ECE 514 Random Processes - Project Report

Final Project

Name: Vikram Pande

Student ID: 200473855

Unity ID: vspande

Introduction:

With the help of MATLAB, we can plot multiple things related to Probability and Statistics. With a single line of functions, we can plot CDFs, PDFs, transform variables, make GUIs. In this Project, we have plotted standard distributions such as Normal, Exponential and Uniform with rejection and routine methods. Their transformations and the concepts of Convergence with GUI.

PART I

1. Simulating Random Variables

• Simulate random variates using both MATLAB routines as well as the rejection method, for with a PDF that is
o Normal with mean= 2 and variance=2
o Uniform on [2, 4]
o Exponential with parameter 2
• Compute the histograms for each of the cases, and estimate the parameters of each of the populations in each of the observation length cases.
• Compare these empirical/computed parameters for each of the populations, to the theoretical ones. How do they compare?
• If they are somewhat different, can you explain these differences? What are they due to?

MATLAB Routine Methods:
To generate any distribution in MATLAB, various in-built functions could be used. In this part, we have created three distributions Normal, Uniform and Exponential, plotted Histograms with in-built functions such as rand, randn, exprand, normrand for different values of samples such as 100,1000,10000 and comparing the parameters mean, variance and standard deviation.

Accept-Reject Methods:
There are many distributions from which it is very difficult to simulate. Moreover, in some cases we are not able to represent the distribution in a usable form such as transformation or a mixture. This is the scene when rejection methods come in picture. The key to this method is to use a simulation wise simpler density g from which the simulation is actually done. There is an instrumentation density called g and thus many target densities f which can be simulated as

$$f(x) \leq Mg(x)$$

Where M is constant.

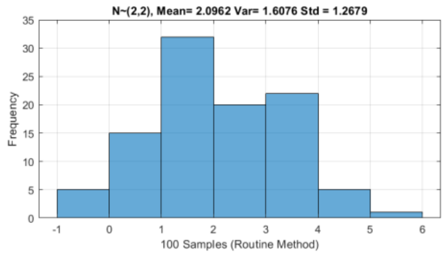
Algorithm:

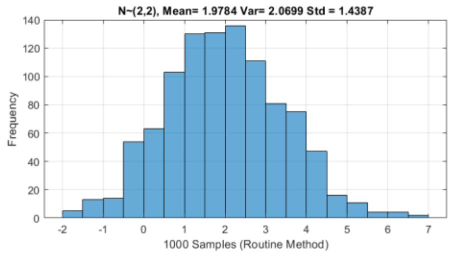
1) Generate $X \sim g$, $U \sim u[0,1]$
2) Accept $Y = X$ if $U \leq f(X)/ Mg(X)$
3) Return to 1. Otherwise.

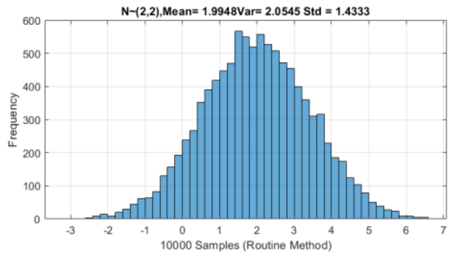
HISTOGRAM PLOTS:

NORMAL DISTRIBUTION:

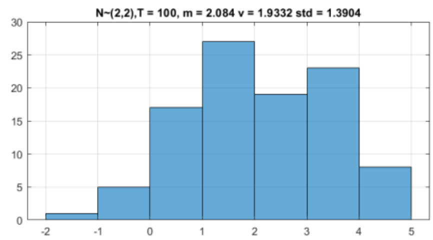
MATLAB ROUTINE METHOD:

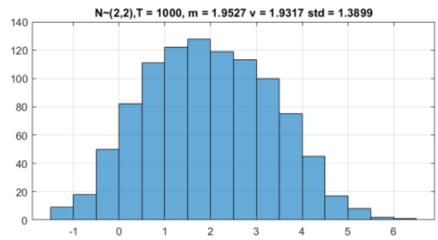


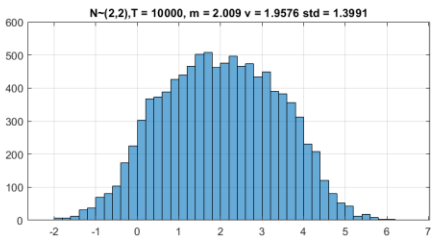




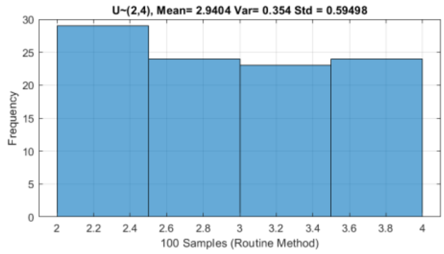
MATLAB REJECTION METHOD:

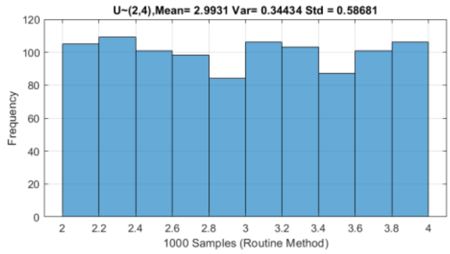


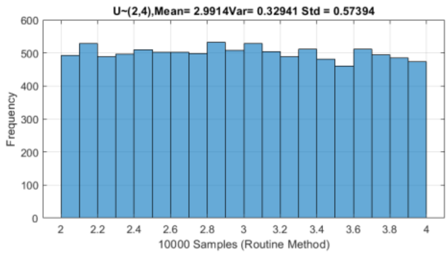




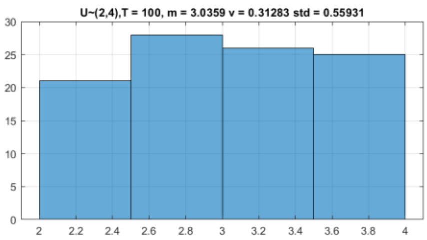
UNIFORM DISTRIBUTION:
MATLAB ROUTINE METHOD:

