

# Part One: A simulation exercise

Samir N. Hag Ibrahim

7/12/2020

## PART ONE: SIMULATION EXERCISE

### General Overview

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.

### Instructions and Requirements

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

illustrate the followings: 1. Show the *sample mean* and compare it to the *theoretical mean* of the distribution. 2. Show how *variable the sample* is and compare it to the *theoretical variance* of the distribution. 3. Show that the *distribution* is approximately *normal*.

### 1. Show the sample mean and compare it to the theoretical mean of the distribution.

Set the variables as defined

```
lambda <- 0.2
n <- 40
nosim <- 1000
set.seed(1234)
simData <- matrix(rexp(n*nosim, rate = lambda), nosim)
```

Now, calculate the Mean of the simulated data (*for code see appendix*)

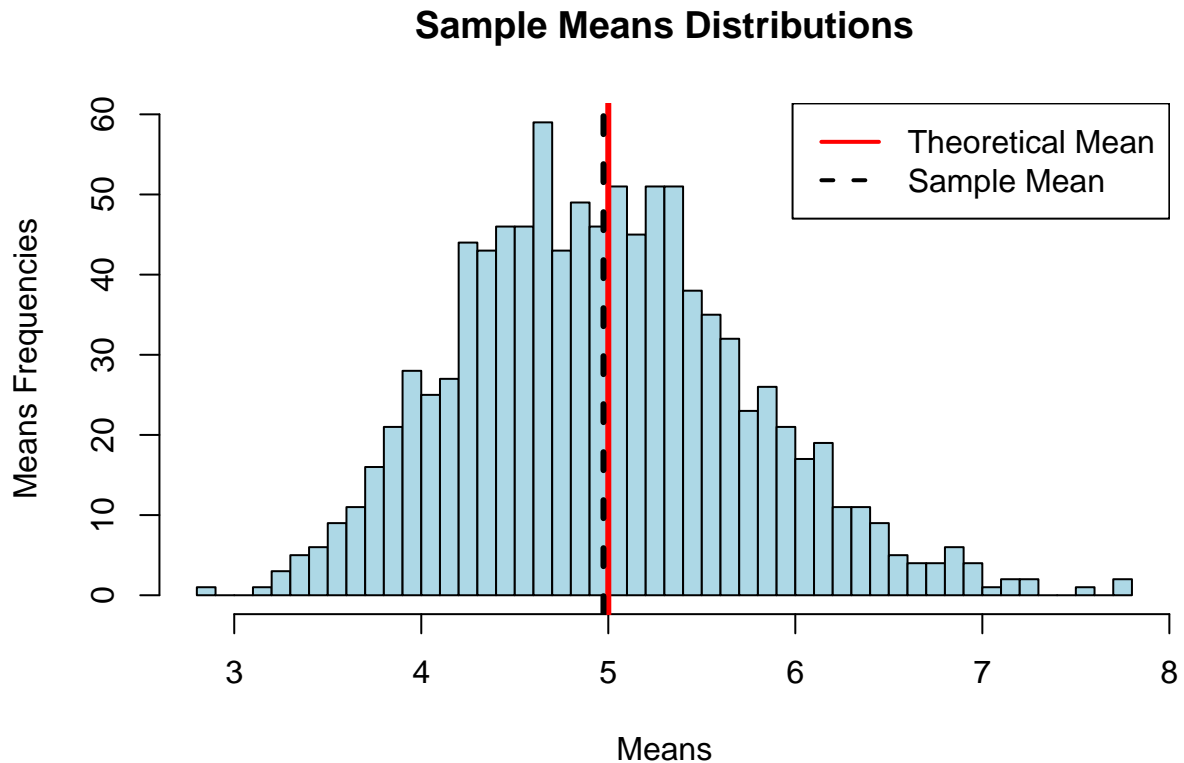
```
sampleMean # sample mean
```

```
## [1] 4.974239
```

```
theorMean # theoretical mean
```

```
## [1] 5
```

the sample means' distribution was found to be centered at 4.9742388, where the theoretical mean is 5 also see the plot below, indicating that both the sample mean and the theoretical mean are almost identical. (*for plot code see appendix*)



## 2. Show how variable the sample is and compare it to the theoretical variance of the distribution.

in this part, the sample variance will be compared with the theoretical variance reported.

first, calculating the sample variance of the means and the theoretical variance (*for code see appendix*)

```
sampleVar
```

```
## [1] 0.5949702
```

```
theorVar
```

```
## [1] 0.625
```

Accordingly, the sample variance of the means was found to be 0.5949702, where the theoretical variance was found to be 0.625 which are “almost identical”.

