Part - I

1. Simulating Random Variables

- Simulate random variates using both MATLAB routines as well as the rejection method, for $\{X_i\}_{i=1,\ldots T}$, $T = 100, 1000, 10000$ with a PDF that is:
  i. Normal with mean = 2 and variance = 2
  ii. Uniform on [2, 4]
  iii. Exponential with parameter 2

  For the simulation of the random variables using MATLAB routines, the function 'makedist' is used for generating normal, uniform and exponential distributions.
  For the simulation of random numbers using the rejection method a uniform distribution U(0, 1) is used as the base envelope distribution. The uniform distribution is scaled in accordance with the maximum possible values of the target distributions. And the width of U(0, 1) is also expanded in accordance with the target distribution's range. The following envelope functions are used:
  1. For N(2, 2) - U(0, 1) is scaled with $1/(2 * \sqrt{(\pi)})$ in the y-axis and expanded to the range of [-5, 9]
  2. For U(2, 4) - U(0, 1) is scaled with 1/(2) in the y-axis and expanded to the range of [2, 4]
  3. For exp(2) - U(0, 1) is scaled with 2 in the y-axis and expanded to the range of [0, 4]

- Compute the histograms for each of the cases and estimate the parameters of each of the populations in each of the observation length cases.
  1. Normal with mean = 2 and variance = 2

Normal distribution simulated using MATLAB routine and rejection method are illustrated below for the values of T = 100, 1000 and 10000.

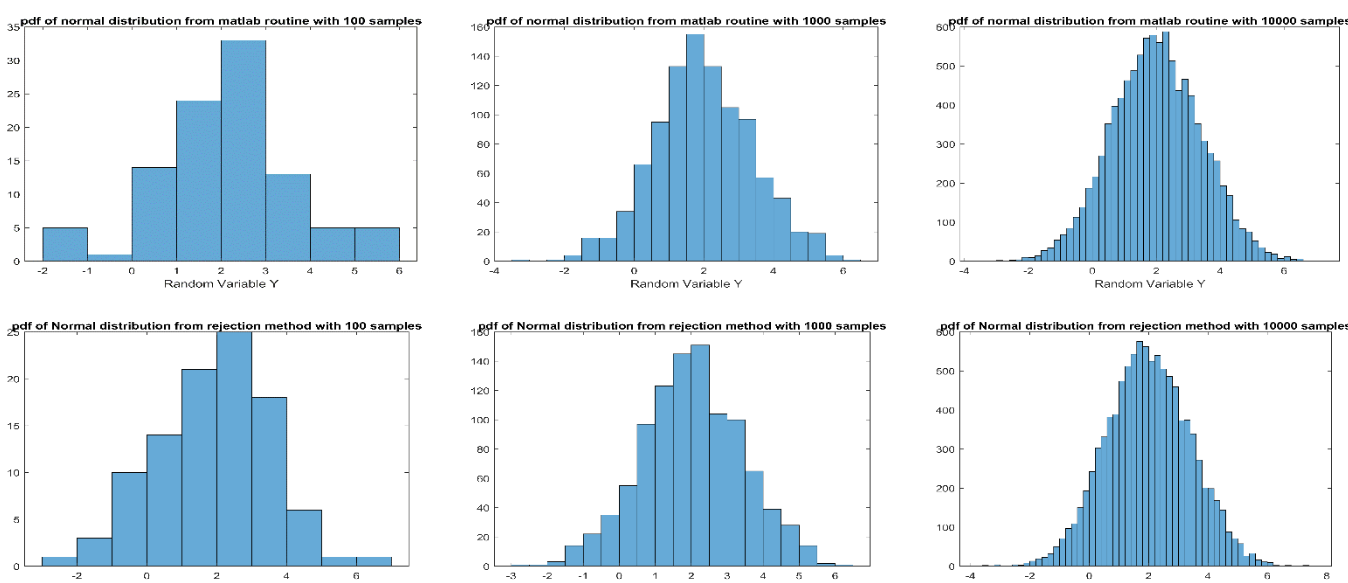

Figure 1. Simulation of Normal variates

Number of Variates T	Mean (MATLAB Routine)	Mean (Rejection Method)	Variance (MATLAB Routine)	Variance (Rejection Method)
100	2.1002	1.9201	2.2296	2.7013
1000	1.9951	2.0323	2.0245	1.9734
10000	2.0031	1.9902	1.9663	2.0175

Table 1. Parameters of Simulated Normal Variates

2. Uniform on [2, 4]

Uniform distribution simulated using MATLAB routine and rejection method are illustrated below for the values of T = 100, 1000 and 10000.

