The Exponential Distribution and the Central Limit Theorem

Summary

The objective of this project is investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT). The exponential distribution can be simulated in R with the function rexp(n, lambda), where lambda is the rate parameter. The mean of the exponential distribution is 1/lambda and the standard deviation is also 1/lambda. To investigate the exponential distribution, it will be used n = 40 and lambda = 0.2, with 1000 simulations. To achieve the objective of this project, it will be follow the next steps:

_ Show the sample mean and compare it to the theoretical mean of the distribution

_ Show the sample variance and compare it to the theoretical variance of the distribution.

_ Show that the distribution is approximately normal.

Data Processing

```
# Establishing the variables
l <- 0.2
n <- 40
s <- 1000
# Establishing a seed number to make reproducible the results
set.seed(29)
```

Using the matrix() and rexp() functions it was created a matrix that contains the one thousand simulations

```
simulated_matrix <- matrix(rexp(n*s, l), s, n)
```

Results

Sample Mean Comparison

Using the apply() function, it was calculated the mean of each exponential distribution of 40 numbers

```
simulated_mean <- apply(simulated_matrix, 1, mean)
```

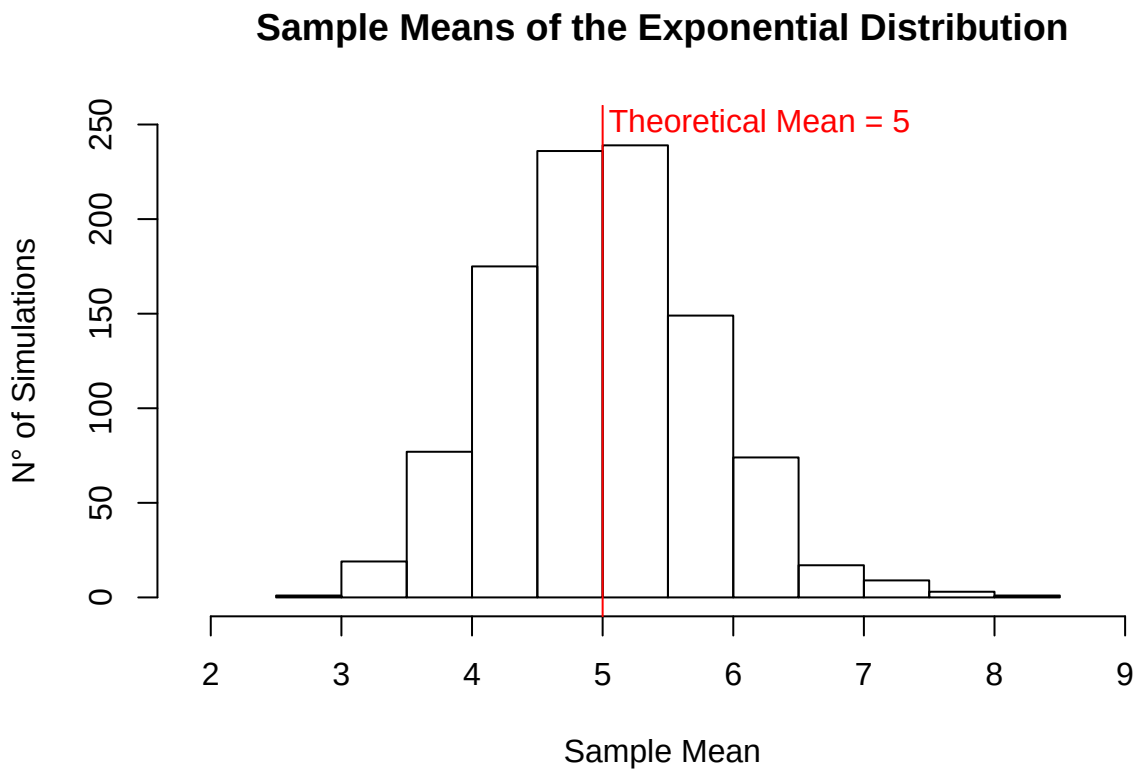
Then it was calculated the sample and theoretical mean

```
# Calculating the sample and theoretical mean
sample_mean <- mean(simulated_mean)
theory_mean <- 1/l
print(c("Sample mean" = sample_mean, "Theoretical mean"  = theory_mean))
```

```
##      Sample mean Theoretical mean
##         4.996344         5.000000
```

