Statistical Inference Course Project: Simulation Exercise

David Garza

11/12/2020

Overview

This project will investigate the population mean and standard deviation of the exponential distribution against sample means and standard deviations computed from Monte Carlo simulation of the exponential distribution.

The second objective is to show that the distribution of the sample means of the simulation fits a normal distribution even though the exponential distribution itself is not normally distributed. Please see appendix 1 for more information on the exponential distribution.

Simulations

The exponential distribution will be simulated using the R function **rexp(n,λ)**. Where n is the number of observations to be generated and λ is the rate. For this particular simulation, n = 1,000 and λ=0.2. For more information on the **rexp()** function please see appendix 2.

The general method for the simulations is:

1. Call **rexp(n,λ)** 40 times
2. Store the values into a vector
3. Convert the vector into a matrix where each row is one simulation of 1,000 observations

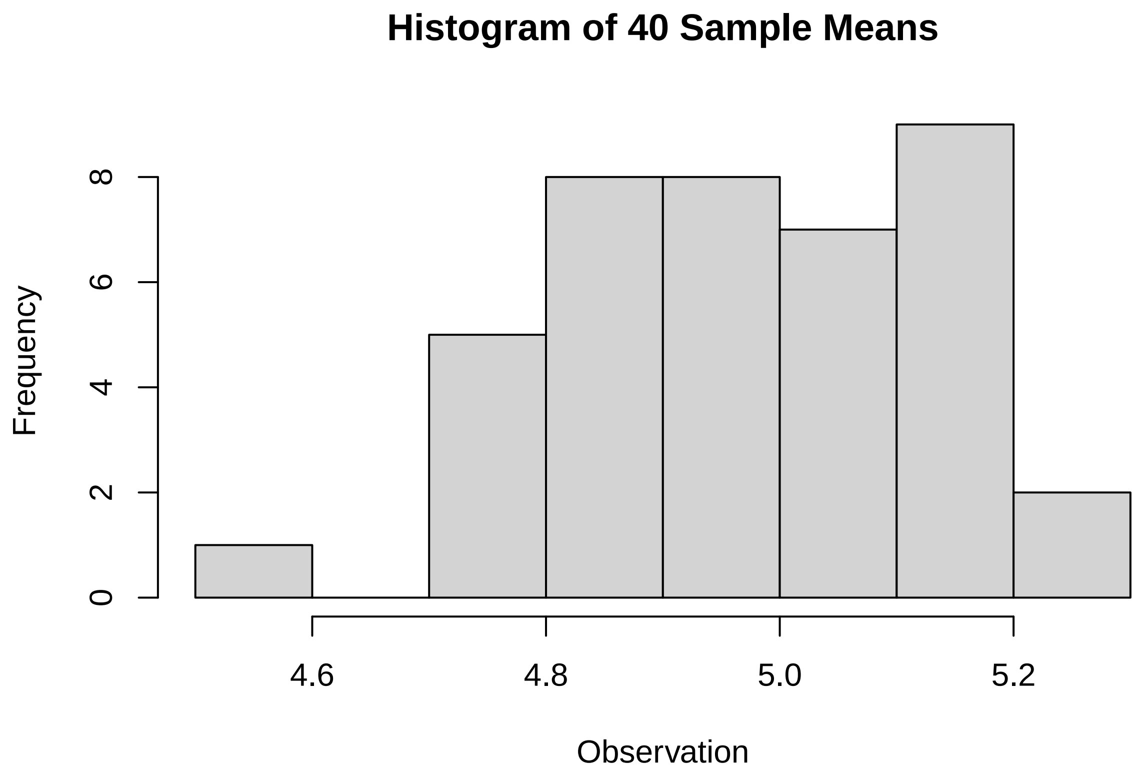
The R code to simulate and store the data is:

```
obs<-NULL
for(i in 1:k) obs<-c(obs,rexp(n,lambda))
obs<-matrix(obs,k,n)
```

Sample Mean versus Theoretical Mean

The theoretical mean of the exponential distribution is $1/\lambda$. With a λ of 0.2 the expected μ is 5.

The distribution of the sample means is below.



