CS215 Assignment 1

Dhananjay Raman
Shantanu Welling

August 2022

Contents

1 Problem 1

1.1 Laplace Distribution

Parameters are: location parameter $\mu := 2$ and scale parameter $b := 2$.

1. Probability Density Function

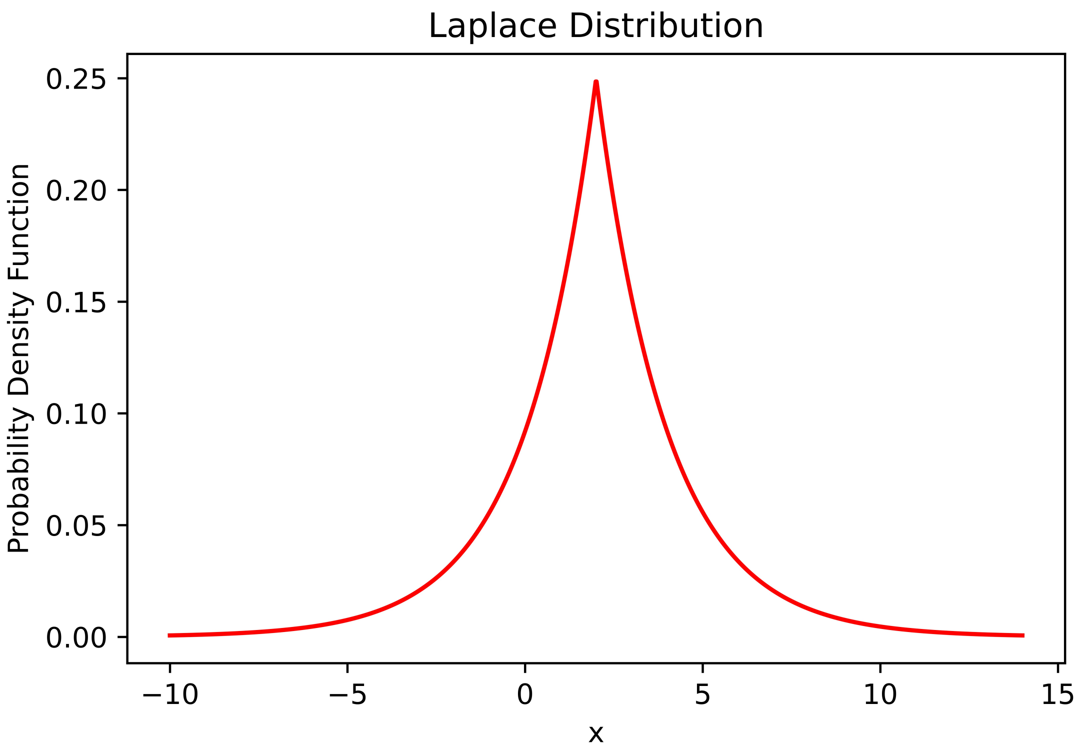

Figure 1: PDF for Laplace Distribution

2. Cumulative Distribution Function

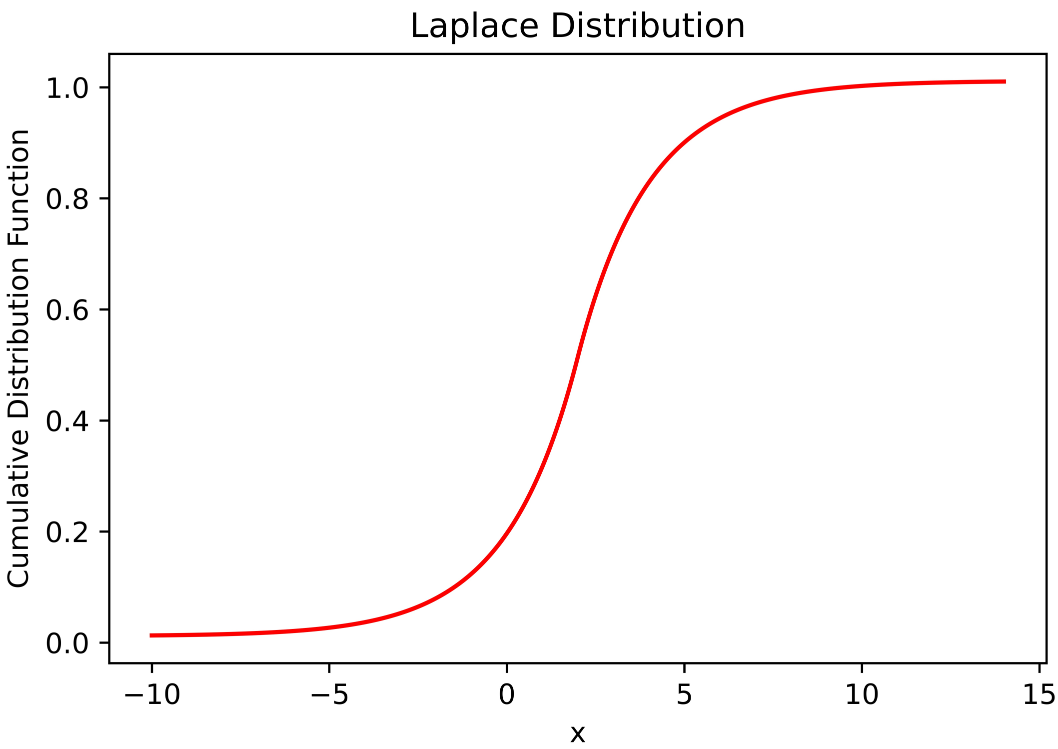

Figure 2: CDF for Laplace Distribution

3. **Variance** of Laplace Distribution is 8.0002, true value is 8. Error is 0.0002.

1.2 Gumbel Distribution

Parameters are: location parameter $\mu := 1$ and scale parameter $\beta := 2$.