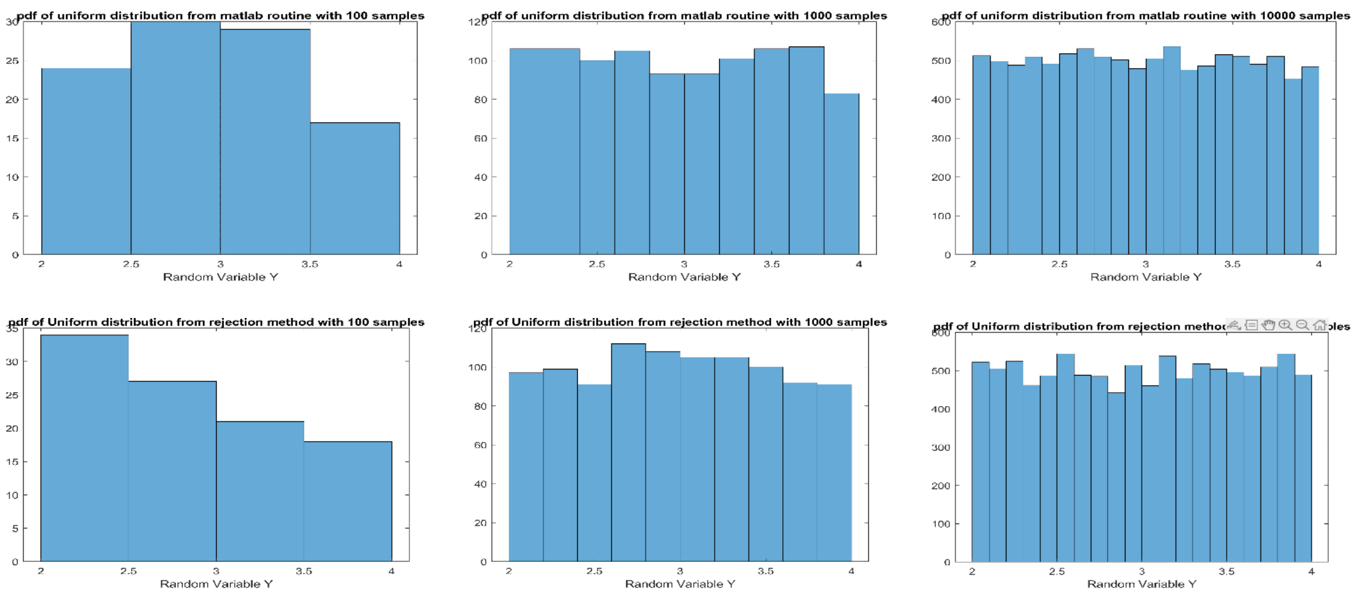

Figure 2. Simulation of Uniform variates

Number of Variates T	Mean (MATLAB Routine)	Mean (Rejection Method)	Variance (MATLAB Routine)	Variance (Rejection Method)
100	2.9476	2.8666	0.2055	0.3265
1000	2.9827	2.9925	0.3314	0.3187
10000	2.9933	3.0021	0.3303	0.3377

Table 2. Parameters of Simulated Uniform Variates

3. Exponential with parameter 2

Number of Variates T	Mean (MATLAB Routine)	Mean (Rejection Method)	Variance (MATLAB Routine)	Variance (Rejection Method)
100	0.4388	0.5236	0.1974	0.2638
1000	0.4922	0.4935	0.2573	0.2395
10000	0.5046	0.5012	0.2560	0.2526

Table 3. Parameters of Simulated Exponential Variates

Exponential distribution simulated using MATLAB routine and rejection method are illustrated below for the values of T = 100, 1000 and 10000.

