Statistical Inference Course Project: Part 1

lcombs

02/05/2016

Simulating an Exponential Distrubution:

Overview

In this project, I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The results will illustrate via simulation the properties of the distribution of the mean of the exponentials.

I will: 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.

```
library(magrittr)
```

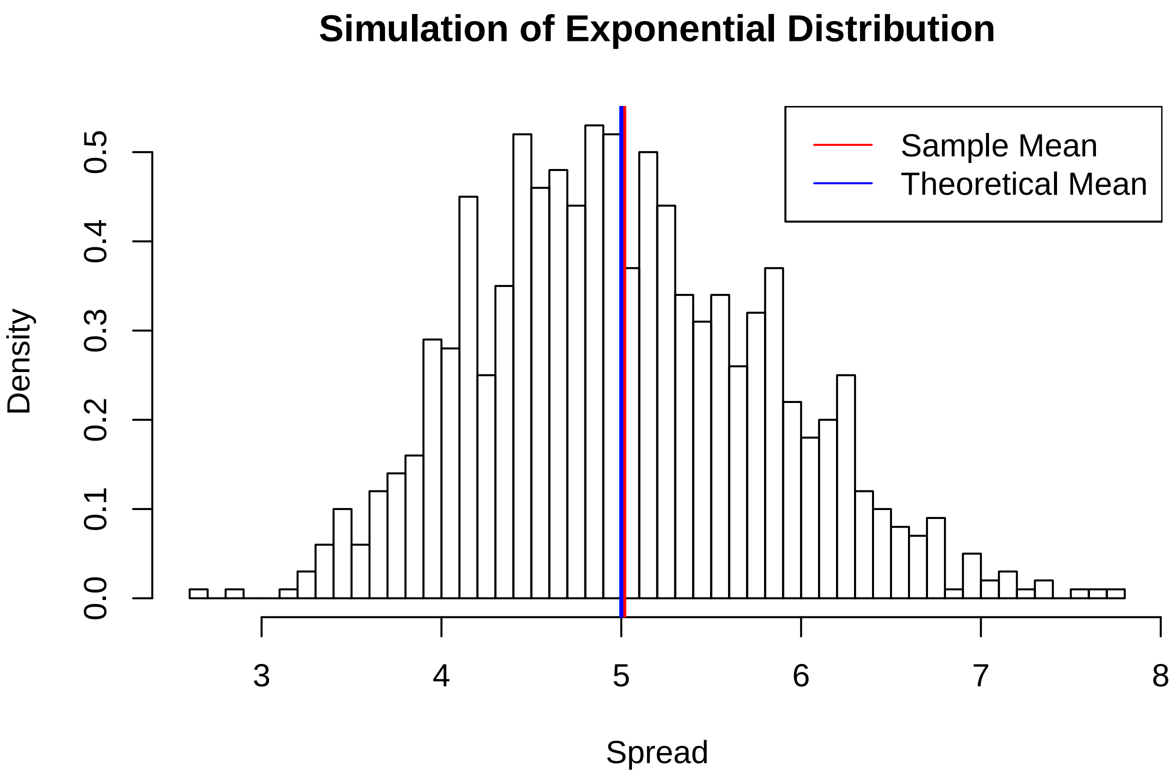
Simulations

In order to carry out the simulations, we set lambda equal to .02. We will carry out 1,000 simulations by collecting the mean of 40 exponentials each time. The resulting means are saved to the variable mns.

```
lambda <-
        .2

n <-
        40

number_simulations <-
        1000

# set random seed for reproducibility
set.seed(2)

mns <-
        NULL

for(i in 1:number_simulations){
        mns <- c(mns, rexp(n, lambda) %>%
                         mean())
}
```

