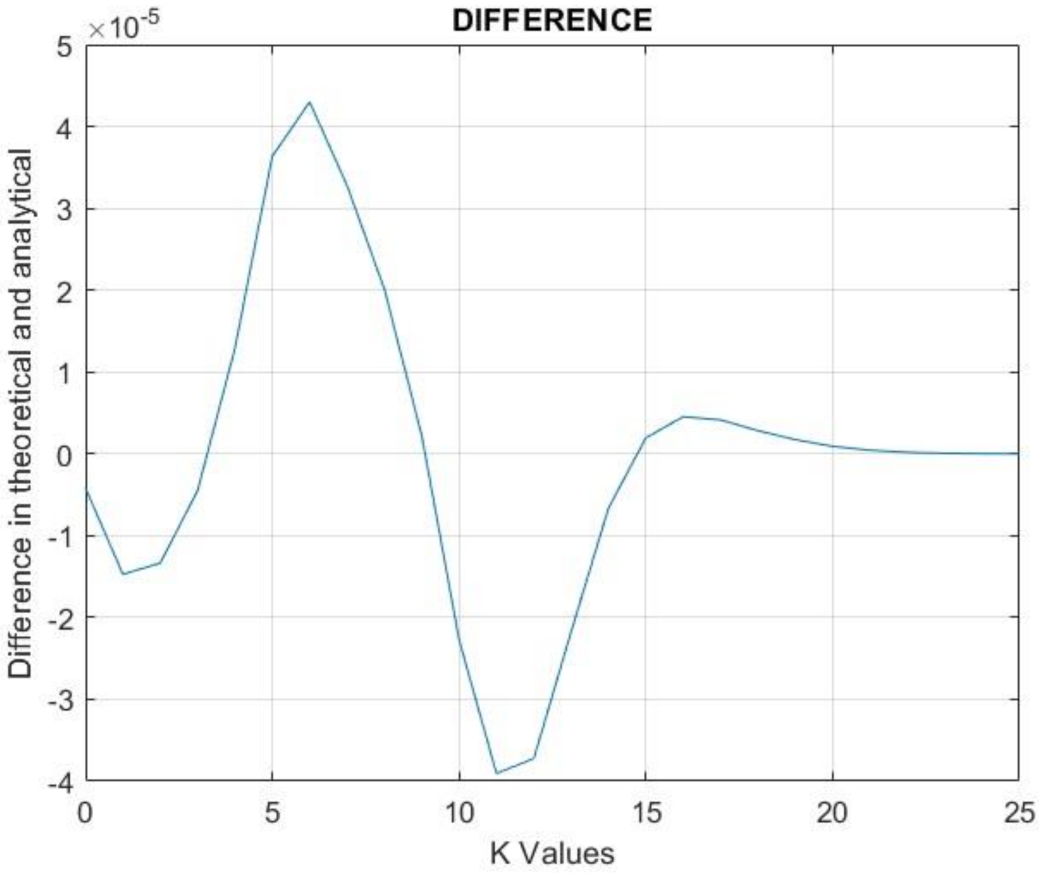


Both the plots on the same graph:



As we can clearly see both the plots are very close and it is quite hard to see the differences between the two. So, I made a plot of the differences between the values at each point.

Plot of the differences:



The values of Z analytically: Values of P(Z=k) for k = 0 to 25

Where value in Column k is equal to P(Z=k-1)

Columns 1 through 10

0.0009 0.0064 0.0223 0.0521 0.0912 0.1278 0.1490 0.1490 0.1304 0.1014

Columns 11 through 20

0.0710 0.0451 0.0263 0.0142 0.0071 0.0033 0.0015 0.0006 0.0002 0.0001

Columns 21 through 26

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

The values of Z theoretically: Values of P(Z=k) for k = 0 to 25

Where value in Column k is equal to P(Z=k-1)

Columns 1 through 10

0.0009 0.0064 0.0223 0.0521 0.0912 0.1277 0.1490 0.1490

0.1304 0.1014

Columns 11 through 20

0.0710 0.0452 0.0263 0.0142 0.0071 0.0033 0.0014 0.0006

0.0002 0.0001

Columns 21 through 26

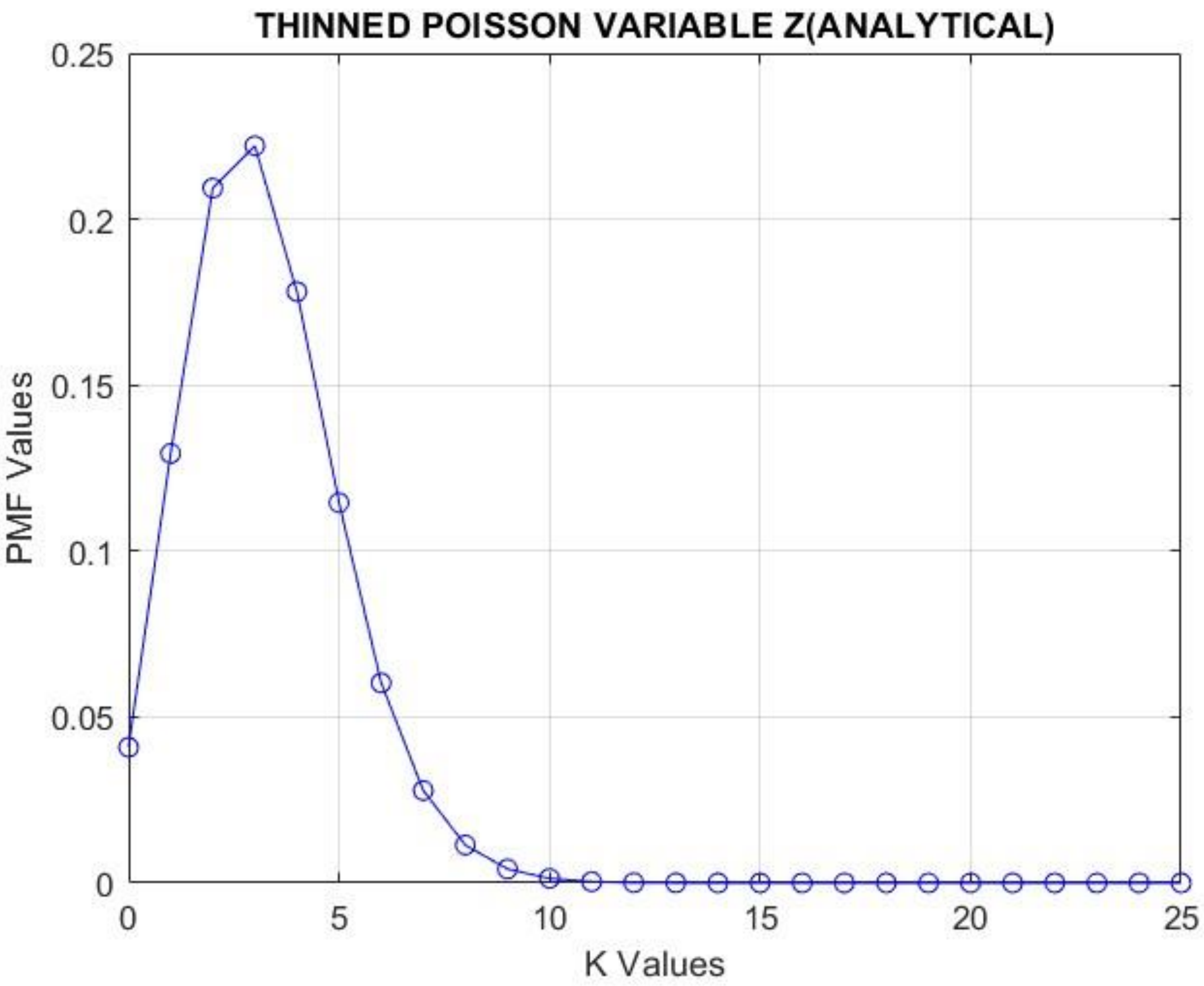
0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

As we can clearly observe, the values obtained from the theoretical and the analytical data is almost identical and the error is in the order of 10^{-5}.

