Coursera Statistical Inference Course Project

Investigating The Exponential Distribution

November 2015

Anthony L.

Overview

This report is going to investigate the exponential distribution and how it compares to a normal distribution using R. Based on the Central Limit Theorem (CLT), the arithmetic means of a large sample generated from the exponential distribution will be approximately normally distributed.

Section 1 - Simulations

The rexp() function in R is used to generated simulations of the exponential distribution. The sample size is defined as 40 and the lambda parameter of the exponential distribution is defined as 0.2. The number of simulation that will be performed is 1000.

```
mns=NULL
sample <- 40
lambda <- 0.2
nosim <-1000
for (i in 1 : nosim) mns = c(mns, mean(rexp(sample,lambda)))
```

Section 2 - Sample vs. Theoretical

Comparison of Means

The sample mean generated from the simulations is:

```
mean(mns)
```

```
## [1] 4.985587
```

The theoretical mean of an exponential distribution is defined as 1 divided by the lambda parameter. This value is:

```
1/lambda
```

```
## [1] 5
```

The difference between the two are very small and are a result of Monte Carlo simulations. If the number of sample size increased, the expected difference between the sample mean and the theoretical mean will decrease.

```
mean(mns) - 1/lambda
```

```
## [1] -0.01441277
```

Comparison of Variance

The sample variance generated from the simulations is:

```
var(mns)
```

```
## [1] 0.6410462
```

The theoretical variance of an exponential distribution is defined as 1 divided by the lambda parameter squared divided by the sample size. This value is:

```
1/lambda^2/sample
```

```
## [1] 0.625
```

The difference between the two are very small and are a result of Monte Carlo simulations. If the number of sample size increased, the expected difference between the sample variacne and the theoretical variance will decrease.

```
var(mns) - 1/lambda^2/sample
```