The sample mean is computed by finding the mean for each row of the matrix *obs* and then finding the average of averages. This can all be done in one step using t.test and pulling the appropriate items. The function **apply()** computes the means for each row of the matrix and returns a vector of means.

```
results<-t.test(apply(obs,1,mean),alternative ="two.sided",mu=5)
results$estimate
```

```
## mean of x
##  4.974239
```

```
results$conf.int
```

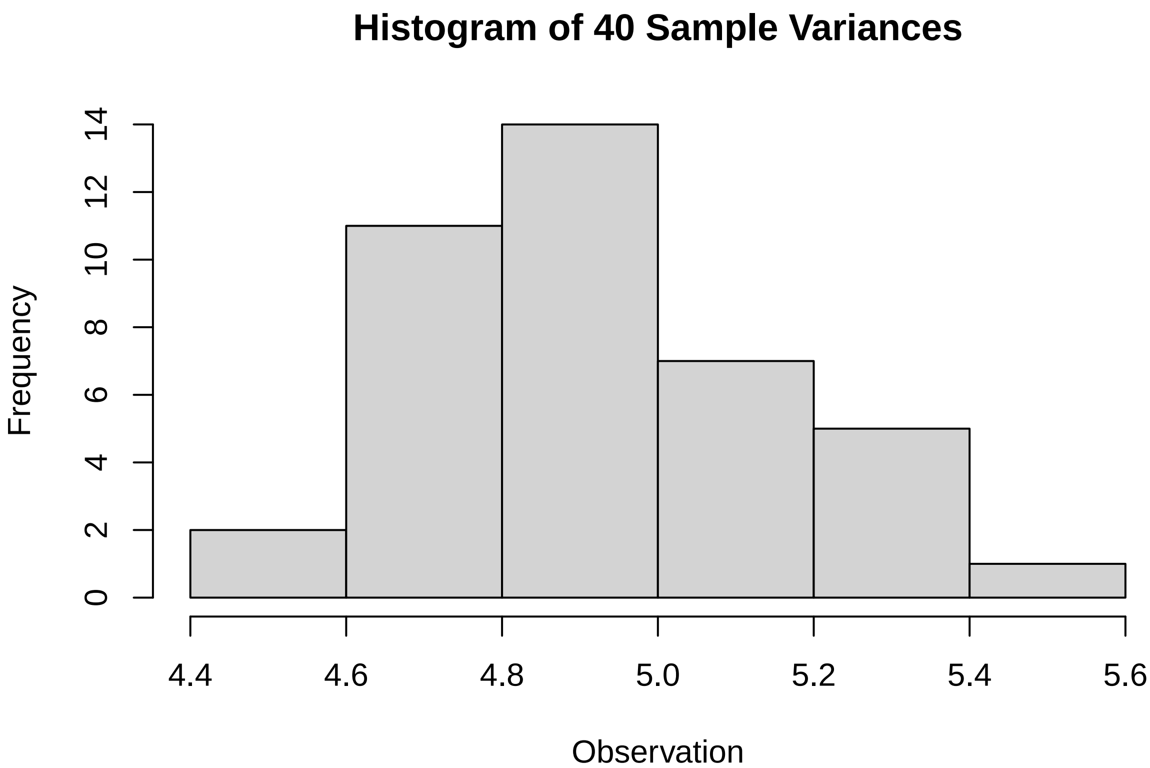
```
## [1] 4.922266 5.026211
## attr(,"conf.level")
## [1] 0.95
```

The estimate is very close to the expected value of 5 and falls within the 95% confidence interval. Therefore we can not reject H_o, where H_o: $\mu = 5$.

Sample Variance versus Theoretical Variance

The theoretical variance of the exponential distribution is $1/\lambda$. With a λ of 0.2 the expected σ is 5.

The distribution of the sample variances is below.



The sample variance is computed by finding the variance for each row of the matrix *obs* and then finding the average of variances. This can all be done in one step using t.test and pulling the appropriate items. The function **apply()** computes the variance for each row of the matrix and returns a vector of variances.

```
results<-t.test(apply(obs,1,sd),alternative ="two.sided",mu=5)
results$estimate
```

```
## mean of x
##   4.92844
```

```
results$conf.int
```

