EE511 – Project #6 Muthulakshmi Chandrasekaran

November 13, 2017 USC ID 4486180242

## 1.1 Problem Statement

To generate continuous random variables. To define $\{X_i\}$ to be independent and identically exponentially distributed random variables with $E[X_i]=\frac{1}{\lambda}=1$. To define the random variable $S_n$ to be the sum of these IID random variables as $S_n=\sum_{i=1}^{n} X_i$

Part A : To generate values for $X_i$ using simulations and generate a series of samples for $S_n$; To compute the sample mean and sample variance for $n=5$; To compare to the analytical value. To plot a histogram approximation of the probability density function.

Part B: To calculate $S_n$ as $S_n=\frac{-\ln(\prod_{i=1}^{n} U_i)}{\lambda}$ where $U_i \sim Uniform(0,1)$. To generate a series of samples for $S_n$ using this approach. To compute mean and sample variance for $n=5$ and compare to part A.

## 1.2 Mathematical Basis

### 1.2.1 Continuous Random Variables

A continuous random variable is a random variable that can take on an uncountably infinite number of possible outcomes. Some continuous distributions include Exponential, Gamma, Gaussian etc. The probability density function of a continuous random variable X is given by

$$f_X(x)=\frac{dF_X(x)}{dx},$$

Where $F_X(x)$ denotes the cumulative distribution function of X. Also,

$$F_X(x)=\int_{-\infty}^{x} f_X(y)dy$$

A continuous random variable $X$ is said to have an exponential distribution with parameter $\lambda > 0$, shown as $X \sim exponential(\lambda)$, if its PDF is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & otherwise \end{cases}$$

The cumulative distribution function of an exponential Random variable is given by

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

The mean of the Exponential Random Variable is given by $E[X] = \frac{1}{\lambda}$

Variance of the Exponential Random Variable is given by $Var[X] = \frac{1}{\lambda^2}$.

### 1.2.2 Independent and Identically Distributed Random Variables

If $X_1, X_2, X_3, \ldots, X_N$ are iid random variables with mean μ and variance $\sigma^2$, then the sum of the random variables is given by

$$S = X_1 + X_2 + X_3 + \cdots + X_N$$

The mean of the IID Sum is given by

$$E[S] = N\mu$$

The variance of the IID Sum is given by

$$V[S] = N\sigma^2$$

## 1.3 Simulation in Matlab

### 1.3.1 Part A – Generation of continuous Random Variables and calculation of $S_n = \sum_{i=1}^{n} X_i$

**<u>Description of the Code :</u>**

Here exponential random variable is generated by means of exprnd function in Matlab. This function generates an exponential random variable with the specified mean. The command exprnd(c) generates an exponential random variable with mean c. A screenshot of generated exponential random variable with mean 1 is shown in Figure 1.

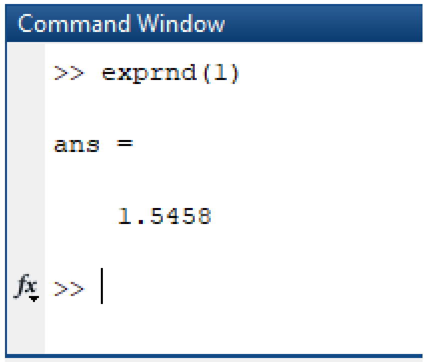


Figure 1 exprnd function

The value for the number of iid random variables is obtained. Based on that value, exponential random variable is generated each time using the exprnd command. Then the sum of all generated exponential random variables is taken. This is $S_n$. This is repeated for many trials (say 10000). Then sample mean and sample variance are calculated . Sample Mean for a set of $N$ elements $X_1, X_2, X_3, \dots, X_N$ is calculated as

$$Sample\ Mean\ \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Sample Variance is calculated as

$$Sample\ Variance\ = \frac{1}{N-1}\sum_{i=1}^{N} (X_i - \bar{X})^2$$

Then the actual mean and actual variance is calculated using the formula for iid random variables and then compared.

**<u>Observations:</u>**

In the first model of the code, the exponential random numbers are generated using exprnd function, as shown in Figure 1. The experiment is repeated for 100,1000 and 10000 trials. The resulting values and the histogram for trials =100 is shown in Figure2.

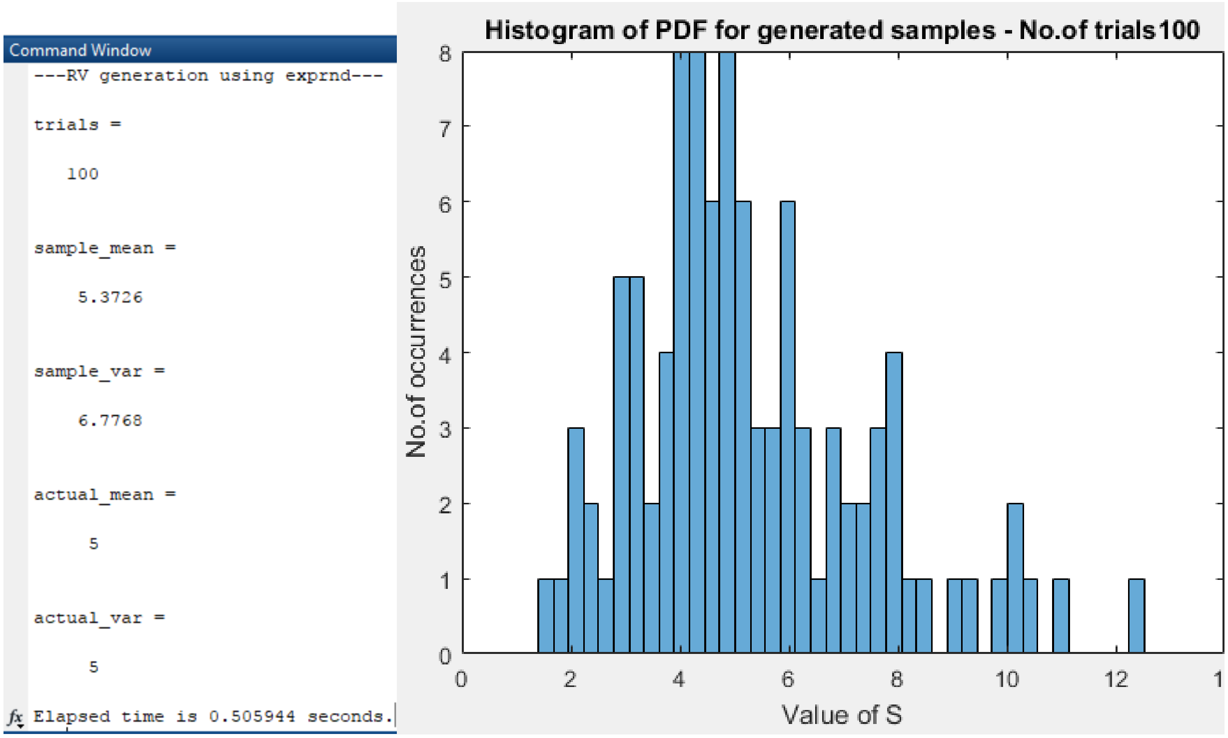


Figure 2 Result obtained when trials =100 – RV generated using exprnd command

The resulting values and the histogram for trials=1000 is shown in Figure3.

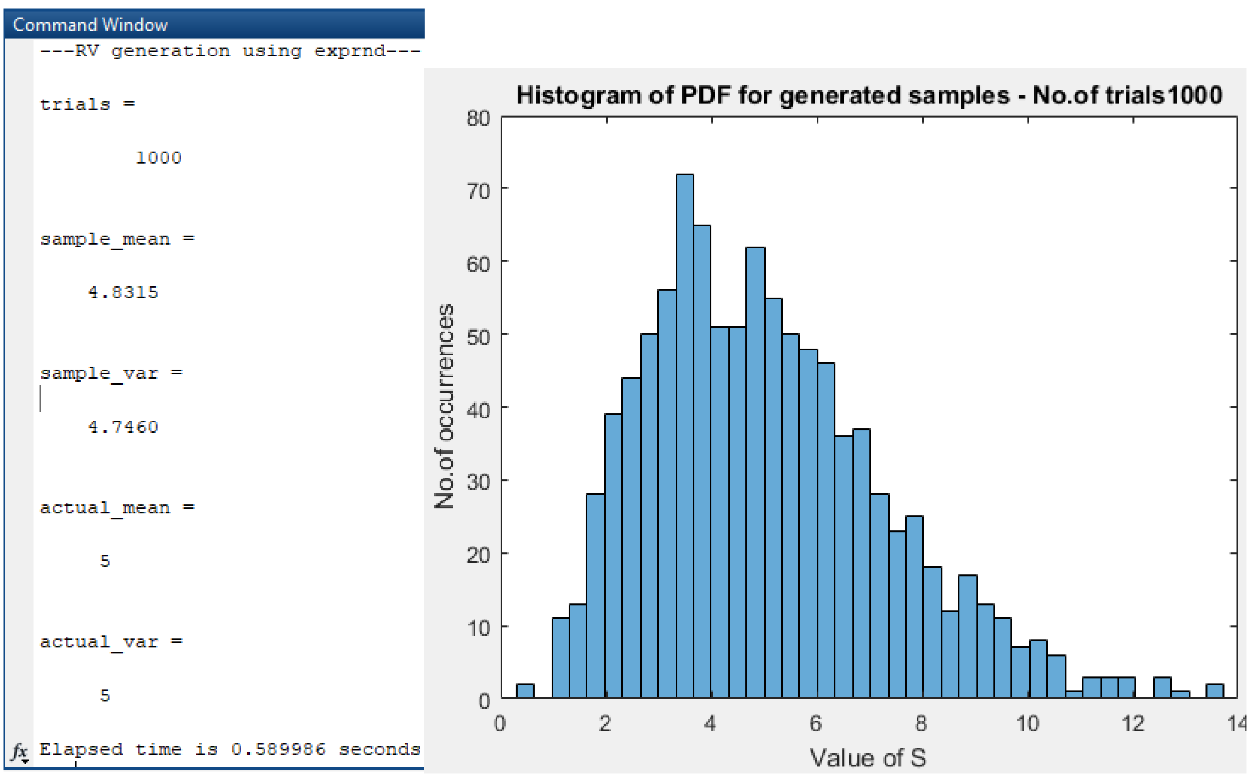


Figure 3 Result obtained when trials =1000 – RV generated using exprnd command

The resulting values and the histogram for trials=100000 is shown in Figure 4.

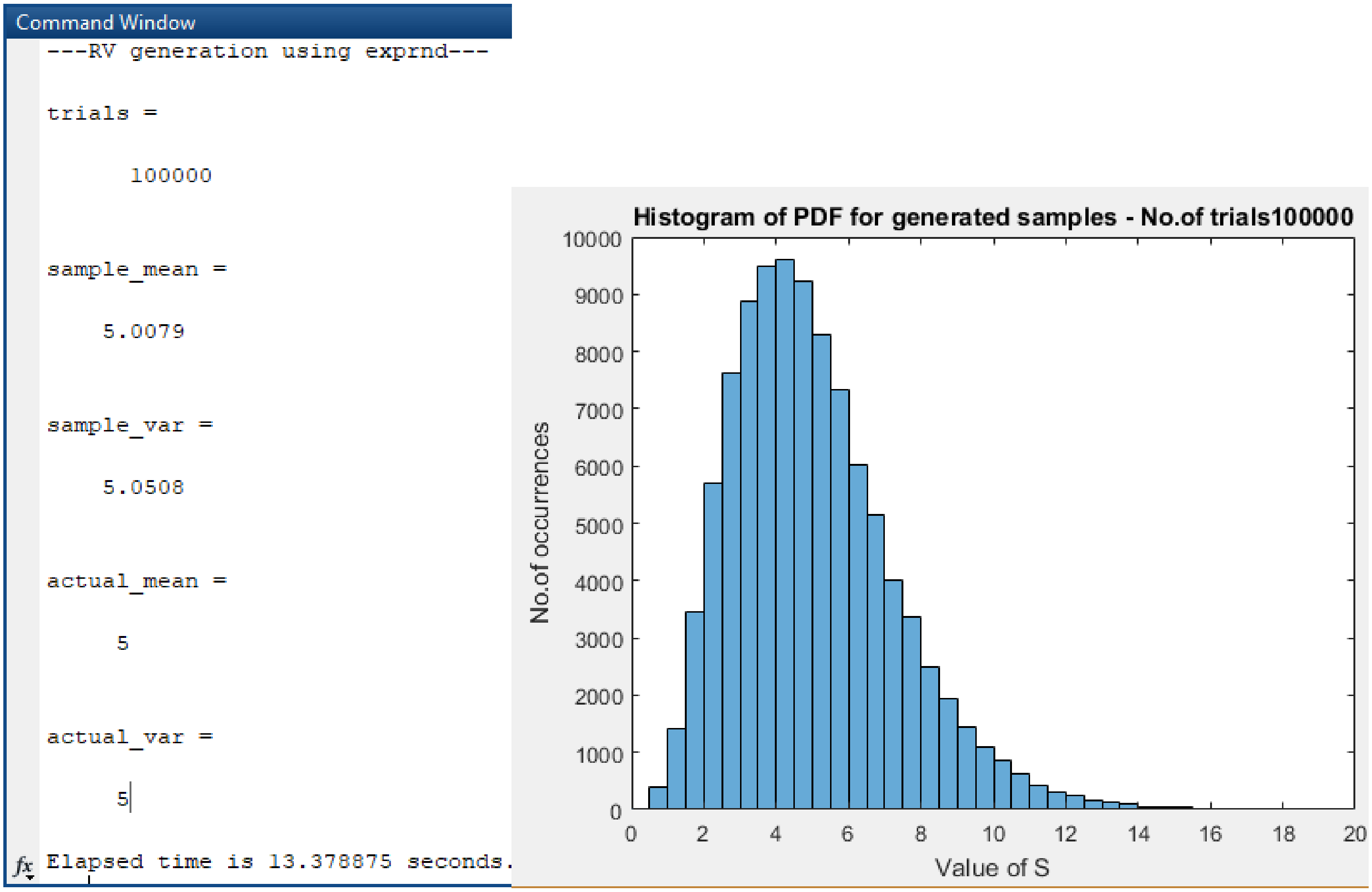


Figure 4 Result obtained when trials =100000 – RV generated using exprnd command

It is found that the distribution of $S_n$ is 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:

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

**<u>Discussion of the results of Part A:</u>**

i. This part deals with the generation of $S_n$ as the sum of exponential random variables as $S_n = \sum_{i=1}^{n} X_i$ . Exponential random variables with mean = 1 are generated. The generation of exponential random variables is carried out using $exprnd$ command .

ii. The simulation is done for many trials. The sample mean, sample variance, actual mean and actual variance are calculated. Histogram plots are also plotted.

iii. Since $S_n = \sum_{i=1}^{n} X_i$ in this part, the $S_n$ follows Erlang-n distribution. The mean and variance of the Erlang distribution is given by

$$Actual\ Mean\ E[S] = \frac{n}{\lambda} = \frac{5}{1} = 5$$

$$Actual\ Variance\ V[S] = \frac{n}{\lambda * \lambda} = \frac{5}{1 * 1} = 5$$

iv. In this case, it is found that as the number of trials increases, the pdf of the samples generated becomes closer to the curve of Erlang distribution.
In Figure 2, when the simulation is done for 100 trials, the pdf is uneven, giving no idea of what distribution it follows. As the number of trials increases to 100000, the graph almost takes the shapes of Erlang – k distribution. This is shown in Figures 2 to Figure 4.

v. When calculating the sample mean and sample variance, it is found that when the number of trials is lesser, the sample mean and variance differ greatly from the actual mean and variance. When the number of trials increases, they become closer to the actual mean and variance, as shown in Table 1.

Table 1 Comparison of Sample Mean and Sample Variance

| **Obtained Values** | | **Trials=100** | **Trials =1000** | **Trials =100000** |
|---|---|---|---|---|
| Actual Mean =5 | Sample Mean | 5.372 | 4.83 | 5.0079 |
| Actual Variance =5 | Sample Variance | 6.776 | 4.74 | 5.0508 |

vi. The time taken for each simulation to run is also calculated. From Figure 2 to Figure 4 the time for each simulation is compared in Table 2.

Table 2 Comparison of Run times

| **No.of trials** | **Trials =100** | **Trials =1000** | **Trials = 100000** |
|---|---|---|---|
| Time(seconds) | 0.5059 | 0.58998 | 13.37 |

### 1.3.2 Part B – Generation of samples using more efficient way

**<u>Description of the Code :</u>**

Here the exponential random variables are generated using the inversion methodFirst a random number of known distribution is generated. In this case, a Uniform random number in the range (0,1) is generated using the rand command. Then the distribution function is inverted to find a sample of the desired distribution. Now, that sample follows the desired exponential distribution. Then sample mean, sample variance, actual mean and actual variance are calculated, and analysed.

The sum of iid random variables is taken as the log of the product of the generated random variables, as

$S_n = \frac{-\ln(\prod_{i=1}^{n} U_i)}{\lambda}$ where $U_i \sim Uniform(0,1)$.

The histogram plot of the PDF for the generated samples is also plotted. The total time for running the code is also calculated.

Figure 5 shows the results of $S_n$ by inversion method for 100 trials.

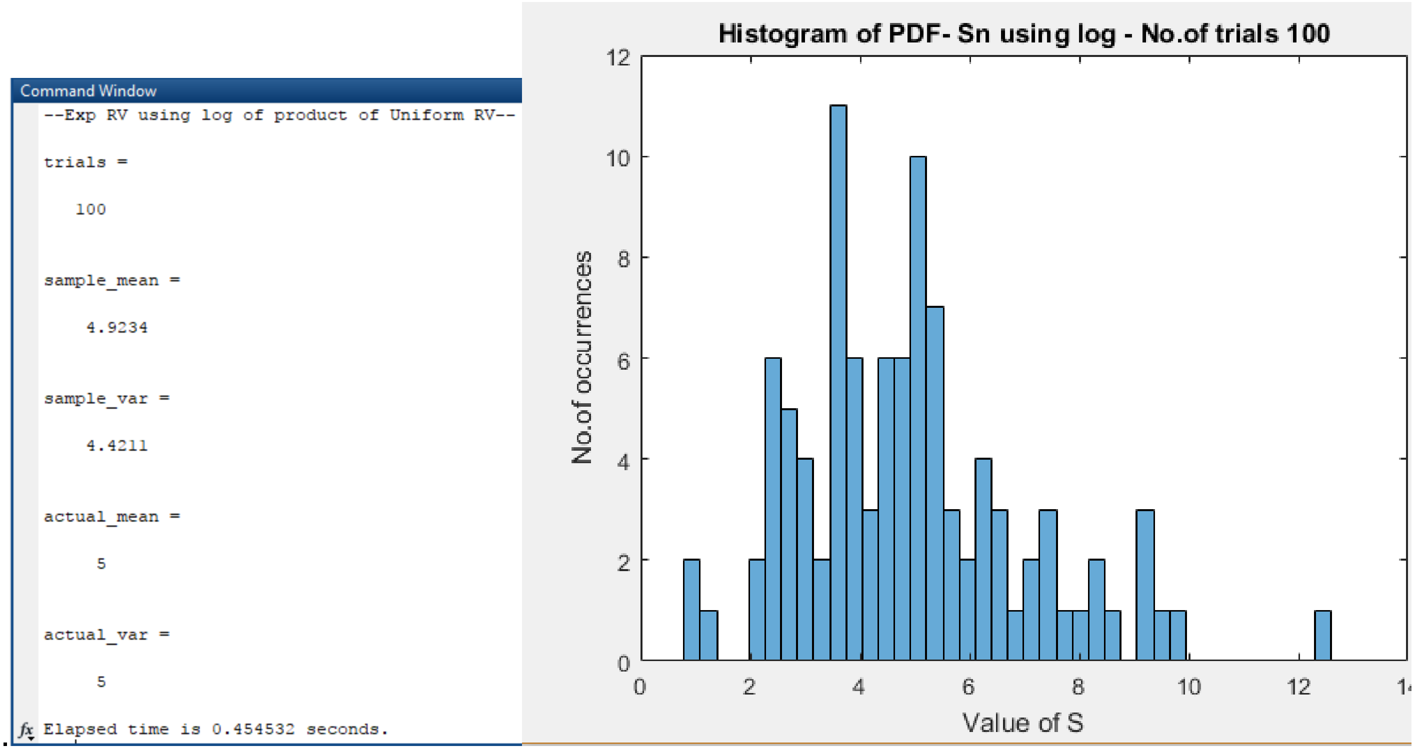


Figure 5 Results of $S_n$ by inversion method for 100 trials.

Figure 6 and Figure 7 shows the results of $S_n$ by method 2 for 1000 trials and 100000 respectively.

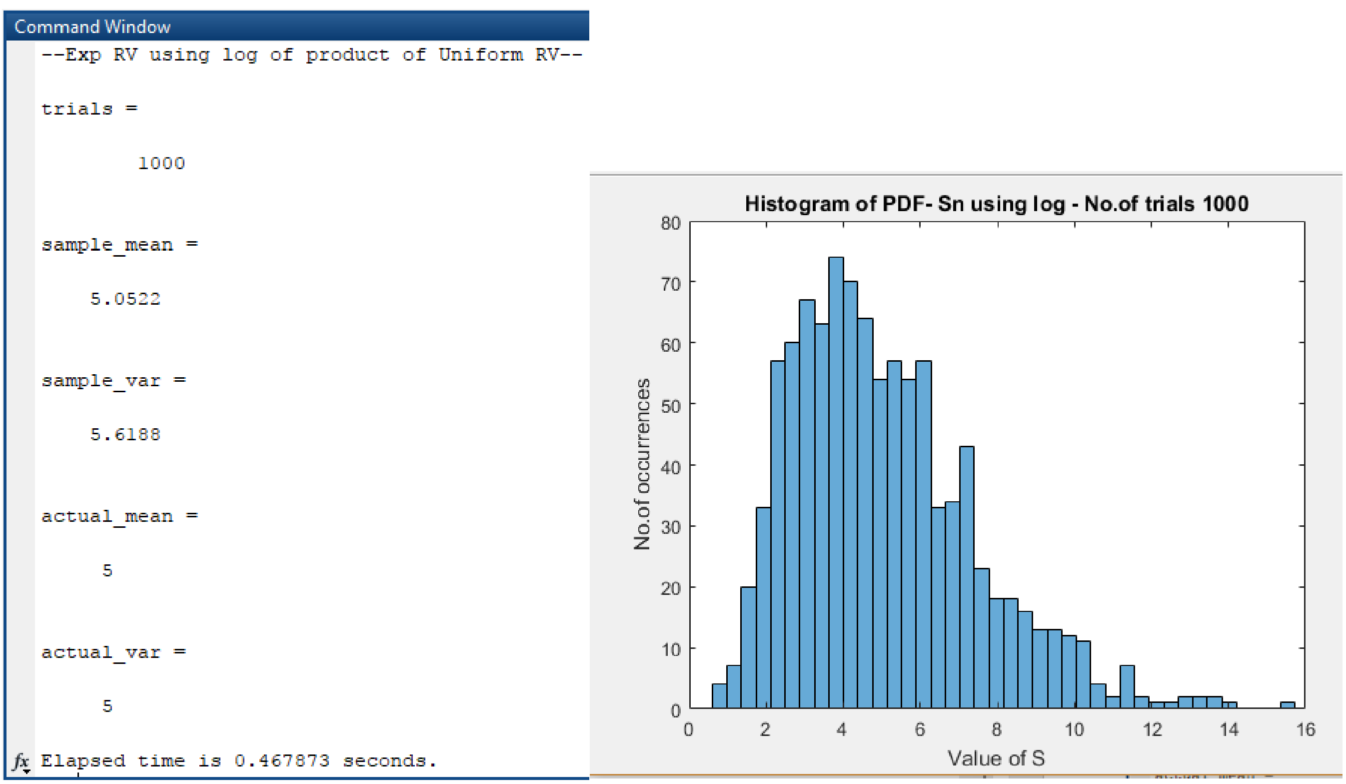


Figure 6 Results of $S_n$ by inversion method for 1000 trials.

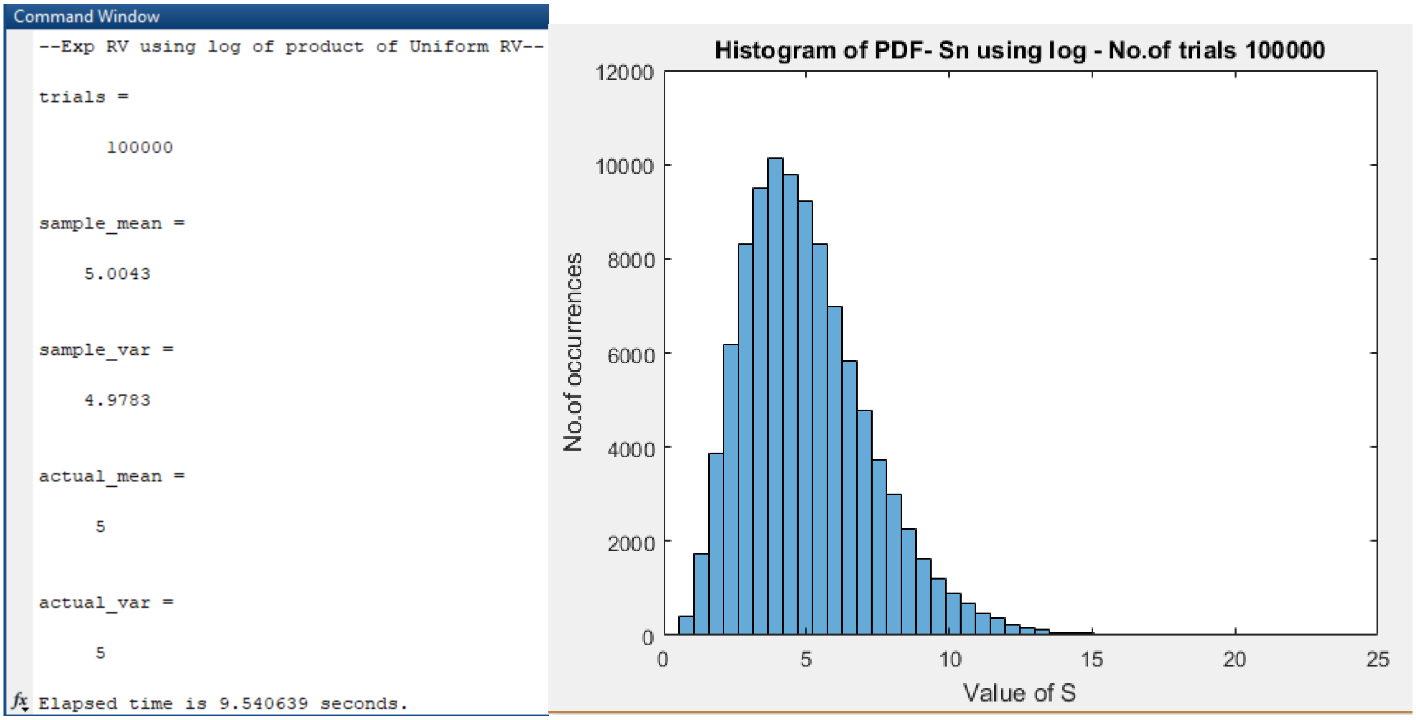


Figure 7 Results of $S_n$ by inversion method for 100000 trials.

**<u>Discussion of the results of Part B:</u>**

i. This part deals with the generation of $S_n$ using Inversion Method i.e. as the log product of Uniform variables as $S_n$ as $S_n = \frac{-\ln(\prod_{i=1}^{n} U_i)}{\lambda}$ where $U_i \sim Uniform(0,1)$.

Exponential random variables with mean = 1 are generated.

ii. The simulation is done for many trials. The sample mean, sample variance, actual mean and actual variance are calculated. Histogram plots are also plotted.

iii. $S_n$ follows Erlang-n distribution. The mean and variance of the Erlang distribution is given by

$$Actual\ Mean\ E[S] = \frac{n}{\lambda} = \frac{5}{1} = 5$$

$$Actual\ Variance\ V[S] = \frac{n}{\lambda * \lambda} = \frac{5}{1 * 1} = 5$$

iv. Even in this case, it is observed that as the number of trials increases, the pdf of the samples generated becomes closer to the curve of Erlang distribution. As in Figure 5, the curves doesn't fit any distribution. This is an uneven curve. When the sample size is increased by increasing the number of trials, the curve reaches a closer approximation to Erlang as in Figure 6 and Figure 7.

v. When calculating the sample mean and sample variance, it is found that when the number of trials is lesser, the sample mean and variance differ greatly from the actual mean and variance. When the number of trials increases, they become closer to the actual mean and variance, as shown in Table 3.

Table 3 Comparison of Sample Mean and Sample Variance

| | | **Trials =100** | **Trials =1000** | **Trials =100000** |
|---|---|---|---|---|
| Actual Mean =5 | Sample Mean | 4.923 | 5.0522 | 5.0043 |
| Actual Variance =5 | Sample Variance | 4.421 | 5.6188 | 4.9783 |

vi. The time taken for each simulation to run is also calculated. From Figure 5 to Figure 7 the time for each simulation is compared in Table 4.

Table 4 Comparison of Run times

| **No.of trials** | **Trials =100** | **Trials =1000** | **Trials = 100000** |
|---|---|---|---|
| Time(seconds) | 0.4545 | 0.4678 | 9.64 |

## 1.4 Results

In the project, the IID random sum is generated by two methods(in general), using Exponential and Uniform Random Variables. Considering the calculation of Random Sum by the generation of exponential random variables using exprnd function(Part A) and by inversion method(Part B), it is found that the inversion method generates computationally efficient samples. This is shown in the following Table 5, where the sample mean, sample variance and the time of simulation is compared for 100000 trials of the experiment. The actual mean and actual variance is 5.

Table 5 Comparison of two methods for 100000 trials

| Parameter | RV Generation using exprnd command | RV generation using Inversion Method |
|---|---|---|
| Sample Mean | 5.0079 | 5.0043 |
| Sample Variance | 5.0508 | 4.9783 |
| Simulation Time(second) | 13.378 | 9.64 |

Thus the inversion method is more efficient for generation of iid sums.

## 1.5 Reference


1. Sheldon M Ross, "Simulation", 5th edition
2. www.mathworks.com


## 1.6 Matlab Codes

**Part A – Using exprnd command**

```
clc;clear all;close all;
disp('---RV generation using exprnd---');tic
N=5;
trials=100
%Each Xi is exponentially distributed Random variable with mean 1.
mean_exp=1;
lambda=1/mean_exp;
S=[];
for j=1:1:trials
X=[];
for i=1:1:N
    %generate every Xi as an exponential RV with mean 1
    X=[X exprnd(mean_exp)];
end
%Define S to be the sum of the generated RVs
S=[S sum(X)];
end

%Computation of Sample Mean for S
k=sum(S);
sample_mean=k/trials
dif=[];
for i=1:1:trials
```

```
    dif=[dif (S(i)-sample_mean)^2];
end
sample_var=(sum(dif))/(trials-1)

%Computataion of actual mean and actual variance for iid
actual_mean=N*lambda
actual_var=N*(lambda^2)

histogram(S,40)
xlabel('Value of S');ylabel('No.of occurrences');
title(['Histogram of PDF for generated samples - No.of trials',num2str(trials)]);
toc
```

## Part B – using Inversion Method

```
%Exponential RV generation using log of product of Uniform RV
clc;clear all;close all;
disp('--Exp RV using log of product of Uniform RV--');
tic;
N=5;
trials=100000
%generation of a series of samples for S
S=[];
for j=1:1:trials
y=[];
%Generate a random variable x
x=[];
for i=1:1:N
x=[x rand(1)];
end
p=prod(x);
exp_mean=1;
lambda=1/exp_mean;
% S is the ;
S=[S (-1/lambda)*log(p)];
end


%Computation of Sample Mean for S
k=sum(S);
sample_mean=k/trials
dif=[];
for i=1:1:trials
    dif=[dif (S(i)-sample_mean)^2];
end
sample_var=(sum(dif))/(trials-1)

%Computataion of actual mean and actual variance for iid
actual_mean=N*lambda
actual_var=N*(lambda^2)
histogram(S,40)
xlabel('Value of S');ylabel('No.of occurrences');
title(['Histogram of PDF- Sn using log - No.of trials ',num2str(trials)])toc
```