Simulation Exercise - Peer Grading

Nikhil Chandra

October 2, 2018

Overview

In this project we 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. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

Simulations

Sample Mean vs Theoretical Mean

We are going to run 1000 simulations in order to create a data set. Each simulation is going to contain 40 observations and the exponential distribution function will be set to rexp(40, 0.2).

values: lambda = 0.2, n = 40, simulations = 1000

```
          # Load required library
          library(ggplot2)

          # Set the below constants
          lambda = 0.2
          n = 40
          nosim = 1000

          set.seed(756)
```

The below code does the simulations and then plots the data:

```
            exp_sim <- function(n, lambda)
            {
                      mean(rexp(n,lambda))
            }

            sim <- data.frame(ncol=2,nrow=1000)
            names(sim) <- c("Index","Mean")

            for (i in 1:nosim)
            {
                      sim[i,1] <- i
                      sim[i,2] <- exp_sim(n,lambda)
            }
```

Mean for n = 1000

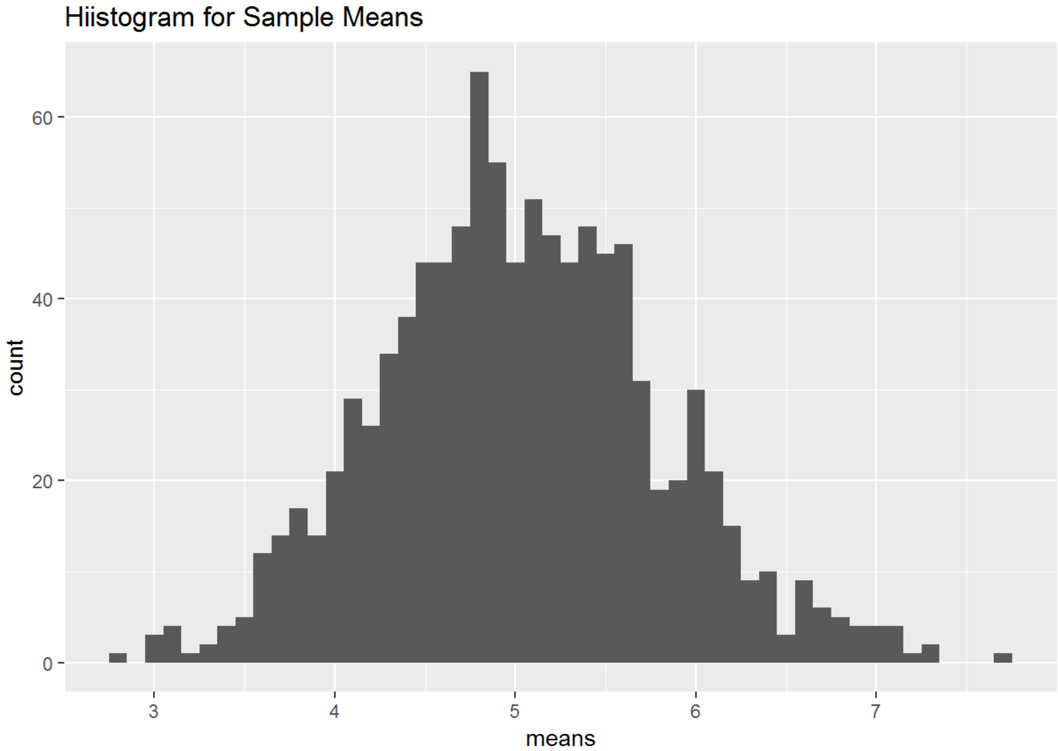
```
        sample_mean <- mean(sim$Mean)
        sample_mean
```

```
## [1] 4.972894
```

Histogram for - sample means

```
exp_dis <- matrix(data=rexp(n * nosim, lambda), nrow=nosim)
exp_dis_means <- data.frame(means=apply(exp_dis, 1, mean))


ggplot(data = exp_dis_means, aes(x = means)) +
        geom_histogram(binwidth=0.1) +
        ggtitle("Hiistogram for Sample Means") +
        scale_x_continuous(breaks=round(seq(min(exp_dis_means$means), max(exp_dis_means$means), by=1)))
```