1. Probability Density Function

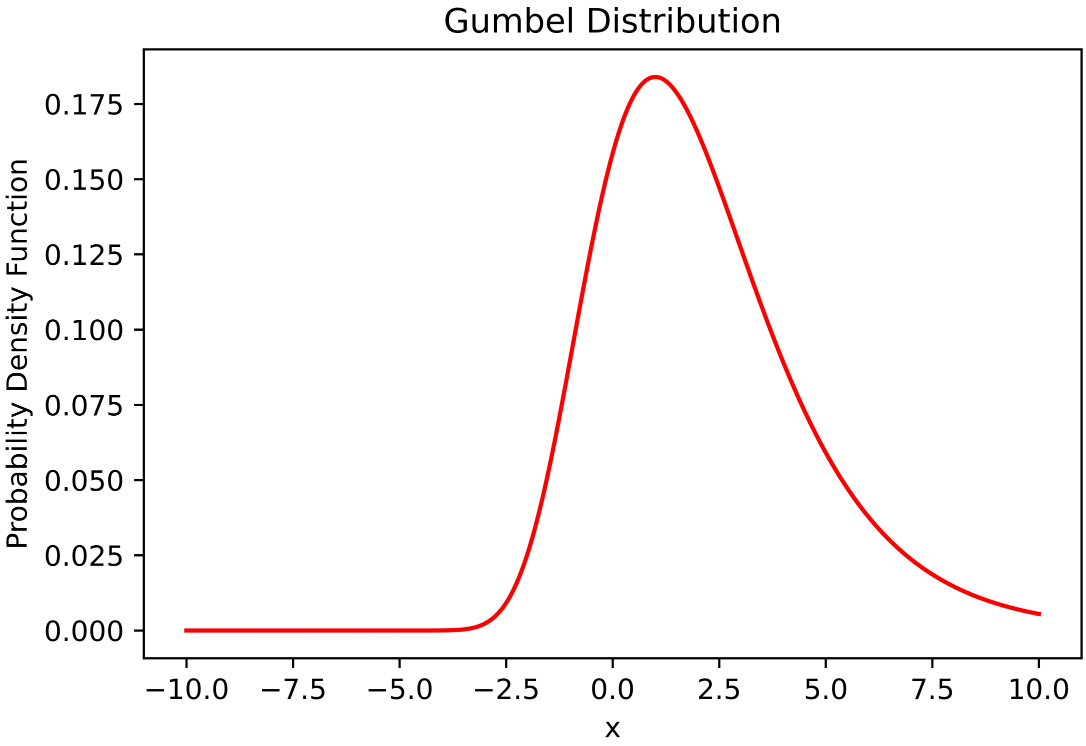

Figure 3: PDF for Gumbel Distribution

2. Cumulative Distribution Function

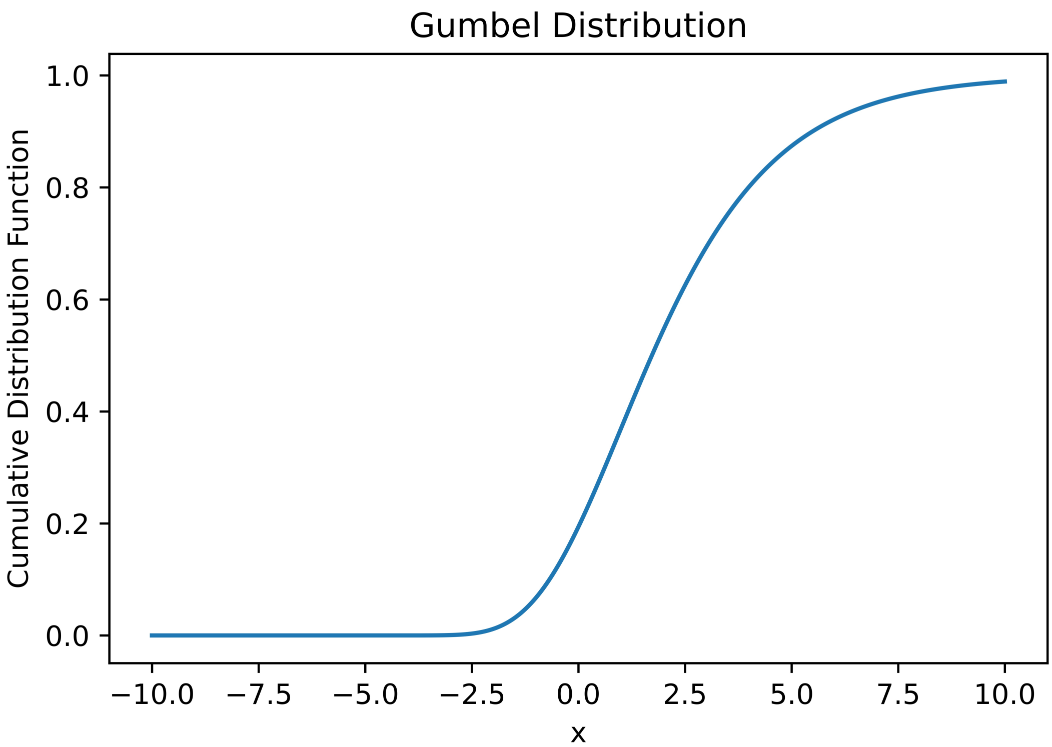

Figure 4: CDF for Gumbel Distribution

3. **Variance** of Gumbel Distribution is 6.5796, true value is 6.5797. Error is 0.0001.

1.3 Cauchy Distribution

Parameters are: location parameter $x_0 := 0$ and scale parameter $\gamma := 1$.