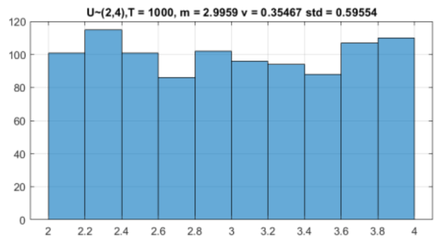


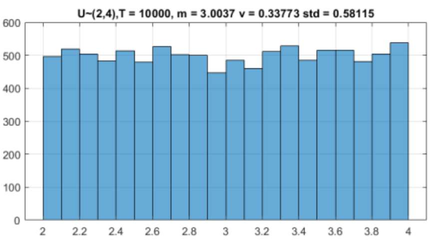




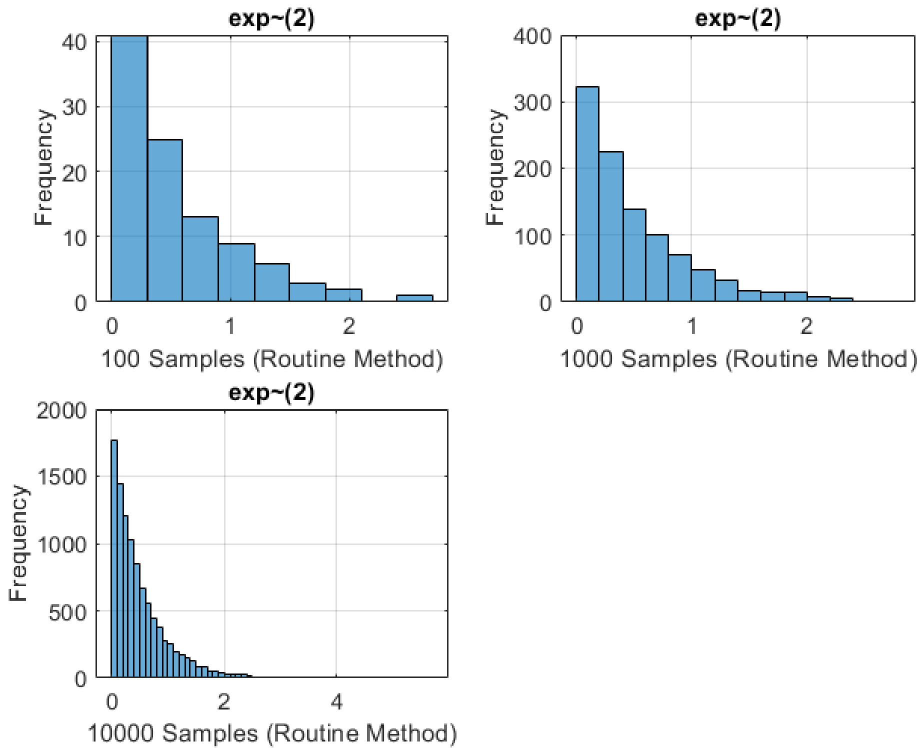
MATLAB REJECTION METHOD:



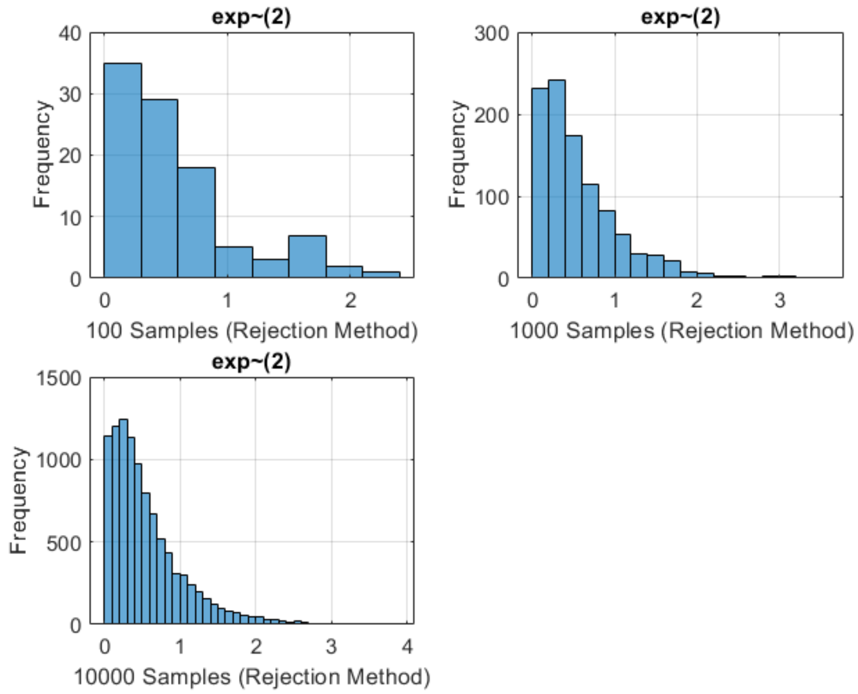




EXPONENTIAL DISTRIBUTION:
MATLAB ROUTINE METHOD:



MATLAB REJECTION METHOD:



RESULTS:

Mean, Variance and Standard Deviation for different samples using both Routine and Rejection Method.