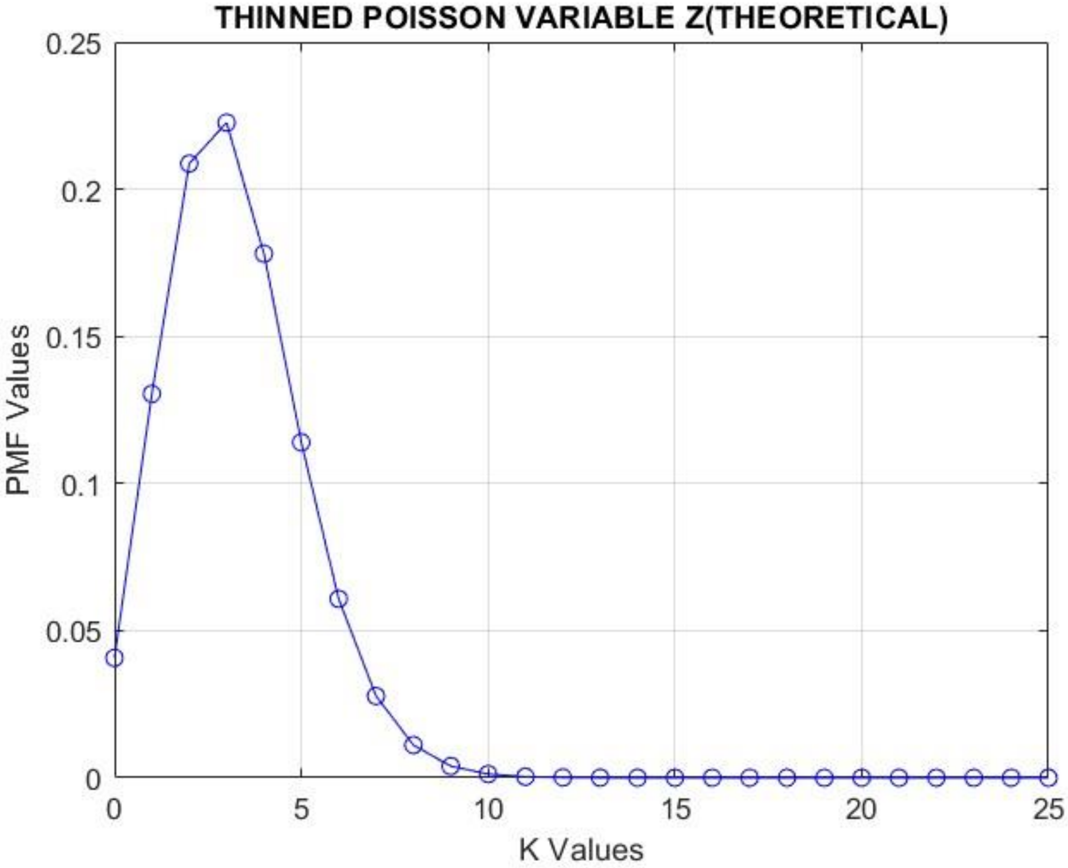
Now implementing a thinning process on the random variable Y

- Lambda Y = 4
- Probability Parameter = 0.8

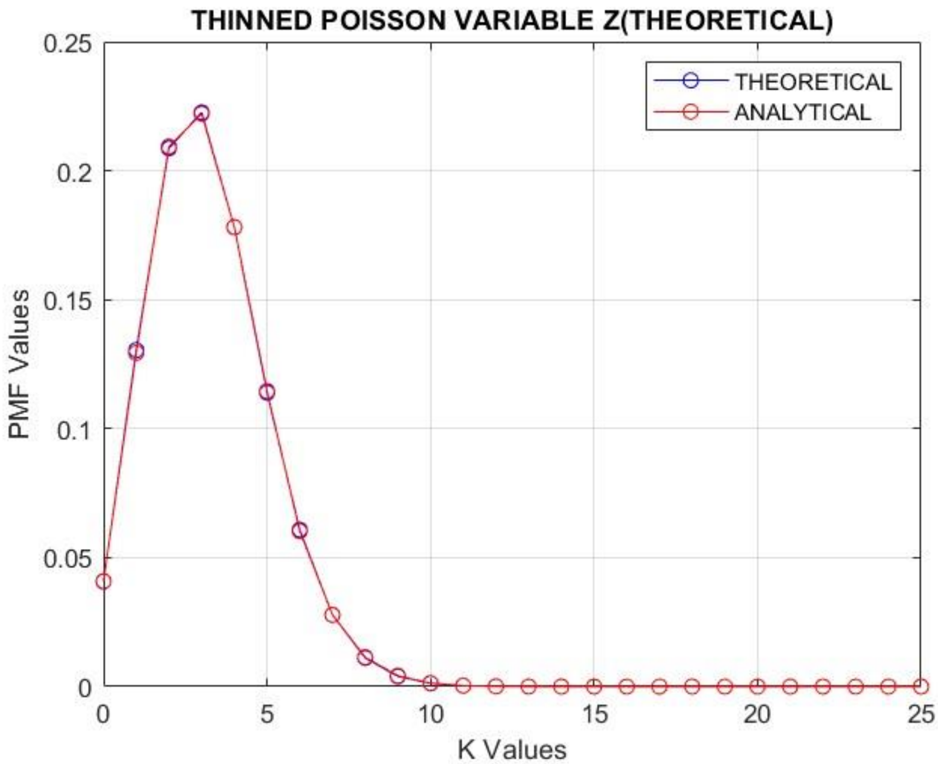
Plot of the thinned random variable analytically:



Plot of the thinned variable theoretically:

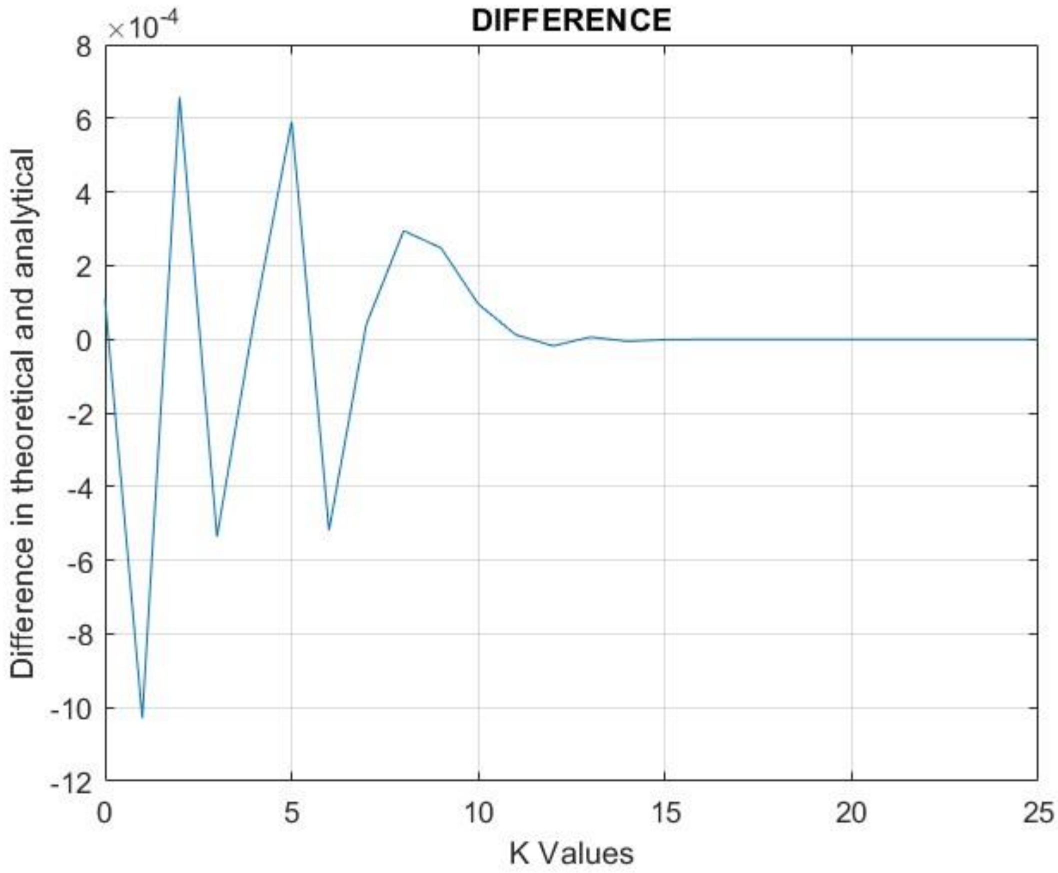


Comparing both the plots on the same graph:



As we can clearly see both the plots are very close and it is quite hard to see the differences between the two. So, I made a plot of the differences between the values at each point.