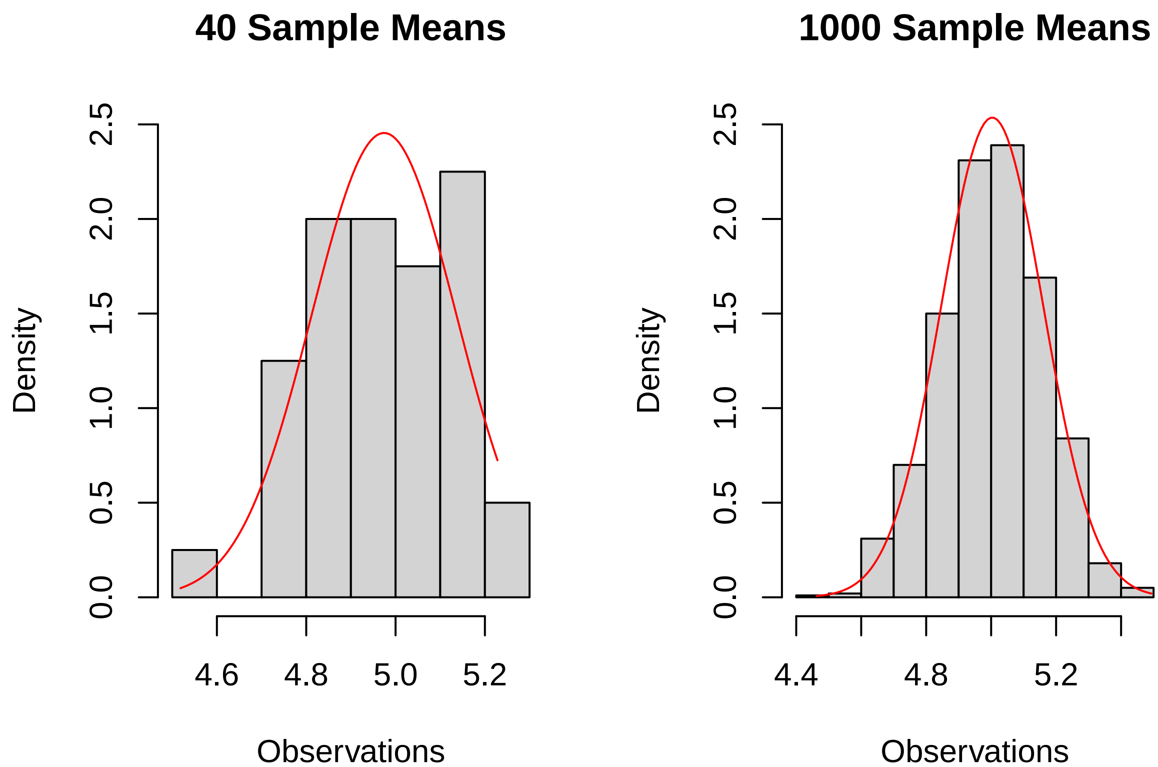
```
## [1] 4.857688 4.999192
## attr(,"conf.level")
## [1] 0.95
```

The estimate is very close to the expected value of 5 and falls within the 95% confidence interval. Therefore we can not reject $\mathrm{H_o}$, where $\mathrm{H_o}$: $\sigma = 5$.

Distribution

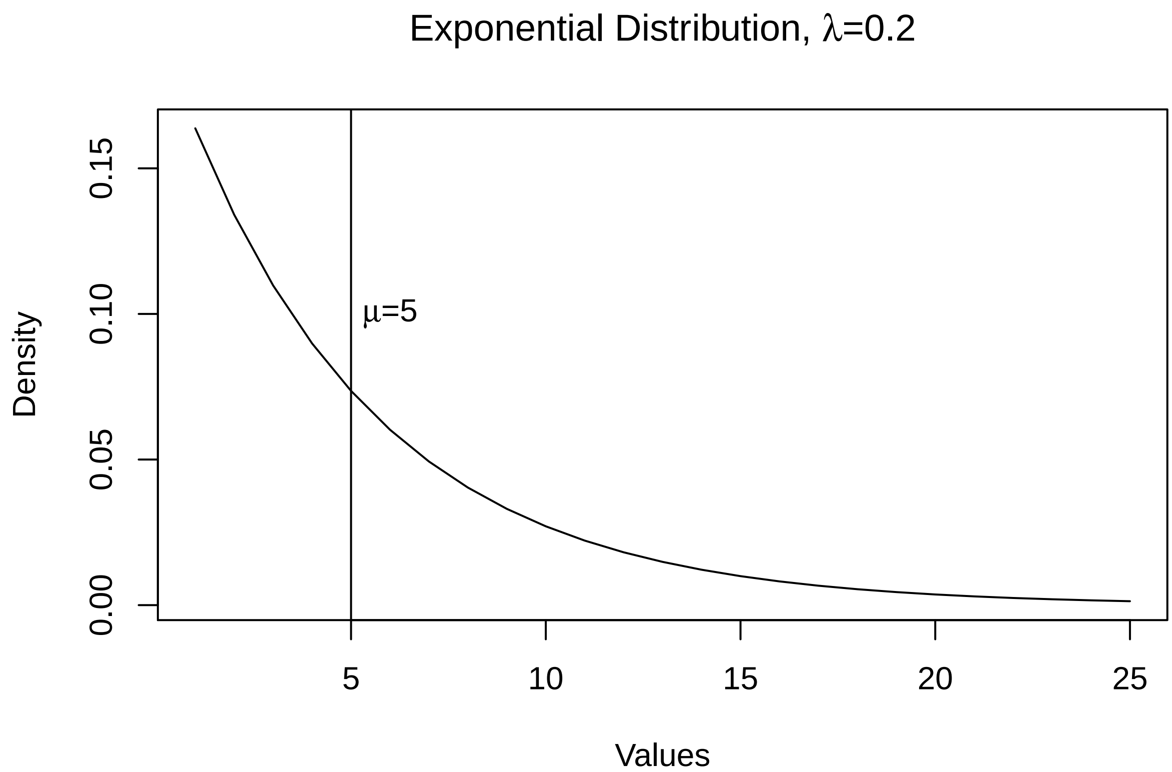
The asymptotic argument will be used to determine if the distribution of sample means is normal. It is assumed as the sample size increases then the distribution of the sample means will fit a normal curve.