Sample Mean versus Theoretical Mean

The expected mean \(\mu\) of a exponential distribution of rate \(\lambda\) is

\(\mu= \frac{1}{\lambda}\)

```
theor_mean <- 1/lambda
theor_mean
```

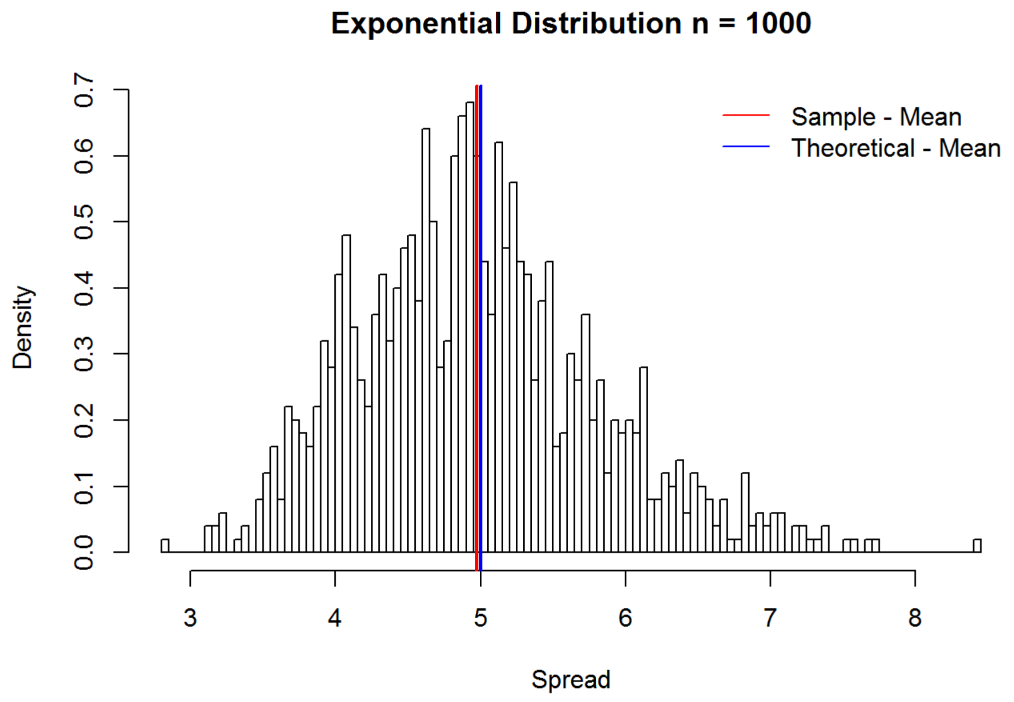
```
## [1] 5
```

The calculated simulation mean of 4.972894 is very near to the theoretical value of 5. We can also say that the the average sample mean of 4.97 and expected theoretical mean of 5 are close enough.

Histogram plot - exponential distribution n = 1000

```
hist(sim$Mean,
        breaks=100,
        prob=TRUE,
        main="Exponential Distribution n = 1000",
        xlab="Spread")
                abline(v = theor_mean,
                        col= 4,
                        lwd = 2)
                abline(v = sample_mean,
                        col = 2,
                        lwd = 2)

        legend('topright', c("Sample - Mean", "Theoretical - Mean"),
                        lty=c(1,1),
                        bty = "n",
                        col = c(col=2, col=4))
```



Sample Variance vs. Theoretical Variance

Now, we need to compare the variance of the sammple means of 1000 simulations and the theoritical variance.

The standard deviation σ of the exponential distribution of rate λ is

$\sigma = \frac{1/\lambda}{\sqrt{n}}$

The variance Var of standard deviation σ is

$Var = \sigma^2$

This gives the following: σ^2=Var(samplemeans)×N.

```
            sample_var <- var(sim$Mean)
            theor_var <- ((1/lambda)^2)/n
```

The theoretical variance of the population is given by σ^2=(1/lambda)2.

```
            sample_var
```

```
## [1] 0.6912115
```

