Statistical Inference - Simulation Exercise

Jerome Locson

1/22/2019

Overview

This project is part of the Data Science Specialization - Statistical Inference course offered by Johns Hopkins University through Coursera. This simulation exercise will investigate the exponential distribution in R, **using the rexp(n, lambda) function**, in comparison to the Central Limit Theorem. After comparing, the results are plotted in a histogram to compare both results and visually show how close the values are.

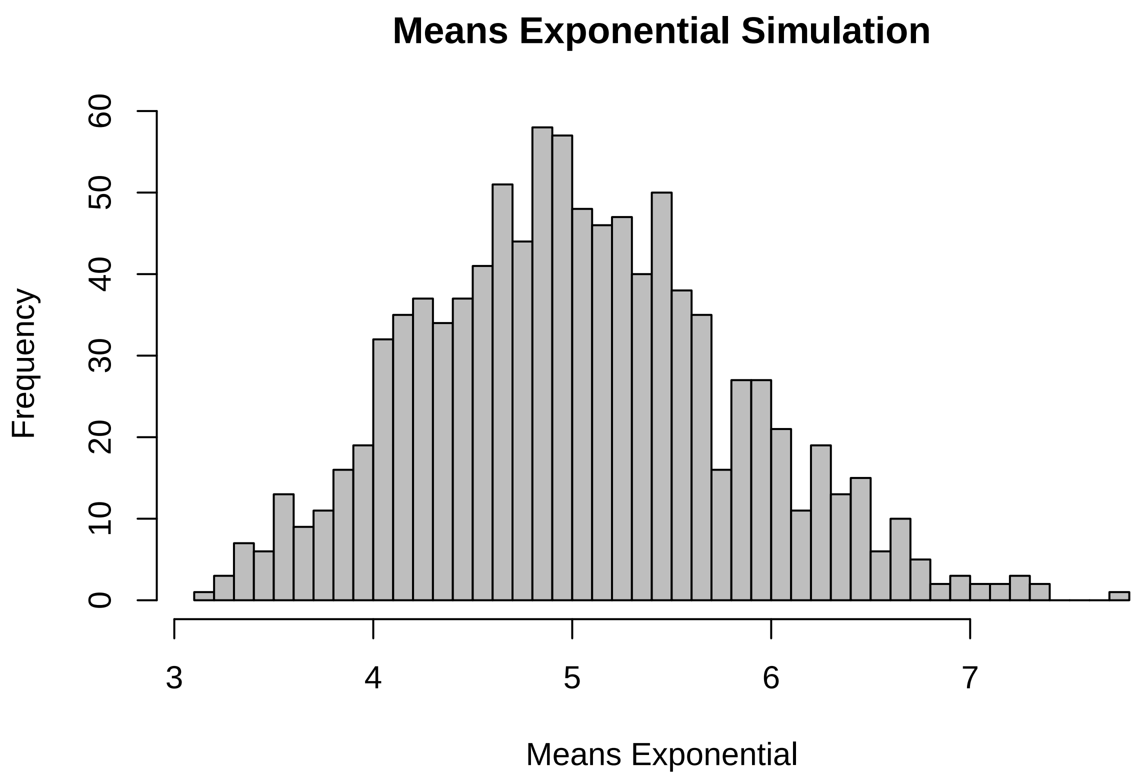
Simulations

Below are the given values to simulate this exercise. Setting seed is set to encourage reproducibility of this simulation.

```
set.seed(1983)          #Setting seed for reproducibility
lambda <- 0.2          #lambda value for all simulations
n <- 40                #Number of exponentials
size <- 1000           #Simulation size
```

We now have the values set above, it is now ready to run the exponential distribution function and replicate the given **size** and get the mean exponentional. Finally, a graph is shown below for the result.

```
sample <- replicate(size, rexp(n, lambda))
meansExp <- apply(sample, 2, mean)
hist(meansExp, breaks=n, main="Means Exponential Simulation",
     col = "gray", xlab="Means Exponential")
```



Sample Mean versus Theoretical Mean

Given the Means Exponential Simulation result above, we will now compare the sample mean versus the theoretical mean. We will still use the same values as set above.

```
sampleMean <- mean(meansExp)
sampleMean
```