NORMAL DISTRIBUTION $N\sim(2,2)$

Theoretical Values:

Mean = 2

Variance = 2

Samples	T = 100				T = 1000				T = 10000			
Method	Routine	Rejection	|d1|	|d2|	Routine	Rejection	|d1|	|d2|	Routine	Rejection	|d1|	|d2|
Mean	2.0962	2.084	0.0962	0.084	1.9784	1.9527	0.0216	0.0473	1.9948	2.009	0.0052	0.009
Variance	1.6076	1.9332	0.3924	0.0668	2.0699	1.9317	0.0699	0.0683	2.0545	1.9576	0.0545	0.0424
Deviation	1.2679	1.3904	0.7321	0.6096	1.4387	1.3899	0.5613	0.6101	1.4333	1.3991	0.5667	0.6009

UNIFORM DISTRIBUTION $U\sim(2,4)$

Theoretical Values:

Mean = $(a+b)/2$ = $(2+4)/2$ = 3

Variance = $(b-a)^2/12 = 1/3 = 0.333$

Samples	T = 100				T = 1000				T = 10000			
Method	Routine	Rejection	|d1|	|d2|	Routine	Rejection	|d1|	|d2|	Routine	Rejection	|d1|	|d2|
Mean	2.9404	3.0359	0.0596	0.0359	2.9931	2.9959	0.0069	0.0041	2.9914	3.0037	0.0086	0.0037
Variance	0.354	0.31283	0.021	0.02017	0.34434	0.35467	0.01134	0.02167	0.32941	0.33773	0.00359	0.00473
Deviation	0.59498	0.55931	0.09498	0.05931	0.58681	0.59554	0.08681	0.09554	0.57394	0.58115	0.07394	0.08115

EXPONENTIAL DISTRIBUTION $Exp\sim(2)$

Theoretical Values:

ʎ = 2

Mean = $1/ʎ = ½ = 0.5$

Variance = $1/ʎ^2 = ¼ = 0.25$

Samples	T = 100				T = 1000				T = 10000			
Method	Routine	Rejection	|d1|	|d2|	Routine	Rejection	|d1|	|d2|	Routine	Rejection	|d1|	|d3|
Mean	0.5483	0.5717	0.0483	0.0717	0.5049	0.5672	0.0049	0.0672	0.5049	0.566	0.0049	0.066
Variance	0.2576	0.2532	0.0076	0.0032	0.2356	0.2533	0.0144	0.0033	0.2576	0.2512	0.0076	0.0012
Deviation	0.5075	0.5032	0.0075	0.0032	0.4854	0.5032	0.0146	0.0032	0.5075	0.5012	0.0075	0.0012

COMPARISON OF VALUES:
The values of Mean, Variance and Standard Deviation are calculated for three different distributions such as Normal, Uniform and Exponential with different number of samples 100, 1000 and 1000 using both MATLAB Routine and Rejection method. Also, theoretical values are also calculated.

The values are written in the tables above. Taking the absolute values of the differences between the theoretical values and observed routine and rejection values.

$$|d1| = X - Xrout \ where \ X = given \ values, d1 = difference \ between \ routine \ and \ given \ values$$

$$|d2| = X - Xrej \ where \ X = \ given \ values, d2 = difference \ between \ rejection \ and \ given \ values$$

MEANS:
It is observed that for Normal Distribution the ideal theoretical value for mean is 2. The closest value we got is by rejection method with 1000 samples which is having less deviation in mean. It is also observed that overall, if we apply routine methods (in-built functions) to get means, we get proper mean values. In case of rejection method, rejection method is implemented manually according to the given parameters, even if we are getting results, it is less accurate than routine method in some places and vice-versa. This could be related to Law of Large Numbers. This means, if we increase the number of samples of random variables then according to LLN we will get the mean values closer to the theoretical/ideal mean. The exact same thing is happening here. Thus, we can conclude that at 1000 samples we are getting more accurate scores. The best closest value of mean in Normal Distribution is 2.009 with lowest deviation from the theoretical value of 0.009 for rejection method and also it is 0.0052 for Routine method with mean 1.9948.

For the remaining two distributions, the same pattern can be observed, the lowest deviation between the theoretical mean and observed mean for UNIFORM distribution is 3.0037 with deviation and 0.0037 for EXPONENTIAL, mean is 0.5049 with lowest deviation 0.0049 with routine method.

VARIANCE:
In all the three distribution, we can see that there are maximum errors i.e. deviation between theoretical and observed values of standard deviation for 100 samples. As we increase the number of samples, the deviation is getting minimum. For both the routine and rejection methods with 10000 samples we are getting minimum deviation and variance closer to the theoretical value. We can say that this also follows pattern of Law of Large Numbers.

EXPLAIN DIFFERENCES:
The main difference is related to the number of samples. There is a slight difference between the routine and rejection methods but a significant difference is observed when we change the samples. As we increase N i.e. the number of samples, we can see that the mean and variance are getting closer to the theoretical values. The most suitable method to trust would be routine method (in-built) MATLAB functions as they produce reliable results and if we keep the values of seed constant everywhere, we get same results and hence easier to calculate and compare. Accept-Reject method produces results same as routine for more samples. However, some times there is a possibility that it will accept some different random number as it accepts a sample by generating uniform random variables $g(x)$ and comparing the generated distribution with the distribution we want $f(x)$, for this

a constant C is used as a scaling factor. By doing so, it is a possibility that incorrect sample could be accepted resulting in deviation.

2. Transforming Random Variables

• Define Y for the THREE different distributions of X in Q1, and compute the associated histograms for each T.
• By consulting standard probability density functions (PDF), find the closest PDF which matches each of the histograms for each of the T's.
• How does the matching vary with T? How can you explain the variation?

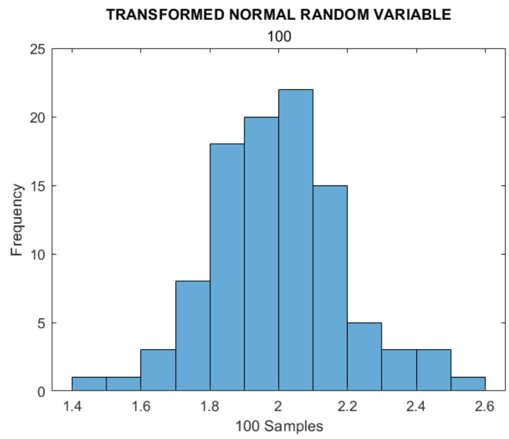
TRANSFORMATION OF RANDOM VARIABLE:
The random variable Y is defined. In the previous part, we generated the PDFs for different number of samples such as 100, 1000 and 10000 for different distributions such as Normal, Uniform and Exponential. Hence, currently there are three populations for random variable Y. Now, Y is transformed such a way that the transformation produces the sum of number of samples divided by the cumulative number of samples. In this way, transformed Y is produced. The nine histograms are plotted below with three different sample sizes.

PROBABILITY DISTRIBUTION FUNCTION MATCHING:
The closest PDF which matches each of the histograms for each of T's would be of Normal Distribution. Because Normal distribution is nothing but a Gaussian distribution. The concept is related to Central Limit Theorem, which again the most important results in probability theory. According to CLT, as we increase the number of samples with finite mean and positive finite variance, then the random variable converges to the standard normal variable. In our case, the sampling distribution of the mean(Y) can be approximated by a normal distribution with the values of population mean and population standard deviation.

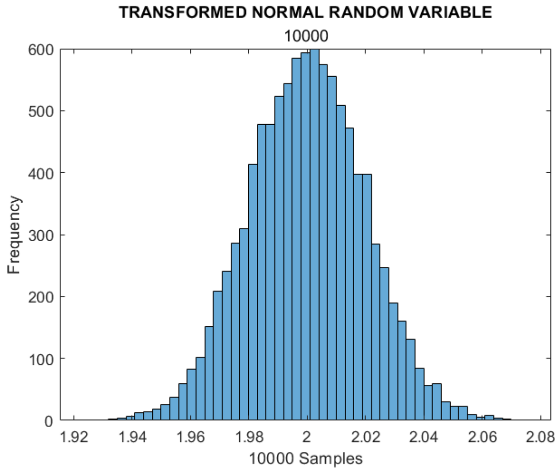
In this part, as we are increasing the value of T repeatedly, the values of samples are 100, 1000, 1000 and when T is large enough i.e. 10000, the distribution is approaching to normal distribution.
We can see the changes in all three distributions, as we transform the original random variable, we see that we are getting a Gaussian (Normal) distribution. Matching is varying with T in such a way that when T is increasing the peak of distributions changes in size and getting a narrow shape.

The changes observed are as following:

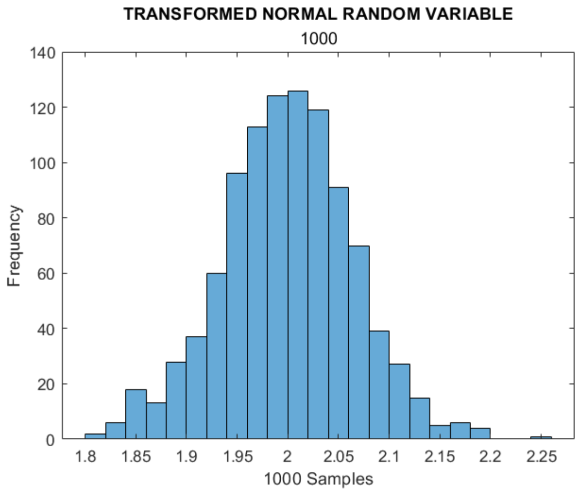
NORMAL DISTRIBUTION:



100 SAMPLES

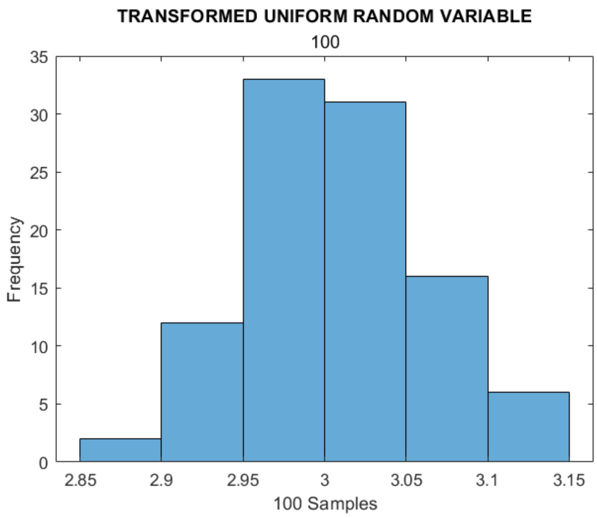


1000 SAMPLES

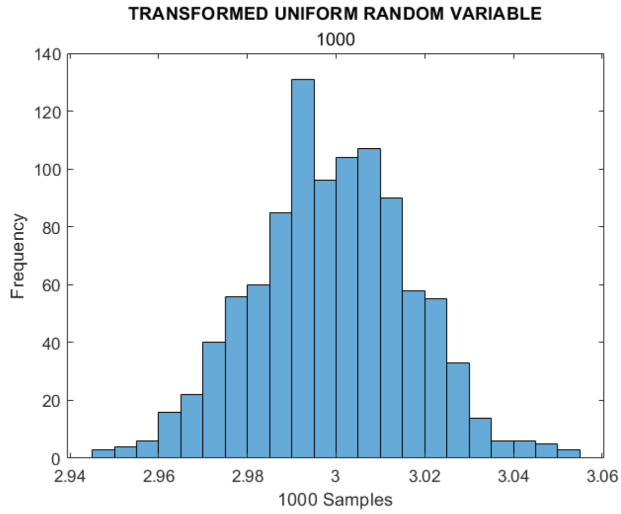


10000 SAMPLES

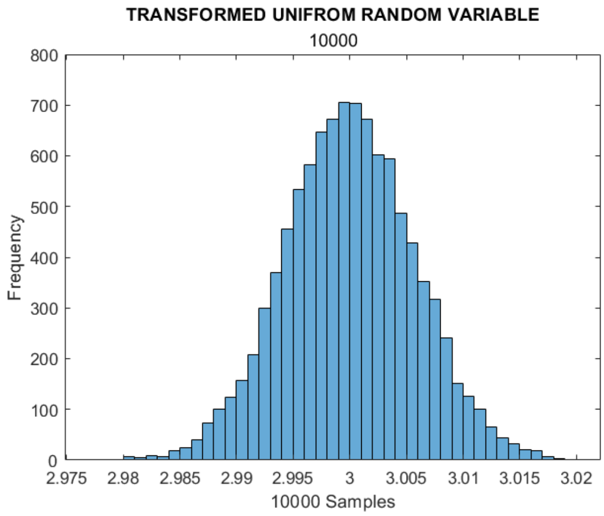
UNIFORM DISTRIBUTION



100 SAMPLES

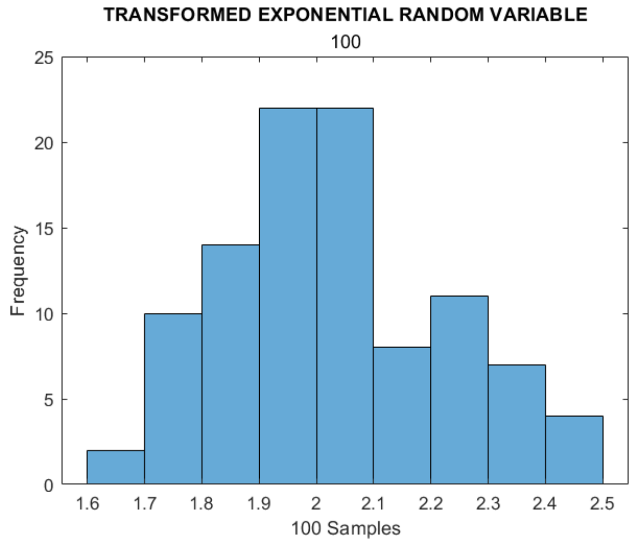


1000 SAMPLES

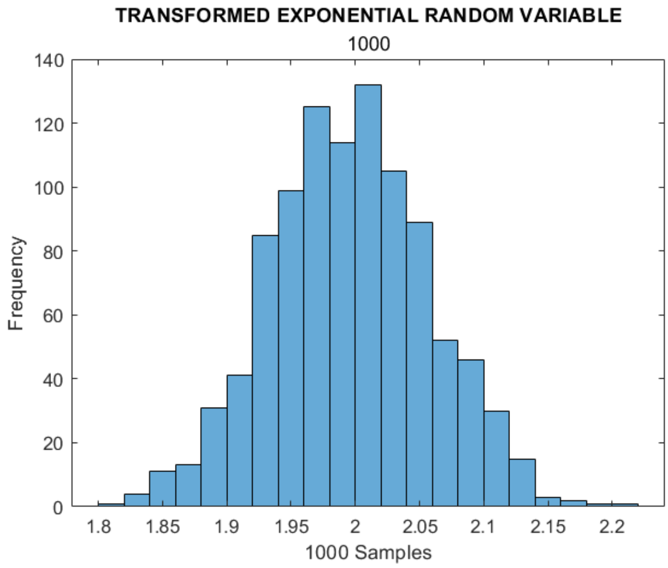


10000 SAMPLES

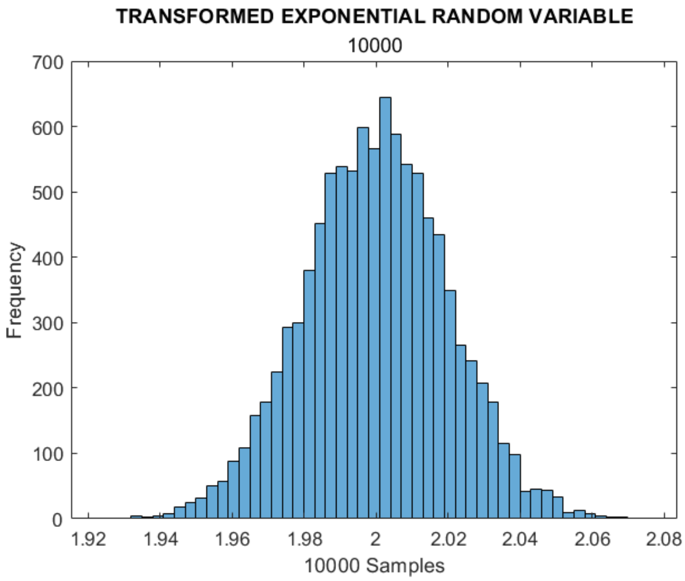
EXPONENTIAL DISTRIBUTION:



100 SAMPLES



1000SAMPLES

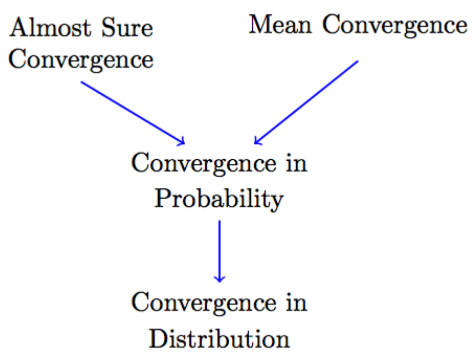


10000 SAMPLES

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.

The GUI is created for the current task showing the proofs of convergence for all three distributions. The strength of convergences is shown below. According to the paper, the parameters are selected. There are 500 realizations, maximum number of samples is 2000 and epsilon value is taken as 0.05. The convergence of the transformed variable is consistent and follows LLN again. In short, the value of normalized mean say M-bar with high probability tends toward the value of mean of the distribution depending upon the samples.



For Normal distribution, it is observed that Y is converging in the mean value of X i.e. $2\ E[X] = 2$. Similarly, for Uniform distribution, it is observed that Y is converging in the mean value of X i.e. $3\ E[3] = 3$ and for Exponential distribution, it is observed that Y is converging in the mean value of X i.e. $0.5\ E[X] = 0.5$.

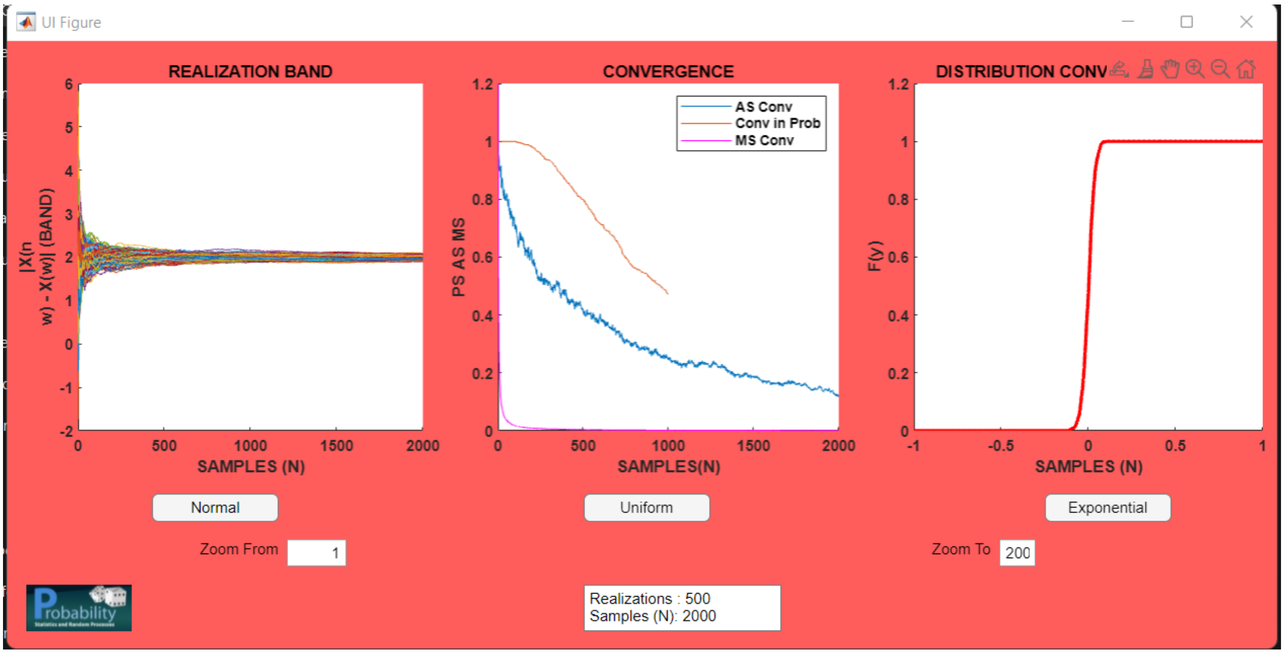
ALMOST SURE CONVERGENCE:
The value of K is taken as 0.5 to ensure sufficient future observations.
We can see from the graphs below that Xw,n is converging into Xn almost surely. The graph is decreasing and touching 0.

MEAN SQUARE CONVERGENCE:
We can see from the graphs below that Xw,n is converging into Xn power of 2.

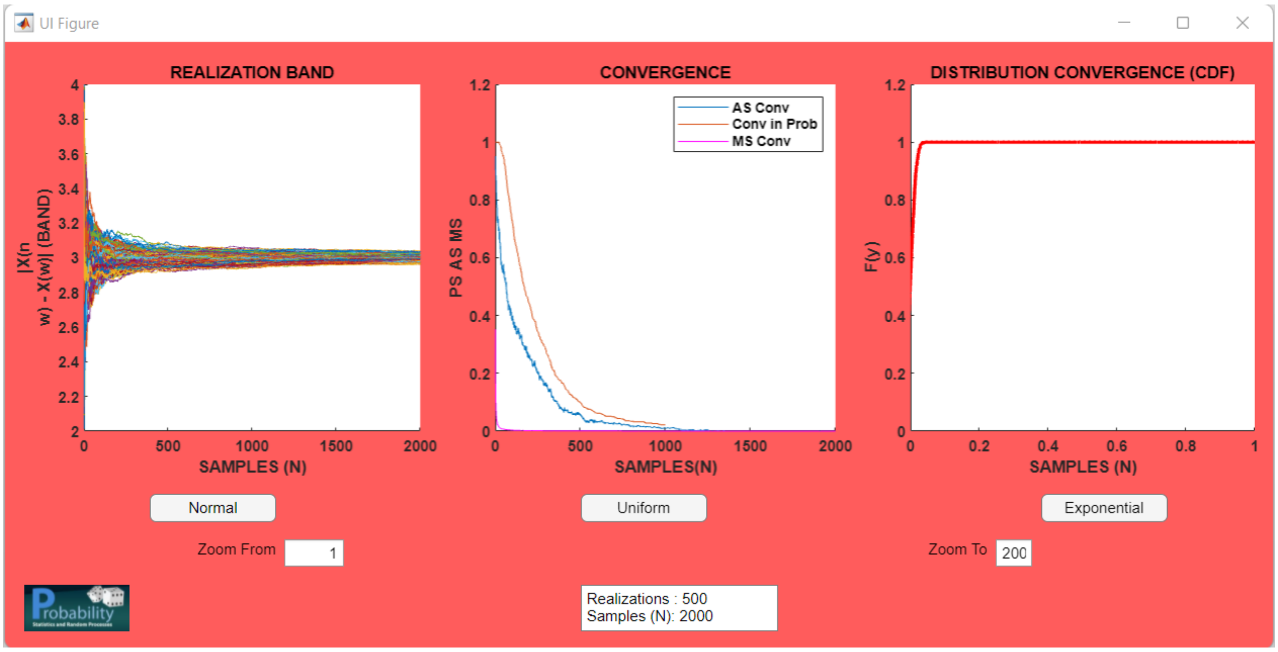
CONVERGENCE IN PROBABILITY:
We can see from the graphs below that Xw,n is converging into Xn probability.

CONVERGENCE IN DISTRIBUTION:
We can see from the graphs below that Xw,n is converging into Xn in distribution.

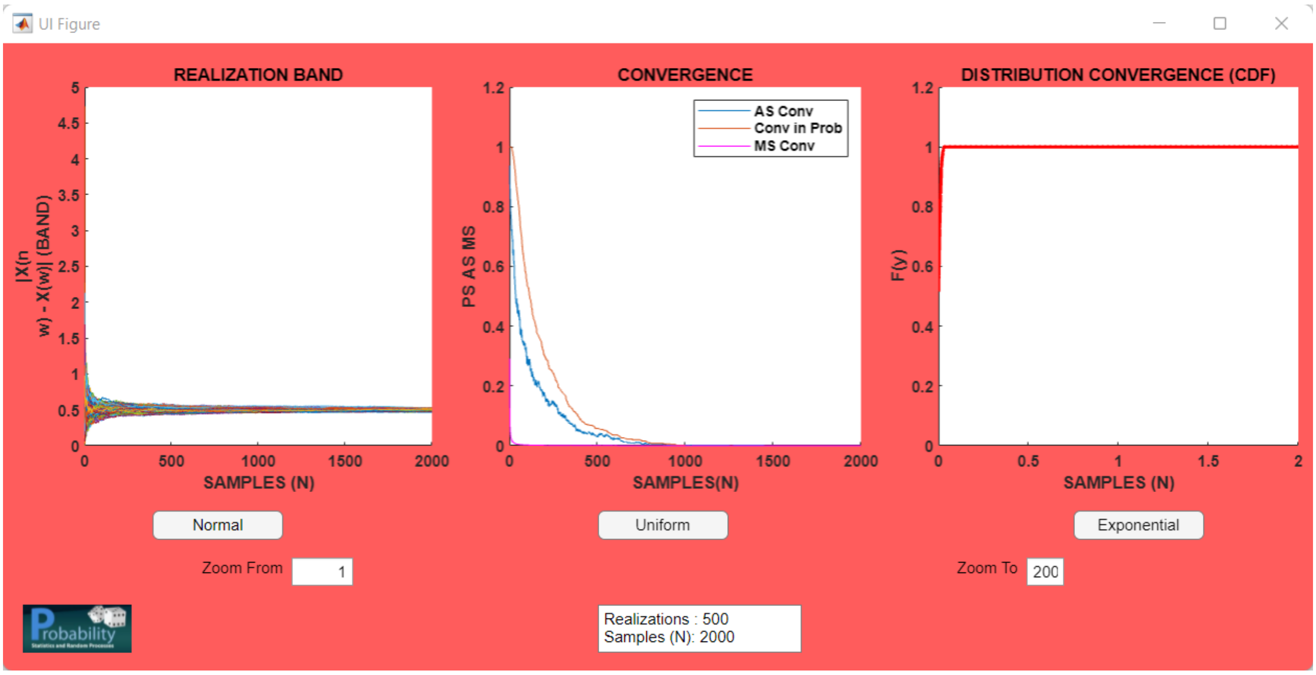
CONVERGENCE NORMAL DISTRIBUTION



CONVERGENCE UNIFORM DISTRIBUTION



CONVERGENCE EXPONENTIAL DISTRIBUTION



PART II

Two jointly continuous random variables with joint PDF

$$f_{XY}(x,y) = \begin{cases} x + \frac{3}{2}y^2, & x < 0,, y \le 1 \\ 0, & otherwise \end{cases}$$

$$U = \binom{X}{Y}$$

1) Correlation and Covariance Matrices of U. (Mathematically)

I) Finding the Marginal PDFs $fx(X)$ and $fy(Y)$.

$Fx(X) = \int_0^1 f_{xy}(x,y)dy = \int_0^1 x + \frac{3}{2}y^2 dy = x + \frac{1}{2}$

$Fy(Y) = \int_0^1 f_{xy}(x,y)dx = \int_0^1 x + \frac{3}{2}y^2 dy = \frac{1+3y^2}{2}$

II) Finding the moments of marginal PDFs (Mean, Variance and Second order Moment)

$E[X], E[X^2]\,, var(X)$:

$E[X] = \int_0^1 x f_x(X)dx = \int_0^1 x\left(x + \frac{1}{2}\right)dx = \frac{7}{12}$

$E[X^2] = \int_0^1 x^2 f_x(X)dx = \int_0^1 x^2\left(x + \frac{1}{2}\right)dx = \frac{5}{12}$

$var(X) = E[X^2] - (E[X])^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$

$E[Y], E[Y^2]\,, var(Y)$:

$E[Y] = \int_0^1 y f_y(Y)dy = \int_0^1 y\left(\frac{1+3y^2}{2}\right)dy = \frac{5}{8}$

$E[Y^2] = \int_0^1 y^2 f_y(Y)dy = \int_0^1 y^2\left(\frac{1+3y^2}{2}\right)dy = \frac{7}{15}$

$var(Y) = E[Y^2] - (E[Y])^2 = \frac{7}{15} - \left(\frac{5}{8}\right)^2 = \frac{73}{960}$

$E[XY]$:

$E[XY] = \iint_0^1 f_{xy}(x,y)dxdy = \iint_0^1 xy\left(x + \frac{3}{2}y^2\right)dxdy = \frac{17}{48}$

$cov(X,Y) = E[XY] - E[X].E[Y] = \frac{17}{48} - (\frac{7}{12})(\frac{5}{8}) = \frac{-1}{96}$

III) Covariance Matrix and Correlation Matrix

The covariance Matrix Cu is given by

$$C_U = \begin{bmatrix} var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix} = \begin{bmatrix} \frac{11}{144} & \frac{-1}{96} \\ \frac{-1}{96} & \frac{73}{960} \end{bmatrix} = \begin{bmatrix} 0.0763888 & -0.0104166 \\ -0.0104166 & 0.07604166 \end{bmatrix}$$

The correlation Matrix Ru is given by

$$R_U = \begin{bmatrix} E[X^2] & E[XY] \\ E[XY] & E[Y^2] \end{bmatrix} = \begin{bmatrix} \frac{5}{12} & \frac{17}{48} \\ \frac{17}{48} & \frac{7}{15} \end{bmatrix} = \begin{bmatrix} 0.4166666 & 0.3541666 \\ 0.3541666 & 0.4666666 \end{bmatrix}$$

2) Cholesky Method: The basic use of Cholesky method is to calculate the decomposition matrix L. It is often used in Monte Carlo Simulations. It is a decomposition of a Hermitian, PD (Positive-Definite) Matrix into the multiplication of lower triangular matrix and respective conjugate response.

   We have now generated a 1000 sample vector series in MATLAB, Xs with Covariance same as of U and estimated $cov(X_s)$ from MATLAB. The MATLAB file is attached and results were as follows: (The code is attached in filename: part2.ml)

```
>> Ex

Ex =

7/12

>> Ey

Ey =

5/8

>> Ex2

Ex2 =

5/12

>> Ey2

Ey2 =

7/15

>> varx

varx =

11/144

>> vary

vary =

73/960

>> Exy
```

```
Exy =

   0.3542

>> covxy

covxy =

-0.010417

>> covariance_matrix

covariance_matrix =

[ 0.076389, -0.010417]
[-0.010417,  0.076042]

>> correlation_matrix

correlation_matrix =

[0.416667, 0.354167]
[0.354167, 0.466667]

>> Xs

Xs =

[ 0.083712, -0.010437]
[-0.010437,  0.081431]

>>
>> estimate

estimate =

[-0.007323,   0.00002]
[  0.00002, -0.005389]

>> Xs1

Xs1 =

[ 0.081381, -0.010513]
[-0.010513,  0.072391]

>> estimate 1

estimate1 =

[-0.004992, 0.000096]
[ 0.000096, 0.003651]

>> Xs2

Xs2 =

[ 0.076976, -0.010753]
[-0.010753,  0.075508]

>> estimate2

estimate2 =
```

[-0.000587, 0.000336]
[0.000336, 0.000534]

We observed the values of $cov(U)$ and $cov(Xs)$. There is a slight difference when Xs was 1000. To estimate Xs, we tried to increase the sample size to 2000 and 4000. The observation is that if we want to improve the estimate of $cov(Xs)$ then after increasing the sample vectors, we can estimate $cov(Xs)$. From the above results, Refer the values of estimate, estimate1, estimate2