Project-01

jose Antonio (joseantonio@me.com)

github: Project

I- TITLE: Simulating Exponential Distribution

II - OVERVIEW:

In this project we use the exponential distribution in R and compare it with the Central Limit Theorem. The following criteria will be used for this project:

- mean of exponential distribution = 1/lambda
- standard deviation = 1/lambda
- lambda = 0.2 for simulations
- investigate distribution of 40 exponentials (will need to do 1000 simulations)

The problems to se solved are:

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.

Libraries

```
getwd()

## [1] "/Volumes/Dados/Coursera/0-A-STATISTIC-INFERENCE"

library(knitr)
library(markdown)
library(rmarkdown)
library(plyr)
library(stats)
library(ggplot2)
library(tidyr)
```

III - SIMULATIONS

Answering the questions using R and explanations

1 - Sample Mean versus Theoretical Mean

```
set.seed(1) ## reproducible results
lambda <- 0.2 # rate for simulations
```

```r
n <- 40 # number of samples
simulations <- 1:1000 # suggested
meansd <- 1/0.2 # mean of exponential distribution
sdexp <- 1/0.2 # standard deviatin of exponential distribution
```

1.1 - Sample Mean

```r
sample_mean <- data.frame(simcol=sapply(simulations, function(simcol)
        {mean(rexp(n,lambda))}))
```

1.2 - Center Sample Distribution

```r
mean(sample_mean$simcol)

## [1] 4.990025
```

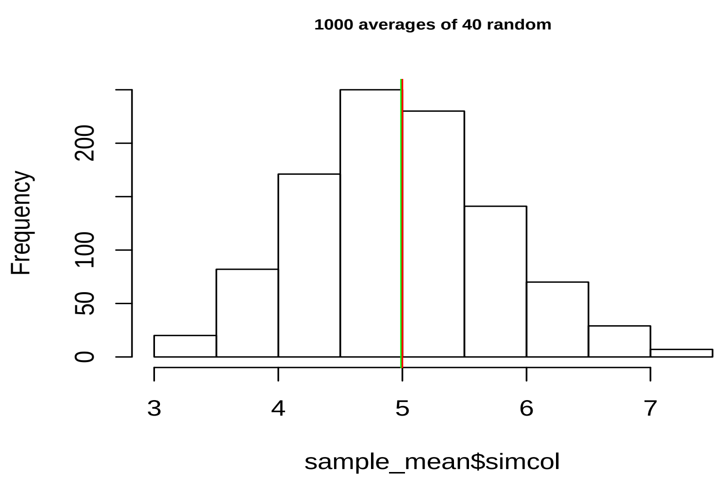
1.3 - Center Theoretical Distribution

```r
1/lambda

## [1] 5
```

- THE SAMPLE MEAN (4,99) IS VERY CLOSE TO THEORETICAL MEAN (5)

1.4 - Histogram

```r
hist(sample_mean$simcol, main = "1000 averages of 40 random", cex.main=0.7)
abline(v=mean(sample_mean$simcol),col="green")
abline(v=1/lambda,col="red")
```



2 - Sample Variance versus Theoretical Variance

2.1 - Calculating the standard deviation (sample and theorical distribution)

```r
sd(sample_mean$simcol, na.rm = TRUE) # sd of sample distribution

## [1] 0.7817394

(1/lambda)/sqrt(40) # sd of theoretical distribution
```

```
## [1] 0.7905694
```

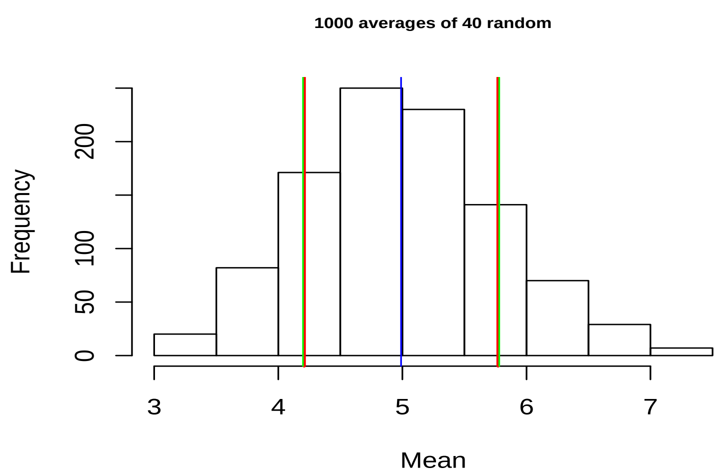
2.2 - Calculating Variance (sample and theorical distribution)

```
## Variance of sample distribution
var(sample_mean$simcol)

## [1] 0.6111165

## Variance of theoretical distribution
((1/lambda)/sqrt(40))^2

## [1] 0.625
```

- THE SAMPLE SD IS CLOSE TO THEORETICAL SD. THE VARIANCE IS SIMILAR

2.3 - Histogram

```
hist(sample_mean$simcol, main = "1000 averages of 40 random", xlab = "Mean",
cex.main=0.7)
abline(v=mean(sample_mean$simcol),col="blue")
## Standard Deviation
abline(v=mean(sample_mean$simcol)+ sd(sample_mean$simcol), col="red", lw=2)
abline(v=mean(sample_mean$simcol)- sd(sample_mean$simcol), col="red", lw=2)
## Theoretical SD
abline(v=mean(sample_mean$simcol)+sqrt(sdexp^2/n), col = "green")
abline(v=mean(sample_mean$simcol)-sqrt(sdexp^2/n), col = "green")
```

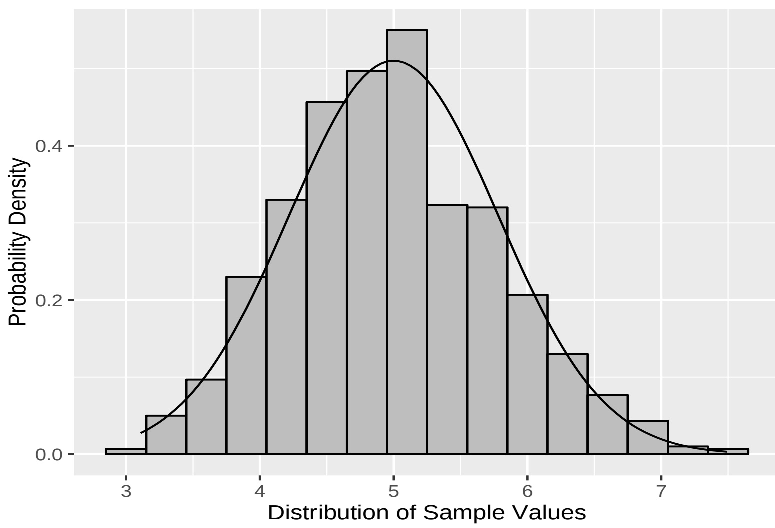


3 - Distribution

Comparing large collection or random exponentials (1000) against large coolection of averages (40). Show that the distribution is approximately normal.

3.1 -

```
g <- ggplot(data = sample_mean, aes(x = simcol))
g <- g + geom_histogram(aes(y = ..density..), fill = I('grey'), binwidth = 0.
30, color = I('black'))
g <- g + xlab("Distribution of Sample Values") + ylab("Probability Density")
g <- g + stat_function(fun = dnorm, args = list(mean = 5, sd = sd(sample_mean
$simcol)))
g
```



- As the overlay in Graphic, the distribution appears Normal.