Plot of the differences:



The values of Z analytically: Values of P(Z=k) for k = 0 to 25

Where value in Column k is equal to P(Z=k-1)

Columns 1 through 9

0.0409 0.1294 0.2094 0.2221 0.1782 0.1146 0.0603 0.0278 0.0114

Columns 10 through 18

0.0042 0.0014 0.0004 0.0001 0.0000 0.0000 0.0000 0.0000 0.0000

Columns 19 through 26

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

The values of Z theoretically: Values of P(Z=k) for k = 0 to 25

Where value in Column k is equal to P(Z=k-1)

Columns 1 through 10

0.0408 0.1304 0.2087 0.2226 0.1781 0.1140 0.0608 0.0278 0.0111 0.0040

Columns 11 through 20

0.0013 0.0004 0.0001 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

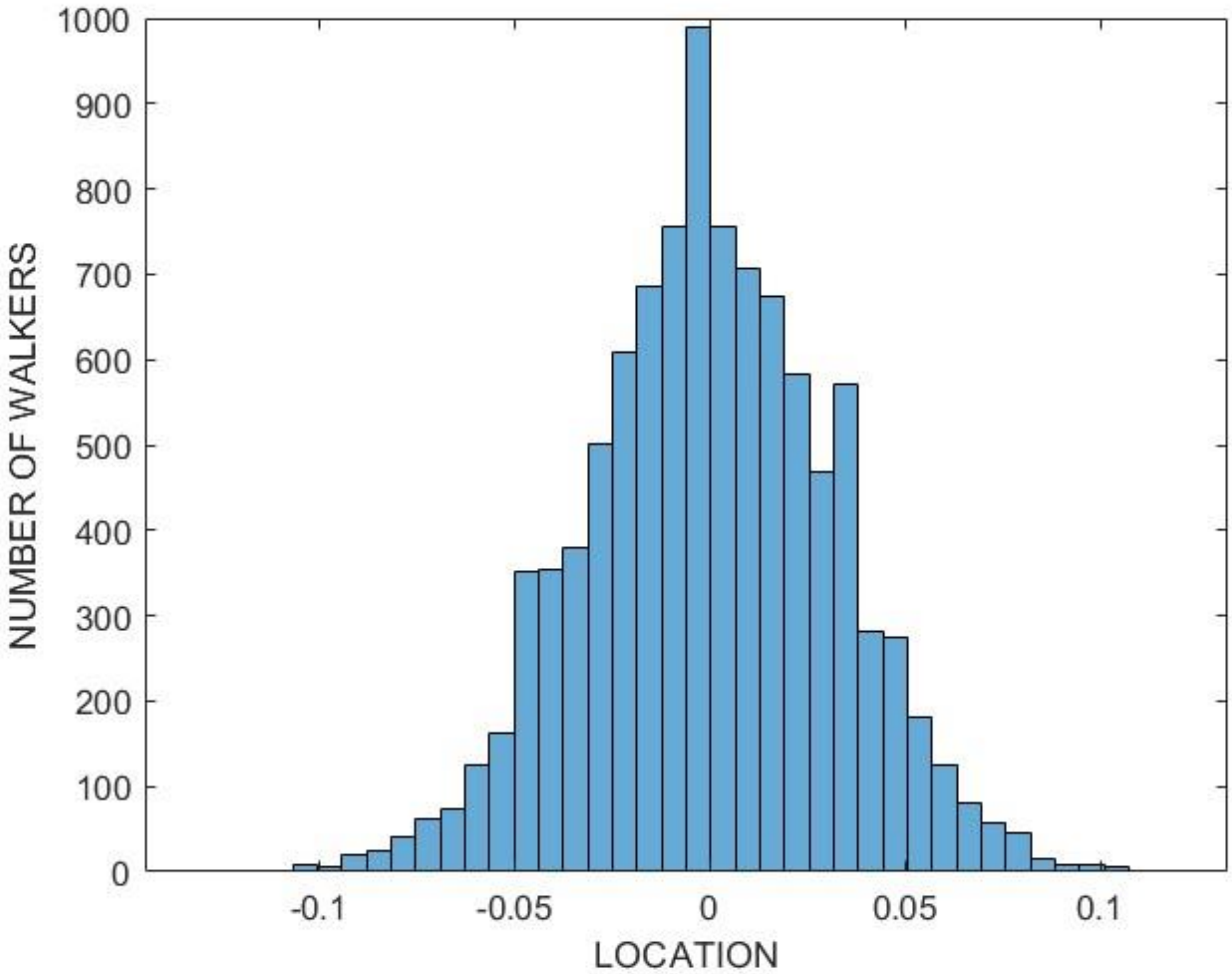
Columns 21 through 26

0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

As we can clearly observe, the values obtained from the theoretical and the analytical data is almost identical and the error is in the order of 10^{-4}.

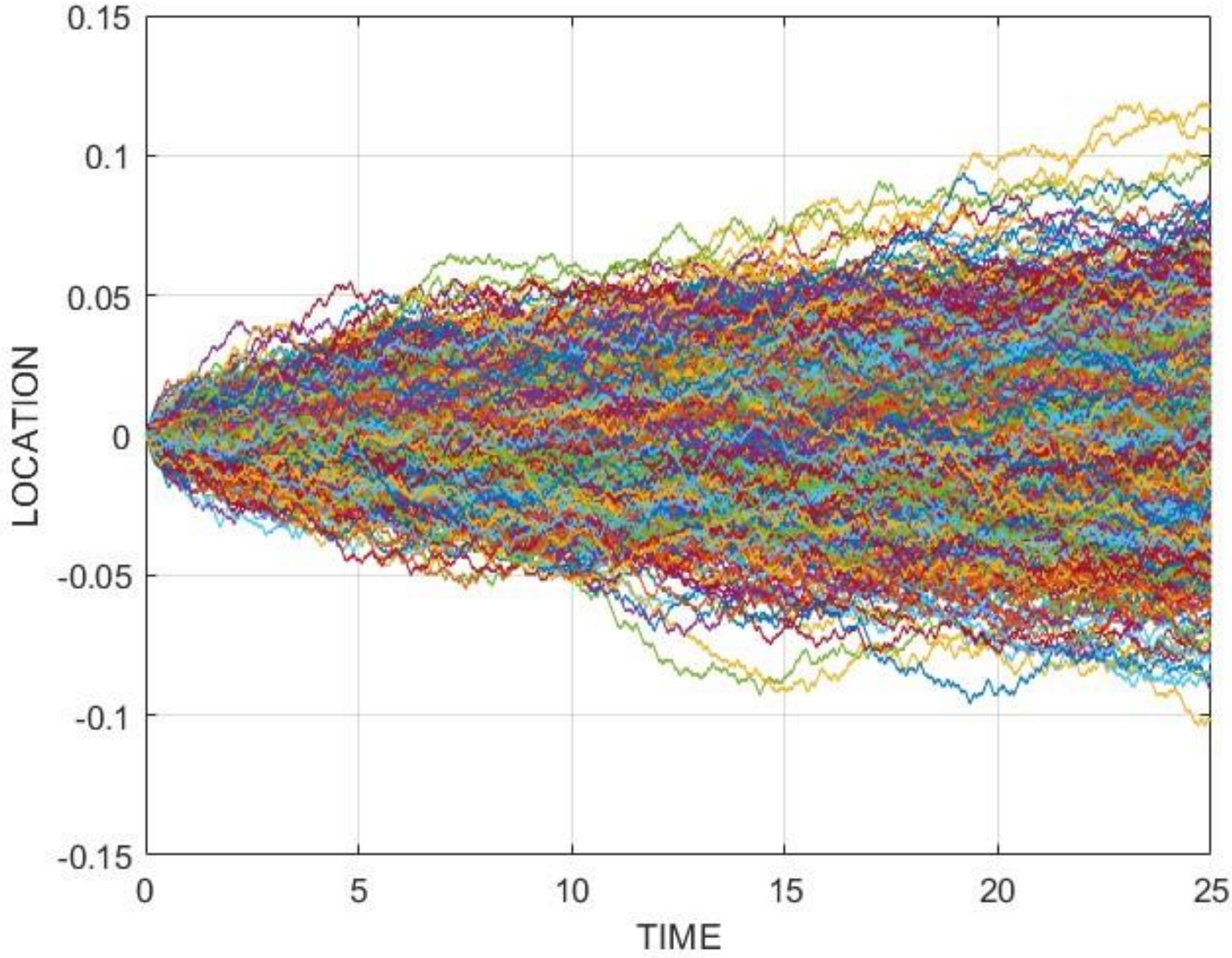
Q3. We had to simulate 10000 random walkers