Finally it was made a histogram to show graphycally the distribution of the sample mean and its relation with the theoretical mean

```r
# Ploting the simulated mean with the sample mean and theoretical mean included
hist(simulated_mean, main = "Sample Means of the Exponential Distribution",
     xlab = "Sample Mean", ylab = "N° of Simulations",
     xlim = c(min(simulated_mean) - 1, max(simulated_mean) + 1), ylim = c(0,250))
abline(v = theory_mean, col = "red", lwd = 1)
text(6.2, 252, "Theoretical Mean = 5", col = "red")
```



The histogram show that the one thousand simulations have a mean very close to the theoretical mean

Sample Variance Comparison

It was calculated the variance of sample means and the theoretical variance of the exponential distribution

```r
# Showing the sample and theoretical variance
sample_var <- var(simulated_mean)
theory_var <- (1/0.2)^2/40
print(c("Sample variance" = sample_var, "Theoretical variance"  = theory_var))
```

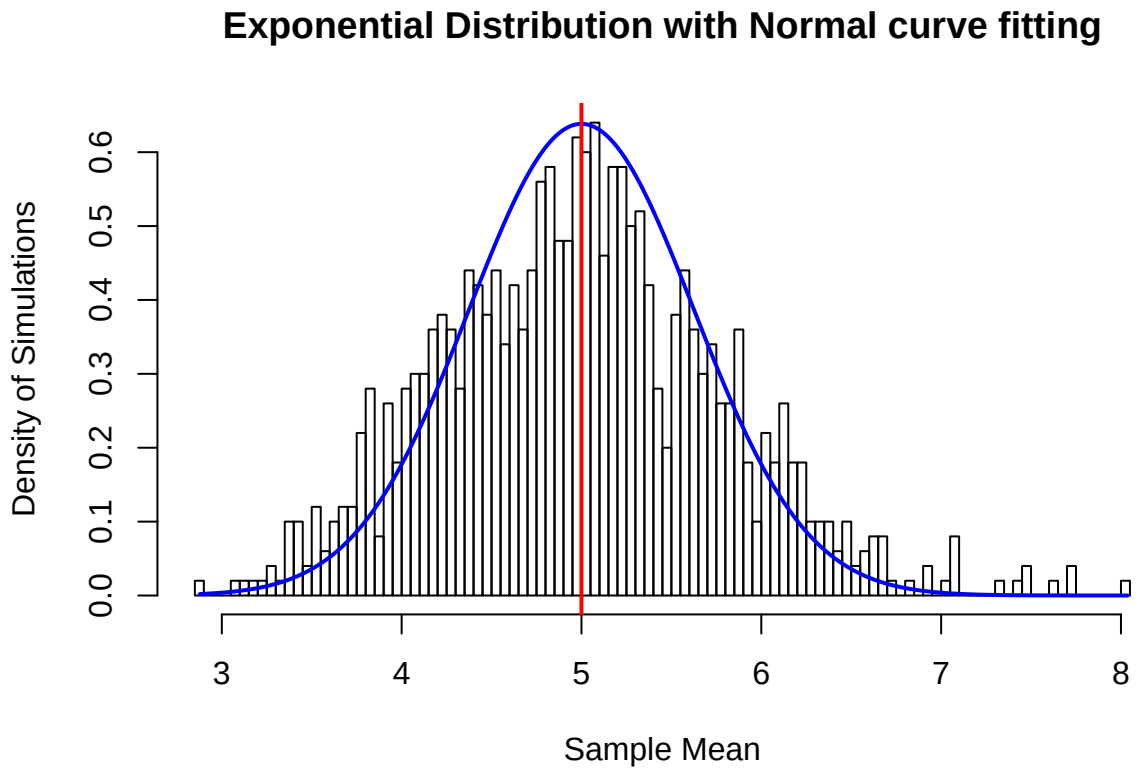
```
##      Sample variance Theoretical variance
##            0.6104458            0.6250000
```

