*Statistical Inference Course Project - Part*1

Giovanny Jimenez Gomez

November 24, 2020

1. A simulation exercise.

In this project you 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. 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. Note that you will need to do a thousand simulations.

1. Simulations.

The sample mean or empirical mean and the sample covariance are statistics computed from a collection of data on one or more random variables.

We are begin by assigning the next relevant parameters:

```
lambda <- 0.2       # It's labmda.
n.exp <- 40         # The number of exponentials.
n.sims <- 1000      # The number of simulations.
set.seed(2018)
```

Then, we proceed to run the simulations:

```
simulation.data <- as.data.frame(replicate(n.sims, rexp(n.exp, lambda)))
```

As a result we obtained a data frame with 40 rows and 1000 columns; each row corresponding to a simulation.

2. Sample Mean versus Theoretical Mean.

Now us find the the sample means:

```
data.means <- apply(simulation.data, 2, mean)
```

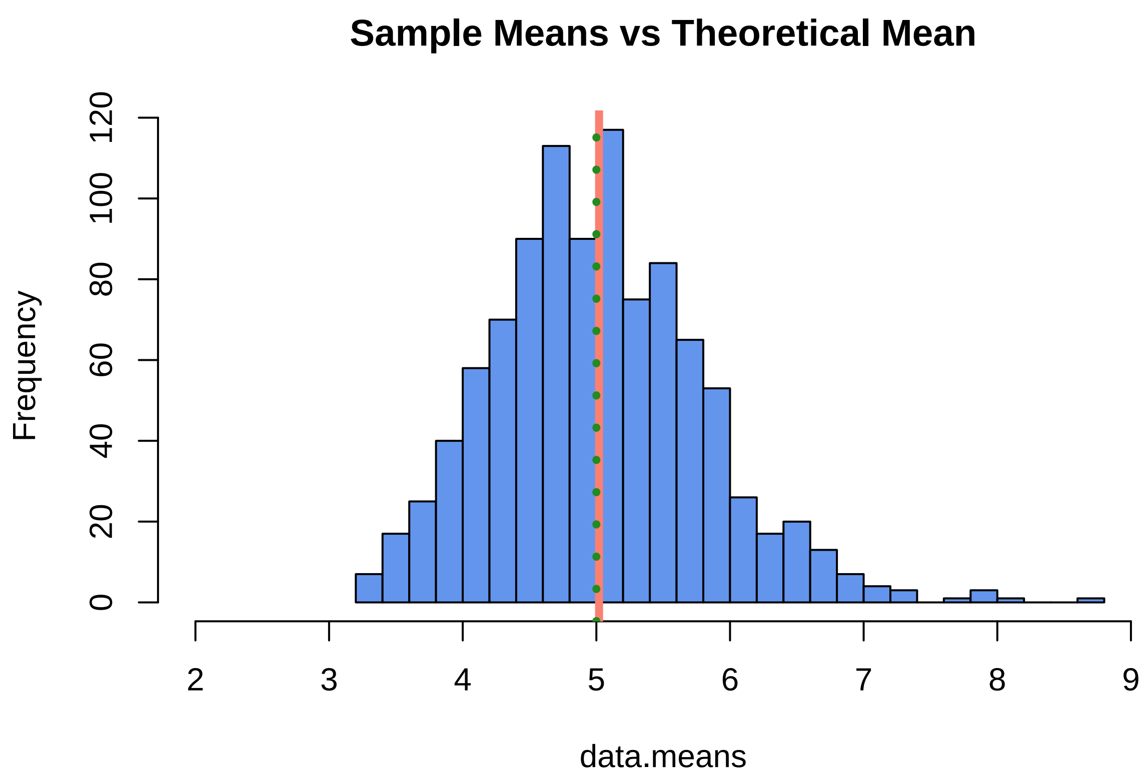
The theoretical mean of exponential distribution is $1/lambda$. In our case, this corresponds to $\mathbf{1/(2/10) = 10/2 = 5}$. The simulation sample mean is:

```
mean(data.means)
```

```
## [1] 5.020107
```

Eventually we are create a plot in order to compare the theoretical mean and the simulations mean:

```r
hist(data.means, breaks=30, xlim = c(2,9), main="Sample Means vs Theoretical Mean", col = "cornflowerbl
abline(v=mean(data.means), lwd="4", col="salmon")    # Simulation Mean
abline(v=5, lwd="4", col="forestgreen",lty="dotted")       # Theoretical Mean
```