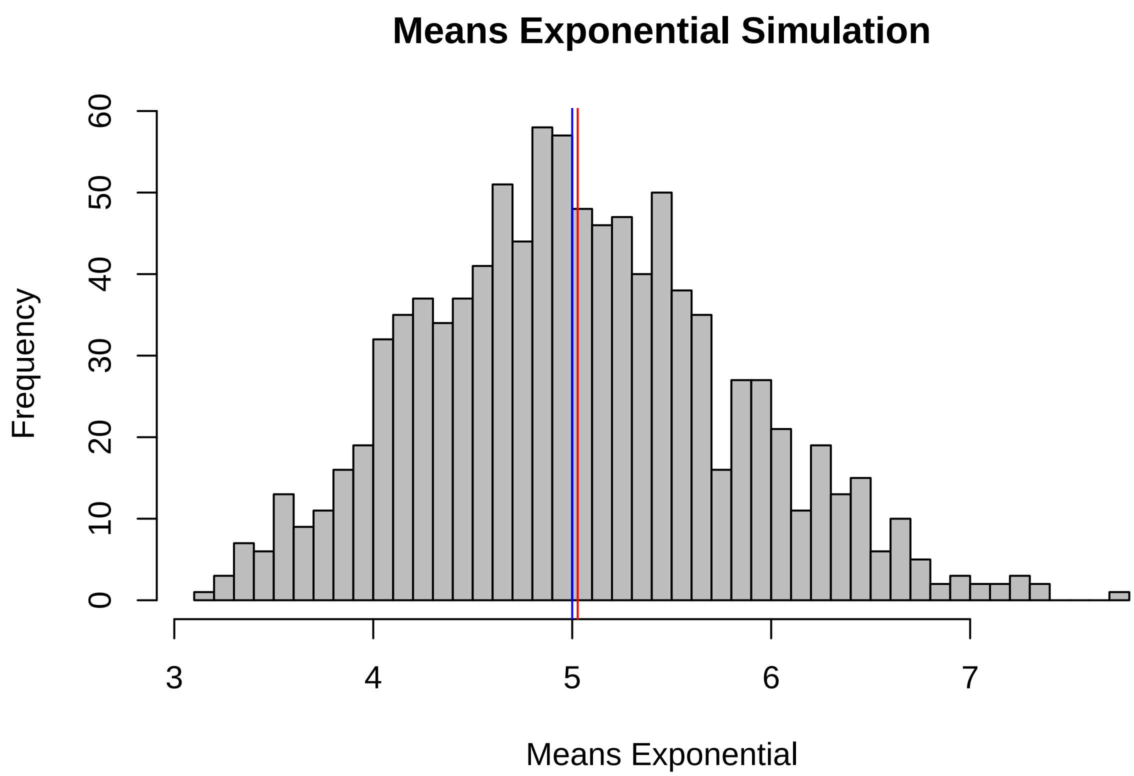
```
## [1] 5.027677
```

```
theoreticalMean <- 1 / lambda
theoreticalMean
```

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

As shown above, the theoretical mean **5** is very close to the sample mean **5.0276769**. The red line plotted on the graph above is the sample mean.

```
hist(meansExp, breaks=n, main="Means Exponential Simulation",
     col = "gray", xlab="Means Exponential")
abline(v=mean(meansExp), lwd="1", col="red")
abline(v=mean(theoreticalMean), lwd="1", col="blue")
```



Sample Variance versus Theoretical Variance

Now let's see how similar the sample variance and theoretical variance are:

```
sampleVariance <- var(meansExp)
sampleVariance
```

```
## [1] 0.6174444
```

```
theoreticalVariance  <- (1 / lambda)^2 / (n)
theoreticalVariance
```

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

Similarly, we can see that the theoretical variance **0.625** is almost the same as to the sample variance **0.6174444** using the sample means above.

```
sampleStandardDeviation <- sd(meansExp)
sampleStandardDeviation
```

```
## [1] 0.7857763
```

```
theoreticalStandardDeviation  <- 1/(lambda * sqrt(n))
theoreticalStandardDeviation
```

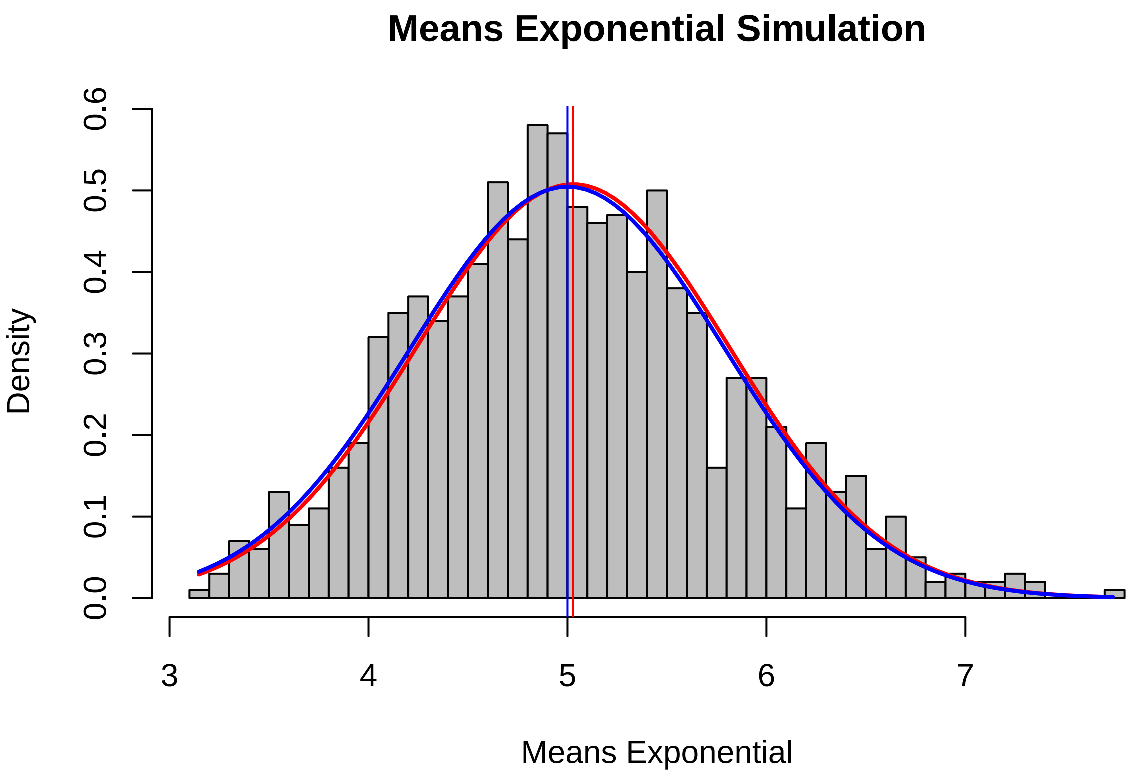
```
## [1] 0.7905694
```

It is also evident that the theoretical standard deviation **0.7905694** versus **0.7857763** is so close.

Distribution

Let's explore if the distribution is normal. Below in the graph, we have a bell curve for both distiribution using both sample and theoretical standard deviation, in blue and red colors respectively.

```
hist(meansExp,breaks=n,prob=TRUE,main="Means Exponential Simulation",
     col = "gray", xlab="Means Exponential")
x1 <- seq(min(meansExp), max(meansExp), length=100)
y1 <- dnorm(x1, mean=sampleMean, sd=sampleStandardDeviation)
lines(x1, y1,lwd=2, col="red")
abline(v=mean(meansExp), lwd="1", col="red")
x2 <- seq(min(meansExp), max(meansExp), length=100)
y2 <- dnorm(x2, mean=theoreticalMean, sd=theoreticalStandardDeviation)
lines(x2, y2,lwd=2, col="blue")
abline(v=mean(theoreticalMean), lwd="1", col="blue")
```



The graph above shows that the distribution of means is a normal distribution, based on the Central Limit Theorem, with independent random variables. It is assumed that as the size of sample increases, the closer the curves to the standard normal distribution.