Simulating Exponential Distribution with R in Comparsion with Central Limit Theorem

Eric Calasans de Barros

December 11, 2017

1 Introduction

In this project we are make an experiment to verify the central limit theorem(CLT), one of the most important theorems of Statistics, by the simulation of 40 random variables with **exponential distribution** in a thousand times repeating test. For each test we take the mean and, at the end, we verify the mean distribution.

2 Methodology

2.1 Load Packages and Build Data Frame

First we load the necessary packages for the task, *dplyr* and *ggplot2*. Given $\lambda = 0.2$(parameter for exponential distribution) and $n = 40$(number of samples per test), we have the following code:

```
library(dplyr)

## Warning:  package 'dplyr' was built under R version 3.4.2

library(ggplot2)

#Parameters for exponential distribution
lambda <- 0.2
samplesPerTest <- 40
numTests <- 1000
```

```
#Matrix of tests
dataSim <- rexp(n = samplesPerTest*numTests,lambda)
testMatrix <- matrix(data = dataSim, nrow = numTests)
```

2.2 Computation of Sample Mean

Now we compute the mean for each line of the matrix as shown below:

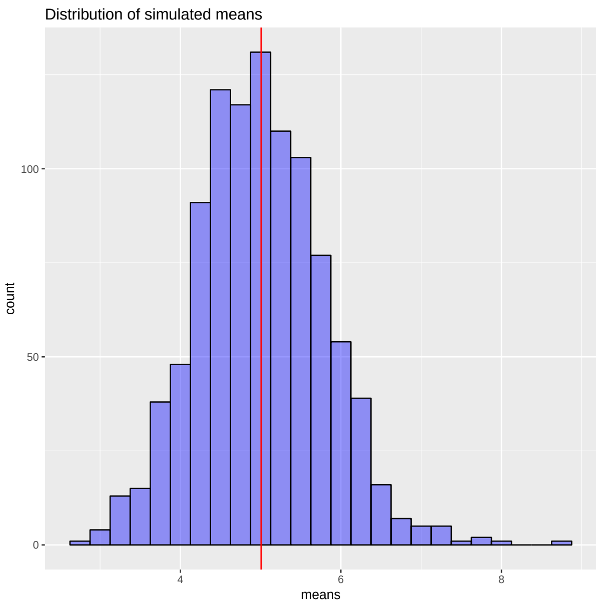
```
meanSamples <- apply(testMatrix,1,mean)
meanSamplesDF <- data.frame(means = meanSamples)
```

And then, we take the mean of the means in result above:

```
##  simMean
## 5.005492
```

Plotting the histogram, we have:

```
g<- ggplot(meanSamplesDF,aes(x = means) ) + geom_histogram(alpha=0.4,
                                binwidth= .25, fill = "blue", col = "black") +
  geom_vline(xintercept = meanSimulation, color="red", size = 0.5) +
  ggtitle("Distribution of simulated means")
g
```



As we can see, the simulated mean, by CLT approaches the theorical mean given by $\frac{1}{\lambda} = 5$.

2.3 Computation of Sample Variance

In this we compute the sample variance in CLT and compare it with the theorical given by $(\frac{\frac{1}{\lambda}}{\sqrt{n}})^2$.

```
varSamples <- meanSamplesDF  %>% select(means)  %>% unlist() %>%  sd()
varSamples <- varSamples ^ 2
varSamples

## [1] 0.6222146
```

If $(\frac{\frac{1}{\lambda}}{\sqrt{n}})^2 = 0.625$ we can observe the confirmation of CLT with respect to variance.

2.4 Normal Distribution

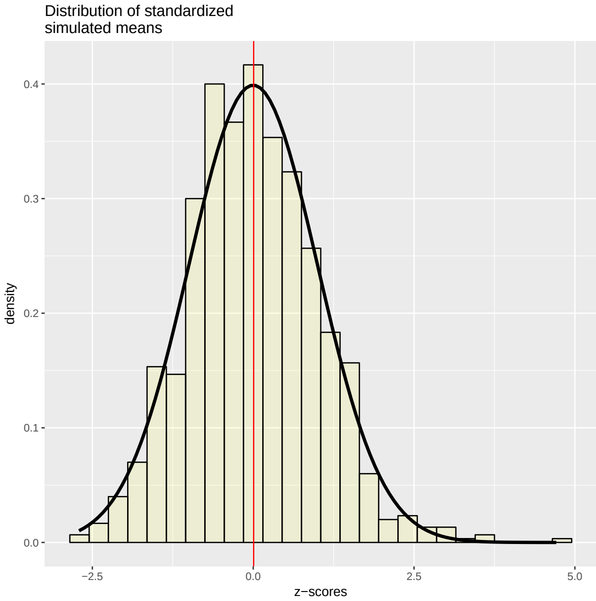
The CLT states that as long as n grows the mean distribution of normalized random variables approaches standard normal distribution(mean $=$ 0 and variance $=$ 1). That's we intend to show now. The code below show the procedures:

```
zScore <- (meanSamplesDF$means - (1/lambda))
zScore <- zScore/(1/lambda/sqrt(samplesPerTest))
zMean <- mean(zScore)
zMean

## [1] 0.006947234
```

Then we plot:

```
zPlot <- ggplot(as.data.frame(zScore), aes(x = zScore)) +
   geom_histogram(alpha=0.1,  binwidth = 0.3, fill="yellow",
                  color="black", aes(y = ..density..)) +
   stat_function(fun = dnorm, size = 1.3) +
   geom_vline(xintercept = zMean, color="red", size = 0.5) +
   ggtitle("Distribution of standardized \nsimulated means") +
   xlab("z-scores")
zPlot
```



Distribution of standardized
simulated means
density
0.4
0.3
0.2
0.1
0.0
−2.5
0.0
2.5
5.0
z−scores