1. Probability Density Function

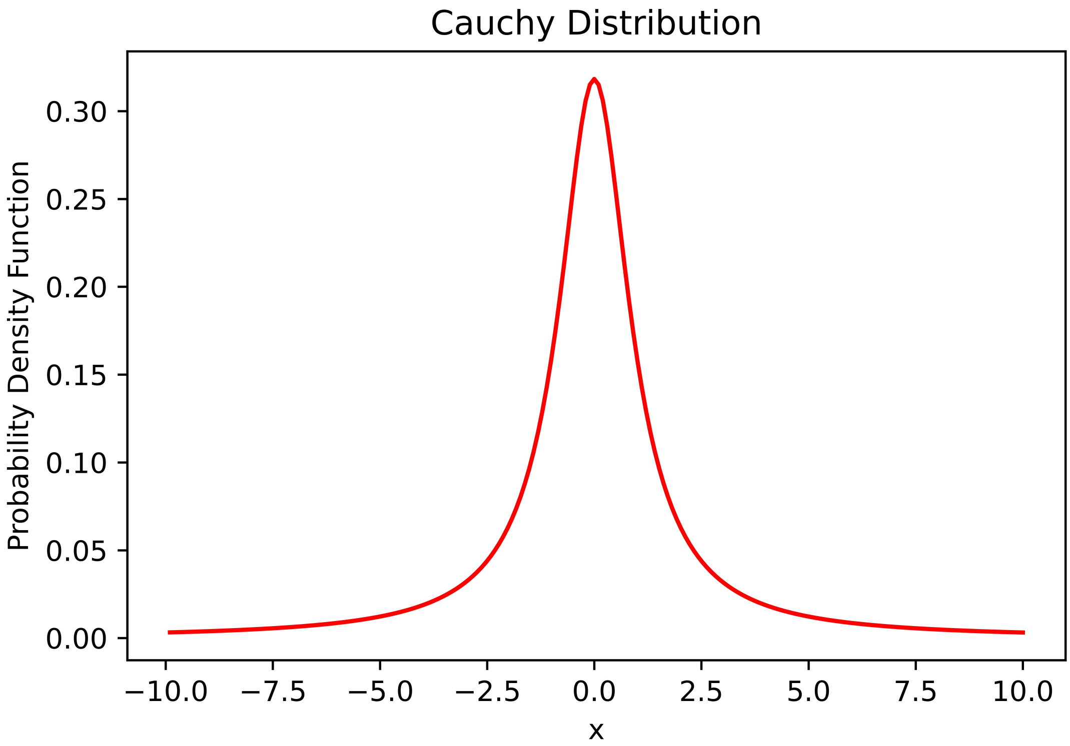

Figure 5: PDF for Cauchy Distribution

2. Cumulative Distribution Function

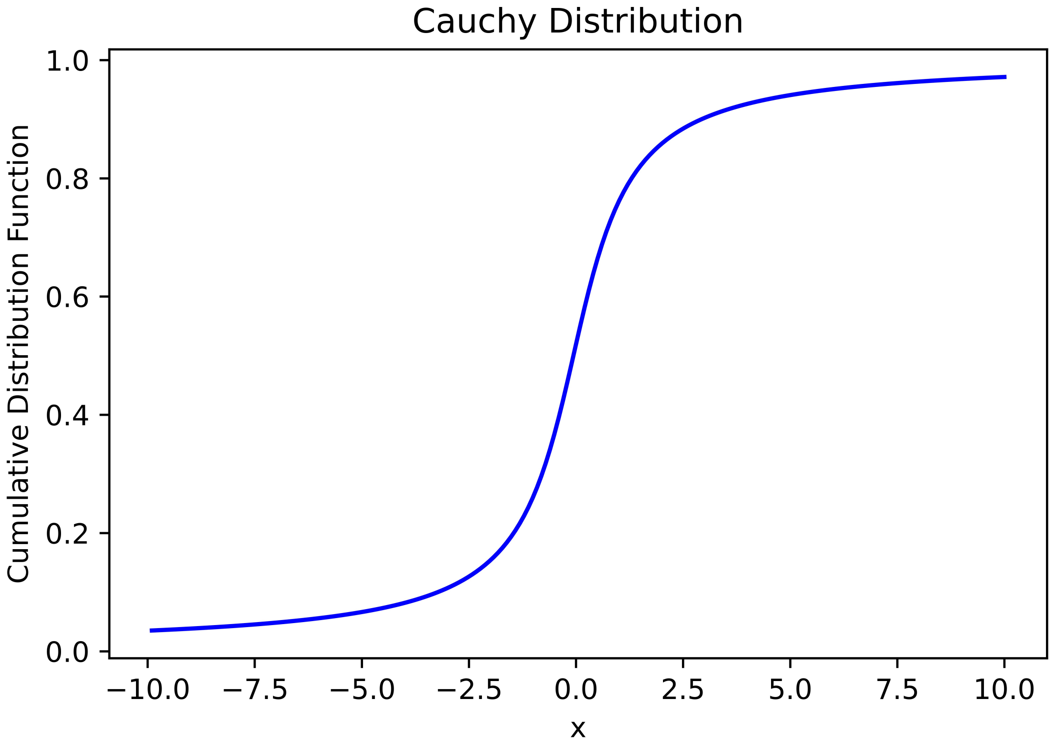

Figure 6: CDF for Cauchy Distribution

3. **Variance:** As the integration limits tend to $\pm\infty$, the variance grows without bounds. Hence the variance of Cauchy Distribution is **undefined**. The true variance is also undefined.

2 Problem 2

2.1 Part A

1. Emperical estimate of PMF of Poisson Distrubution of $Z = X + Y$

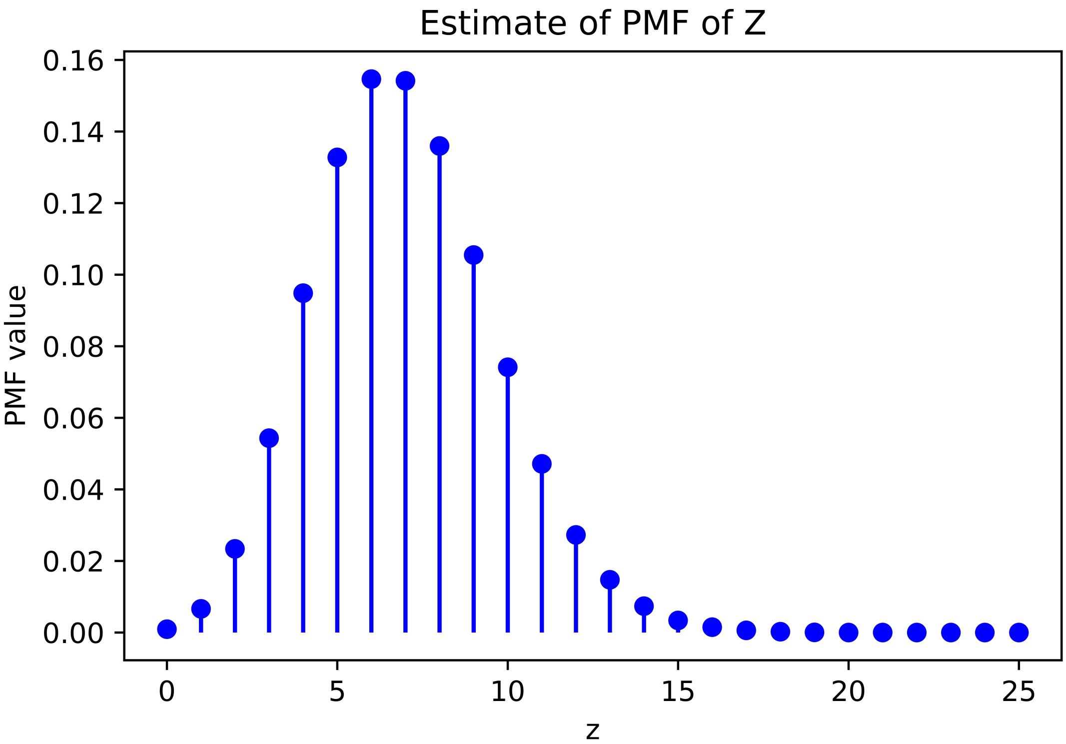

Figure 7: Poisson Distrubution PMF (Empirical)

2. Theoretically, the PMF of random variable $Z = X + Y$, where X and Y have the parameters $\lambda_x = 3$ and $\lambda_y = 4$ will be:

$$P(Z = k) = \frac{e^{-\lambda_x + \lambda_y}(\lambda_x + \lambda_y)^k}{k!}$$
$$= \frac{e^{-7}7^k}{k!}$$

