Course Project (Part 1): Simulation Exercise

Cesar Rene Pabon Bernal

July 24, 2016

Overview

In this report we aim to investigate the exponential distribution in R and compare it with the Central Limit Theorem. We will illustrate, via simulation and associated explanatory text, the properties of the distribution of the mean of 40 exponentials using 1000 simulations.

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. You will investigate the distribution of averages of 40 exponentials using a thousand simulations. Three questions will be asnwered:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Enviroment information

```
library(knitr)
knitr::opts_chunk$set(echo=TRUE, fig.path='part1/', fig.width=10, fig.height=6, cache=TRUE)
set.seed(1000)
```

Simulations

Before we process and examine our data, we must first simuate our parameters using the information provided.

```
lambda <- 0.2
n <- 40
sim <- 1000
sim_matrix <- matrix(rexp(sim*n, rate=lambda), sim, n)
```

Processing & Examining Data

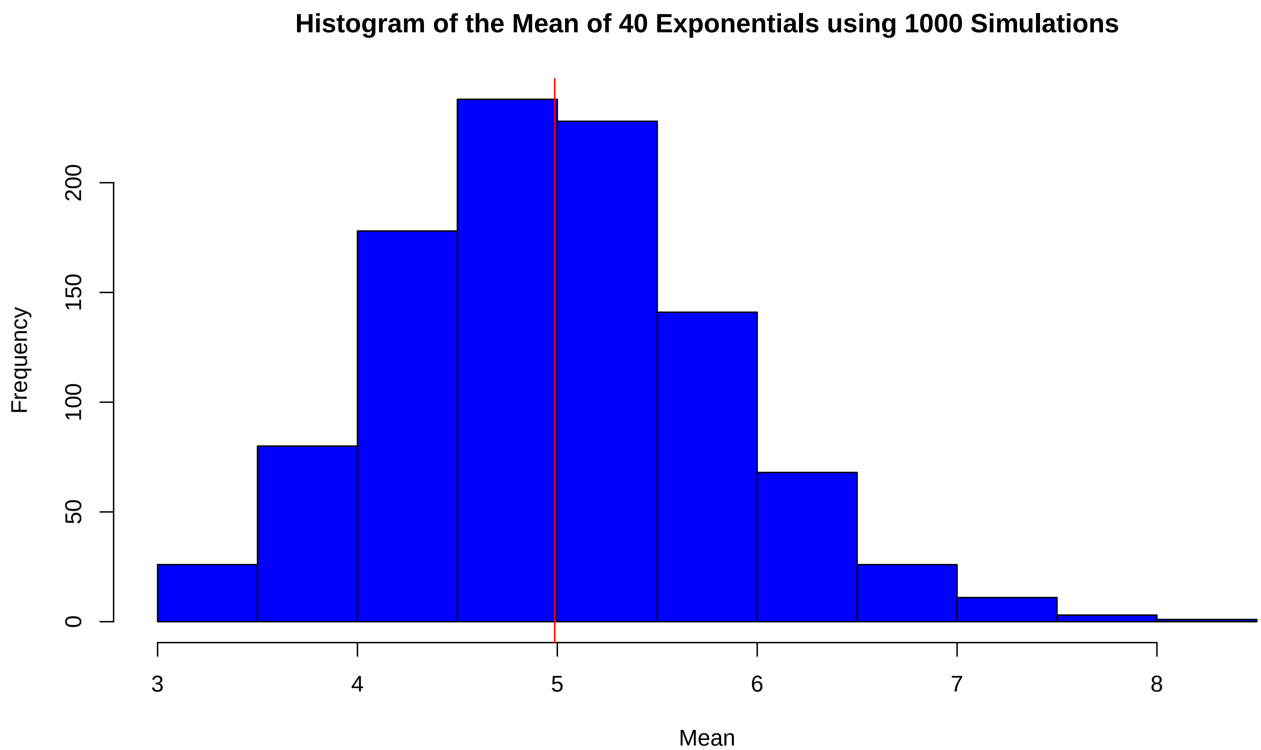
Question 1:Sample Mean versus Theoretical Mean

A. Let's find out what the sample mean is so we can then compare it to our theoretical mean and graph it

```
sim_mean <- rowMeans(sim_matrix)
sample_mean <- mean(sim_mean)
```

```
## Sample Mean:  4.986963
```

```
## Graph
hist(sim_mean, xlab="Mean", ylab="Frequency", main="Histogram of the Mean of 40 Exponentials using 1000
abline(v = sample_mean, col = "red")
```



B. Let's find out what the theoretical mean is

```
theoretical_mean <- 1/lambda
difference_of_means <- (theoretical_mean - sample_mean)
```

```
## Theoretical Mean:  5
```

```
## Differnce between both Means:  0.01303661
```

As we can wee the sample mean is 0.01303661 less than the theoretical mean; quite close.

