Statistical Inference: Project - Part 1

Piotr Brdeja

Sunday, January 25, 2015

Overview

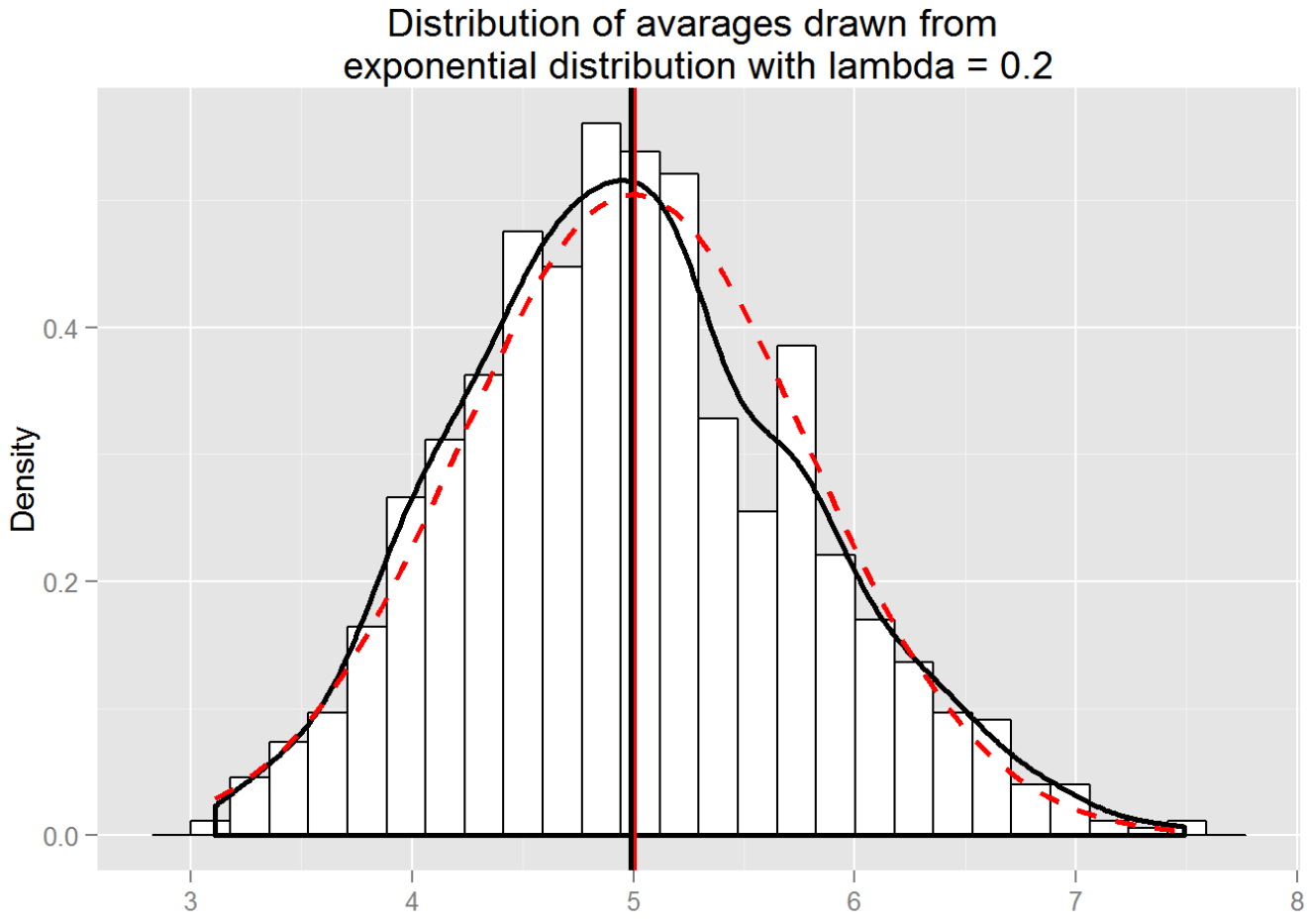
In this project we had to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. We have set lambda = 0.2 for all of the simulations. We investigated the distribution of averages of 40 exponentials.