3. Sample Variance versus Theoretical Variance.

From the Central Limit Theorem, we know the theoretical standard deviation of the mean is $(1/lambda)/sqrt(n)$, and the variance is $(1/lambda)^2/n$.

```r
print(paste("Theoretical variance: ", round( ((1/lambda)/sqrt(n.exp))^2 ,5)))
```

```
## [1] "Theoretical variance:  0.625"
```

```r
print(paste("Sample variance: ", round(var(data.means) ,5) ))
```

```
## [1] "Sample variance:  0.62613"
```

```r
print(paste("Theoretical standard deviation: ", round( (1/lambda)/sqrt(n.exp) ,5)))
```

```
## [1] "Theoretical standard deviation:  0.79057"
```

```
print(paste("Sample standard deviation: ", round( sd(data.means),5)))
```

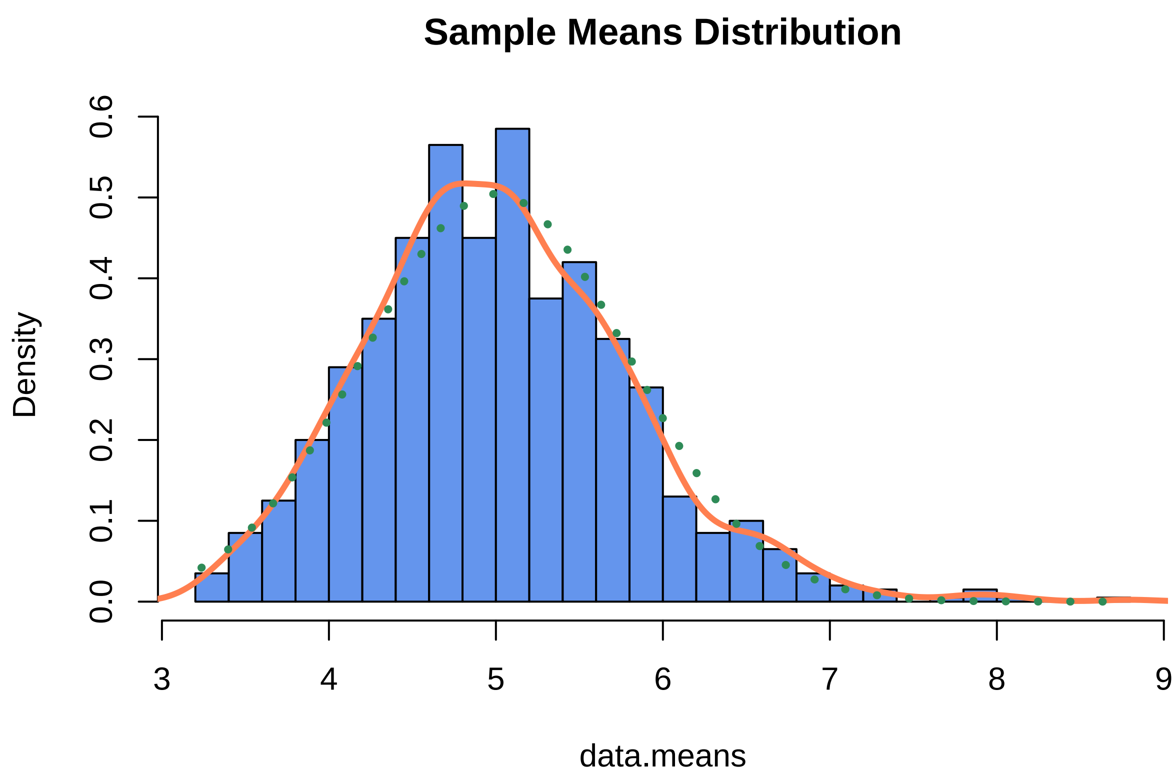
```
## [1] "Sample standard deviation:  0.79129"
```

The results show that variances are very close, as well as standard deviations.

4. Distribution.

Finally, let us investigate if the exponential distribution is approximately normal. We know from the Central Limit Theorem that the averages of samples should follow a normal distribution as samples increase in number.

```
#Plotting of the mean distribution of the samples
hist(data.means, breaks = 30, prob=TRUE, col="cornflowerblue", main="Sample Means Distribution")
lines(density(data.means), lwd=3, col="coral")
#Plotting of the normal distribution line
x <- seq(min(data.means), max(data.means), length=2*n.exp)
y <- dnorm(x, mean=1/lambda, sd=sqrt(((1/lambda)/sqrt(n.exp))^2))
lines(x, y, pch=22, col="seagreen", lwd=4, lty = "dotted")
```



The graph shows that the distribution of the sample means (of our exponential distributions) resembles a normal distribution. Once again, this is secured by the Central Limit Theorem, that states that if we were to increase our number of samples (currently 1000), the distribution would be even closer to the standard normal distribution. The green dotted line above is a normal distribution curve and we can see that it is very close to our sampled curve, which is the red line above.