

statinf_cp1_joannanw

joannanw

Sunday, March 22, 2015

This project investigates the exponential distribution in R and compare it with the Central Limit Theorem. The distribution is made up of the average of 40 exponentials ran 1000 times.

Simulations

Lambda is set to 0.2 for this simulation. The theoretical mean of distribution is first calculated. Following that, a sample simulation of the average of 40 exponentials is ran 1000 times.

```
lambda <- 0.2  
theoreticalMean <- 1/lambda
```

The theoretical Mean is 5.

```
set.seed(1)  
sampleSize <- 40  
sampleRun <- 1000  
data <- NULL  
for (i in 1 : sampleRun) data <- c(data, mean(rexp(sampleSize,lambda)))  
dataMean <- mean(data)
```

The sample mean is 4.9900252.

The theoretical and sample means are very close to each other. This shows that for the large simulation of a distribution, the average centers to its theoretical mean.

```
theoreticalVar <- (1/lambda)^2/sampleSize  
dataVar <- var(data)
```

The theoretical variance is 0.625 and the sample variance is 0.6111165. The variability of the sample is close to the theoretical value.

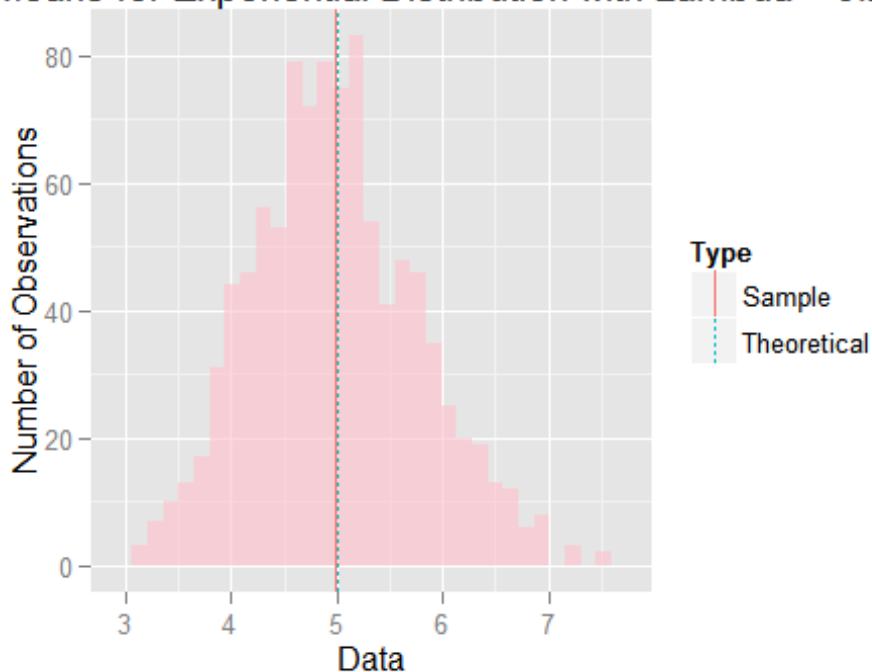
Results

Now let's show the distribution of the sample.

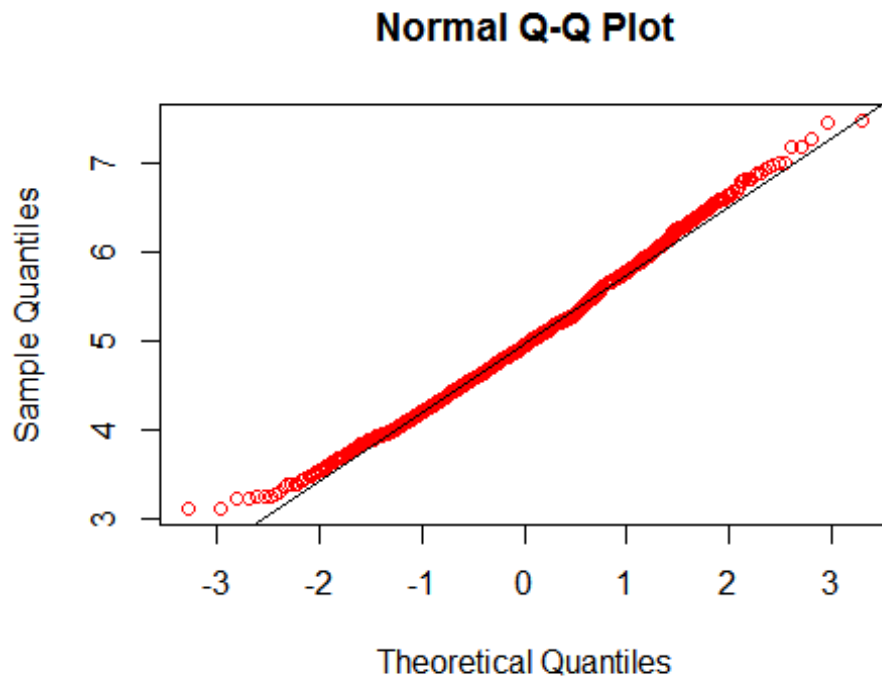
```
library(ggplot2)
sampleData <- data.frame(Type = "Sample", Data = data, stringsAsFactors =
"FALSE")
bothMeans <- data.frame(Type = c("Sample", "Theoretical"), Mean = c(dataMean,
theoreticalMean))
ggplot(sampleData, aes(x = Data)) + geom_histogram(alpha = 0.6, fill =
"pink") + ggtitle("Sample Means for Exponential Distribution with Lambda =
0.2") + scale_y_continuous("Number of Observations") + geom_vline(data =
bothMeans, aes(xintercept = Mean, color = Type, linetype = Type), show_guide
= TRUE)

## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust
this.
```

Means for Exponential Distribution with Lambda = 0.2



```
qqnorm(data, col = "red")
qqline(data, col = 1)
```



The distribution of the sample follows the Central Limit Theorem (CLT) where the arithmetic average and variance of the sample are well-defined and converges to a normal distribution. This is further shown by the Q-Q plot that the sample data lies very closely to the normal distribution.