Statistical Inference Course Project

Part 1: Simulation Exercise

In this project I 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. It is used lambda = 0.2 for all of the simulations. I will investigate the distribution of averages of 40 exponentials.

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

It is calculated the average of 40 exponentials on 1000 simulations

```
nsim <- 1000
lambda <- 0.2
n <- 40

# mi is the mean for the simulation i ( 1 <= i <= nsim)
mi = as.data.frame(matrix(nrow=1000, ncol=1))
for (i in 1:nsim) {
   dist = rexp(n,lambda)
   mi[i,1] <- mean(dist)
}

meand <- mean(mi[,1])
```

Sample mean

```
#The mean of mi
meand
```

```
## [1] 4.998295
```

Theoretical mean of the distribution

```
#The mean of exponential distribution is 1/lambda
1/lambda
```

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

As we can see, the distribution of the sample mean is centered aproximately at the distribution of the population mean. The sample mean is an estimate of the population mean

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

Variance of the sample

```
var(mi[,1])
```

```
## [1] 0.5761118
```

The theoretical variance of the sample mean is (1/lambda)^2/n as the standard deviation is 1/lambda:

```
tvar <- (1/lambda)^2/n; tvar
```

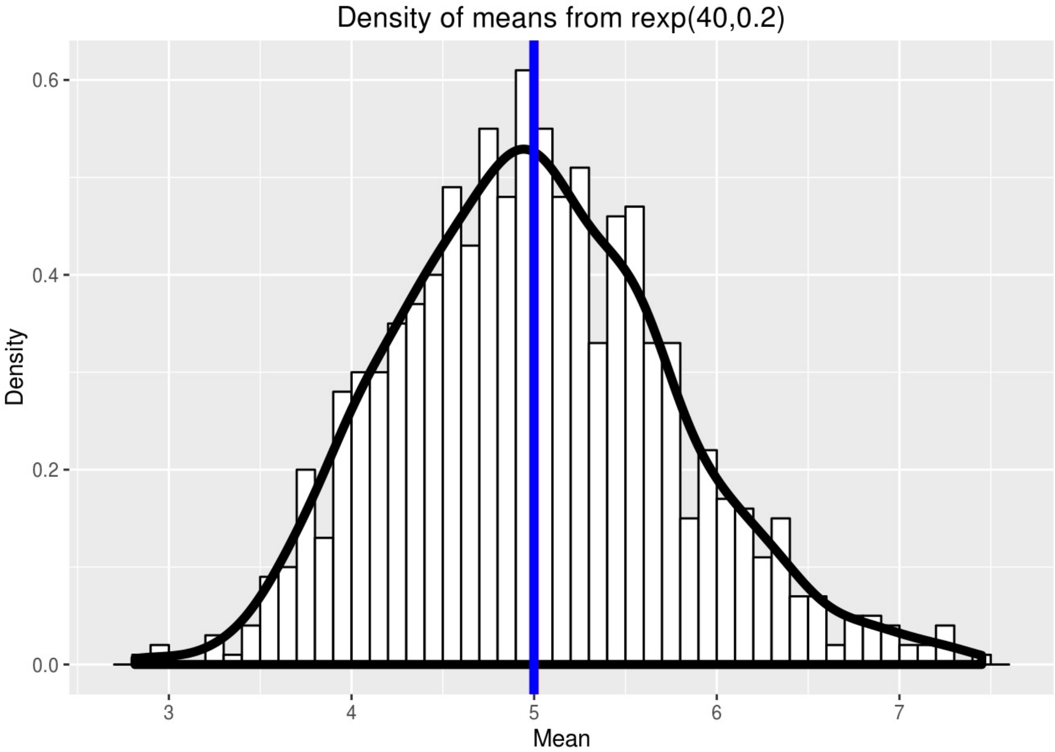
```
## [1] 0.625
```

Both variances are very close, then the distribution of the sample variance is centered around population variance.

3. Show that the distribution is approximately normal

```r
library(ggplot2)

g <- ggplot(mi, aes(x = mi$V1))
g <- g + geom_histogram(fill = "white", binwidth=0.1, aes(y = ..density..), color = "black")
g <- g + geom_density(size = 2)
g <- g + geom_vline(xintercept = 1/lambda, size = 2, color = 'red') # theoretical mean
g <- g + geom_vline(xintercept = meand, size = 2, color = 'blue')          # sample mean
g <- g + labs(title = paste('Density of means from rexp(40,0.2)'))
g <- g + labs(x = "Mean", y = "Density")
g
```



In the figure above, it’s clear that the distribution is approximately normal, since it follows a Gaussian distribution. The red line is the theoretical mean, and the blue one, the sample mean