Peer-graded Assignment: Statistical Inference Course Project | SIMULATION

Roxana Trejos Ramirez

11/7/2020

Part 1: Simulation Exercise Instructions

In this project the student will 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.
- **mean of exponential distribution** is 1/lambda.
- **standard deviation** is also 1/lambda.

Assumptions:

1. Set lambda = 0.2 for all of the simulations.
2. You will investigate the distribution of averages of 40 exponentials.
3. Note that you will need to do a thousand simulations.

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

a. Show the sample mean and compare it to the theoretical mean of the distribution.
b. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
c. Show that the distribution is approximately normal.

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

As a motivating example, compare the distribution of 1000 random uniforms

```
#  Libraries
library(ggplot2)
```

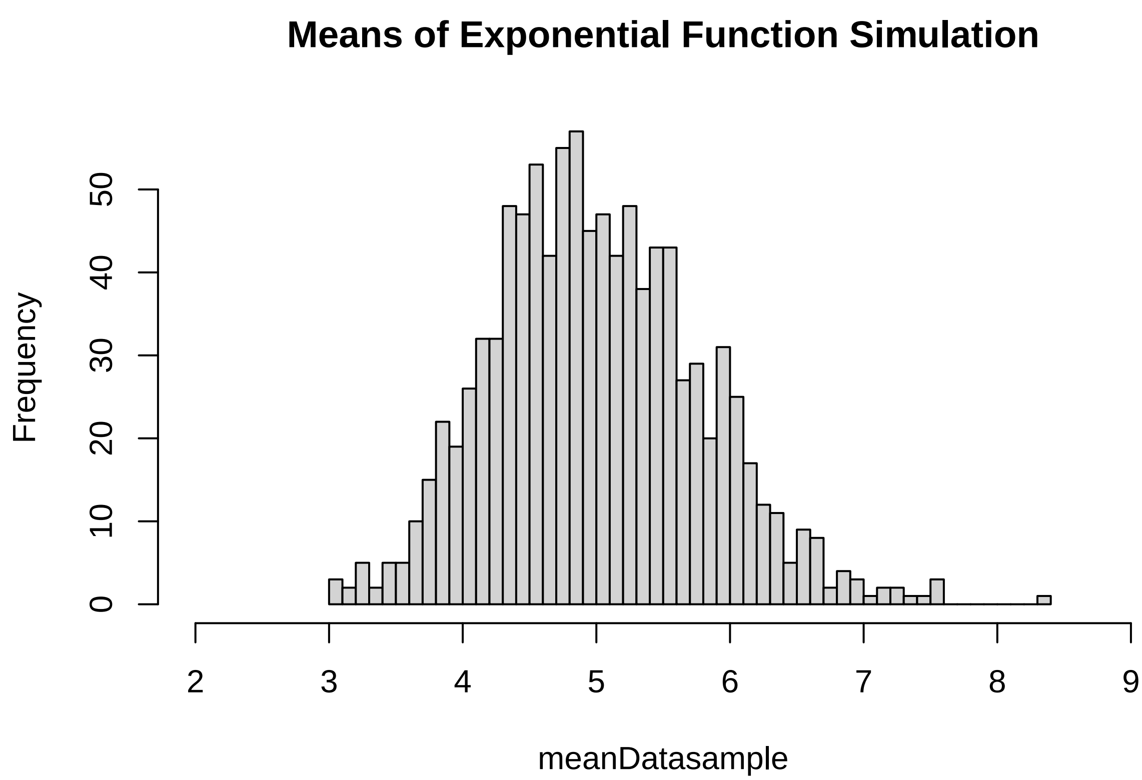
```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```r
library(tinytex)
```

```
## Warning: package 'tinytex' was built under R version 4.0.3
```

```r
set.seed(10000)
lambda <- 0.2 # Lambda
n <- 40 # Number of observations
numberSim <- 1000 # Number of simulations
dataSample <- replicate(numberSim, rexp(n, lambda))
meanDatasample <- apply(dataSample, 2, mean)

# Histogram of the simulated mean values is generated.
# Plot means histogram
hist(meanDatasample,
     breaks = 40,
     xlim = c(2,9),
     main = "Means of Exponential Function Simulation")
```



a. Show the sample mean and compare it to the theoretical mean of the distribution.