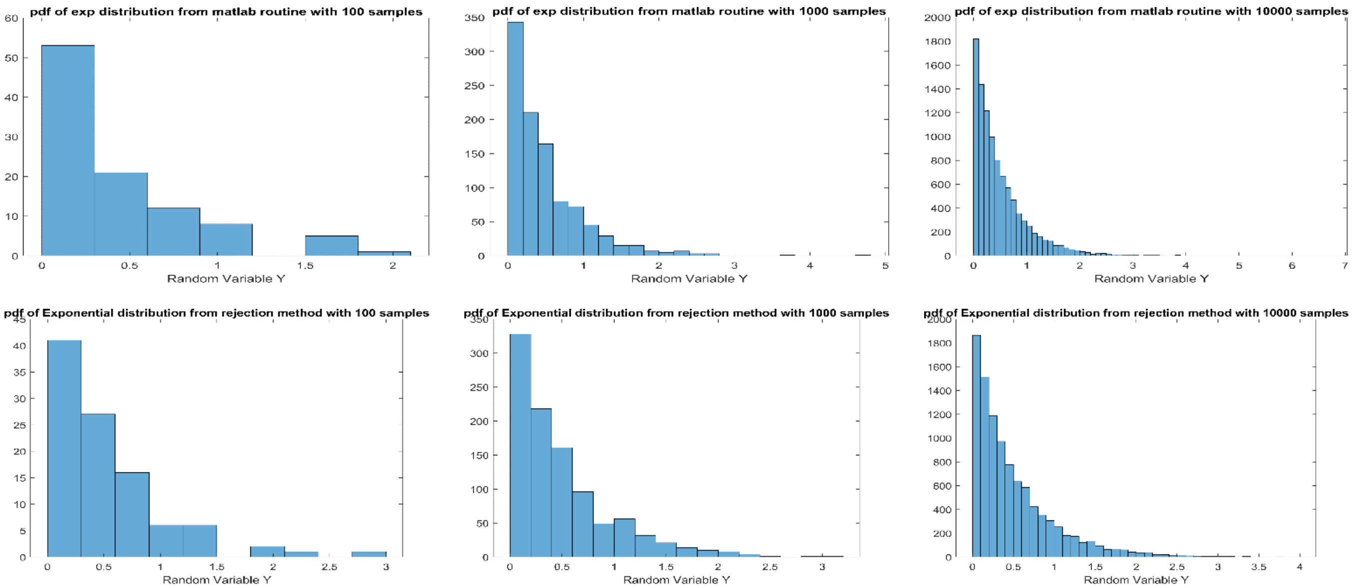

Figure 3. Simulation of Exponential variates

- Compare these empirical/computed parameters for each of the populations, to the theoretical ones. How do they compare?

  Theoretical values:

  1. Normal distribution: Mean = 2, Variance = 2
  2. Uniform distribution: Mean = 3, Variance = 0.33
  3. Exponential distribution: Mean = 0.5, Variance = 0.25

  The computed parameters for each of the distributions is provided in Tables 1, 2 and 3.

  Comparison of the Mean values:

  The deviation of the empirical/experimental mean values from the theoretical values are tabulated in Table 4 for different values of T. It can be observed that the deviation from the theoretical values reduces as we increase the number of samples in the distribution. For example, in case of the normal distribution, the absolute deviation reduces by almost 33 times from 0.1002 to 0.0031 in case of variables simulated using MATLAB routines and reduces by almost 10 times for the rejection based simulated variates when the number of samples is increased from 100 to 10000. Similar behavior can be observed in case of exponential and uniform distributions.

Number of Variates T	Mean Deviation for Normal (MATLAB Routine)	Mean Deviation for Normal (Rejection Method)	Mean Deviation for Uniform (MATLAB Routine)	Mean Deviation for Uniform (Rejection Method)	Mean Deviation for Exponential (MATLAB Routine)	Mean Deviation for Exponential (Rejection Method)
100	-0.1002	0.0799	0.0524	0.1334	0.0612	-0.0236
1000	0.0049	-.00323	0.0173	0.0075	0.0078	0.0065
10000	-0.0031	0.0098	0.0067	-0.0021	-0.0046	-0.0012

Table 4. Deviation of Mean from theoretical values

Comparison of Variance values:

The deviation of the empirical/experimental variance values from the theoretical values are tabulated in Table 5 for different values of T. It can be observed that the deviation from the theoretical values reduces as we increase the number of samples in the distribution. For example, in case of the normal distribution, the absolute deviation reduces by almost 7 times from 0.2296 to 0.0337 in case of variables simulated using MATLAB routines and reduces by almost 40 times for the rejection based simulated variates when the number of samples is increased from 100 to 10000. Similar behavior can be observed in case of exponential and uniform distributions.