Sample Mean versus Theoretical Mean

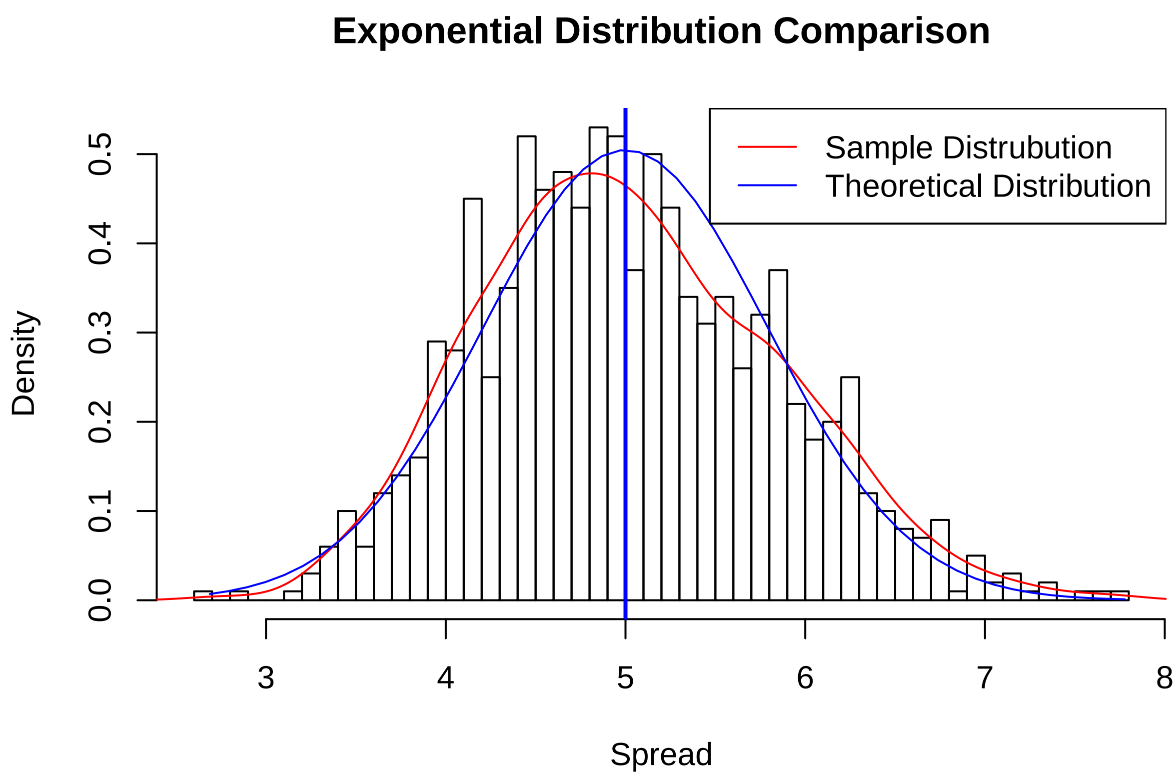
```r
sample_mean <-
        mns %>%
        mean()

theoretical_mean <-
        1/lambda

hist(mns,
     breaks = 50,
     prob = TRUE,
     main = "Simulation of Exponential Distribution",
     xlab = "Spread")
abline(v = sample_mean,
       col = "red",
       lwd = 2)
abline(v = theoretical_mean,
       col = "blue",
       lwd = 2)
legend("topright",
       c("Sample Mean", "Theoretical Mean"),
       lty = c(1, 1),
       col = c("red", "blue"))
```



As the plot shows, the sample mean (5.016) and theoretical mean (5) are close to one another. We can already start to see that the plot looks approximately normal.