The results show that the sample and theoretical variances are very similar

Distribution

Finally it was made a histogram to show graphycally that the sample means follow a normal distribution, as is expected by the CLT.

```r
# Creating a normal distribution to see how similar are with the simulated matrix
hist(simulated_mean, main = "Exponential Distribution with Normal curve fitting",
     xlab = "Sample Mean", ylab = "Density of Simulations", prob = TRUE, breaks = 100)
xfit <- seq(min(simulated_mean), max(simulated_mean), length = 1000)
yfit <- dnorm(xfit, mean = 1/0.2, sd = (1/0.2)^2/40)
lines(xfit, yfit, col = "blue", lwd = 2)
abline(v = theory_mean, col = "red", lwd = 2)
```



Then, it was calculated the 95% confidence intervales of the sample and theoretical distribution showing that they are very close

_ Confidence Interval of the Sample Distribution

```r
print(round(sample_mean + c(-1,1)*1.96*sd(simulated_mean)/sqrt(n),3))
```

```
## [1] 4.754 5.238
```

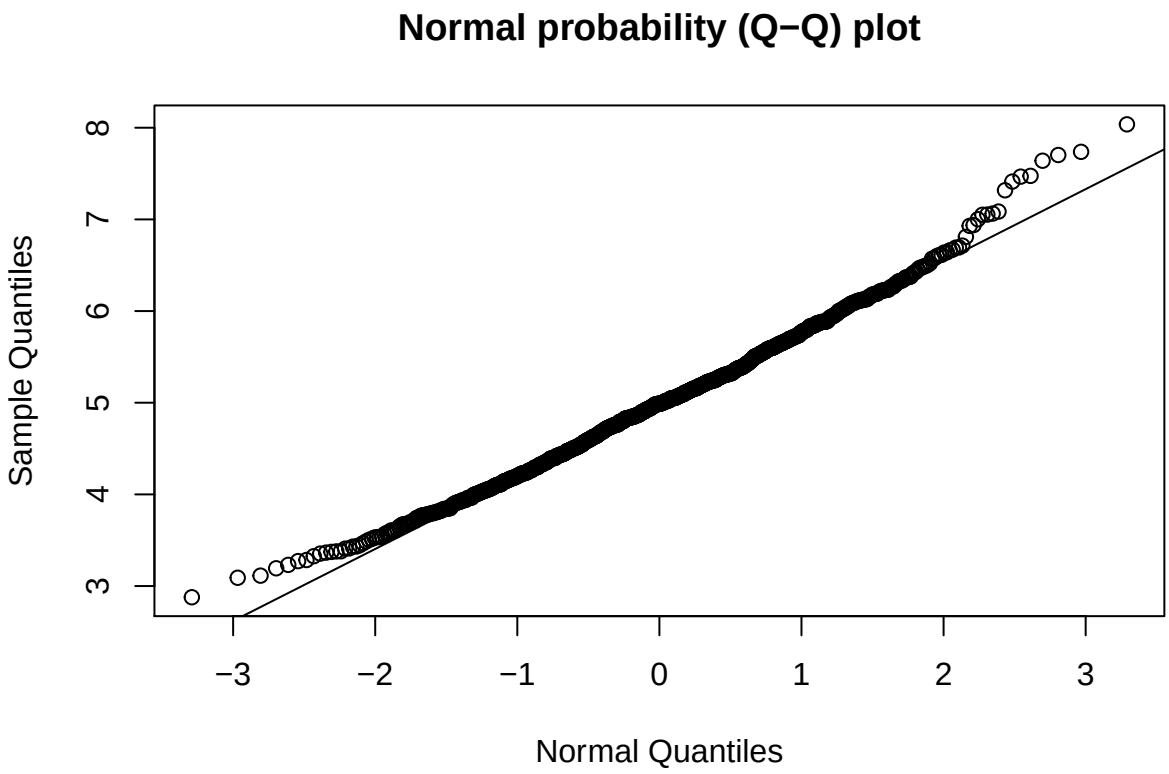
_ Confidence Interval of the Theoretical Distribution

```r
print(theory_mean + c(-1,1)*1.96*sqrt(theory_var)/sqrt(n))
```

```
## [1] 4.755 5.245
```

Finally, it was made a QQplot to compare the sample and normal distribution. The results shows that both distributions are very similar

```r
# Drawing the QQplot
qqnorm(simulated_mean, main ="Normal probability (Q-Q) plot", xlab = "Normal Quantiles")
qqline(simulated_mean)
```



Conclusion

_ The sample means of the exponential distribution are distributed in a normal distribution, corroborating the CLT