Number of Variates T	Variance Deviation for Normal (MATLAB Routine)	Variance Deviation for Normal (Rejection Method)	Variance Deviation for Uniform (MATLAB Routine)	Variance Deviation for Uniform (Rejection Method)	Variance Deviation for Exponential (MATLAB Routine)	Variance Deviation for Exponential (Rejection Method)
100	-0.2296	-0.7013	0.1278	0.0068	0.0526	-0.0138
1000	-0.0245	0.0266	0.0019	0.0146	-0.0073	0.0105
10000	0.0337	-0.0175	0.003	-0.0044	-0.006	-0.0026

Table 5. Deviation of Variance from theoretical values

- If they are somewhat different, can you explain these differences? What are they due to?

  The difference is in accordance with the Strong Law of Large Numbers (SLLN) which states that as the number of samples increases the sample mean moves closer to the mean of the underlying distribution from which it was sampled.

  $$\lim_{n\to\infty} M_n \to m$$

  $$where\ M_n\ is\ the\ sample\ mean\ and\ m\ is\ the\ mean\ of\ the\ underlying\ distribution$$

  There exists a difference due to the possibility of random noise from the process of generating the random variables. If we take rejection method for instance, it depends on acceptance/rejection of a randomly generated variable (whether it lies below the pdf of target distribution or not) and there is a possibility of noise/error in this process due to the randomness of the process.

2. Transforming Random Variables

- Define $Y_j = \frac{1}{T}\sum_i X_i, i = 1, \dots T$ for the three different distributions of X_i in Q1, and compute the associated histograms for $\{Y_j\}_{j=1,\dots T}$ for each T.