3. Emprical and theoretical PMF values for $k = 0, 1, 2, \ldots 25$ are shown here:

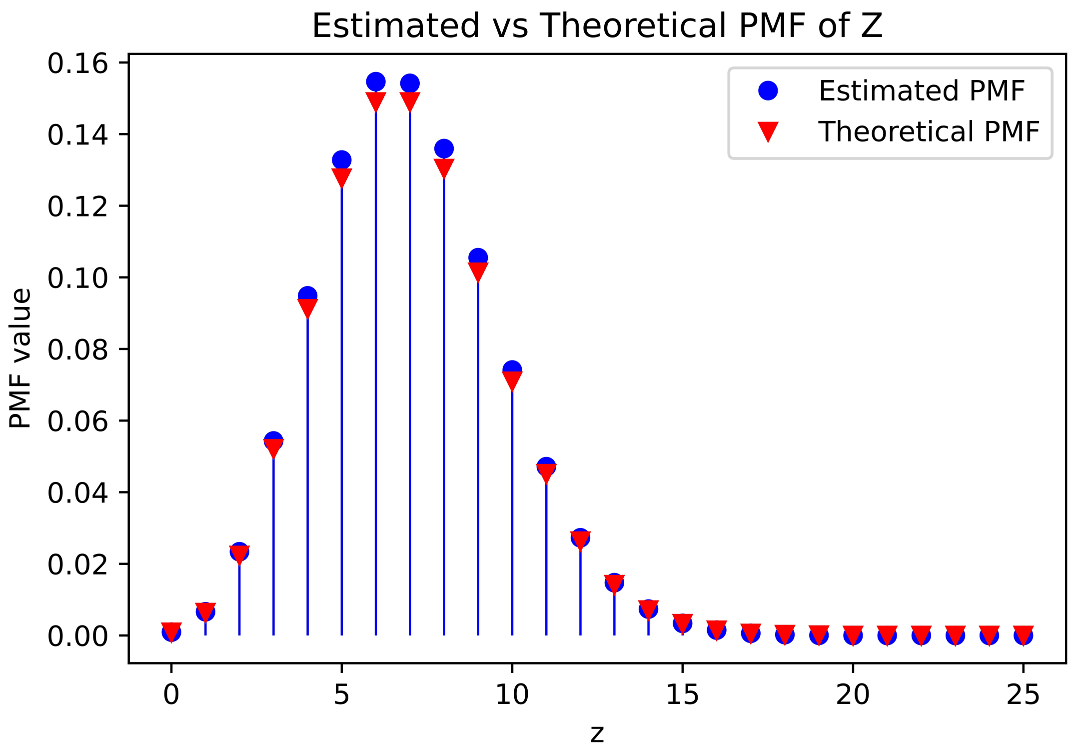

Figure 8: Comparison of Empirical and Theoretical values of Poisson PMF

1. Estimated P(Z=0) is 0.00095, actual is 0.00091
2. Estimated P(Z=1) is 0.00666, actual is 0.00638
3. Estimated P(Z=2) is 0.02341, actual is 0.02234
4. Estimated P(Z=3) is 0.05432, actual is 0.05213
5. Estimated P(Z=4) is 0.09487, actual is 0.09123
6. Estimated P(Z=5) is 0.13279, actual is 0.12772
7. Estimated P(Z=6) is 0.15468, actual is 0.14900
8. Estimated P(Z=7) is 0.15419, actual is 0.14900
9. Estimated P(Z=8) is 0.13599, actual is 0.13038
10. Estimated P(Z=9) is 0.10551, actual is 0.10140
11. Estimated P(Z=10) is 0.07415, actual is 0.07098
12. Estimated P(Z=11) is 0.04716, actual is 0.04517
13. Estimated P(Z=12) is 0.02728, actual is 0.02635
14. Estimated P(Z=13) is 0.01477, actual is 0.01419
15. Estimated P(Z=14) is 0.00738, actual is 0.00709
16. Estimated P(Z=15) is 0.00341, actual is 0.00331
17. Estimated P(Z=16) is 0.00149, actual is 0.00145
18. Estimated P(Z=17) is 0.00062, actual is 0.00060
19. Estimated P(Z=18) is 0.00026, actual is 0.00023
20. Estimated P(Z=19) is 0.00008, actual is 0.00009
21. Estimated P(Z=20) is 0.00002, actual is 0.00003

22. Estimated P(Z=21) is 0.00001, actual is 0.00001
23. Estimated P(Z=22) is 0.00000, actual is 0.00000
24. Estimated P(Z=23) is 0.00000, actual is 0.00000
25. Estimated P(Z=24) is 0.00000, actual is 0.00000
26. Estimated P(Z=25) is 0.00000, actual is 0.00000