## 3. Show that the distribution is approximately normal.

First lets calculate the sd form both sample means and the theoretical means (*for code see appendix*):

```
sampleSD
```

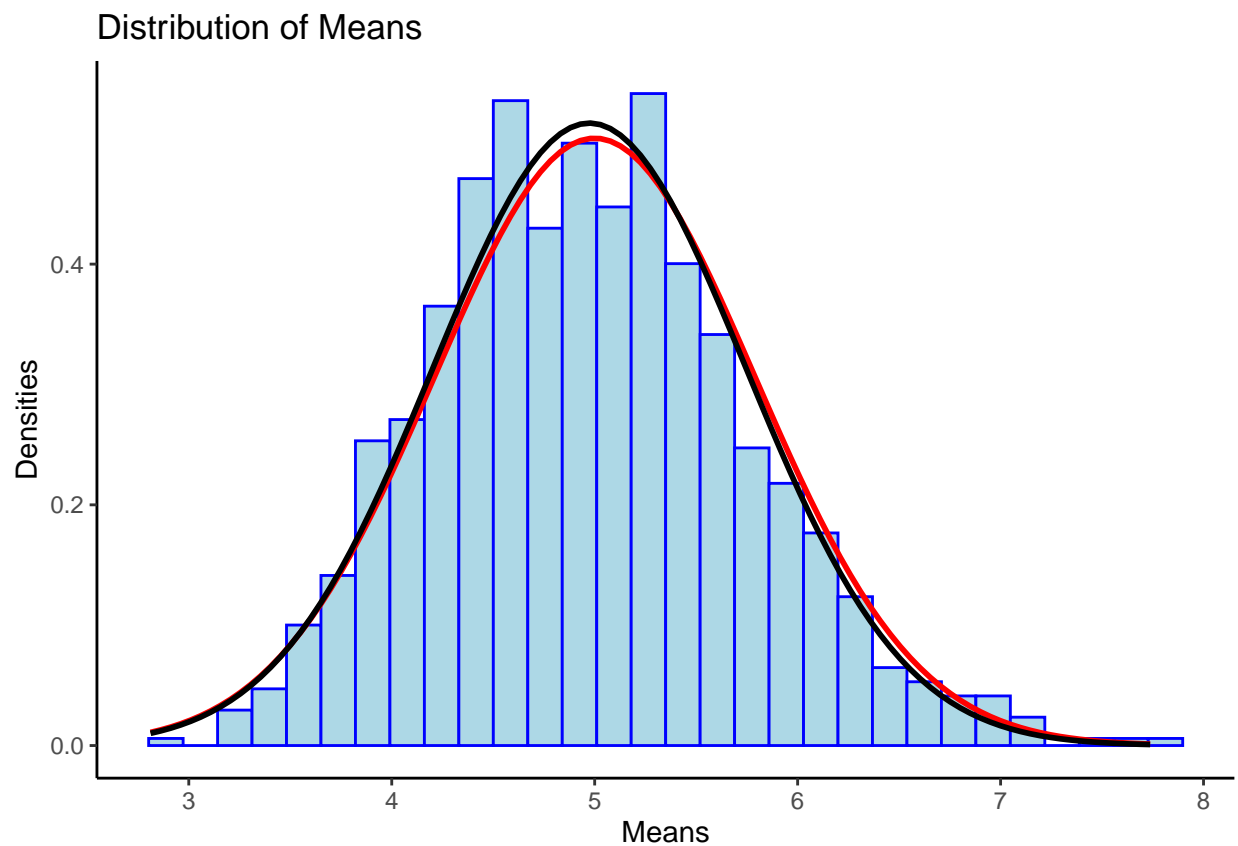
```
## [1] 0.7713431
```

```
theorSD
```

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

we observe that that both SD are close. to get more insight, the below is the distribution plot (*for code see appendix*).

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



from the graph above we notice that the sample means are normally distributed.

## #APPENDIX

Codes used

```
simMeans <- apply(simData, 1, mean) # calculate the sample rows means
sampleMean <- mean(simMeans) # calculating the overall sample Mean
theorMean <- 1/lambda # theoretical Mean
```

```
hist(simMeans, breaks = 40, main = "Sample Means Distributions", xlab= "Means",
     ylab = "Means Frequencies", col = "lightblue")
abline(v = theorMean, lty = 1, lwd = 3, col = "red")
abline(v = sampleMean, lty= 2, lwd = 3, col = "black")
legend("topright", legend = c( "Theoretical Mean", "Sample Mean"),lty = c(1,2) ,
     lwd = 2, col = c("red", "black"))
```

```
sampleVar <- var(simMeans)
theorVar <- (1/(lambda)^2/n)
```

```
sampleSD <- sd(simMeans)
theorSD <- 1/(lambda*sqrt(n))
```

```
data <- data.frame(simMeans)
m <- ggplot(data, aes(x=simMeans))+
  geom_histogram(aes(y = ..density..), colour = "blue", fill = "lightblue")+
  labs(title = "Distribution of Means", x = "Means", y = "Densities")+
  stat_function(fun = dnorm, args = list(mean = theorMean, sd = theorSD),
    color = "red", size = 1)+
  stat_function(fun = dnorm, args = list(mean = sampleMean, sd = sampleSD),
    color = "black", size = 1)+
  theme_classic()
```

DONE \*\*\*\*\*