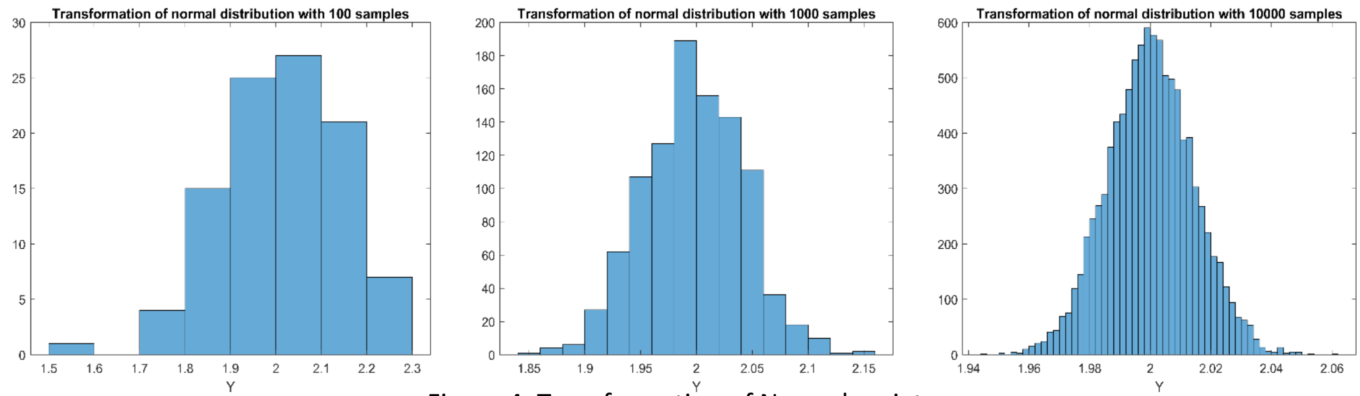

Figure 4. Transformation of Normal variates

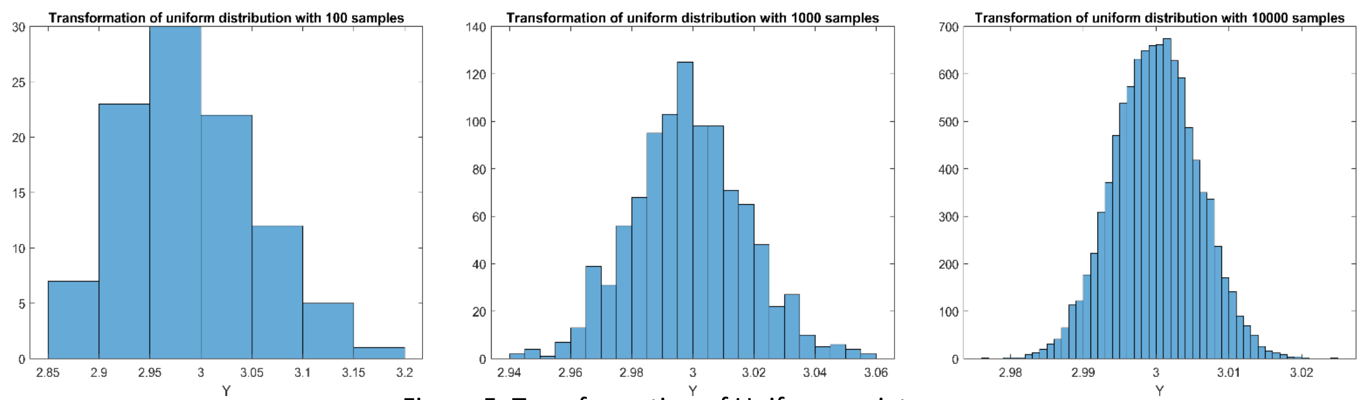

Figure 5. Transformation of Uniform variates

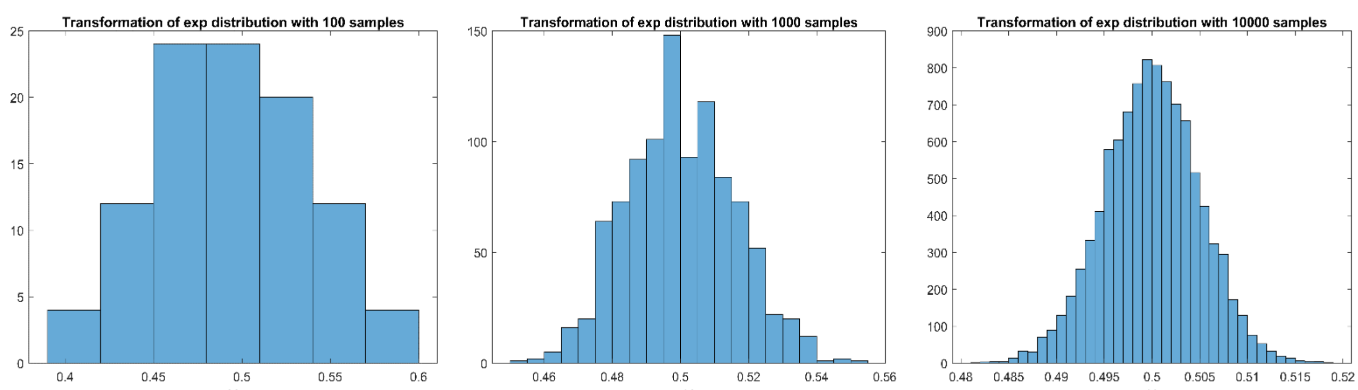

Figure 6. Transformation of Exponential variates

- By consulting standard probability density functions (PDF), find the closest PDF which matches each of the histograms for each of the T's

 The transformation Y_i corresponds to the sample mean of the X_i variables.
 Observing Figures 4, 5 and 6, we can see that all the histograms/density functions are closest to the PDF of a normal distribution. From the definition of the transformation which is the sample mean, the theoretically computed mean and variances are:

$$E[Y_j] = E[M_n] = E[\frac{1}{T}\sum_i X_i]$$

$$= \frac{1}{T}\sum_i E[X_i] = E[X]$$

$$var(Y_j) = var(M_n) = \frac{var(X_i)}{n}$$

 The mean of the set of $\{Y_j\}$ is just the mean of each Y_j and similarly has the same variance as Y_j. We can observe that the experimental results in Table 6 closely follow the theoretical results. For example, when T = 100, the resulting histogram for the uniform distribution is a normal distribution with mean 2.9897 (close to actual mean of 3) and variance 0.0041 (close to var(X)/100, where var(X) is 0.33).

T	100		1000		10000	
	Mean	Variance	Mean	Variance	Mean	Variance
Normal	2.0085	0.0184	1.9976	0.0024	2.0001	2x10^-4
Uniform	2.9897	0.0041	2.9988	3.4 x 10^-4	3	3.4 x 10^-5
Exponential	0.4947	0.0017	0.5001	2.5 x 10^-4	0.5	2.5 x 10^-5

Table 6. Experimental values of mean and variance

- How does the matching vary with T? How can you explain the variation?
 As the number of samples T (T and n are used interchangeably in the report for representing number of samples) is increased, the mean becomes closer to the mean of the underlying distribution. This is in accordance with the strong law of large numbers,

$$\lim_{n\to\infty} M_n \to m, where\ M_n\ is\ the\ sample\ mean\ (Y\ in\ this\ case)\ and\ m\ is\ the\ mean\ of\ X$$