Question 2: Sample Variance versus Theoretical Variance

A. Using variance, let's find out how variable the sample is and compare it to the theoretical variance of the distribution.

```
sample_var <- var(sim_mean)
theoretical_var <- (1/lambda)^2/n
difference_of_variance <- (theoretical_var - sample_var)
```

```
## sample variance:  0.6583551
```

```
## theoretical variance:  0.625
```

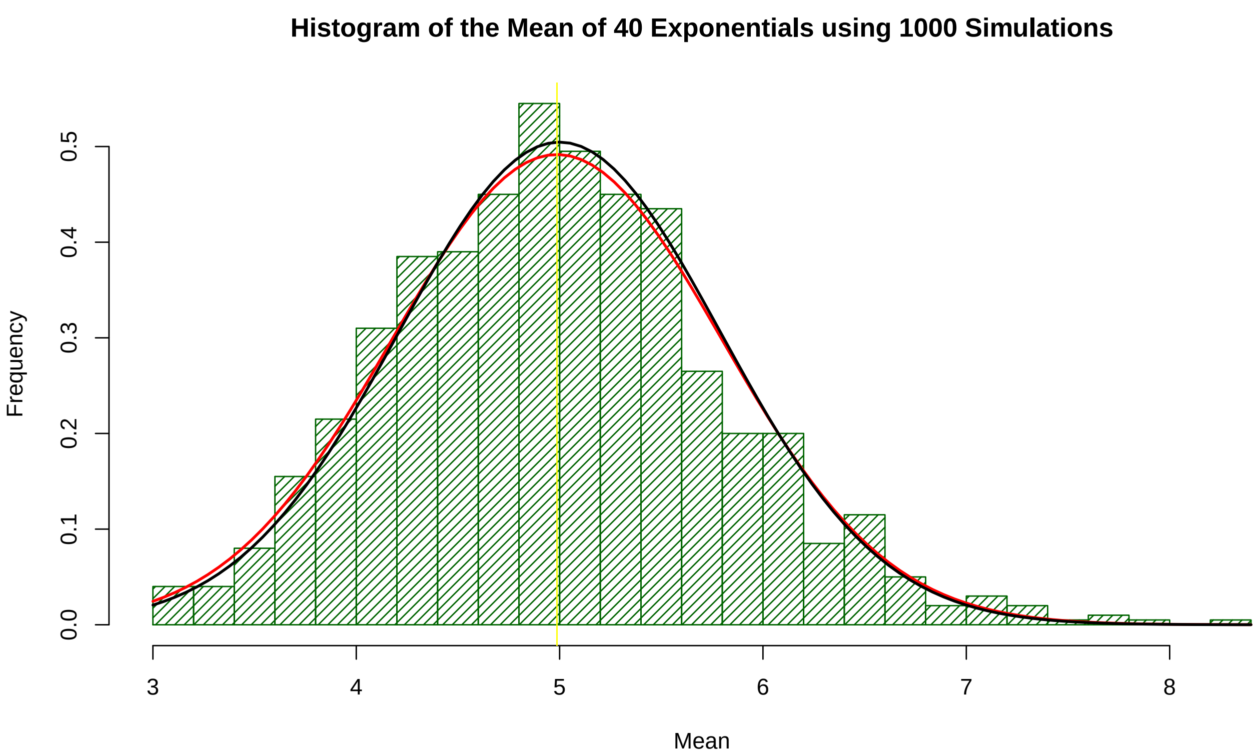
```
## Differnce between both Variances:  -0.03335507
```

As we can see the sample variance is 0.03335507 greater than the theoretical variance; again, very close.

Question 3: Distribution

A. In order to show that the distribution is approximately normal, let's first create and examine an approximate normal distribution for our data.

```
hist(sim_mean, density=20, breaks=20, prob=TRUE, xlab="Mean", ylab="Frequency", main="Histogram of the M
curve(dnorm(x, mean=sample_mean, sd=sqrt(sample_var)), col="red", lwd=2, lty = "solid", add=TRUE, yaxt=
curve(dnorm(x, mean=theoretical_mean, sd=sqrt(theoretical_var)), col="black", lwd=2, add=TRUE, yaxt="n")
abline(v = sample_mean, col = "yellow")
```



As we can see, normal and theoritcla results are very close and therefore can be examined.

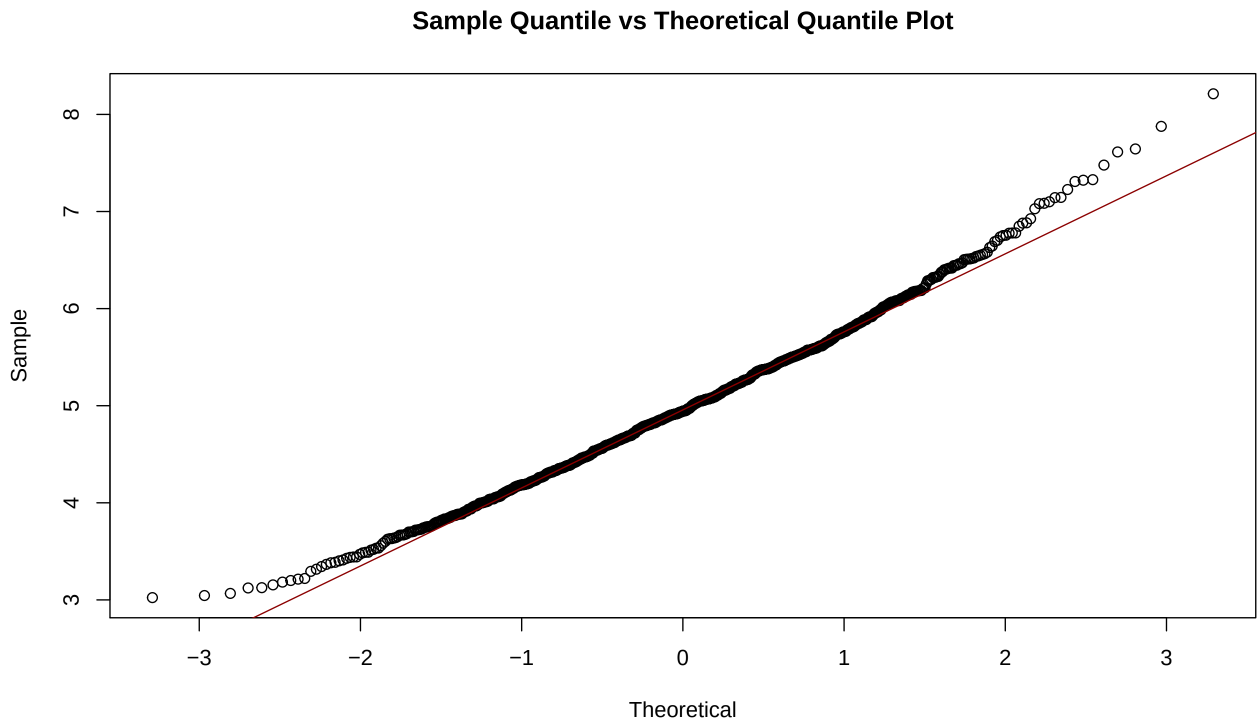
B. In order to reassure our resuls, confidence intervals will be used.

```
Sample_confidence <- round (mean(sim_mean) + c(-1,1)*1.96*sd(sim_mean)/sqrt(n),3)
Theoretical_confidence <- theoretical_mean + c(-1,1)*1.96*sqrt(theoretical_var)/sqrt(n)
```

```
## Sample Confidence Interval:  4.736 5.238
```

```
## Theoretical Confidence Interval:  4.755 5.245
```

C. In order to finalize our comparison, a sample quantile to theoretical quantile examination will be performed



Conclusion

In conclusion, we have observed, that the distribution of the mean of 40 exponentials using 1000 simulations is close to a normal distribution. The sample and theortical results (mean and vairance) exemplified such results through investigation of the exponential distribution and comparisons to Central Limit Theorem (CLT).