Plot of the final location vs frequency of the people in that location:



We can clearly observe that the locations close to 0 are highly populated and very few people can make to very far from 0.

Plot of the location vs time graph for the first 1000 walkers:



We can again observe that the regions near 0 are highly populated and very few people reach regions far from 0.

As we know that for very large sets the random walkers tend to a Gaussian distribution and the true values of mean and variance can be found in terms of steps taken and step length by modelling the distribution of random walkers as a Gaussian distribution.

For such a Gaussian distribution:

True Mean = 0

True Variance = 2*D*T

Where D = (Step Length)^2/2

T = Steps Taken

In this case Step Length = 10^{-3} and Steps Taken = 10^3.

And hence,

True Variance = $2*(10^{-3})^2/2*10^3$

True Variance = 0.001

The Calculated Mean and Variance of the final location of the random walkers and its deviation from the true values are as follows:

True Mean = 0

True Variance = 0.001

Calculated Mean = -1.3860e-04

Calculated Variance = 0.001(Approximate)

Error in mean:

|Calculated Mean - True Mean| = 1.3860e-04

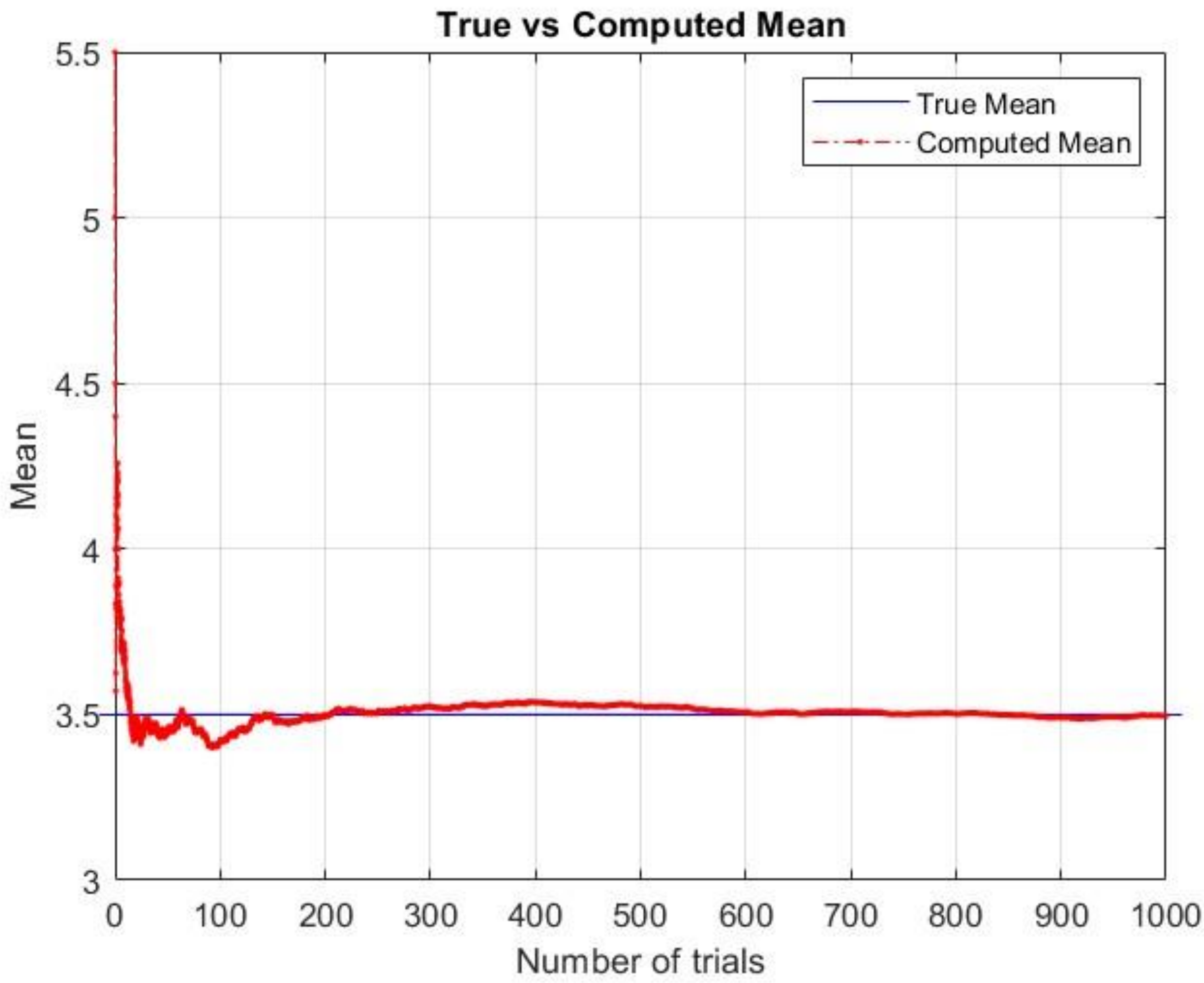
Error in variance:

|Calculated Variance - True Variance| = 4.7684e-06

To show law of large numbers. We considered the random variable X to be the number obtained on rolling a fair die.

The true mean of X = 3.5

Plot of the mean of X for increasing number of trials:



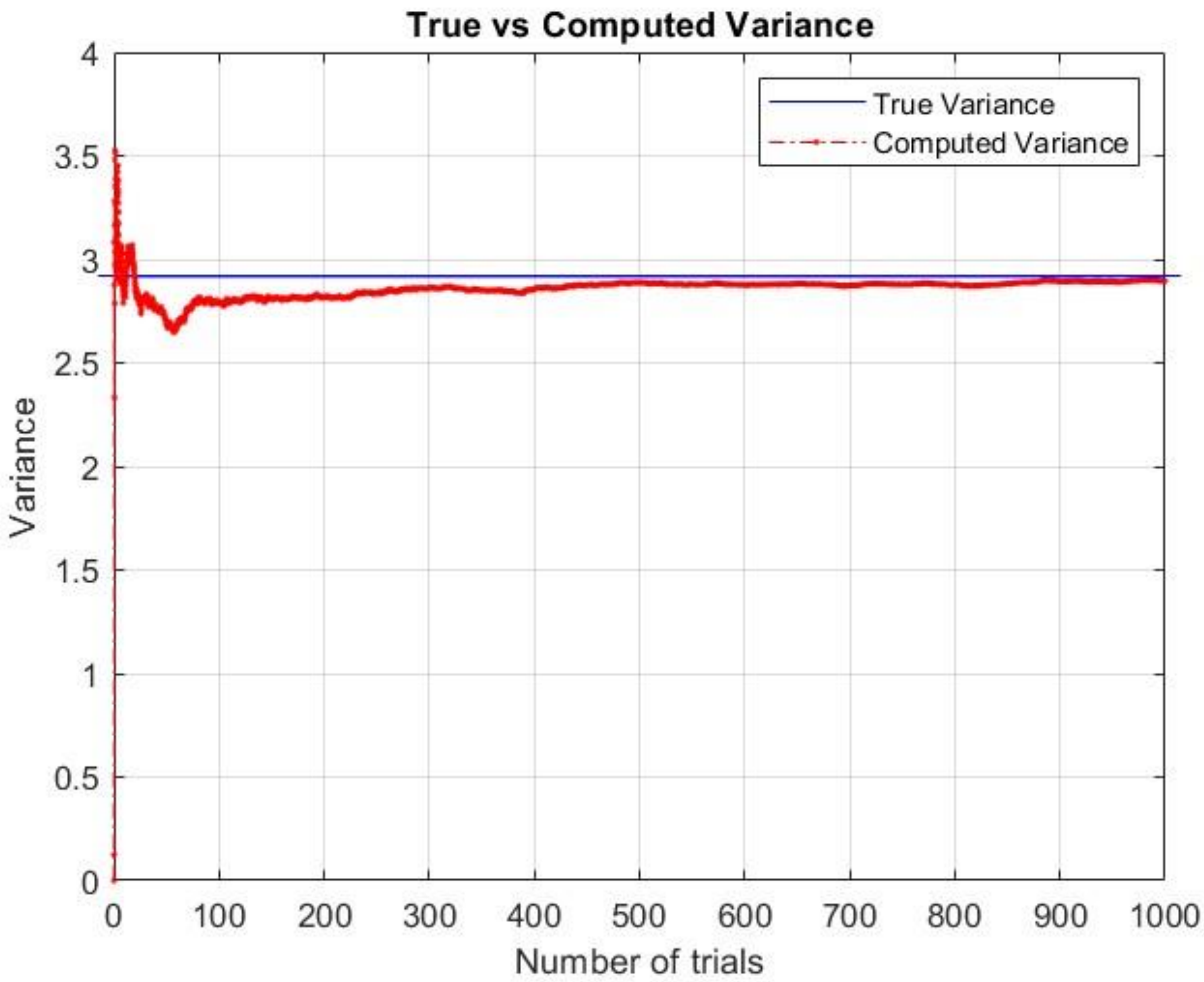
As we can see, the mean started of very far from the true mean(Blue line) and was fluctuating above and below the mean for low number of trials but as the number of trials increased and went beyond 600 the mean started to converge towards

the true mean which is exactly what the law of large number states.

Similarly, for Variance

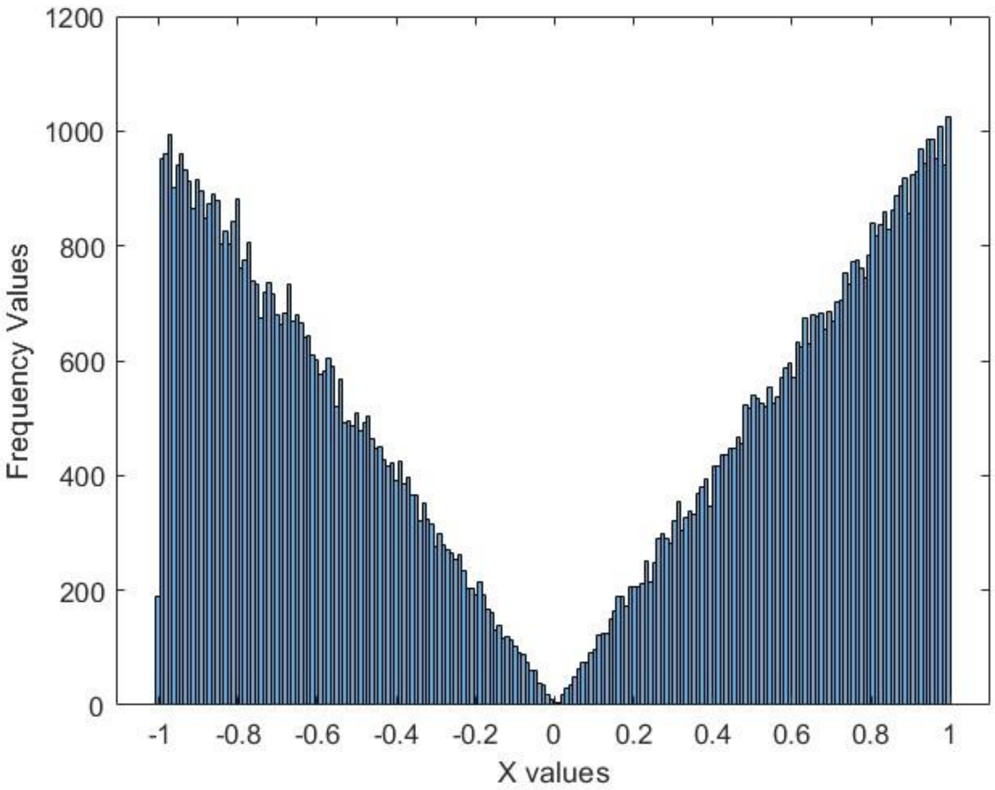
True Variance = 2.91667

Plot of the variance of X for increasing number of trials:



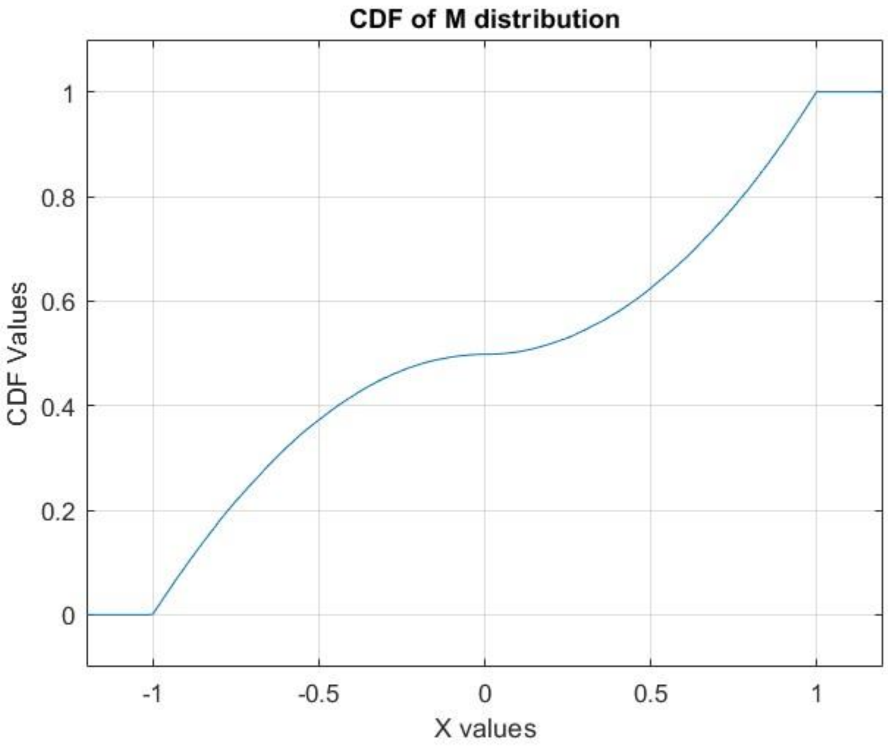
Here also , it is clearly observable that the variance started of very far from the true variance(Blue line) and was fluctuating above and below the mean for low number of trials but as the number of trials increased the variance started to converge towards the true mean.