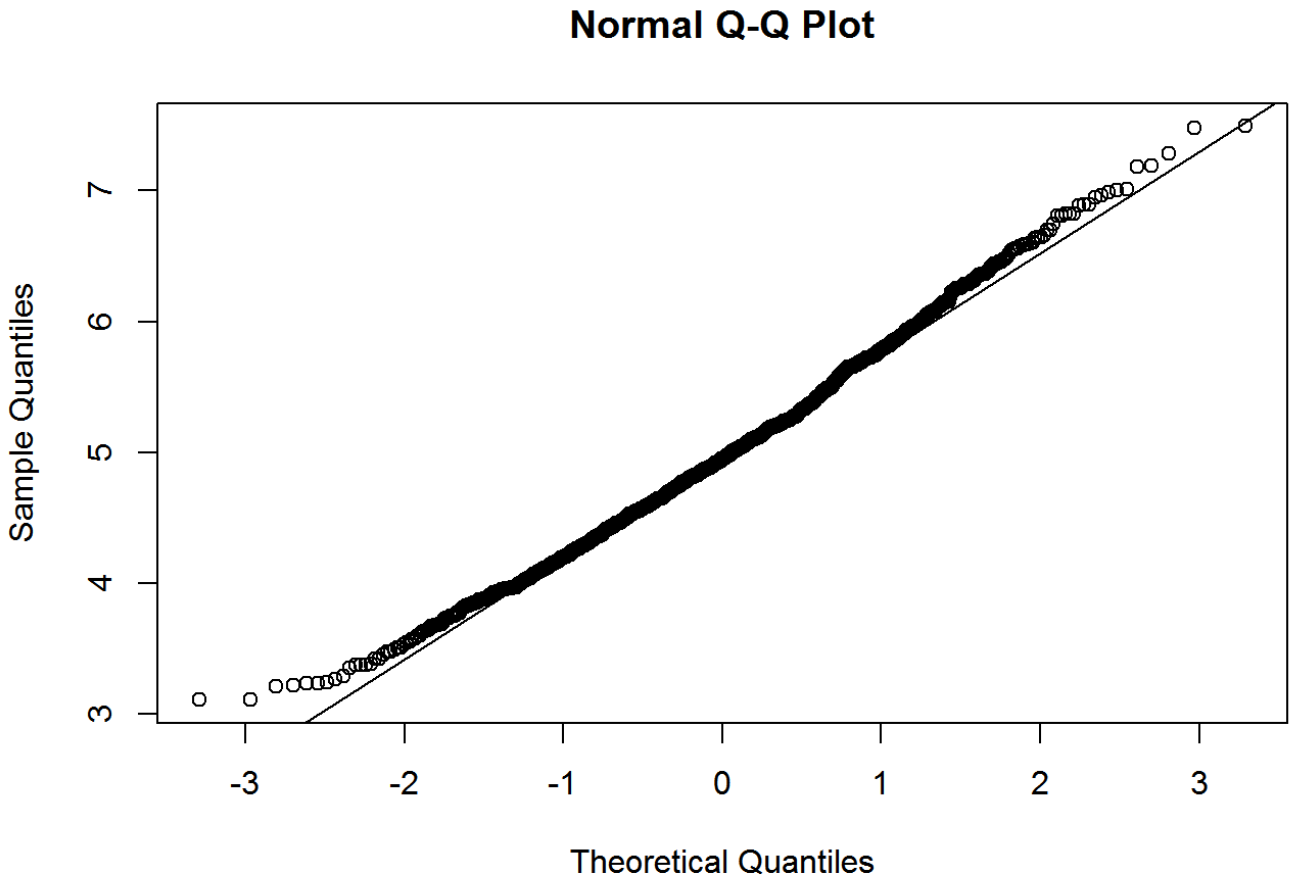
Simulations

We had to simulate 1000 averages of 40 observations drawn from the exponential distribution. The code is as follows:

```
set.seed(1)
lambda <- 0.2
numSim <- 1000
sampleSize <- 40

#Simulating data
data = NULL
for (i in 1 : 1000){
    data = c(data, mean(rexp(40, rate = lambda)))
}
```





Sample Mean versus Theoretical Mean

The theoretical sample mean is centered at lambda^-1 = **5** and the distribution of sample means is centered at mean **4.9900**.

Sample Variance versus Theoretical Variance

The variance of sample means is **0.6111** where the theoretical variance of the distribution is 2/n = 1/(2n) = 1/(0.04 × 40) = **0.625**.

Distribution

This chart from above suggests the normality of distribution according to the Central Limit Theorem. Additionally we can run **Shapiro-Wilk** test, which checks whether or not the distribution is normal, but it **was not in the lectures**. This test shows us that our data is not normally distributed with alfa = 0.05 and p-value = 0.00025.

```
shapiro.test(data$avarage)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$avarage
## W = 0.9935, p-value = 0.0002466
```