 This explains the convergence of the resulting mean towards the mean of the underlying distribution.
 With respect to the resulting pdf, the transformation converges to a normal distribution in accordance with the Central Limit theorem. From central limit theorem we have,

$$\lim_{n\to\infty} Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i - m}{\sigma}\ and\ Z_n \sim N(0,1)$$

 Dividing by n on both sides, we get,

$$\frac{Z_n}{n} = \frac{1}{\sqrt{n}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{X_i - m}{\sigma}\right)$$

$$\frac{Z_n}{n} = \frac{1}{\sqrt{n}}\left(\frac{1}{n\sigma}\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m\right)\right)$$

$$\frac{Z_n}{n} = \frac{1}{\sigma\sqrt{n}}\left(\frac{1}{n}\left(\sum_{i=1}^{n} X_i - nm\right)\right)$$

$$\frac{Z_n}{n} = \frac{1}{\sigma\sqrt{n}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - m\right)$$

$$\frac{Z_n}{n} = \frac{1}{\sigma\sqrt{n}}(Y_i - m)$$

$$Y_i = \frac{Z_n}{n}(\sigma\sqrt{n}) + m$$

$$Y_i = \frac{Z_n}{\sqrt{n}}(\sigma) + m$$

From the scaling and shifting property of the normal distribution,

$$Y_i \sim N(m, \frac{\sigma^2}{n})$$

So, the transformation will be a normal distribution with mean converging towards m and variance $\frac{\sigma^2}{n}$ converging towards 0 and n tends to infinity.

When the sample size is small there is a possibility of sampling values which are farther away from the mean value. As the sample size n increases, the probability of sampling a value farther away from the mean value will be very low because of which the deviation is lesser. This can also be seen by how the standard deviation converges to zero as the sample size increases indicating that majority of the values sampled are closer to the mean (Property of normal distribution).

3. Convergence of Random Variables

Following the paper "Understanding Convergence Concepts: A Visual-Minded and Graphical Simulation Based Approach", establish a demo by GUI of MATLAB to show and answer the following questions based on Y_T in Q2:

$$Y_T \text{ is the sample mean defined as } \frac{1}{T}\sum_i X_i, i = 1, \dots T.$$

- Does Y_T converges in Probability to M? Why?

  From Chebyshev inequality we have,

  $$P(|Y_T - M| > \varepsilon) \leq \frac{{\sigma_{Y_T}}^2}{\varepsilon^2}$$

As indicated in the answer for Question 2, the variance of $Y_T = \frac{var(X_i)}{n}$ and as $n \to \infty$, $P(|Y_T - M| > \varepsilon)$ *tends to* 0 which indicates convergence in probability. So, Y_T converges in probability to M. This is also indicated in the MATLAB GUI based demo by considering a variable $Z_T = Y_T - M$ and shown that as T increases, Z_T converges to zero in probability.

- Does Y_T *almost surely converges to* M? *Why*?

  Y_T almost surely converges to M. Since Y_T is the sample mean, from strong law of large numbers we have that: $\lim_{T\to\infty} M_T \to M$ indicating that as T increases, the number of samples which have a value deviating from M will be zero. In the demo, the variable Z_T as defined above is monitored to count the number of times the value deviates from M by a small margin. It is shown that this number tends to 0 as the number of samples/T is increased.

- Does Y_T *converges in Mean Square to* M? *Why*?
  Y_T converges in mean square to M. From the definition of convergence in Mean square, we have: $E(|Y_T - M|^2) \to 0$ *as* $T \to \infty$
  From strong law of large numbers, we know that Y_T which is the sample mean tends to M as the value of T increases, so the difference $|Y_T - M|^2$ converges to 0.
  This is also indicated in the MATLAB GUI based demo by considering the variable $Z_T = Y_T - M$ and shown that as T increases, $|Z_T|^2$ converges to zero.

- Does Y_T *converges in Law/Distribution to* X? *Why*?
  Y_T does not converge to X in distribution. As derived above, Y_T converges to a normal distribution $N(m, \frac{\sigma^2}{n})$ in accordance with Central Limit Theorem.

