Name: Gunjan Jhawar

Subject code: EE 511

Simulation of Scholastic System

USCID: 1857534264

Project: 6

Aim:

Let $\{X_i\}$ be iid exponentially distributed random variables with $E[X_i] = \frac{1}{\lambda} = 1$.

Define the random variable $S_n = \sum_{i=1}^{n} X_i$. The distribution of S_n is called an Erlang-*n* distribution.

The pdf for S_n can be shown to be:

$$f_{S_n}(x) = \frac{(\lambda x)^{n-1}}{(n-1)!} \lambda e^{-\lambda x}$$

And the CDF can be expressed as:

$$F_{S_n}(x) = 1 - \sum_{j=0}^{n-1} \frac{1}{j!} (\lambda x)^j e^{-\lambda x} \quad \text{(defining } 0! = 1 \text{ for convenience)}$$

In this project, you are asked to generate samples of S_n for $n = 5$ by three different methods

A. By means of a simulation experiment, generate values for X_i and use them to generate a series of samples for S_n. Compute the sample mean and sample variance for $n = 5$ and compare to the analytical value. Also plot a histogram approximation of the probability density function, $f_{S_n}(x)$.

B. A more (computationally) efficient way to generate samples for S_n uses:

$$S_n = \frac{-\ln\left(\prod_{i=1}^{n} U_i\right)}{\lambda}$$

Where $U_i \sim Uniform(0,1)$. (Why does this work?) Use this approach to generate a series of samples for S_n and compute mean and sample variance, and compare to part A

C. Compare the time it takes to generate (10,000, say) samples of S_n using these 3 approaches.

Software Used:

MATLAB

Theory:

The **Erlang distribution** (sometimes called the Erlang-k distribution) was developed by A.K. Erlang to find the number of phone calls which can be made simultaneously to switching station operators.
Erlang's distribution has since been expanded for use in queuing theory, the mathematical study of waiting in lines. It is also used in stochastic processes and in mathematical biology.

The Erlang distribution is a specific case of the Gamma distribution. It is defined by two parameters, k and &u, where:

- k is the shape parameter. This must be a positive integer (an integer is a whole number without a fractional part). In the Gamma distribution, k can be any real number, including fractions.
- μ is the scale parameter. Must be a positive real number (a real number is any number found on the number line, including fractions).
-

The Erlang distribution with shape parameter k simplifies to the exponential distribution. [1]

Statistics of the Erlang Distribution

- Mean: k / λ.
- Variance: k / λ^2. [2]

A conceptually very simple method for generating exponential variates is based on inverse transform sampling: Given a random variate U drawn from the uniform distribution on the unit interval (0, 1), the variate

$$T = F^{-1}(U)$$

has an exponential distribution, where F^{-1} is the quantile function, defined by

$$F^{-1}(p) = \frac{-\ln(1-p)}{\lambda}.$$

Moreover, if U is uniform on (0, 1), then so is $1 - U$. This means one can generate exponential variates as follows:

$$T = \frac{-\ln(U)}{\lambda}.$$

Derivation of exp from U [3]