Q4. Histogram of the random variable X:



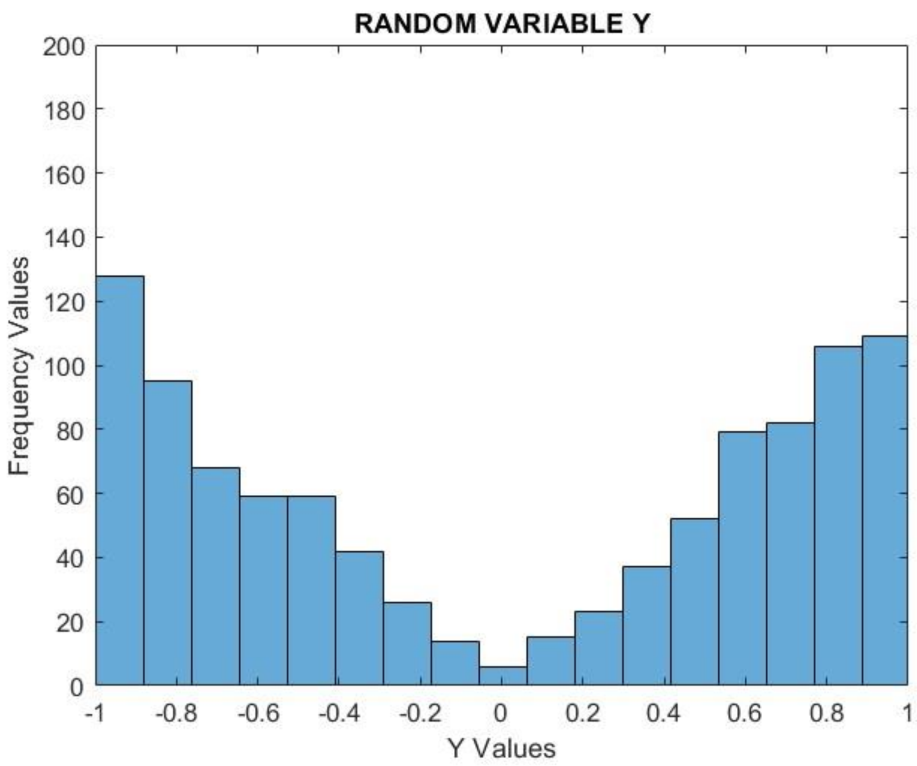
The plot clearly looks like an M – which was specified in the problem statement.

The CDF of the above PDF:

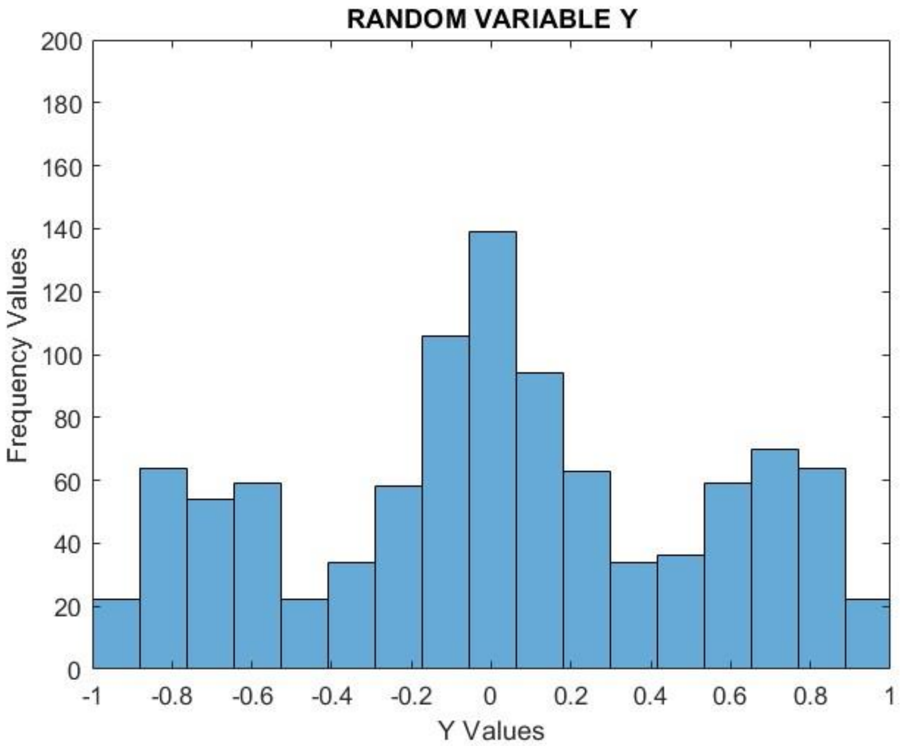


Plots of Y for different values of N:

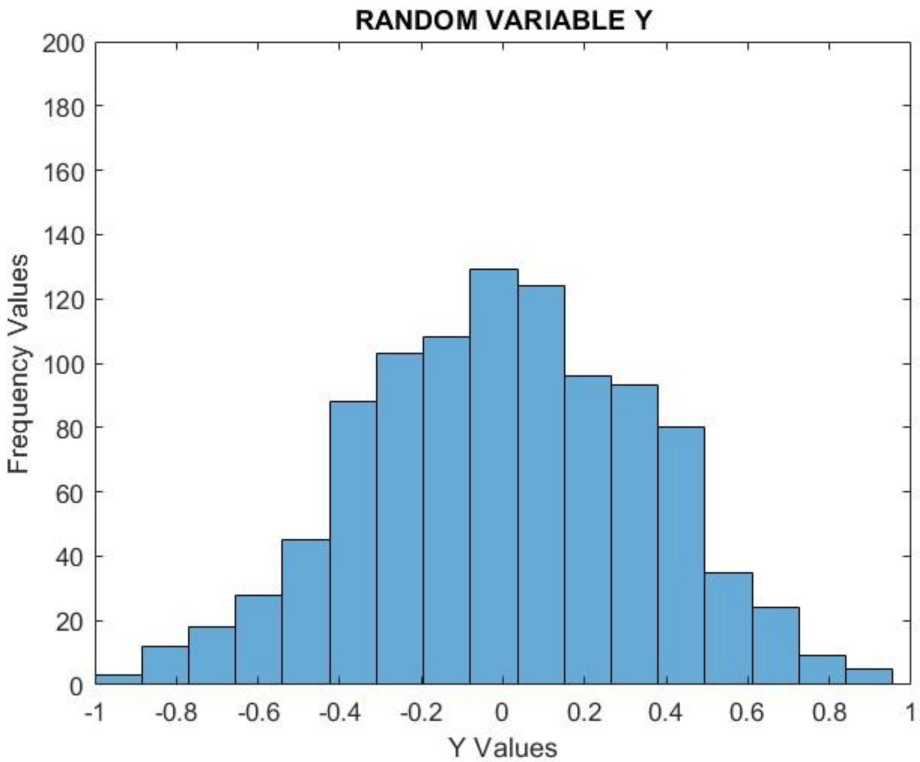
- For N=1



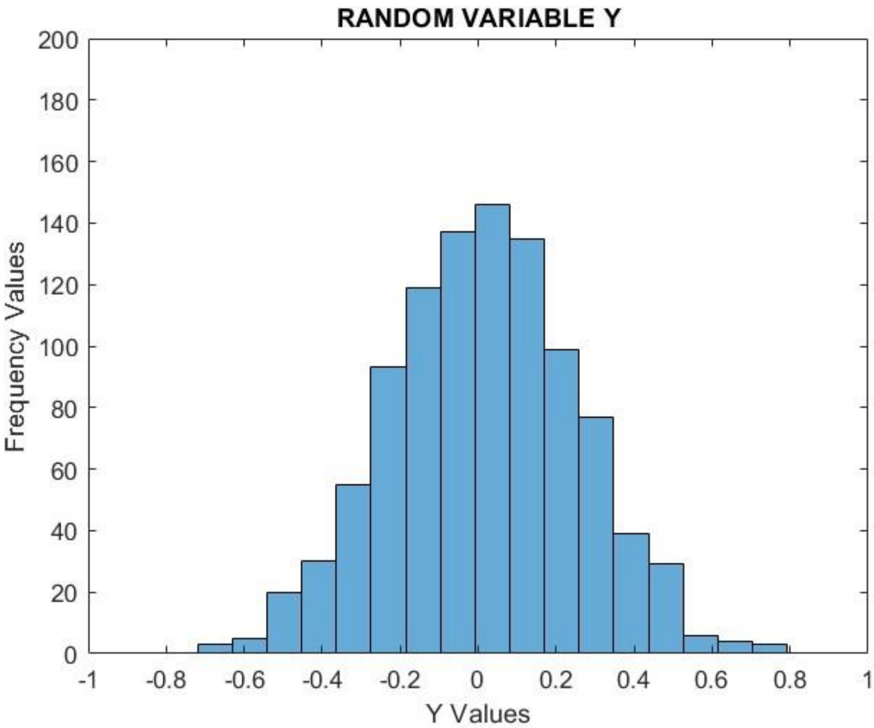
- For N=2



- For N=4

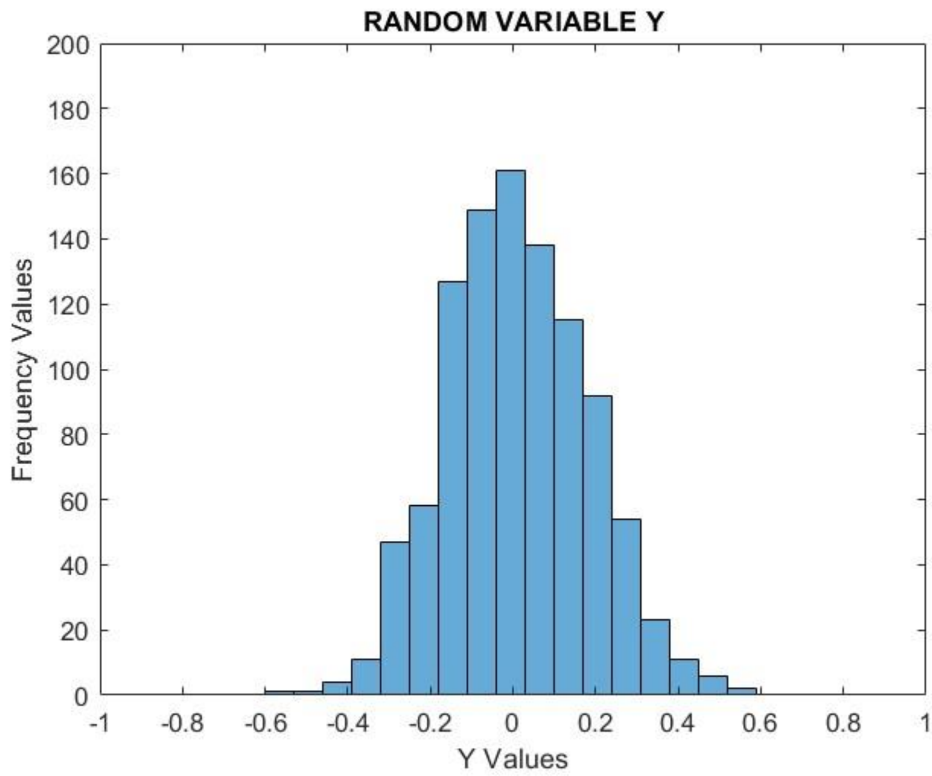


- For N=8

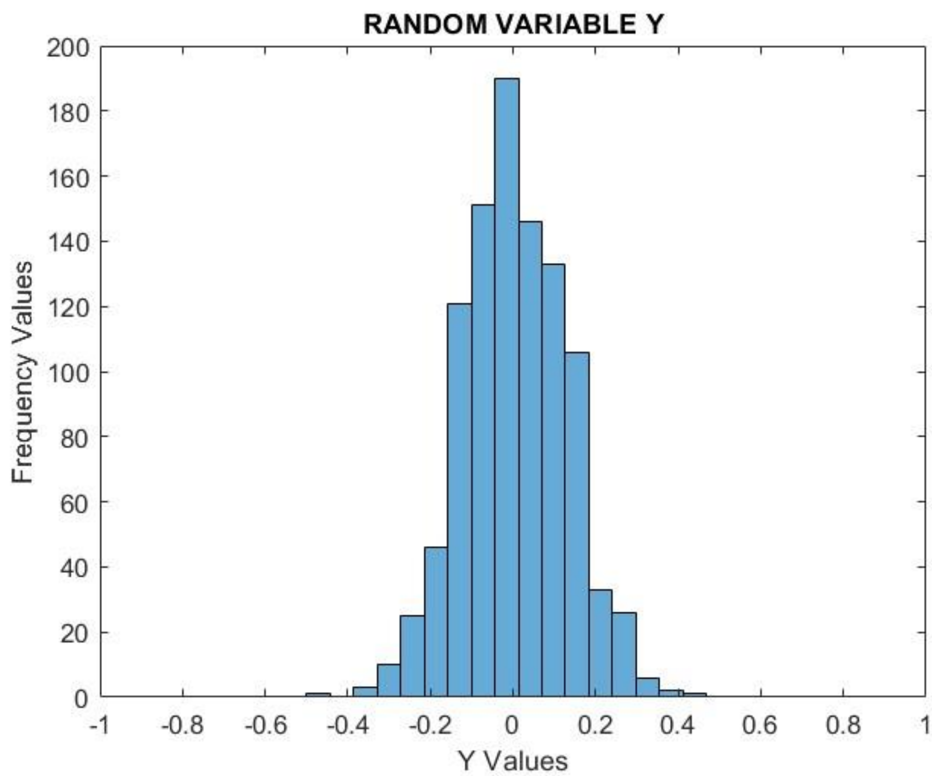


RANDOM VARIABLE Y
Frequency Values
Y Values
200
180
160
140
120
100
80
60
40
20
0
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