To determine whether Y_T is converging to the mean of X_i we can define a variable say $Z_T = Y_T - M$ and prove that Z_T converges to 0 in the different modes of convergence (in probability, mean square, almost sure). In the GUI, the following parameters are taken as input:

1. M (Number of realizations)
2. N (Number of samples per realization)
3. Distribution type (Implemented as a drop-down list with “normal”, “uniform” and “exponential” options)

The following parameters are initialized in accordance with the paper “Understanding Convergence Concepts: A Visual-Minded and Graphical Simulation Based Approach”:

1. $\varepsilon\ (\frac{offset}{deviation} \text{ from Mean threshold})$ – 0.05
2. $k\ \left(\begin{matrix}\text{constant factor indicating the range with respect to } N \text{ for which the almost sure} \\ \text{convergence is checked}\end{matrix}\right)$– 0.5
3. $t\ (\text{The range of values at which the cdf is computed for } Y_T) -$

```
[mean - 1: 0.00001: mean + 1]
```

The theoretical mean values for the distributions as computed above are used. The following are snippets of the demo for the different distributions:

The left most figure illustrates the variation of the realizations as the number of samples increases. It can be observed that they converge towards the mean of X as n increases. The center plot illustrates the convergence in probability (blue), mean square (yellow) and almost sure convergence (red). The rightmost figure represents the Cumulative Distribution Function (CDF) of Y. The CDF of Y is a normcdf.