As can be seen, the distribution of the sample means approaches a normal distribution as the sample size increases.



Appendix

1 Exponential Distribution



The exponential distribution is similar to an exponential decay curve. Smaller values are more probable than larger values. It is defined as $d = \lambda e^{-\lambda x}$. It is clearly not normal. With λ set to 0.2, μ and σ are 5.

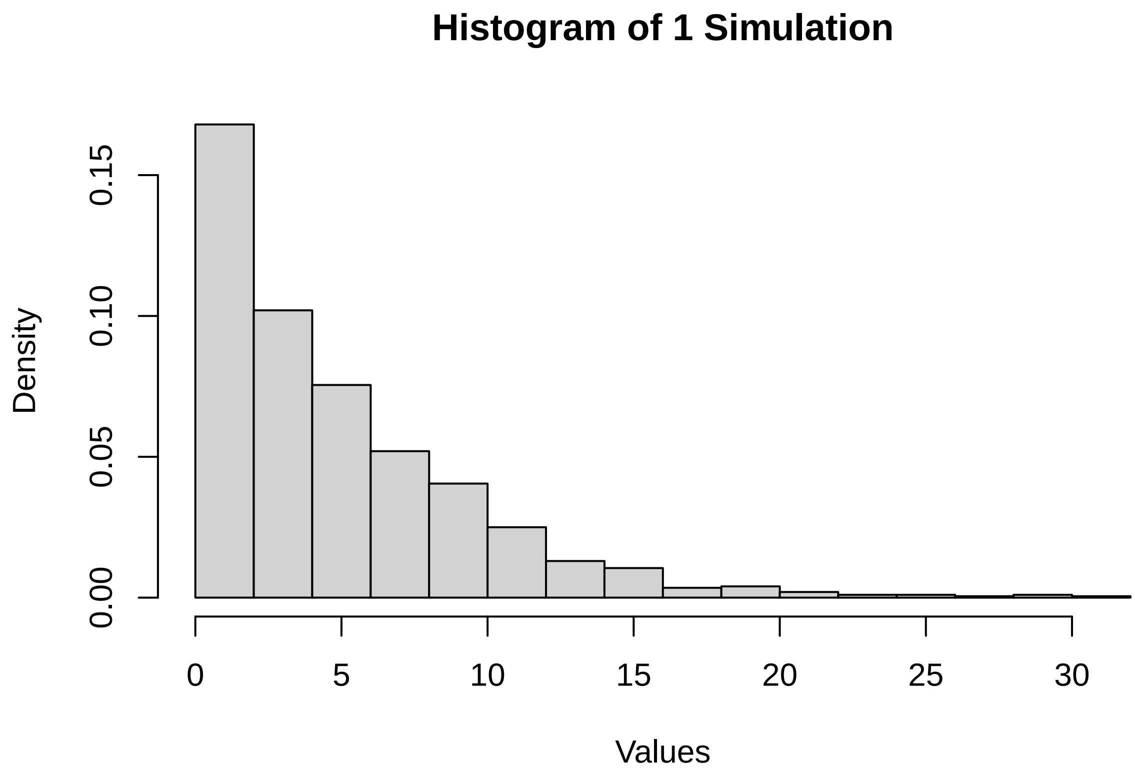
2 rexp() Function

The **rexp(n,λ)** generates random numbers using the exponential distribution (see appendix 1). An example of the output is provided below:

```
head(rexp(n,lambda))
```

```
## [1]  0.5540066  2.9151359 10.8362556  7.2270079  6.3757484  1.4261344
```

Below is the histogram of one simulation of **rexp(n,λ)**. Notice it follows the distribution of the ideal exponential distribution, but there is some noise and variation from the ideal.


Histogram of 1 Simulation
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30
Values