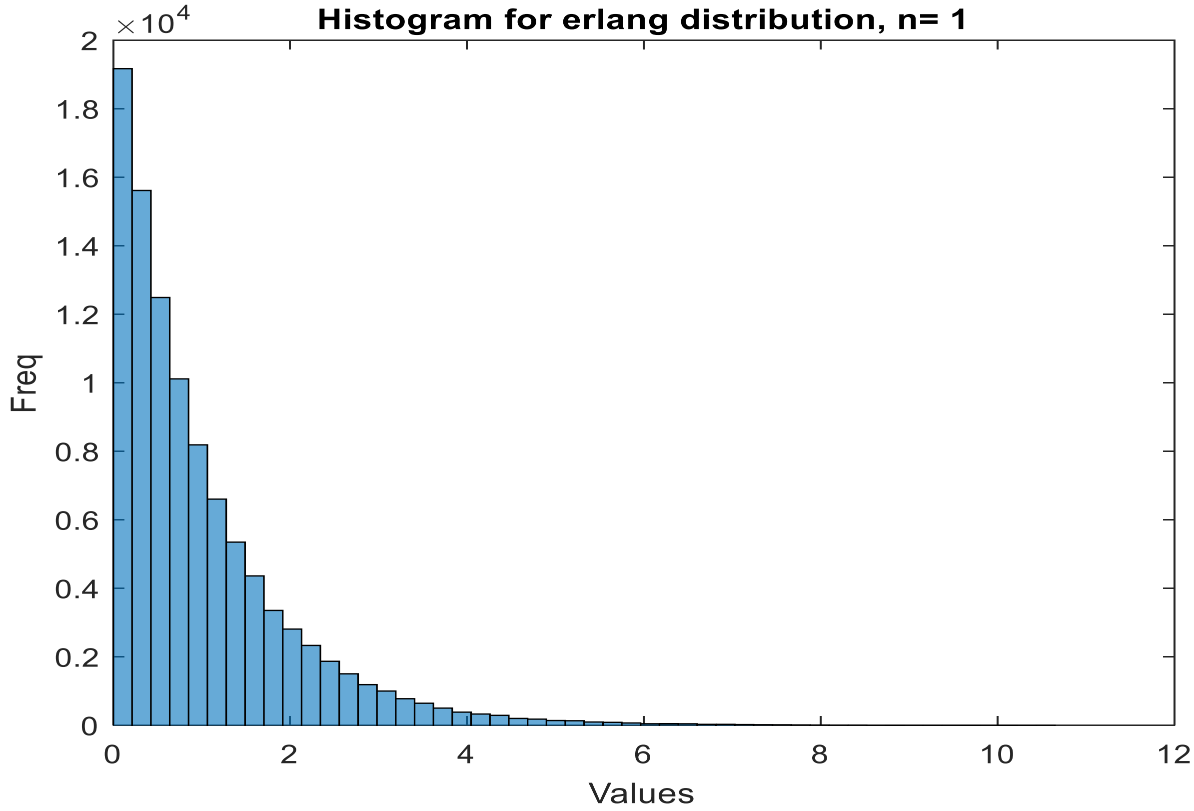
Procedure:

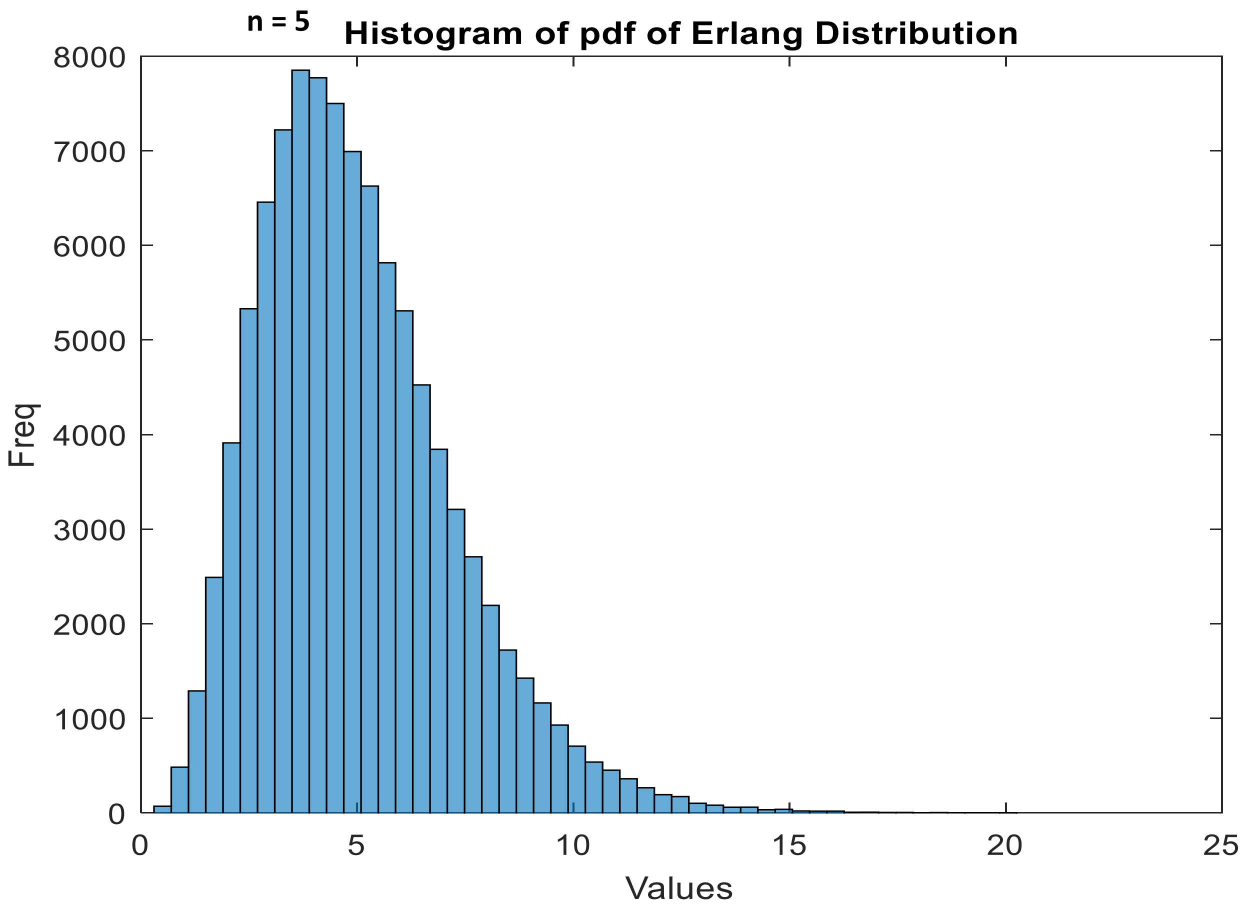
A. Using the above knowledge that taking inverse of x = Uniform Distribution from [0,1] using pseudo random number generator gives back y which belongs to an exponential distribution.
   1. Generating pseudo random number using rand() function in MATLAB. It will be [100000,5] matrix, 5 representing the number of samples for every X(i) and 10,000 representing the number of i.
   2. Assigning lambda value as 1.
   3. Using the function 'expinv' to take the inverse of the uniform random distribution. The resultant will be uniform distribution X.
   4. Defining S which is the sum of iid exponentially random variables also called Erlang Distribution. Summing all random variables X along the column.
   5. Using function 'mean' and 'var' to compute mean and variance of S.

B. The second one is simplification of the first one as explained above in Theory.
   1. Generating pseudo random number using rand() function in MATLAB. It will be [100000,5] matrix, 5 representing the number of samples for every X(i) and 10,000 representing the number of i.
   2. Assigning lambda value as 1.
   3. Using the formula – (log (uniform distribution))/ lambda to compute the exponential distribution X.
   4. Defining S which is the sum of iid exponentially random variables also called Erlang Distribution. Summing all random variables along the column.
   5. Using function 'mean' and 'var' to compute mean and variance of S.

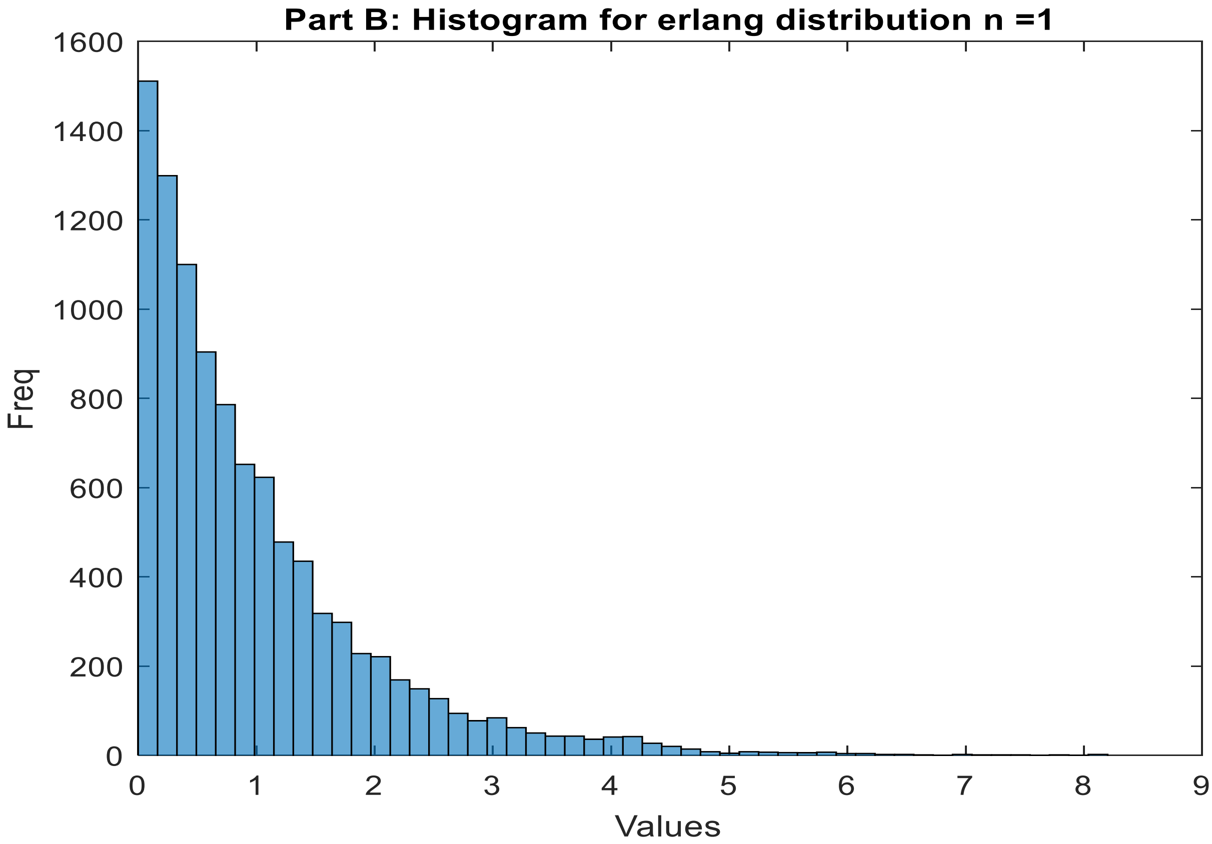
Result:

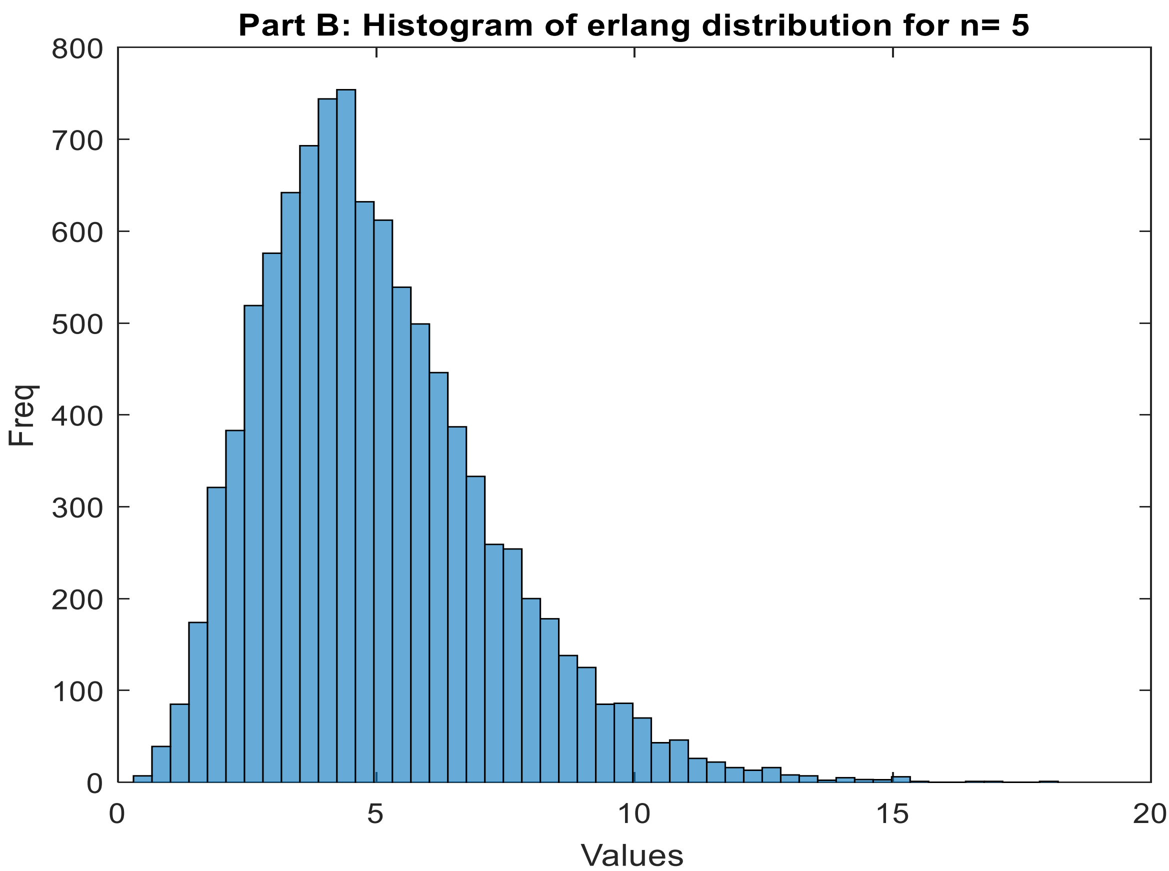
Part A)


Histogram for erlang distribution, n= 1
×10^4
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Freq
0
2
4
6
8
10
12
Values


n = 5
Histogram of pdf of Erlang Distribution
8000
7000
6000
5000
4000
3000
2000
1000
0
Freq
0
5
10
15
20
25
Values

Part B)


Part B: Histogram for erlang distribution n =1
Freq
Values
0
200
400
600
800
1000
1200
1400
1600
1
2
3
4
5
6
7
8
9


Part B: Histogram of erlang distribution for n= 5
Freq
Values
0
100
200
300
400
500
600
700
800
5
10
15
20

Command Window

```
h =

  Histogram with properties:

             Data: [100000x1 double]
           Values: [1x50 double]
          NumBins: 50
         BinEdges: [1x51 double]
         BinWidth: 0.3990
        BinLimits: [1x2 double]
    Normalization: 'count'
        FaceColor: 'auto'
        EdgeColor: [0 0 0]

  Show all properties

Elapsed time is 0.271102 seconds.
Elapsed time is 0.014111 seconds.
fx >>
```

Current Folder | Workspace

Name	Value
h	*1x1 Histogram*
i	10000
j	5
lambda	1
mean_parta	5.0006
mean_partb	5.0115
mu	1
numbins	50
S_A	*100000x1 double*
var_parta	5.0110
var_partb	5.0624
x	*10000x5 double*
X_A	*100000x5 double*
x_result	*10000x1 double*
xmul	*10000x1 double*

Observation and Analysis:

A) We observe that the mean is 5.0006 for part A.
Thus, comparing with the analytical result means k/ (lambda) which for our parameters where k = 5 and lambda = 1.
Mean = 5
So, simulation gives us an approximate result

B) We observe that the variance from part A is 5.0110. Thus, comparing by the analytical result which gives variance to be k/(lambda)^2.
Variance = 5
So, our simulation method gives us an approximate result.

C) We observe that for part B, we get mean as 5.0015 and variance as 5.0624 which are very close to the analytical result.
Mean = 5
Variance = 5

D) When we compare the time taken by Part A and Part B, we observe that Part B takes very less time as compared to part A. Part A takes 0.271102 seconds as compared to part B which takes just 0.01411 seconds. So, **second method is more efficient than the first one**.

Conclusion:

Part B is computationally more efficient as compared to Part A (simulation method).

Reference:

[1] http://www.statisticshowto.com/erlang-distribution/
[2] http://mathworld.wolfram.com/ErlangDistribution.html
[3] https://en.wikipedia.org/wiki/Erlang_distribution