1. `normal: Mean = 2`

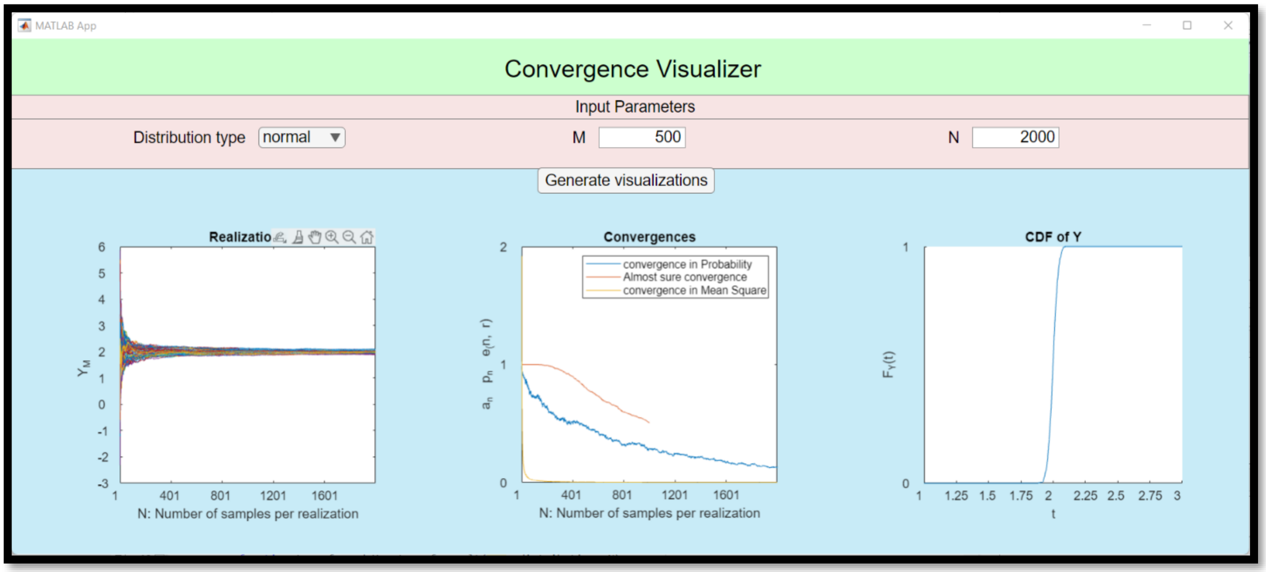

Figure 7. GUI demo for Normal distribution

2. `uniform: Mean = 3`

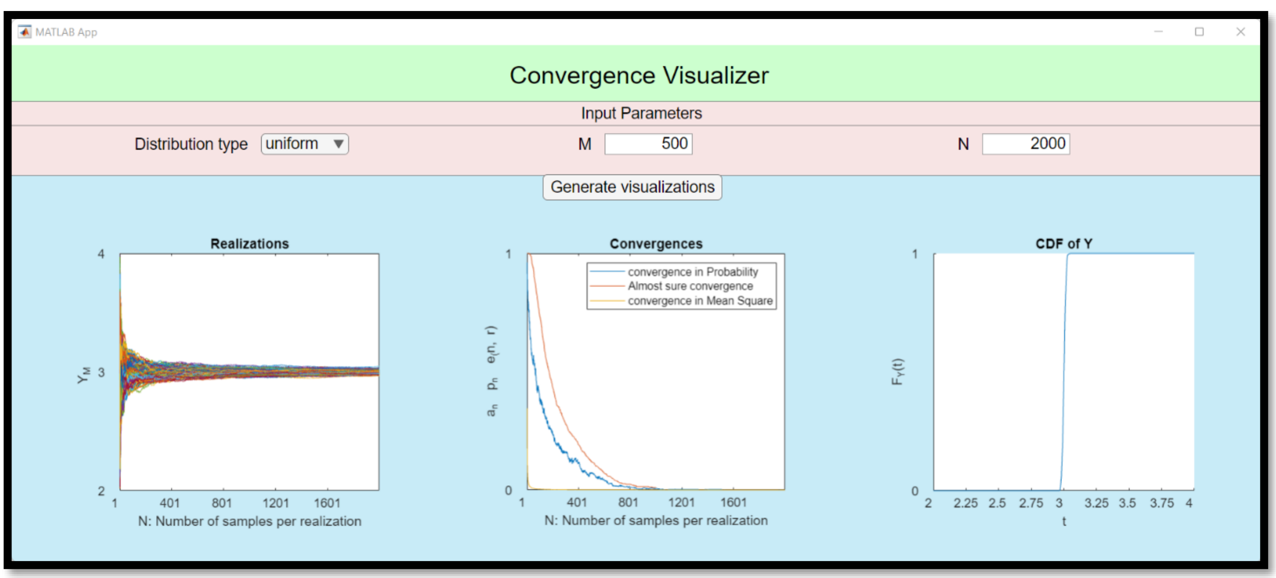

Figure 8. GUI demo for Uniform distribution

```
3.  Exponential: Mean = 0.5
```

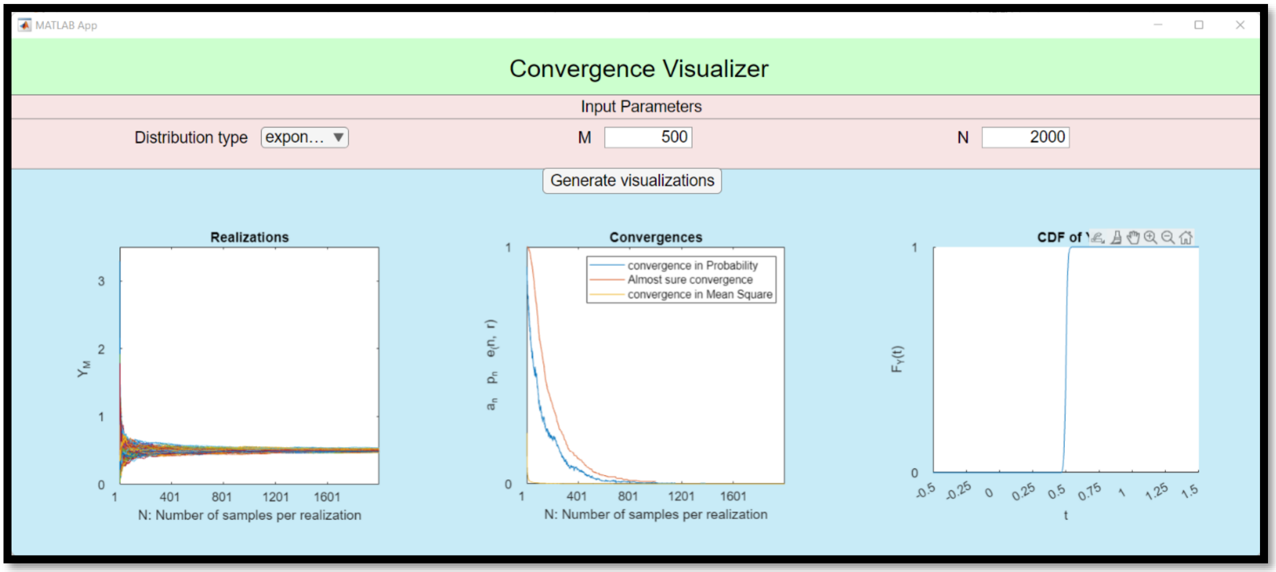

Figure 9. GUI demo for Exponential distribution