2.2 Part B

1. Emperical estimate of PMF of Thinned Poisson Distrubution ($\widehat{P}(Z)$)

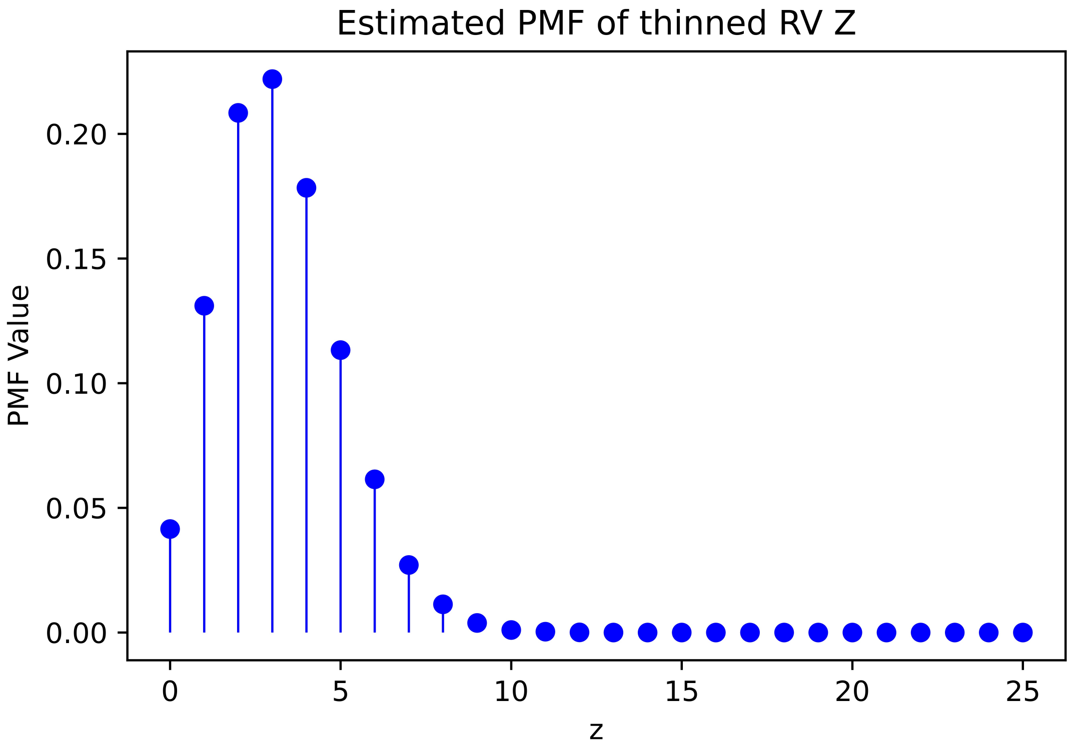

Figure 9: Empirical PMF of Poisson Thinned Random Variable Z

2. Theoretically, the PMF of Thinned Poisson Random Variable Z with parameters $\lambda_y = 4$ and $p = 0.8$ will be:

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

3. Emprical and theoretical PMF values for $k = 0, 1, 2, \ldots 25$ are shown here:

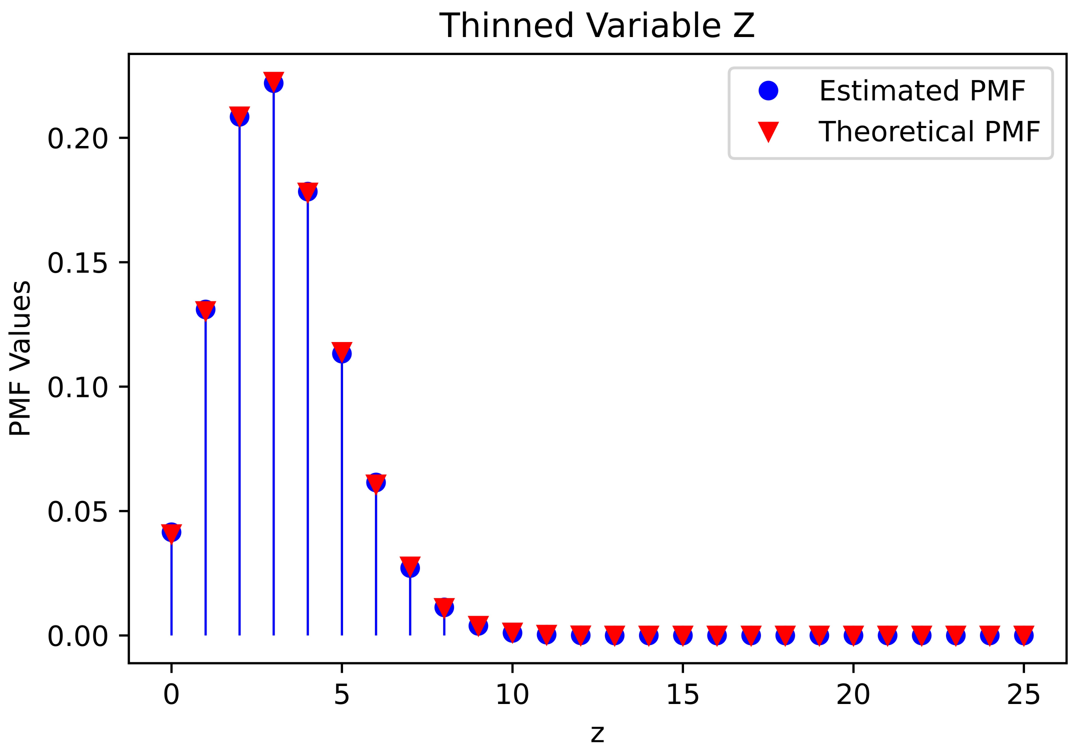

Figure 10: Comparison of Empirical and Theoretical values of "Thinned" Poisson PMF

1. Estimated P(Z=0) is 0.0415, actual is 0.04076
2. Estimated P(Z=1) is 0.13105, actual is 0.13044
3. Estimated P(Z=2) is 0.20846, actual is 0.20870
4. Estimated P(Z=3) is 0.222, actual is 0.22262
5. Estimated P(Z=4) is 0.17839, actual is 0.17809
6. Estimated P(Z=5) is 0.11328, actual is 0.11398
7. Estimated P(Z=6) is 0.06148, actual is 0.06079
8. Estimated P(Z=7) is 0.02711, actual is 0.02779
9. Estimated P(Z=8) is 0.01133, actual is 0.01112
10. Estimated P(Z=9) is 0.00389, actual is 0.00395
11. Estimated P(Z=10) is 0.00106, actual is 0.00126
12. Estimated P(Z=11) is 0.00034, actual is 0.00037
13. Estimated P(Z=12) is 8e-05, actual is 0.00010
14. Estimated P(Z=13) is 2e-05, actual is 0.00002
15. Estimated P(Z=14) is 1e-05, actual is 0.00001
16. Estimated P(Z=15) is 0.0, actual is 0.00000
17. Estimated P(Z=16) is 0.0, actual is 0.00000
18. Estimated P(Z=17) is 0.0, actual is 0.00000
19. Estimated P(Z=18) is 0.0, actual is 0.00000
20. Estimated P(Z=19) is 0.0, actual is 0.00000
21. Estimated P(Z=20) is 0.0, actual is 0.00000

22. Estimated P(Z=21) is 0.0, actual is 0.00000
23. Estimated P(Z=22) is 0.0, actual is 0.00000
24. Estimated P(Z=23) is 0.0, actual is 0.00000
25. Estimated P(Z=24) is 0.0, actual is 0.00000
26. Estimated P(Z=25) is 0.0, actual is 0.00000