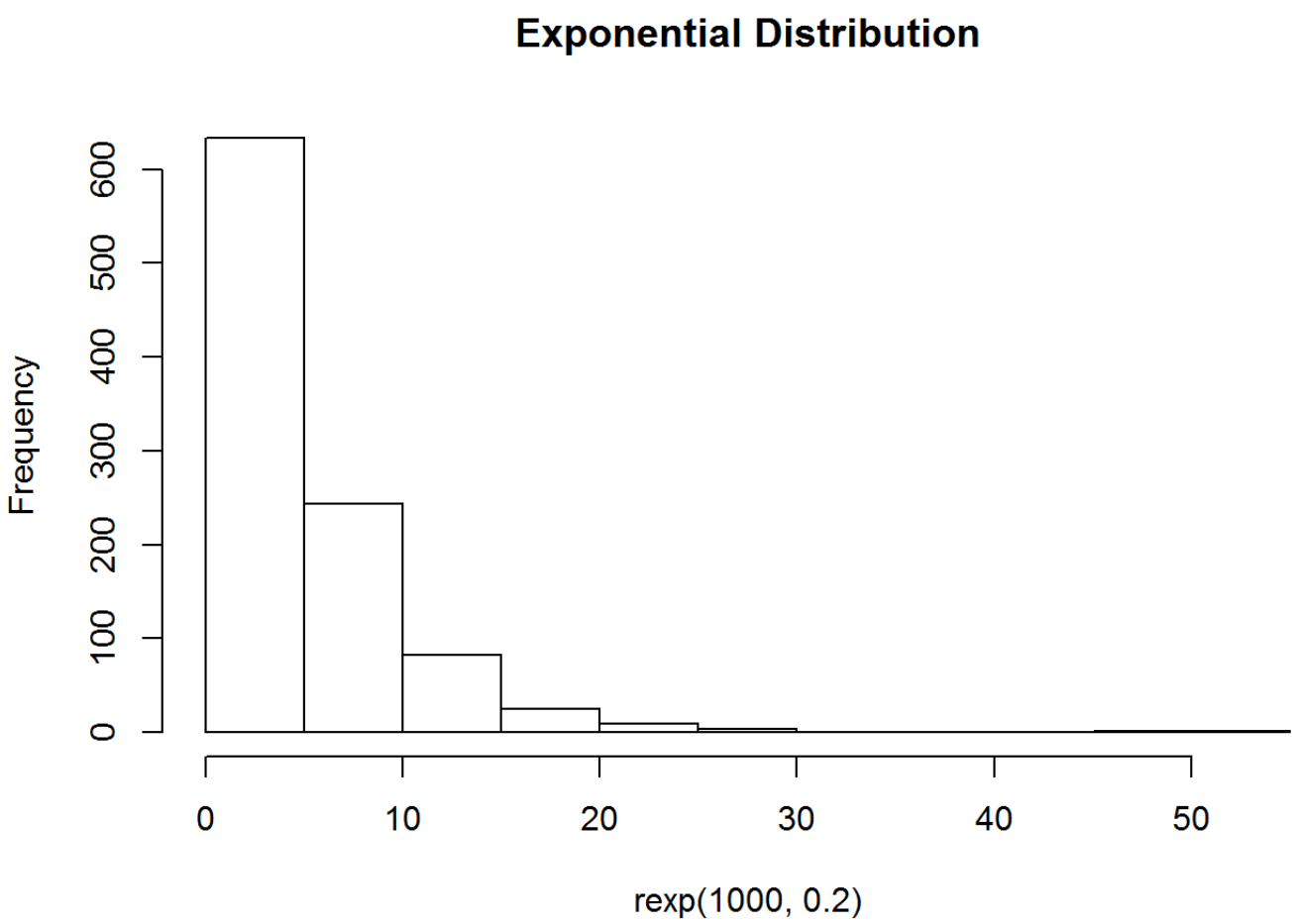
```
## [1] 0.01604615
```

Section 3 - Comparison to Normal Distribution

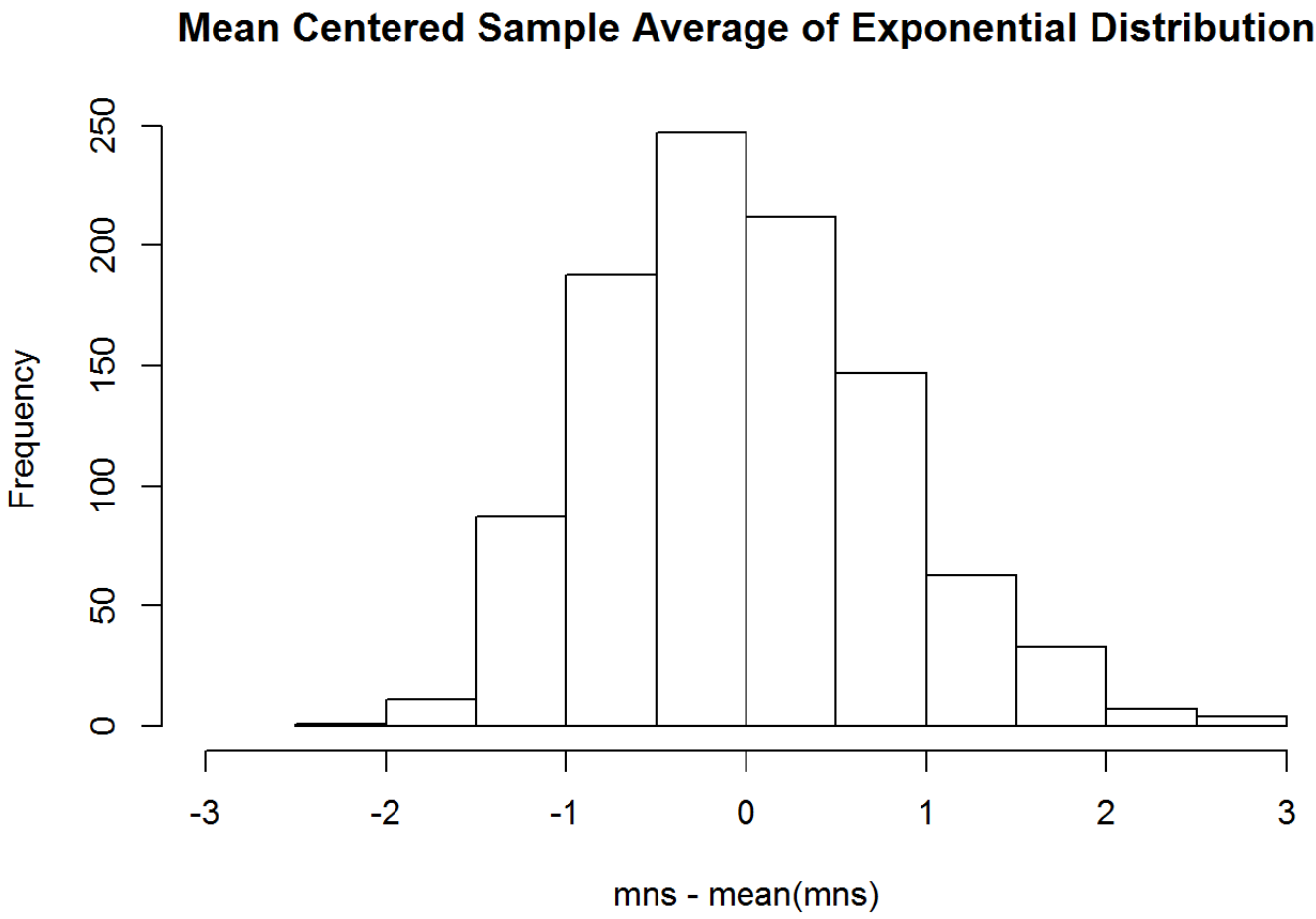
According to the Central Limit Theorem, the arithmetic mean of a large sample of any random distribution with a defined expected value for mean and variance will be approximately normally distributed. The simulations of the exponential distribution generated in this report have a sample size of 40, an expected mean of 1/.2 and an expected variance of 1/.2^2/40.

