Statistical Inference Course Project

Ignacio Ojea

June 2, 2018

Course Project - Part 1

Assignment 1

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.

We begin by assigning the relevant parameters:

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

We then proceed to run the simulations:

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

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

Assignment 2: Sample Mean vs Theoretical Mean

Let us find 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 1/(2/10) = 10/2 = 5. The simulation sample mean is:

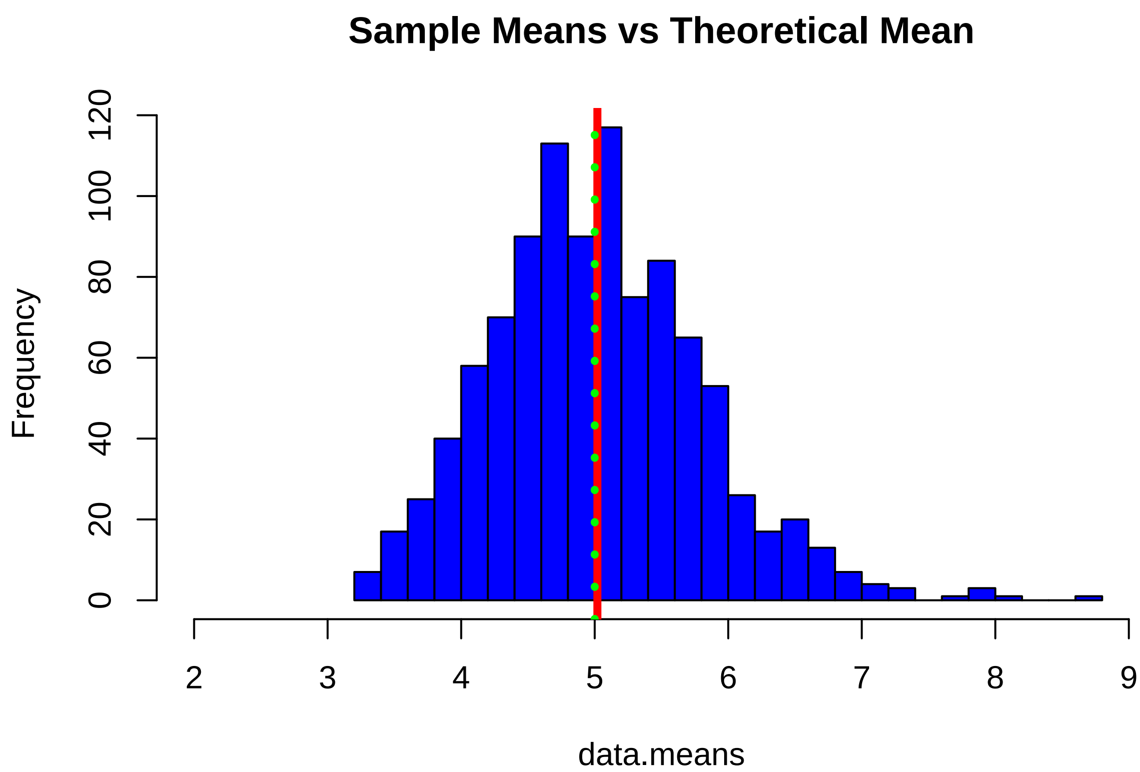
```
mean(data.means)
```

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

Which is very close to the theoretical mean.

Let us now create a plot in order to compare the theoretical mean and the simulations mean:

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



Assignment 3: Sample Variance vs Theoretical Variance From CLT, we know the theoretical standard deviation of the mean is (1 / lambda) / sqrt(n), and the variance is (1 / lambda)^2 / n. Let us now compare:

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

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

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

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

```
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 standar deviations.

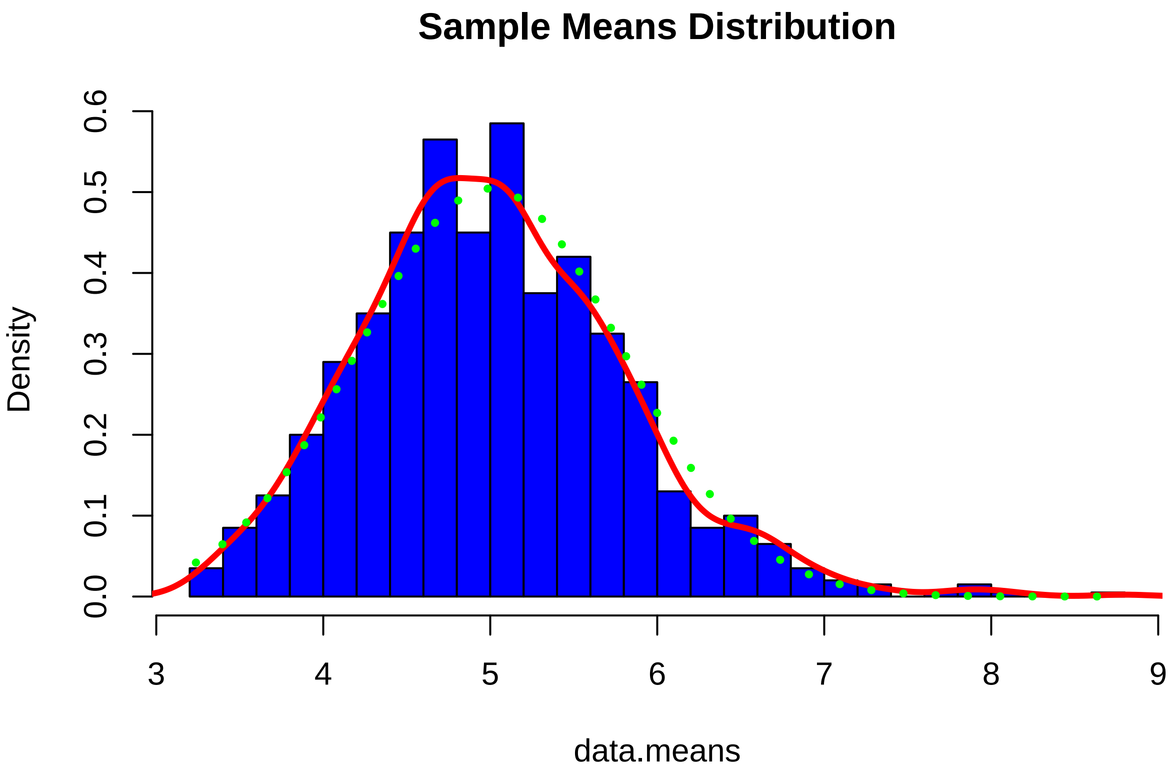
Assignment 2: Approximation to normality

Finally, let us investigate if the exponential distribution is approximately normal. We know from the CLT 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="blue", main="Sample Means Distribution")
lines(density(data.means), lwd=3, col="red")

#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="green", 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.