3 Problem 3

1. Histogram of final locations of all 10^4 random walkers are shown here:

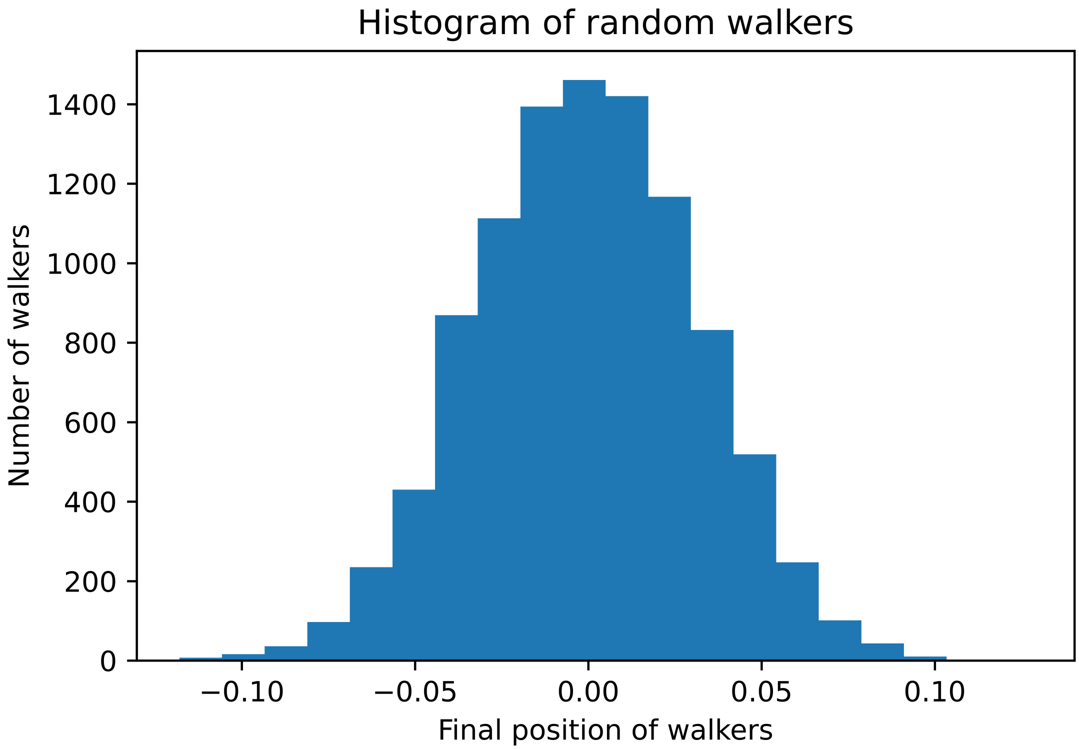

Figure 11: Histogram of Final Positions of Random Walkers

2. Space time curve of paths taken by the first 10^3 walkers shown here:

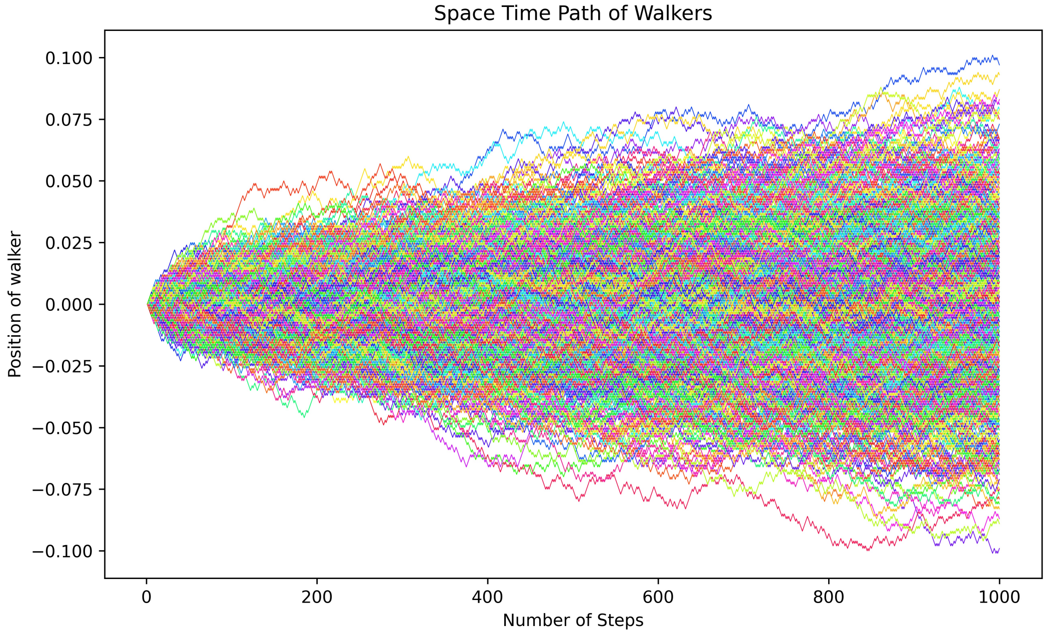

Figure 12: Paths taken by Random Walkers

3. All the code is stored in a .ipynb file in the "code" folder.

4. If $X_1, X_2, \ldots X_n$ are N independent draws from the distribution of X (with mean M), and $\widehat{M} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$, then by the law of large numbers, for all $\epsilon > 0$, as $n \to \infty$,
$$P(|\widehat{M} - M| \geq \epsilon) \to 0$$
$\implies \widehat{M}$ converges to the mean $M = E[X]$ as $n \to \infty$.
Now,
$$\begin{aligned}\widehat{V} &= \frac{\sum_{i=1}^{n}(X_i - \widehat{M})}{n} \\ &= \sum_{i=1}^{n}\frac{X_i^2}{n} + \sum_{i=1}^{n}\frac{\widehat{M}^2}{n} - \sum_{i=1}^{n}\frac{2X_i\widehat{M}}{n} \\ &= \sum_{i=1}^{n}\frac{X_i^2}{n} - \sum_{i=1}^{n}\frac{\widehat{M}^2}{n}\end{aligned}$$
The second term converges to $M^2 = E[X]^2$ (limit of $x * y$ is equal to product of limits of x and y, if the two exist), as $n \to \infty$.
The first term, by central limit theorem on $X_1^2, X_2^2, \ldots X_n^2$, converges to $E[X^2]$, as $n \to \infty$.
Hence, $\widehat{V}$ converges to $V = E[X^2] - E[X]^2$, which is exactly the variance of the distribution of X.

5. Random Walkers Final Position Mean: -0.0006; Variance: 0.0010
6. The value of the true mean M should be 0, and the true variance should be $V = n\epsilon^2$, where n is the number of steps and ϵ is the step size.
   For $n = 10^3$ and $\epsilon = 10^{-3}$, we will have $M = 0$ and $V = 10^{-3}$.
7. Error between the empirically-computed mean and the true mean:
   $$|\widehat{M} - M| = 0.0006$$
   Error between the empirically-computed variance and the true variance:
   $$|\widehat{V} - V| = 0.0000$$

4 Problem 4

1. Well documented code for this is stored in the .ipynb file in the "code" folder.
2. Histogram of distribution following M-shaped PDF:

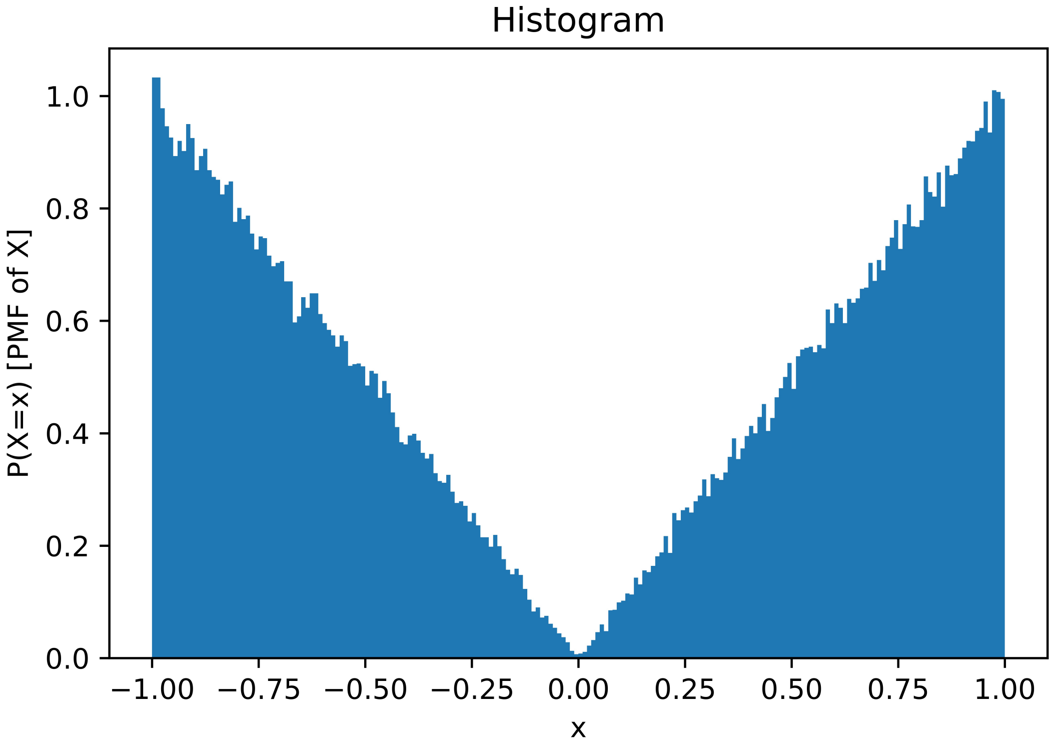

Figure 13: Histogram of M-function distribution

3. CDF of the distribution:

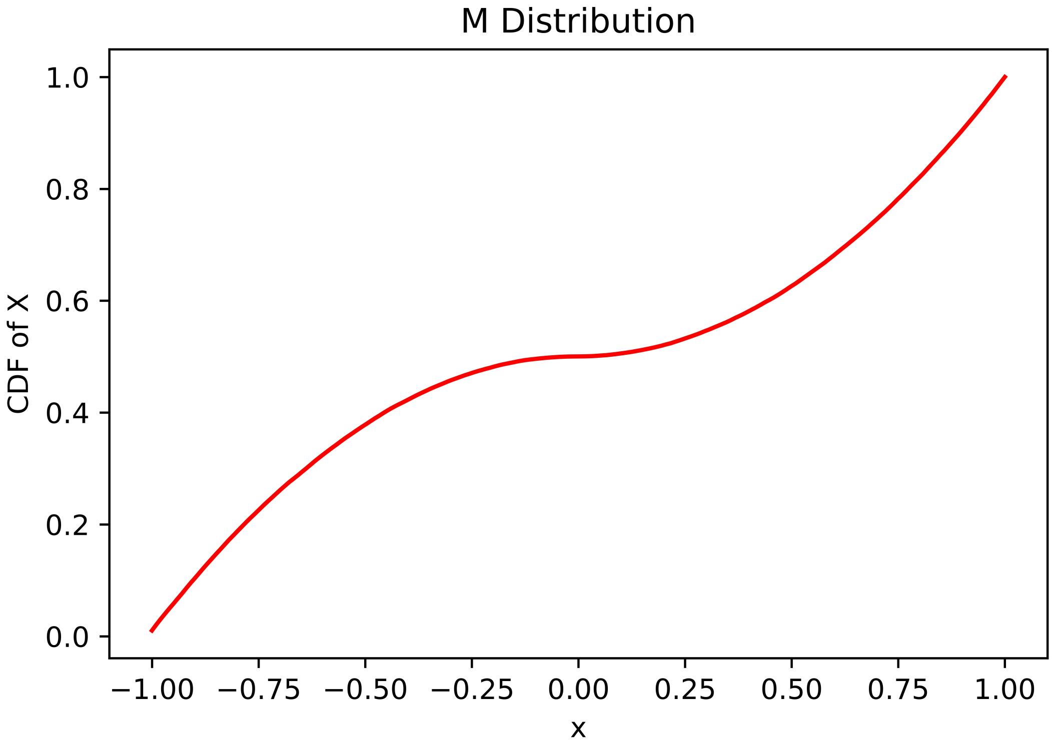

Figure 14: CDF of M-function Distribution

4. Well documented code for this is stored in the .ipynb file in the "code" folder.
5. Histograms of Y_N distributions for $N = 2, 4, 8, 16, 32, 64$ (Number of draws= 10^4):

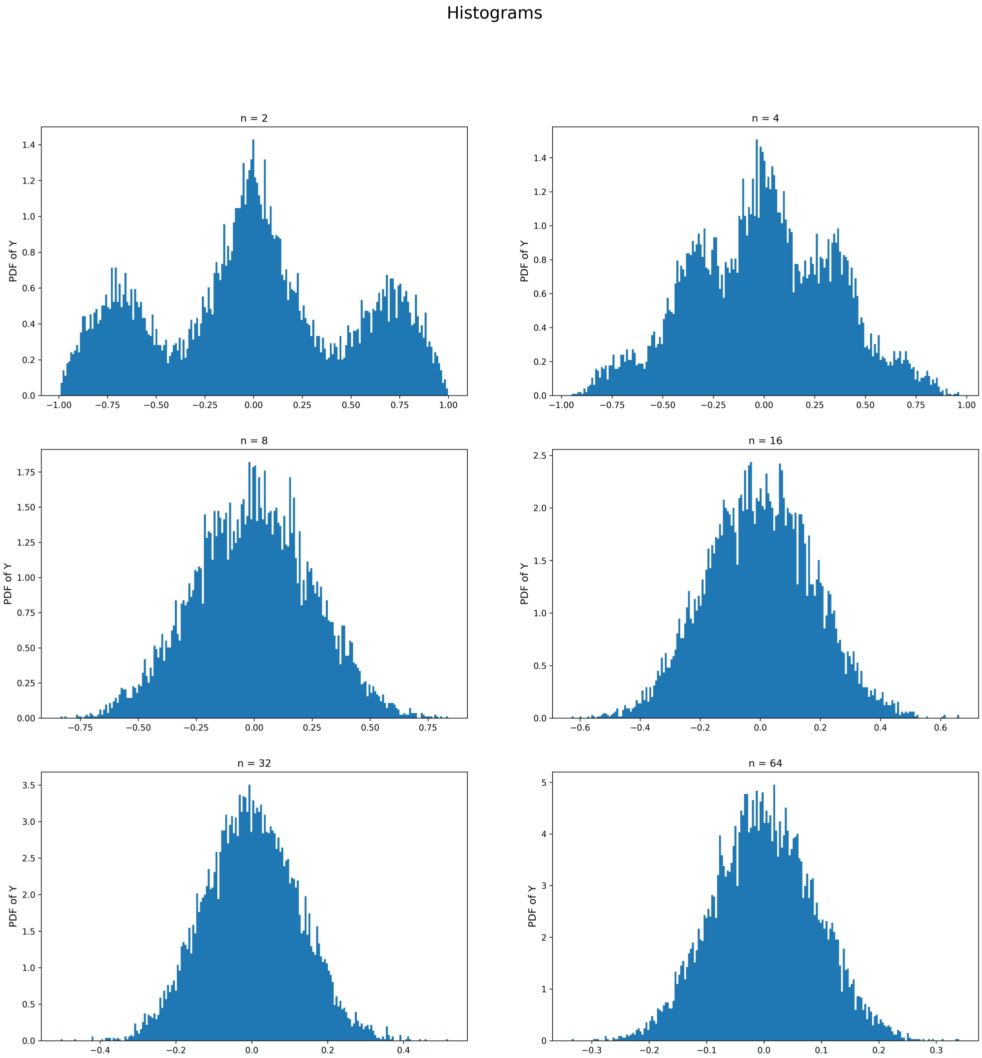

Figure 15: Histograms of M-function mean distributions for $N = 2, 4, 8, 16, 32, 64$

Plots of CDFs of the corresponding distributions (Plotted in same graph):

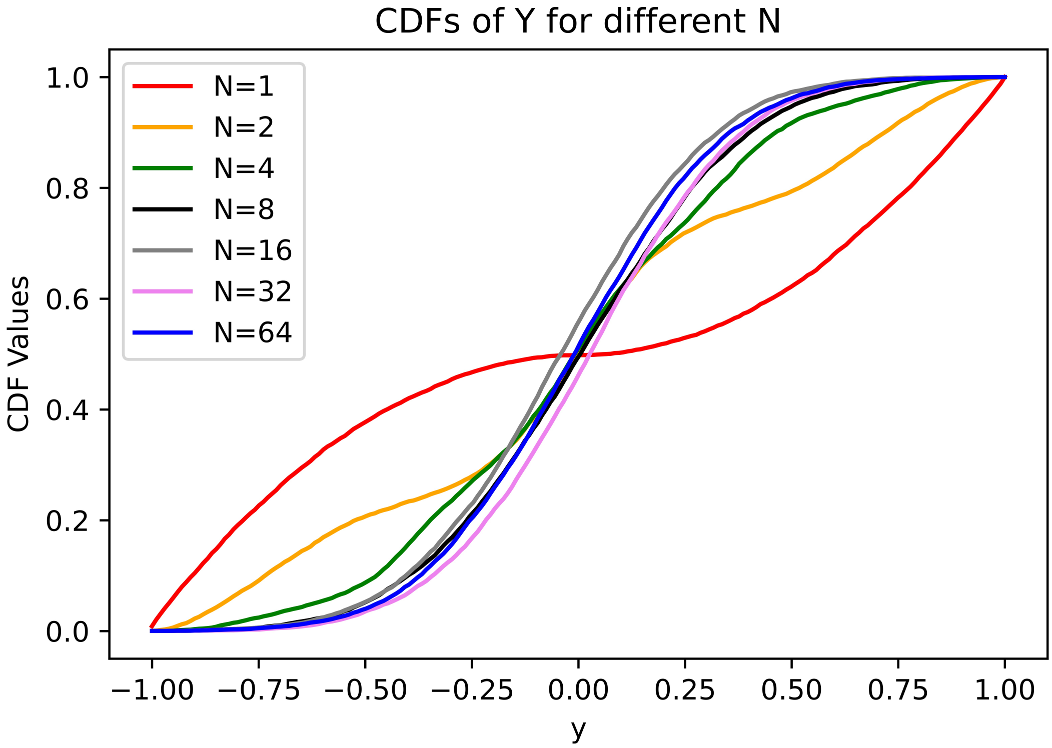

Figure 16: CDFs of M-function mean distributions for $N = 1, 2, 4, 8, 16, 32, 64$