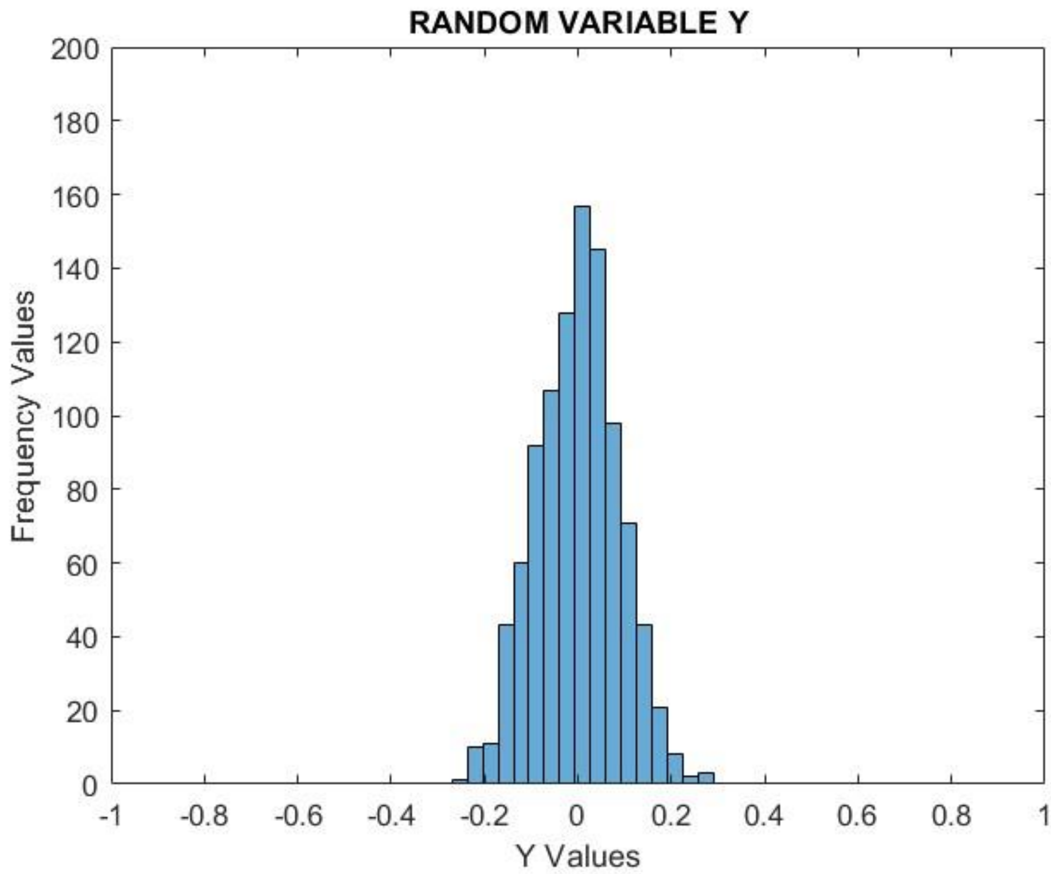
- For N=16



- For N=32



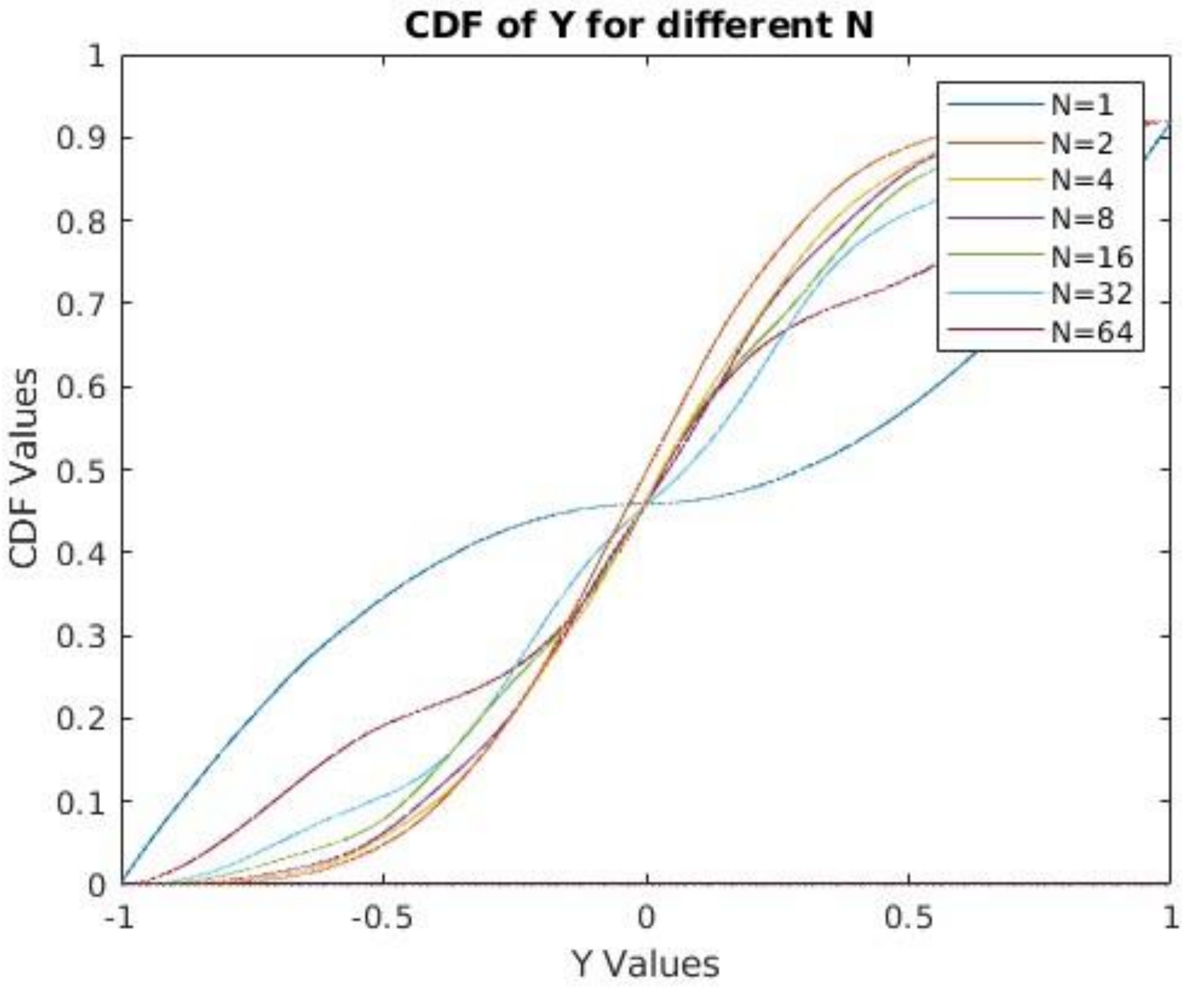
- For N=64



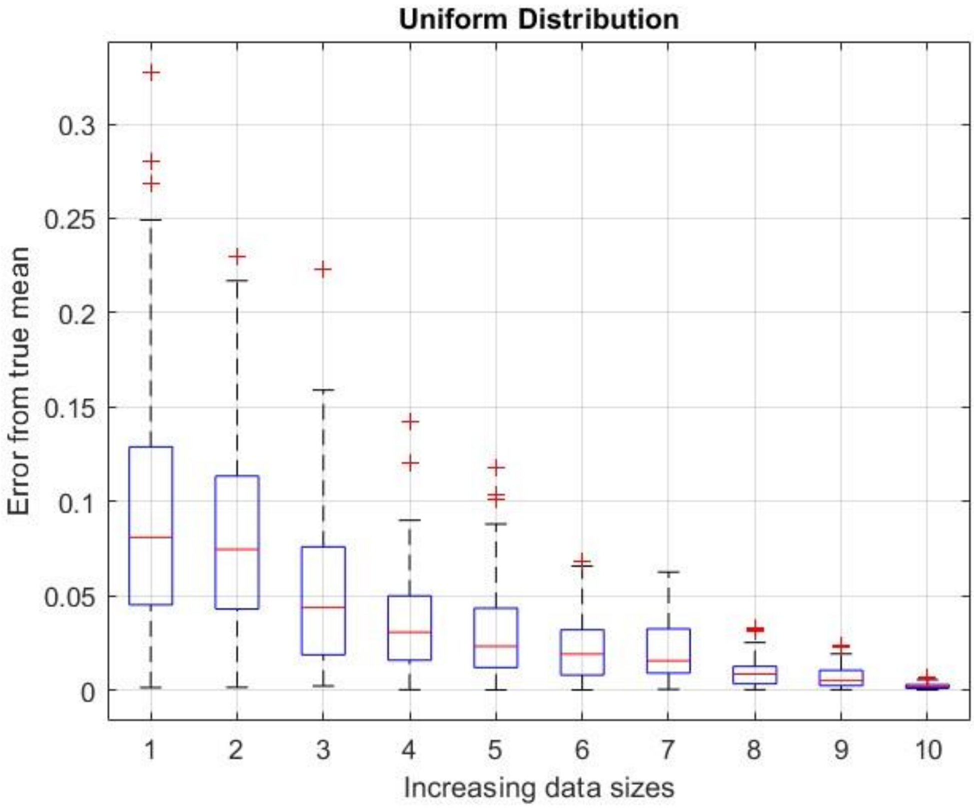
From all these different plots we can observe that as the value of N increases the distribution which was initially in an M-like shape gets converted to a gaussian like curve and upon increasing the N further the gaussian gets shorter in width i.e. the variance keeps decreasing and on N tending to infinity this becomes a delta function.

This is in accordance with the Central Limit Theorem which states exactly this – that any distribution turns into a Gaussian when many observations are averaged.

The CDF of all these plots together:



Q5. Boxplot for a uniform distribution:



Boxplot for a Gaussian distribution: