Random Processes- Project 1

Binoy Thomas, North Carolina State University 11/09/2018

Question 1 – Simulating Random Variables

1.1 Simulate Random variables using Matlab Routine and Rejection Method

Rejection Method

The Acceptance- Rejection method is an algorithm which is used to generate random samples from a probability distribution that is not known to us. The Acceptance- rejection method is better than the other methods because it doesn't need the cumulative distribution nor its inverse to be computed.

If we assume the span of the random variable as (rmin, rmax), then generate X ~ F = U(rmin, rmax) and V ~ U(0,1) independently. The random variable Y is given as,

$$Y = X : V \leq \frac{g(X)}{M * f(X)}$$
$$where$$
$$M = \sup \left\{ \frac{g(X)}{f(X)} \right\}$$

Matlab routine

The Matlab routine to generate random variables includes the use of Matlab functions like rand,normrnd and exprnd to generate sequences of pseudorandom numbers.

1.2 Computing the histograms and estimating the parameters

The desired signal F(X) can be any signal which is present in the range of interest. The number of bins can be changed to observe better resolution of the histogram. Figure 1 Shows the PDFs of normal distribution with mean as 2 and variance as 2. Figure 3 has the Uniform distribution on [2,4] and Figure 2 shows the exponential distribution with parameter 2. All the distributions are plotted using the Matlab routine and the rejection method. The number of bins is assumed to be 50. The parameter values have been computed and given in the legends of the figures.

The PDF can be computed from the histogram as,

$$g(x) = \frac{h(y : argmin_y\, |y - x|)}{T\delta}$$

where y takes discrete values in the range (r_{min}, r_{max}) in steps of $\delta = [(r_{max} - r_{min})/Nbins]$ (Nbins is the number of histogram bins). Moreover, we shall assume that the histogram bins are uniformly spaced. g(x) is obtained for the samples generated by the rejection method and the Matlab routine.