Sample Variance versus Theoretical Variance

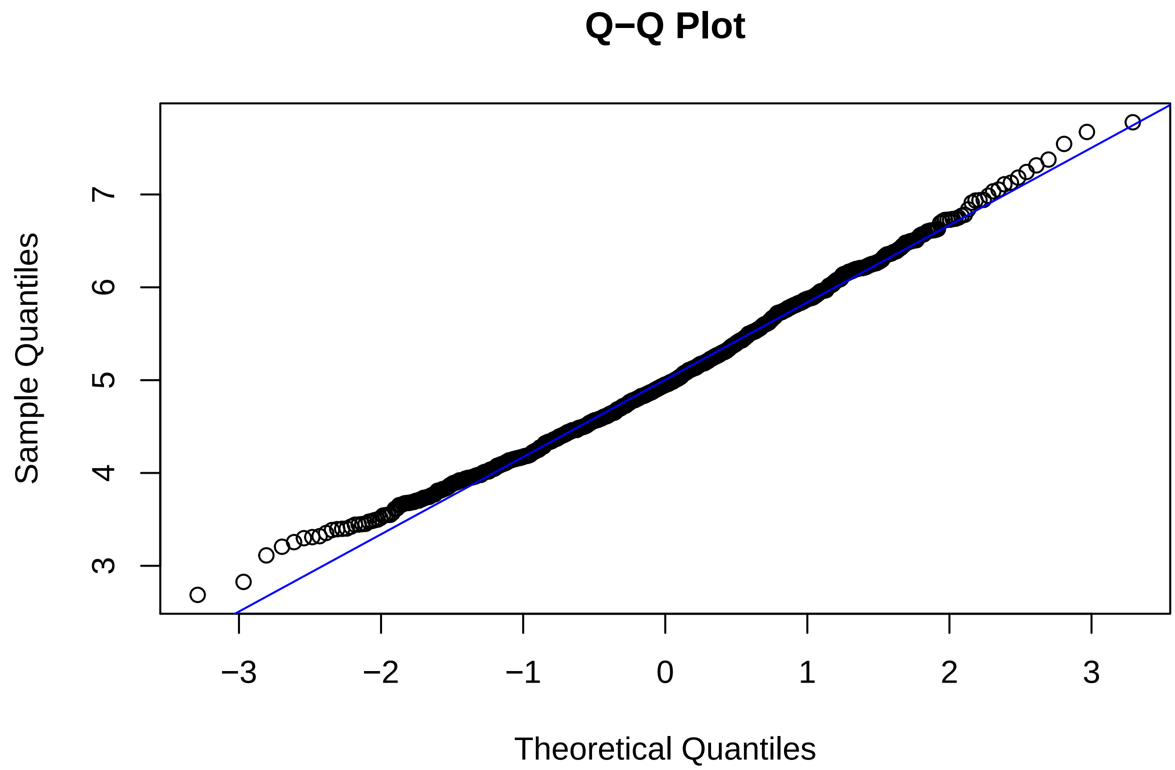
```
sample_variance <-
        var(mns)

theoretical_variance <-
        ((1/lambda) ^2) / 40

x_values <-
        seq(min(mns),
            max(mns),
            length = 50)
y_values <-
        dnorm(x_values,
              mean = 1/lambda,
              sd = 1/lambda/sqrt(40))

hist(mns,
     breaks = 50,
     prob = TRUE,
     main = "Exponential Distribution Comparison",
     xlab = "Spread")

lines(mns  %>% density(), col = "red")
lines(x_values, y_values, col = "blue")

abline(v = theoretical_mean,
       col = "blue",
       lwd = 2)
legend("topright",
       c("Sample Distrubution", "Theoretical Distribution"),
       lty = c(1, 1),
       col = c("red", "blue"))
```



As the plot shows, the sample distribution and theoretical distribution possess simular shapes. The density functions make it easier to compare the distributions holistically. The variance of the sample is 0.669; while the theoretical variance is 0.625.

Distribution is approximately normal.

```
qqnorm(mns,
       main = "Q-Q Plot")
qqline(mns,
       col = "blue")
```



The q-q plot above shows the difference between the sample distribution and the normal distribution. Here, we can see that the theoretical and observed quartiles are very similar. The sample difference most at the ends where in both cases the points are larger than what we would expect from a perfectly normal distribution.