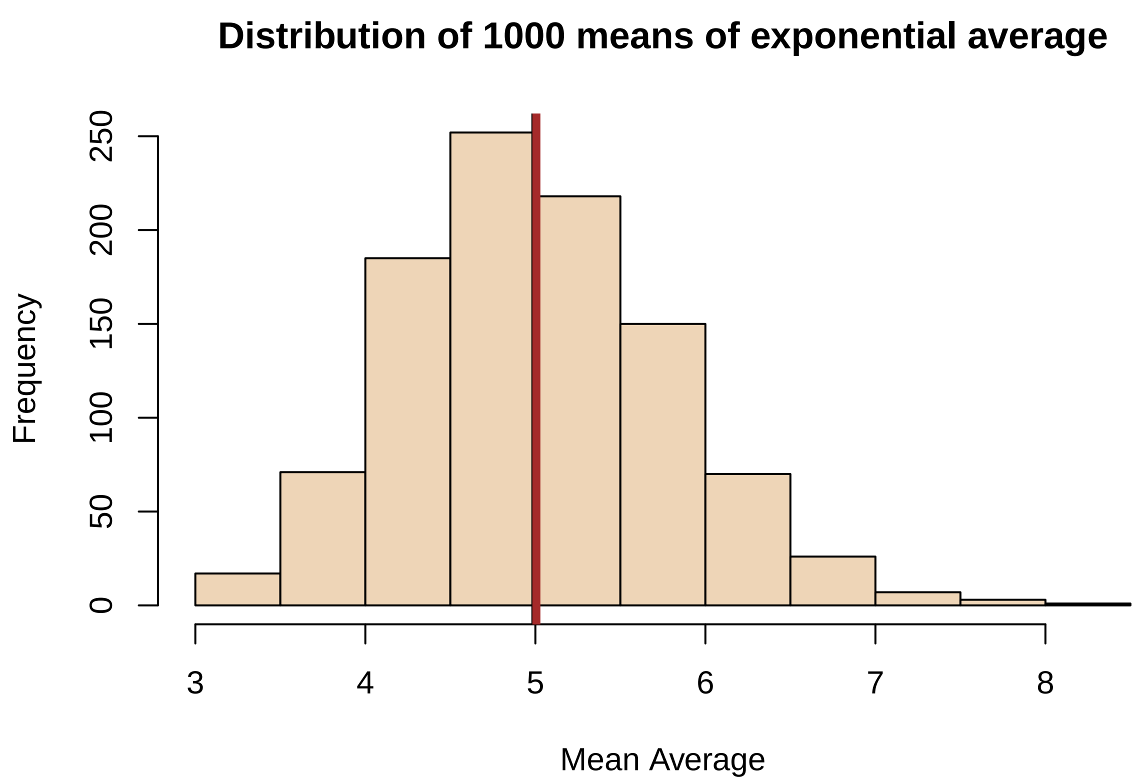
```r
# Sample Mean
sampleMean <- mean(meanDatasample) # Mean of sample means
print (paste("Sample Mean = ", sampleMean))
```

```
## [1] "Sample Mean =  5.00599048186961"
```

```r
# Theoretical Mean
theoreticalMean <- (1/lambda)
print (paste("Theoretical Mean = ", theoreticalMean))
```

```
## [1] "Theoretical Mean =  5"
```

```r
# Histogram shows differences
hist(meanDatasample,
     col="bisque2",
     xlab = "Mean Average",
     main="Distribution of 1000 means of exponential average")
abline(v = theoreticalMean, col="black", lwd = "4")
abline(v = sampleMean, col="brown", lwd = "4")
```



b. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
variance <- sd(meanDatasample)
variance
```

```
## [1] 0.7832666
```

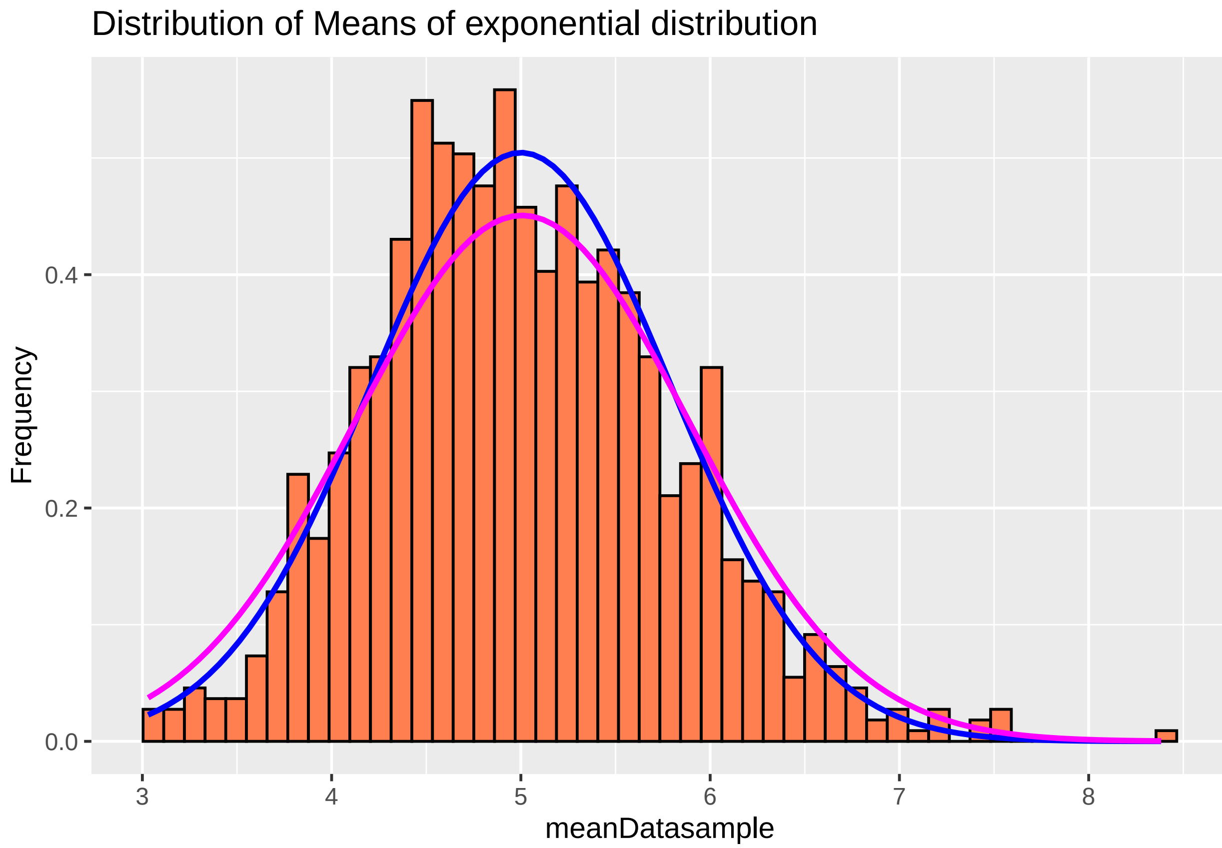
And the theoretical is

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

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

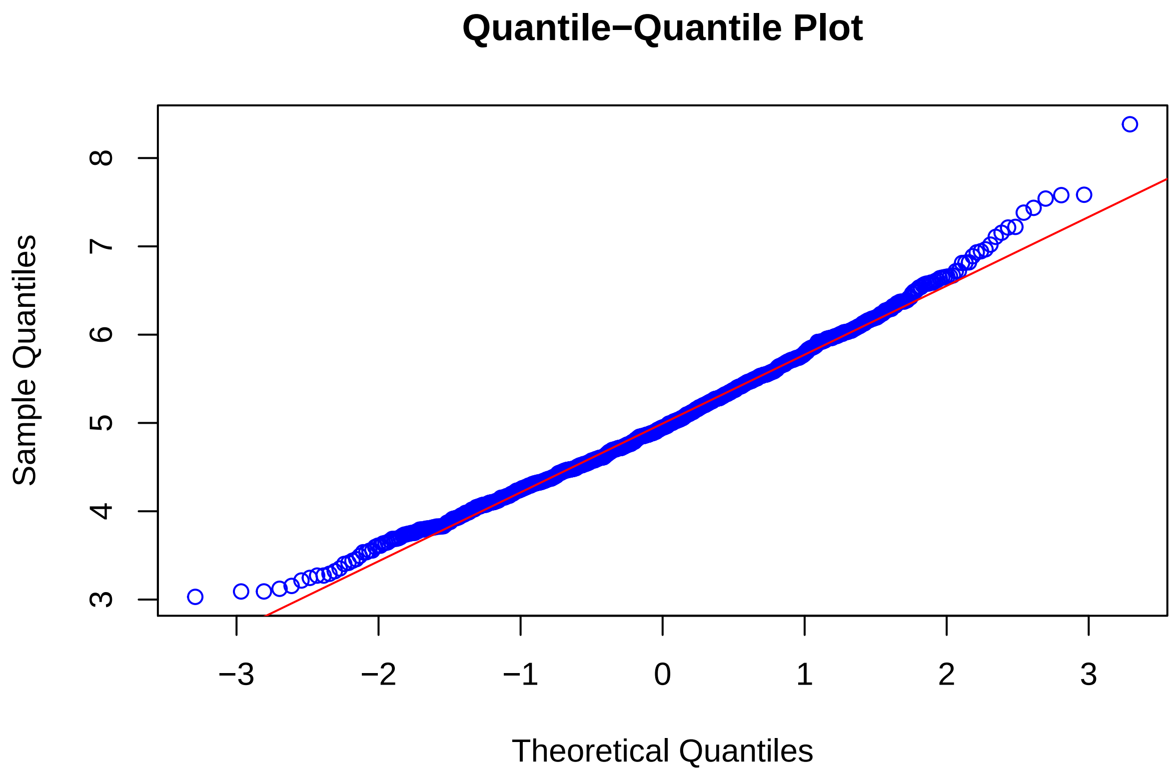
The theoretical and sample variance are quite similar

```
df <- data.frame(meanDatasample)
ggplot(df,aes(x = meanDatasample)) +
  geom_histogram(aes(y=..density..), bins=50, colour="black",fill="coral") +
  labs(title="Distribution of Means of exponential distribution", y="Frequency") +
  stat_function(fun=dnorm,args=list( mean=1/lambda, sd=sqrt(theoreticalVariance)),
                color = "blue",
                size = 1.0) +
  stat_function(fun=dnorm,args=list( mean=mean(meanDatasample),
                                     sd=sqrt(variance)),
                color = "magenta",
                size = 1.0)
```



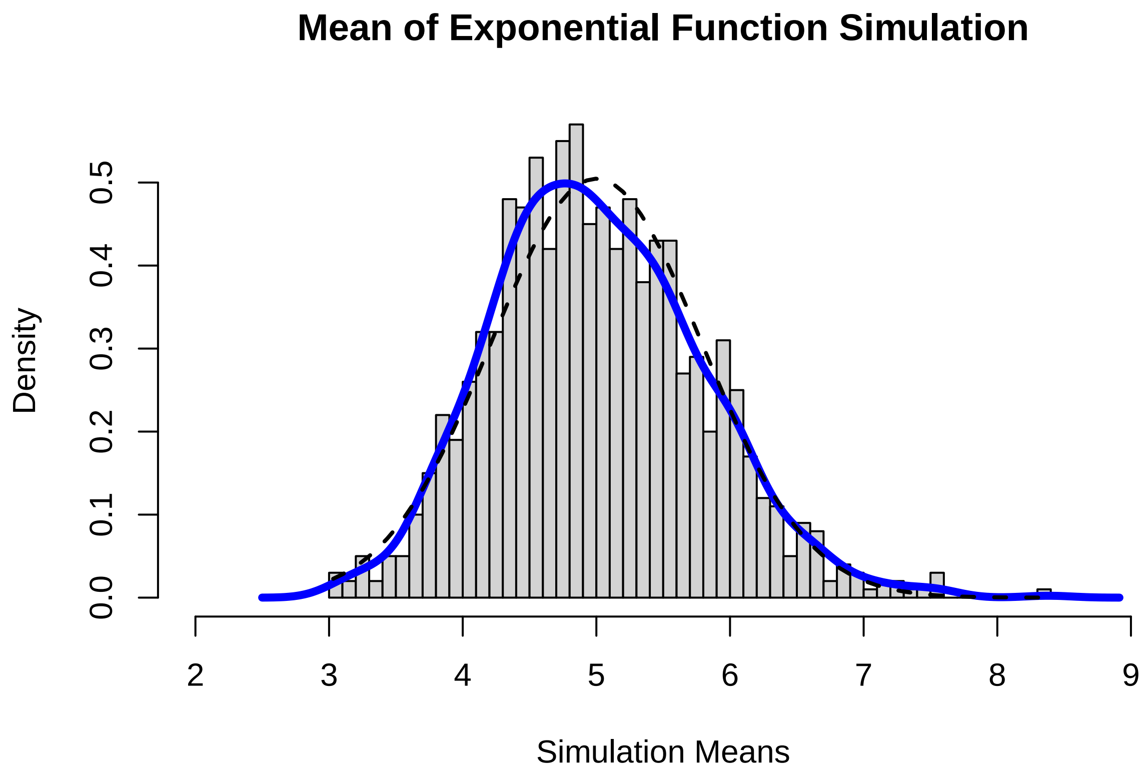
c. Show that the distribution is approximately normal.

```
# Quantile-Quantile plot
qqnorm(meanDatasample, main ="Quantile-Quantile Plot", col = "blue")
qqline(meanDatasample, col = "red")
```




```r
# Histogram with distribution curve included
hist(meanDatasample,
     prob=TRUE,
     main = "Mean of Exponential Function Simulation",
     breaks = 40,
     xlim = c(2,9),
     xlab = "Simulation Means")
lines(density(meanDatasample), lwd=4, col="blue")

# Normal distribution line
x <- seq(min(meanDatasample), max(meanDatasample), length = 2*n)
y <- dnorm(x, mean = 1/lambda, sd = sqrt(((1/lambda)/sqrt(n))^2))
lines(x,y, pch = 20, lwd = 2, lty = 2)
```



Conclusions:

Having over 1000 simulations of 40 observations:

1. Sample mean and theoretical mean of the distribution both are very close.
2. Variance sample is very similar to the theoretical variance.
3. Both graphics can show you that the distribution is normal.