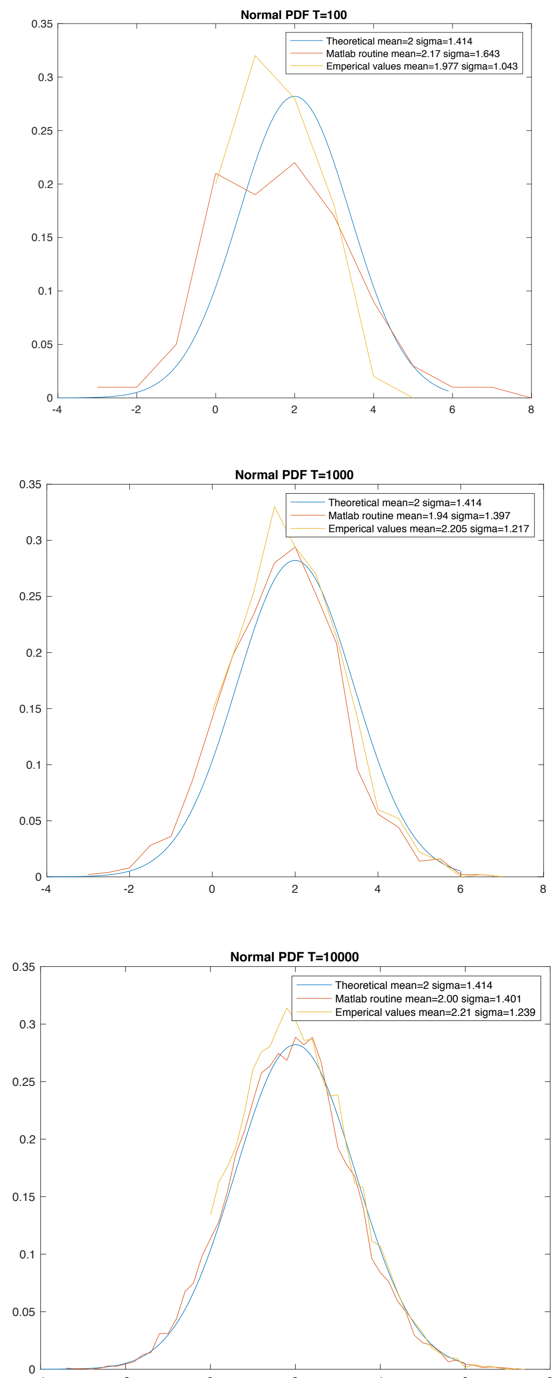

Figure 1: Normal Distribution with mean =2 and variance=2

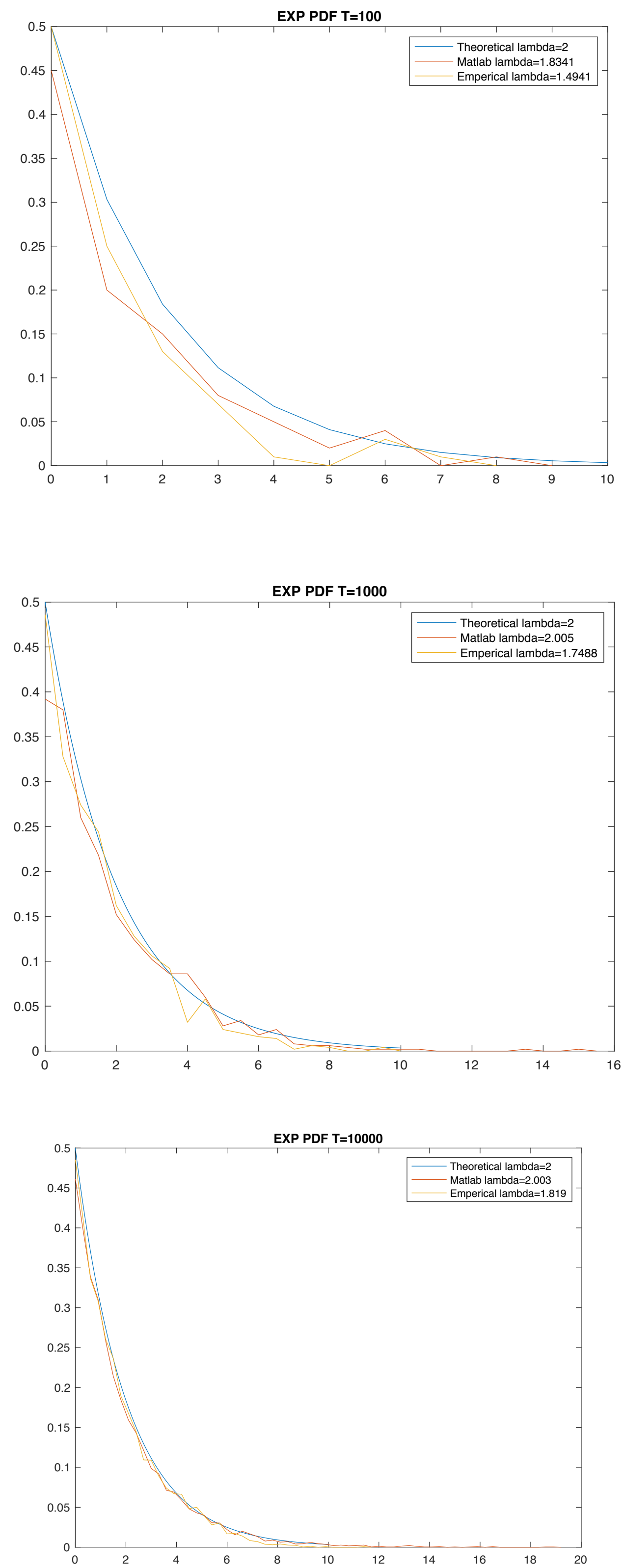

Figure (2): Exponential distribution with parameter 2

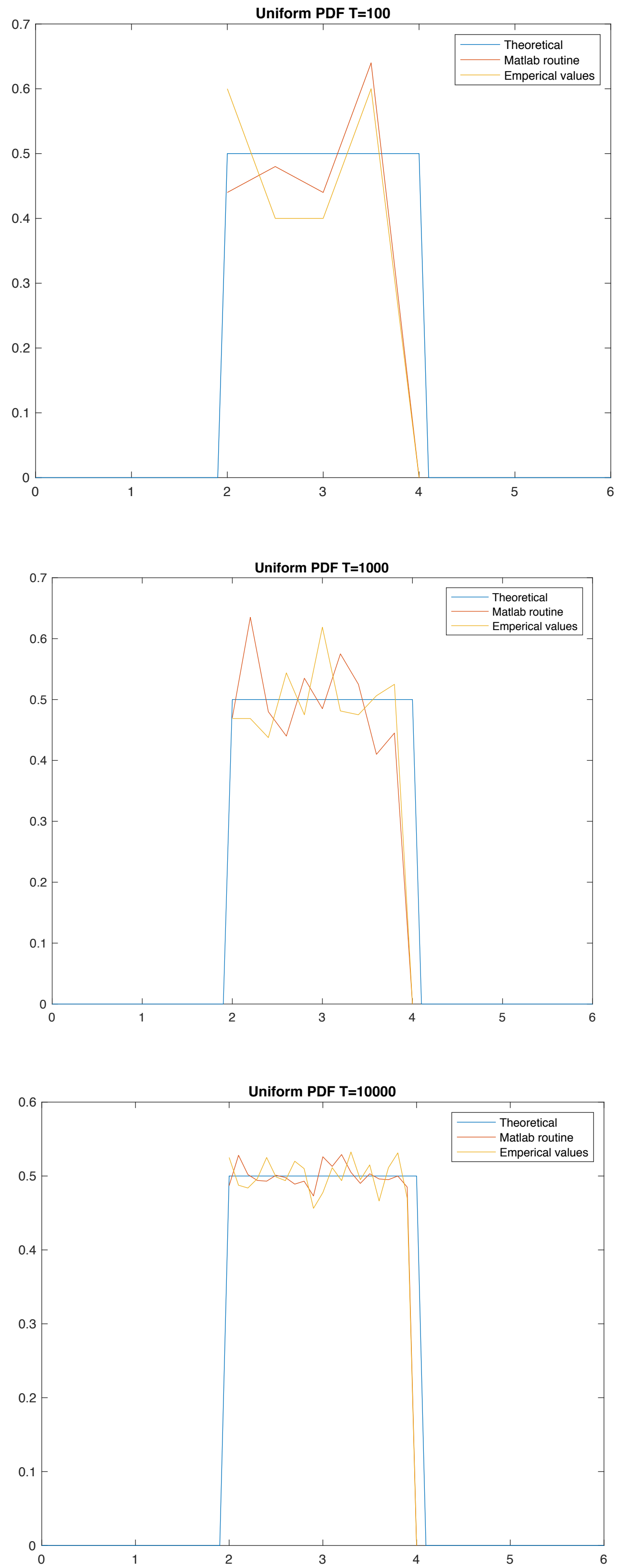

Figure (3): Uniform distribution between[2,4]

1.3 Comparison of the Emperical and Theoretical parameters

The plots of the probability density functions approach the theoretical values as the value of T increases. As T grows to 10000, the empirical plot almost overlaps the theoretical plot. We need to have a unified parameter to compute the error in the estimation. The error in estimation can be obtained by using$\Delta\tau_{emp} = \tau_{emp} - \tau_{th}$and similarly for the Matlab values as $\Delta\tau_{mat} = \tau_{mat} - \tau_{th}$. The mean square error can be computed for each of the cases and it is shown that the mean square error decreases as the values of T increases. For comparison we use the parameters as mean and variance for the Gaussian and Uniform cases and the parameter λ for the exponential case.

1.4 Reason behind differences in the parameters

The value of T is 100,1000 and 10000. T signifies the sample points to represent the probability density function. The PDF generated for fewer samples will experiences some disparities compared to the theoretical values. Hence, as T increases the computed values will approach the theoretical values.

Question 2 – Transformation of Random Variables

2.1 Transform Xi from part 1.1

The random variables obtained in Part 1.1 are added together for each of the distributions and then divided by twice the value of its index. This results in most of the values to be obtained at values close to one. As the value of T increases the distribution becomes more densely populated together.

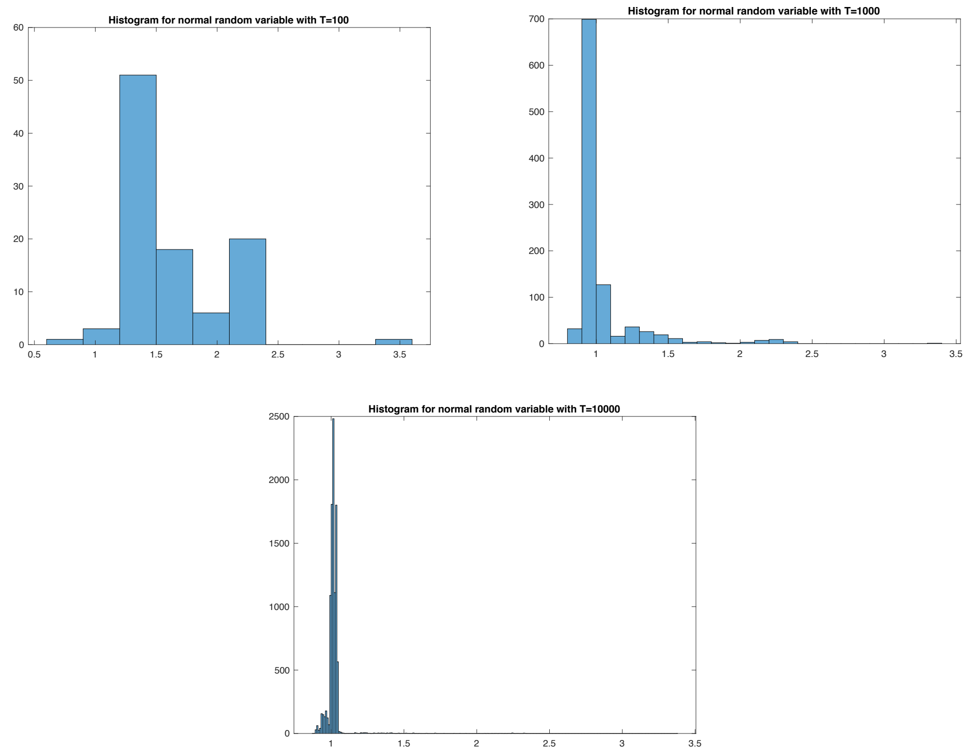

Figure (4): Histogram plot for normal random variables

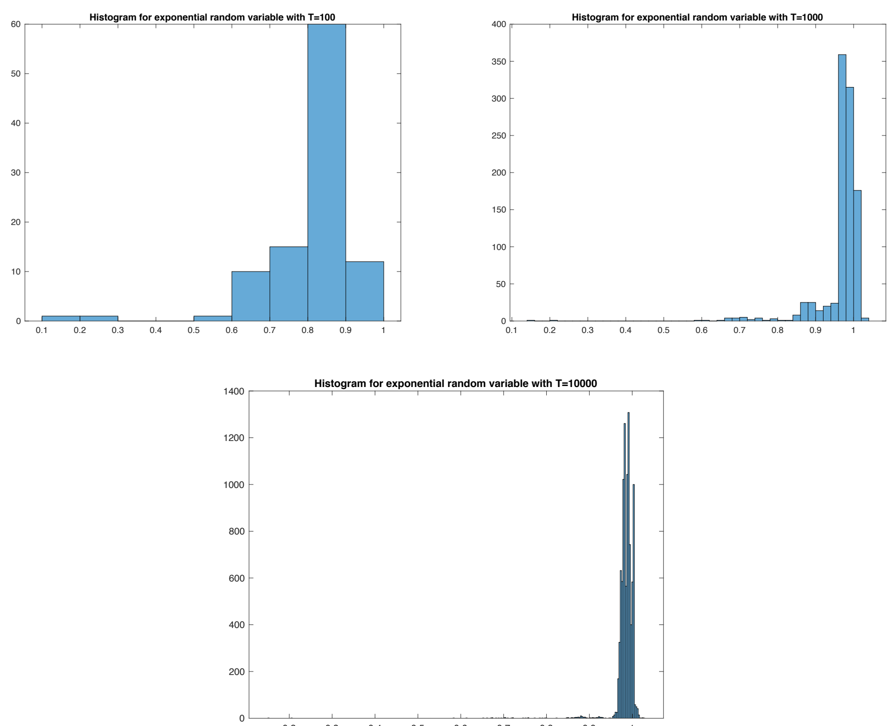

Figure (5): Histogram plot for exponential random variables

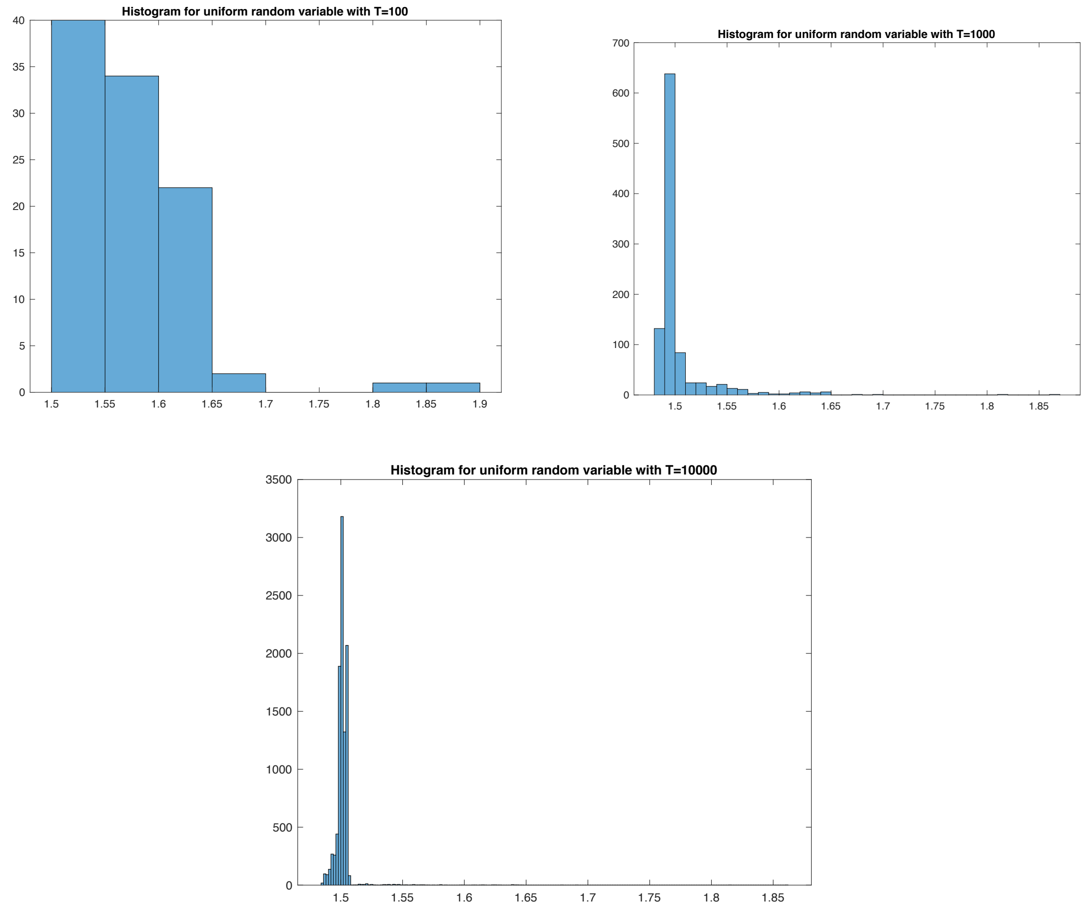

Figure (6): Histogram plot for uniforml random variables

2.2 Determination of closest PDF

The PDF obtained from part 2.1 is a function with values which are highly populated at the value one. The histogram for T=10000 shows that there are almost 8000 in the bin at 1. This shows that the value converges to 1 and the value of T increases. This is proved by the demonstration in Question 3. The standard