The 3 histograms below show the exponential distribution, the averages of a sample of exponential distribution (mean centered), and a normal distribution. The exponential distribution at first does not appear normally distributed. However, when the averages of a sample of the distributions are taken, it becomes very clear that the output approximately follows a normal distribution. The sample satisfies the conditions of the CLT and as a result the output is approximately normally distributed.

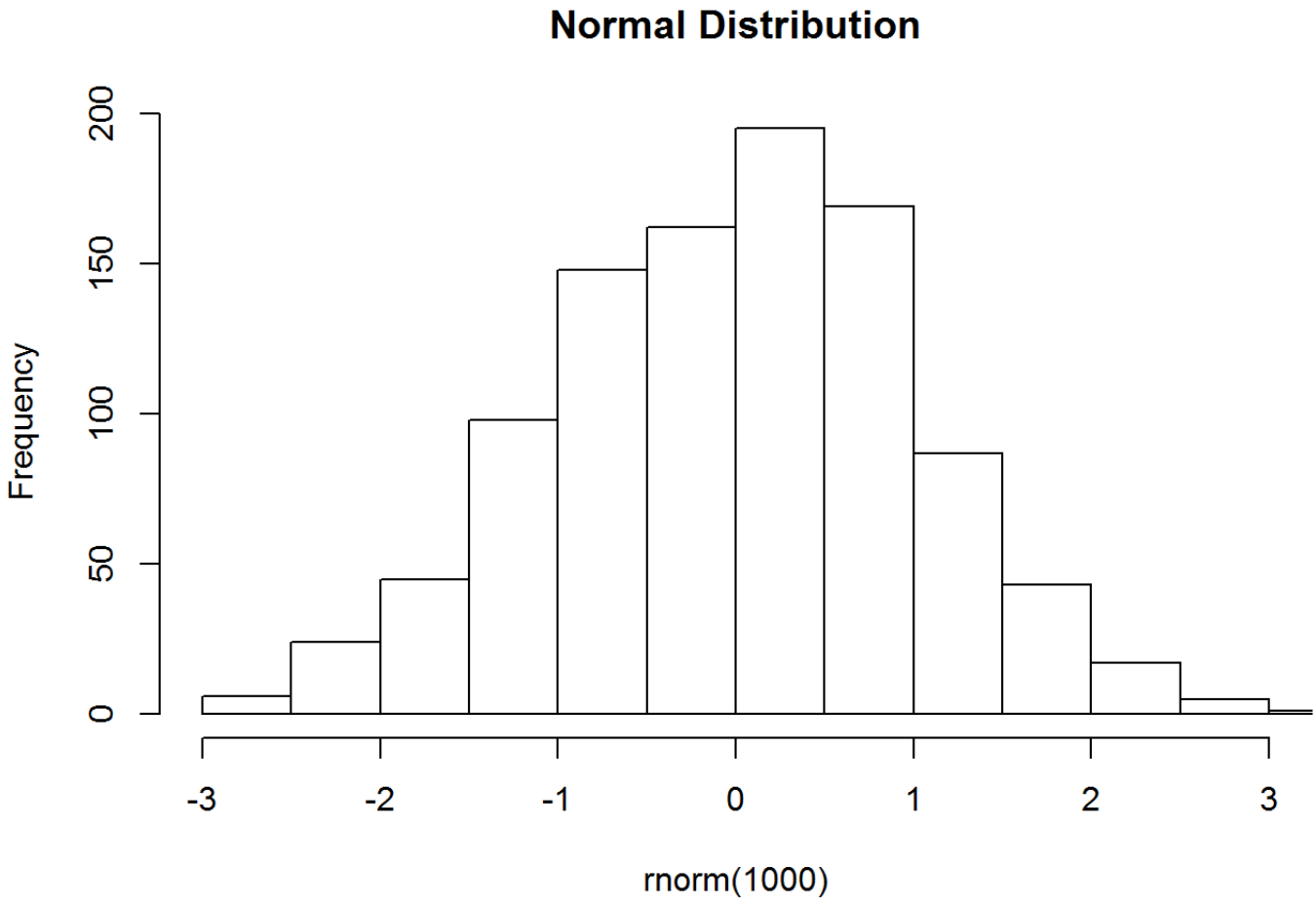
```
hist(rexp(1000,.2),
     main="Exponential Distribution")
```




```
hist(mns-mean(mns),
     xlim = c(-3,3),
     main="Mean Centered Sample Average of Exponential Distribution")
```




```
hist(rnorm(1000),
     xlim = c(-3,3),
     main="Normal Distribution")
```


Normal Distribution
Frequency
0
50
100
150
200
-3
-2
-1
0
1
2
3
rnorm(1000)