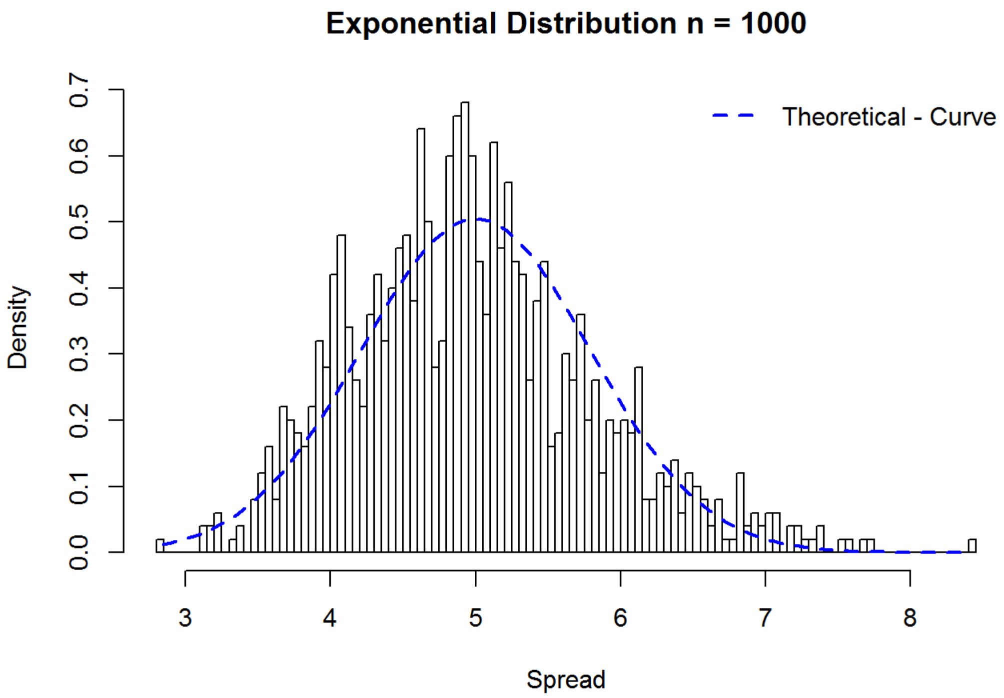
```
            theor_var
```

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

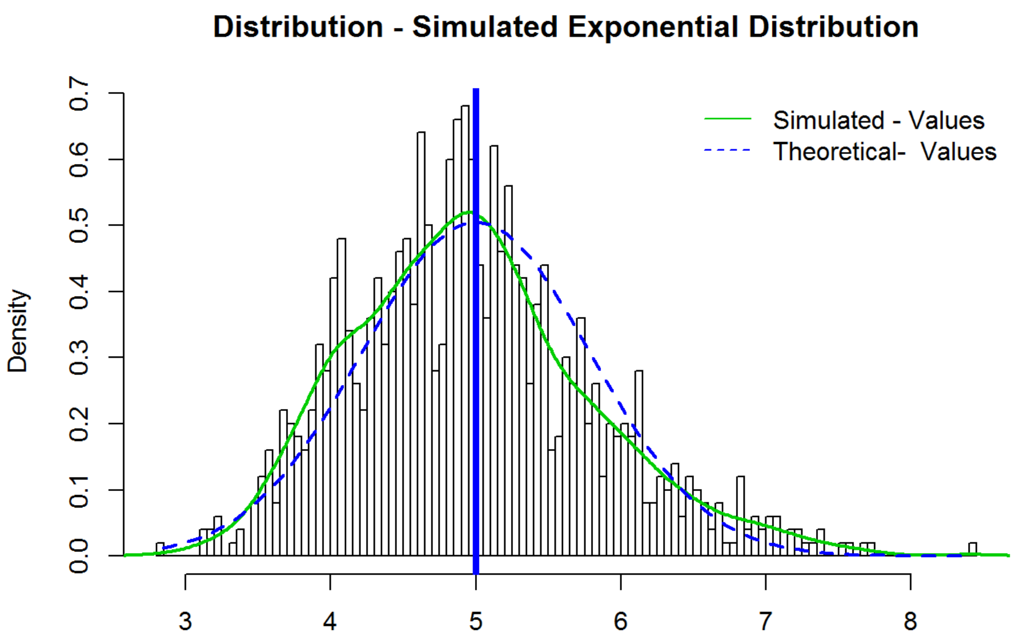
With this we can conclude that the variances are very close.

Sample Distribution vs. Theoretical Distribution

```
        hist(sim$Mean,
                    breaks = 100,
                    prob = TRUE,
                    main ="Exponential Distribution n = 1000",
                    xlab ="Spread")
                    xfit <- seq(min(sim$Mean), max(sim$Mean), length = 100)
                    yfit <- dnorm(xfit, mean = 1/lambda, sd = (1/lambda/sqrt(40)))
                    lines(xfit, yfit, pch=22, col=4, lty=2, lwd=2)
                    legend('topright', c("Theoretical - Curve"), bty="n", lty=2,lwd=2, col=4)
```




```
hist(sim$Mean,
              breaks = 100,
              prob = TRUE,
              main = "Distribution - Simulated Exponential Distribution", xlab="")
              lines(density(sim$Mean), col=3, lwd=2)
              abline(v = 1/lambda, col = 4, lwd=4)
              xfit <- seq(min(sim$Mean), max(sim$Mean), length = 100)
              yfit <- dnorm(xfit, mean = 1/lambda, sd = (1/lambda/sqrt(40)))
              lines(xfit, yfit, pch=22, col=4, lty=2, lwd=2)
              legend('topright', c("Simulated - Values", "Theoretical-  Values"),
              bty="n", lty=c(1,2), col=c(3,4))
```



We can say that with the given lambda, the mean of exponential distribution overlaps with normal distribution.

Conclusion: We can conclude that Distribution is more or less

normal

The values of variances and sample mean are close to the expexted theoritical mean. This is suggestive of ‘Normality’. Also the below plot displays the that match is closer of theoritical quantiales with actual quantiales. Colelctively all of the above prove that the Distribution is Normal.

```
            qqnorm(sim$Mean,
                            main ="Normal Q-Q Plot")
                            qqline(sim$Mean,
                            col = "4")
```