density function which matches the histograms is the normal density function. We can use the function histfit and fitdist and find out the normal probability density function which most closely matches the histograms in 2.1.

2.3 Variation of matching with T

From the figure 4,5,6 It is observed that the shape of the empirically derived PDF gets closer and closer to the value 1. The number of bins is fixed to 50. T signifies the sample points to represent the PDF. The PDF obtained with a smaller number of samples will have some differences as compared to the theoretical values. To evaluate these differences we use the Mean Square Error (MSE). The MSE is inversely proportional to the sample size T.

Question 3 – Convergence of Random Variables

Convergence of random variables concerns itself with the limiting behavior of the sequence of random variables X_1 ,X_2 , X_3 .....X_n. It looks into what happens if we gather more and more data together. There are basically four types of convergence

1. Convergence in Probability
2. Almost sure convergence
3. Convergence in r^{th} mean
4. Convergence in Law

A GUI was created using app designer in Matlab which can be used to pedagogically view how the variables behave. It consists of a plot whose behaviors change depending on the value of the number of samples T and the value of epsilon(ε)

3.1 YT→0(Convergence in probability)

Y_T is obtained and when we add the random variables and divide it by twice the index of the random variable. For observing convergence in probability, we assume a parameter epsilon ε which is the deviation. A sequence is said to be convergent in probability if

$$p_n = P[|Y_T| > \varepsilon] \to 0 \; as \; n \to \infty$$

Through figure…, we can clearly see that the sequence Y_T converges to and as the values of n goes on increasing, the deviation from the value where it converges goes on decreasing. The band that is chosen i.e. the values of ε can be considerably narrow.

For experimentation, we take the values of T to be 100,1000 and 10000 and the value of ε to be 0.01.