5 Problem 5

1. Box-and-whisker plot of distribution of error across $M = 100$ trials, plotted for dataset sizes $N = 5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4$ (for **Uniform Distribution**):

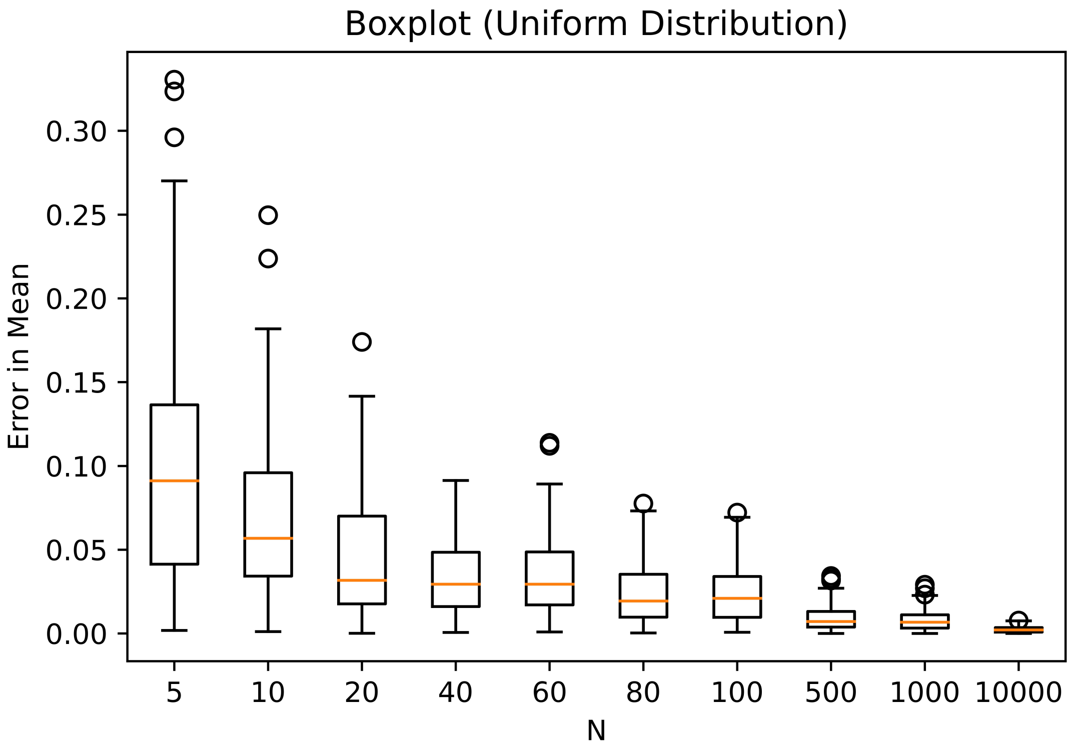

Figure 17: Boxplot of Error distribution across all trials for various N (Uniform Distribution)

2. Box-and-whisker plot of distribution of error across $M = 100$ trials, plotted for dataset sizes $N = 5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4$ (for **Gaussian Distribution** with parameters $\mu := 0$ and $\sigma^2 := 1$):

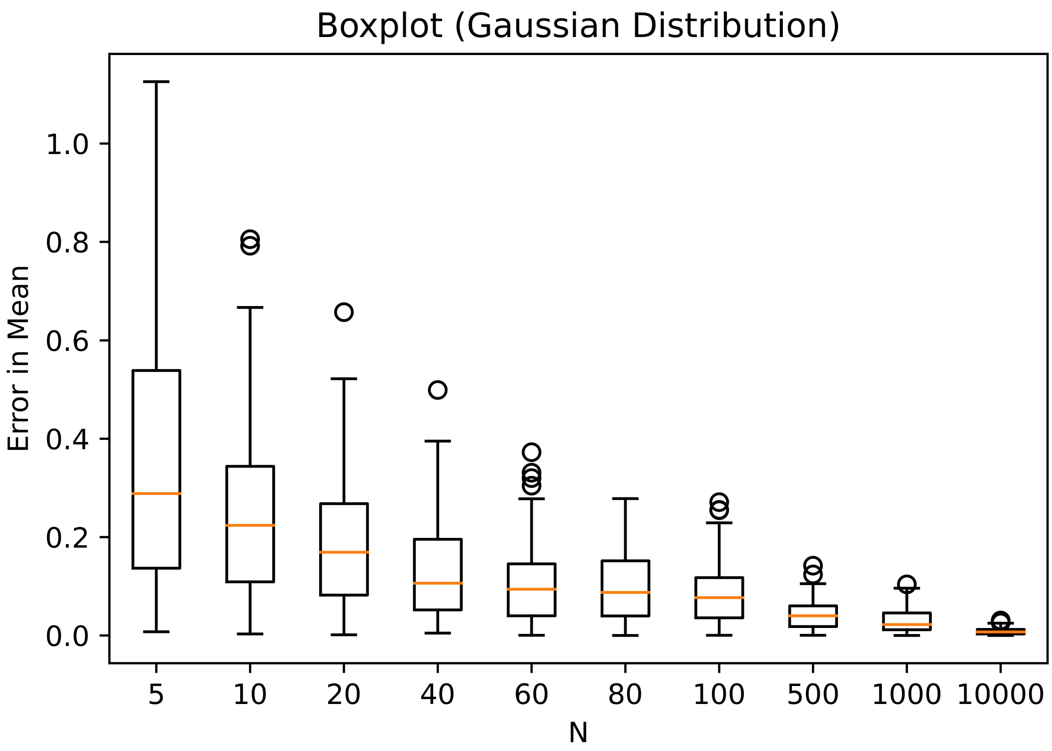

Figure 18: Boxplot of Error distribution across all trials for various N (Gaussian Distribution)

3. Interpretation from graphs:
   As N increases, the distribution of error shrinks. The error decreases and the estimated mean approaches the true mean. This is in accordance with the Weak Law of Large Numbers.
   In a sample of independent and identically distributed random variables, as the sample size grows larger, the sample mean will tend toward the population mean.

4. Well documented code for both parts are stored in the .ipynb file in the "code" folder.