Exponential Distribution vs CLT in R

Jackie Milhans

July 24, 2015

Overview

In this project we investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. We investigate the distribution of averages of 40 exponentials. Note that we do a thousand simulations.

Simulations

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

We simulate the data using a 1000 x 40 matrix. With apply(), we get 1000 means of each of the 40 samples.

```r
#load ggplot
library(ggplot2)

#Set Seed for Reproducible Simulation
set.seed(99)

numSim <- 1000
lamda <- 0.2
n <- 40

theoMean <- 1/lamda
simData <- matrix(rexp(numSim*n, lamda), numSim)
xbar <- apply(simData, 1, mean)
```

Sample Mean versus Theoretical Mean

Here we compare the sample mean versus the theoretical mean.

```r
#Empirical Mean
sampleMean <- mean(xbar)
sampleMean
```
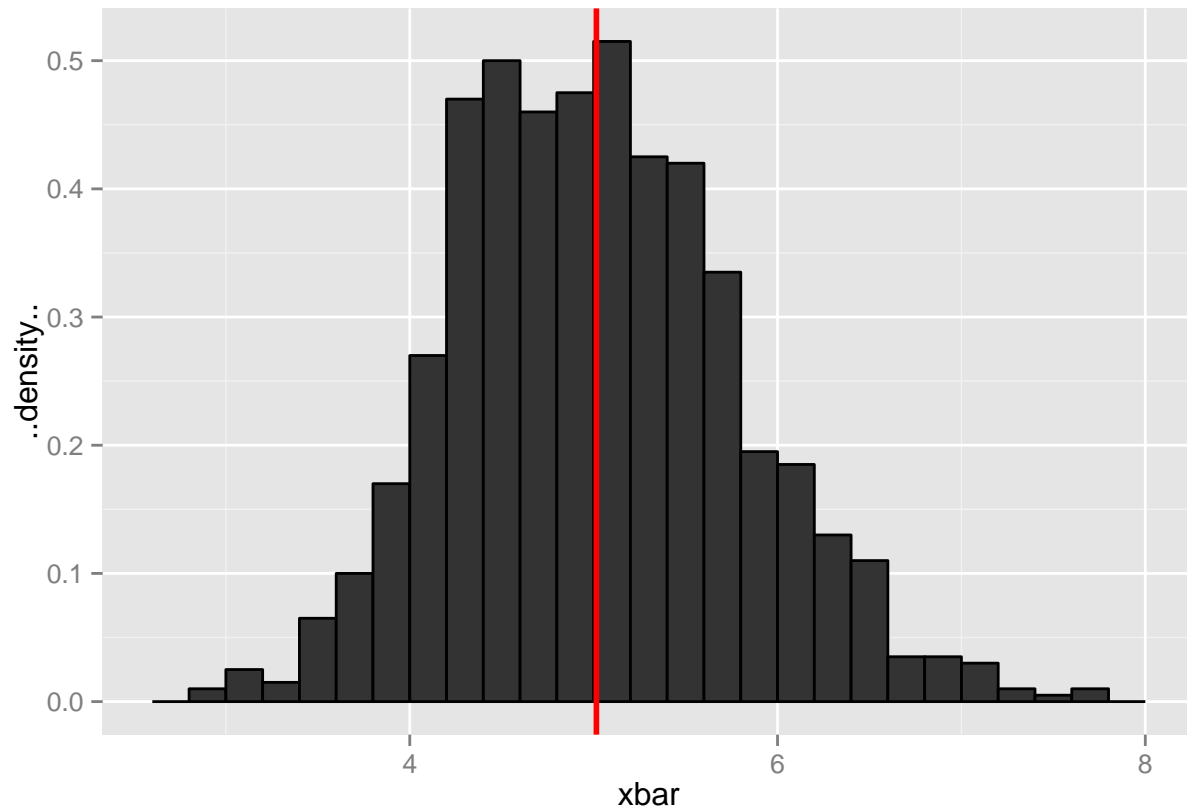
```
## [1] 5.014808
```

```r
#Theoretical Mean
theoMean
```

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

The sample mean is similar to the theoretical mean of the distribution.

The plot below shows the average of the means by the red line. This appears close to the theoretical mean of 5.

```
ggplot() + aes(xbar, y=..density..) +
        geom_histogram(,binwidth=.2, color="black") +
        geom_vline(xintercept = mean(xbar), colour = 'red',size=1)
```


Sample Variance versus Theoretical Variance:

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
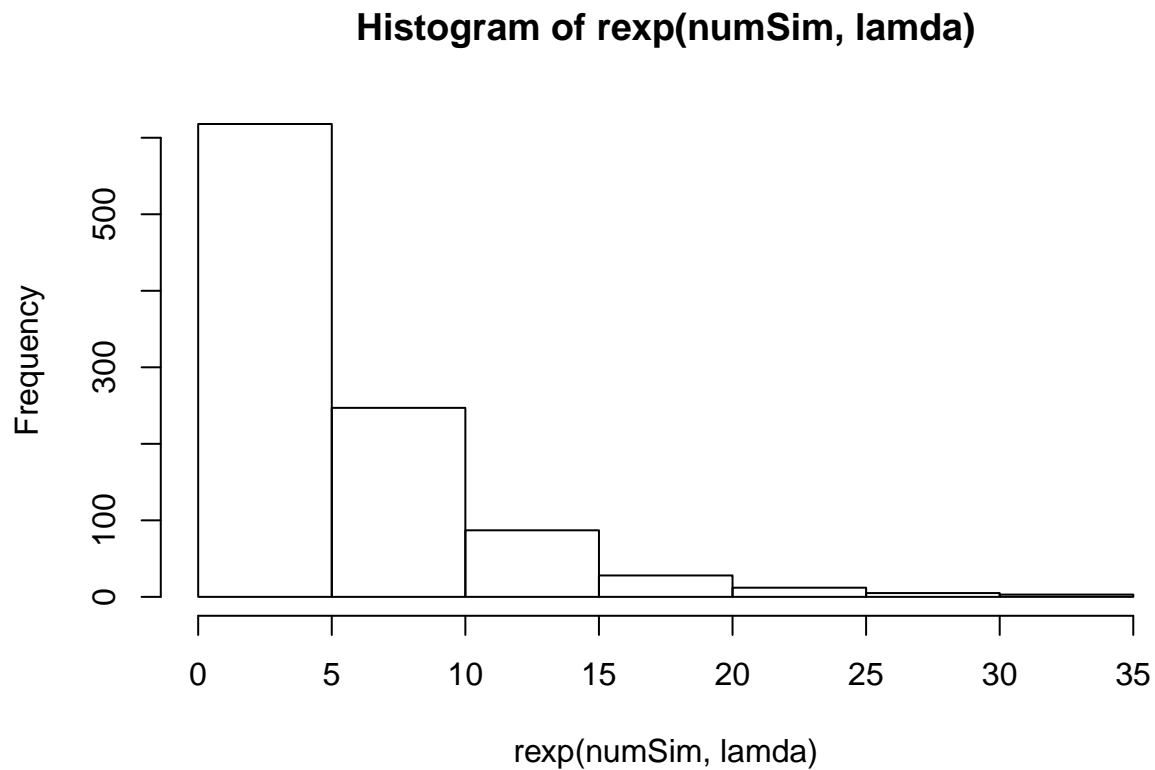
```
## [1] 0.7685519
```

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

The sample variance is 0.7685519 and is below the theoretical variance, 0.7905694. The standard deviation of the exponential distribution and the theoretical variation are not as close as expected.

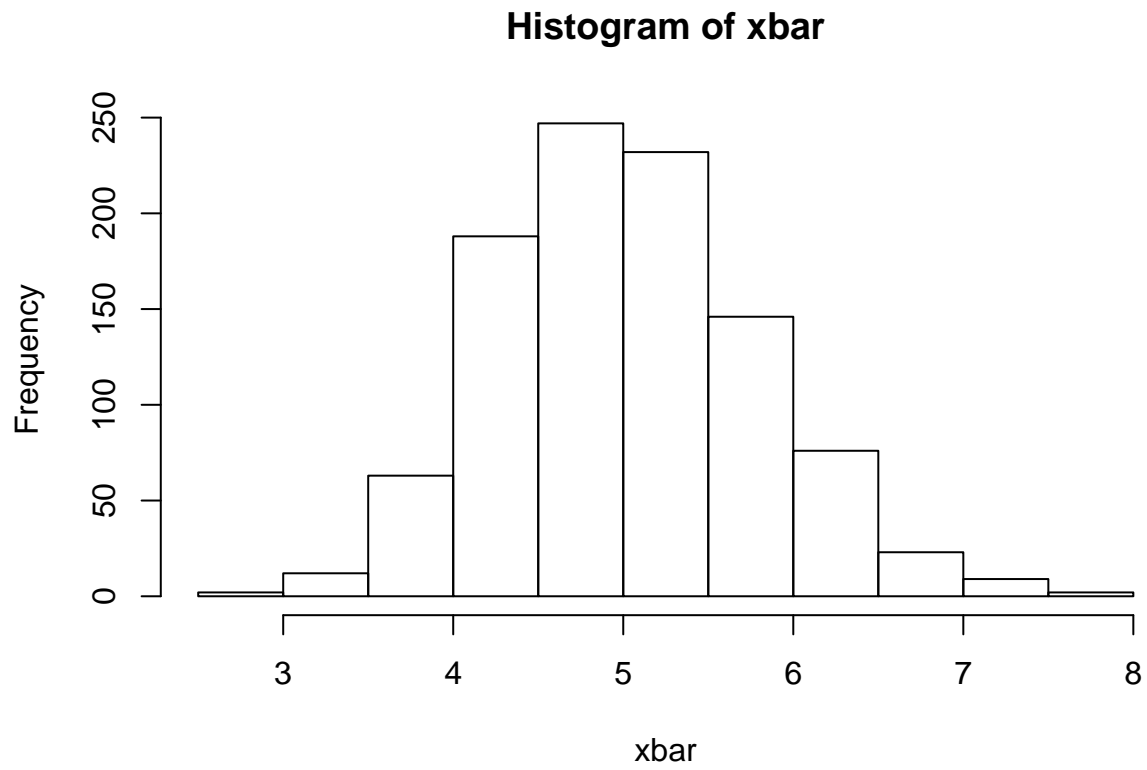## Distribution

Check if the exponential distribution is normal.

```r
hist(rexp(numSim, lamda))
```

Histogram of rexp(numSim, lamda)

From this plot, we can see that it is not a normal distribution.

Here, we check the sample data:

```r
hist(xbar)
```

Histogram of xbar

We can see that the sample data is a normal distribution, in accordance with the central limit theorem.