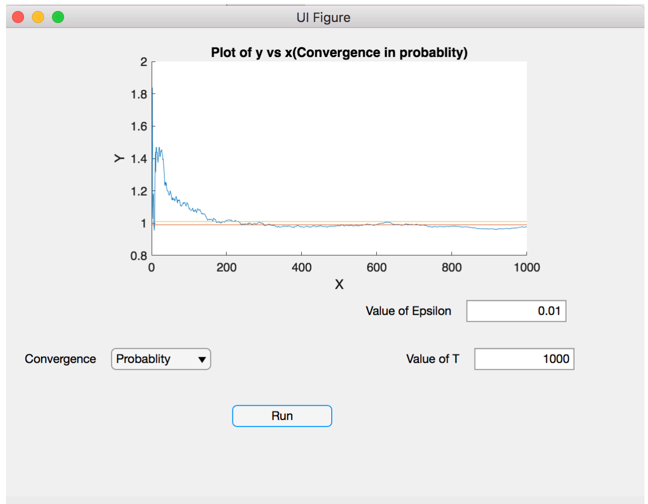

Figure (7): Matlab GUI output for Convergence in Probablity

Value of T=10000

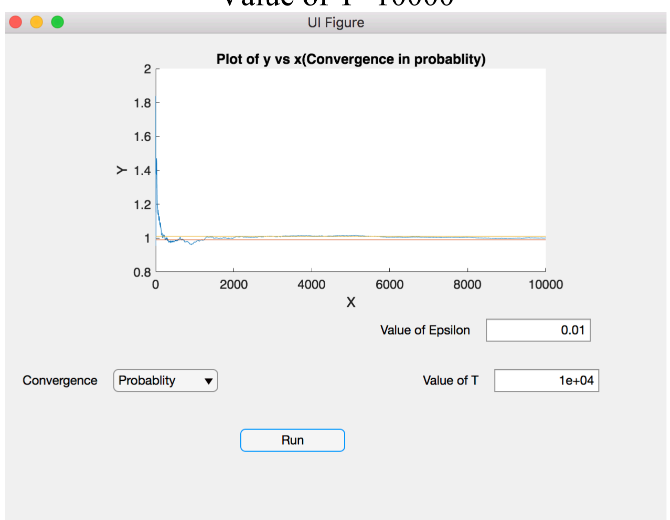

Figure (8): Matlab GUI output for Convergence in Probablity(T=10000)

As the value of T increases the number of points which are outside the bar of epsilon decreases and hence a higher number of sample points will ensure that the graph converges to the desired value

3.2 $Y_T \rightarrow 0$(Almost sure convergence)

We can say that Y_T converges almost sure to 0 if

$$P[\omega; Y_T = 1] \rightarrow 0 \; as \; n \rightarrow \infty$$

This is observed from the implementations in figure…. The difference between convergence in probability and almost sure convergence is that for almost sure convergence there must exist a value k such that k≥n such that the value of Y_k tends to the value 1. This can be proved by comparing the values for T=1000 and T=10000. The value of Y_T converges to 1 for large values of n and there always exists a value k where we can obtain convergence

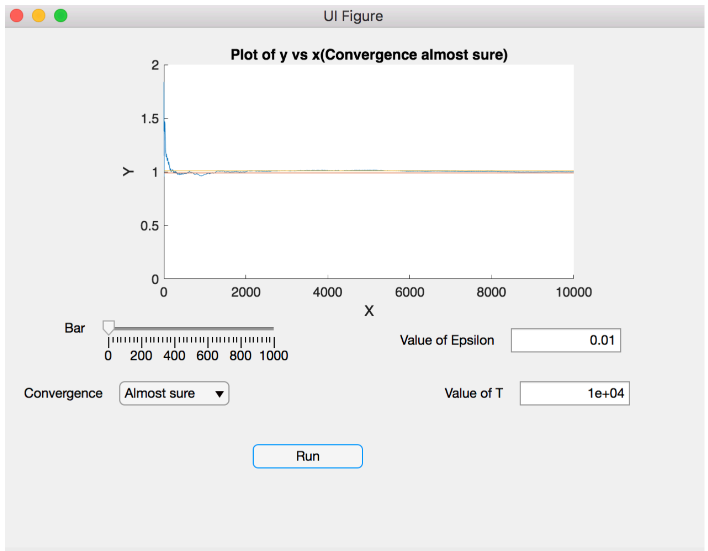

Figure (9): Matlab GUI output for Almost Sure Convergence(T=10000)

This is observed by observing the figure … Hence, we can say that Y_T almost surely converges to 0.

3.3 $Y_T \rightarrow 0$(2nd mean convergence)

We can say that this is true iff

$$\lim_{n\to\infty} E[|Y_T|^r] = 0$$

To prove that Y_T is convergent in the 2nd mean we consider the square of Y_T. This would be a value close to 0 because the expectation is calculated by multiplying the index of the value and the square of the value. We can say that since this value is very close to 0, the 2nd mean convergence of Y_T converges to 0

3.4 $Y_T \rightarrow X$(Convergence in Law)

In order to prove this, we have to consider the cdf of Y_T. We can say that Y_T converges in Law to X if and only if

$$\lim_{n\to\infty} F_{Y_T}(x) = F_X(x)$$

The mind visualization techniques is useful to prove this statement. The values of Y_T are computed by summing the values of X and then dividing it by its index. Thus Y_T contains the values of X with a little work done on them. The PDF of Y_T will contain the vales of X. Hence if we cumulatively add the values of Y_T it will be the same as X for very large values of n.

References

[1] Athanasios Papoulis and S Unnikrishna Pillai. Probability, random variables, and stochastic processes. Tata McGraw-Hill Education, 2002.