Statistical Inference Course Project - Part 1

Li Xin
23 December, 2015

Overview

This is the first part of project for the statistical inference class. In this part, a simulation execise will be conducted to explore inference.

In this project the exponential distribution will be investigated in R and it will be compared 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. Set lambda = 0.2 for all of the simulations. The distribution of averages of 40 exponentials will be investigated. Note that there will be a thousand simulations.

Simulations

Below codes will perform the simulations of data.

```
data <- 1000
lambda <- 0.2
count <- 40

#set seed to allow reproductivity
set.seed(384252)

data_mean = NULL
for (i in 1 : 1000) data_mean = c(data_mean,mean(rexp(40,0.2)))
```

Sample Mean versus Theoretical Mean

Comparison of sample mean data_mean with theoretical mean 1/lambda.

```
mean_sample <- mean(data_mean)
mean_sample
```

```
## [1] 4.97905
```

```
mean_th <- 1/lambda
mean_th
```

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

From the above observation, we can find the sample mean **4.9790501** is close to the theoretical mean **5**.

Sample Variance versus Theoretical variance

Comparison of sample variance and theoretical variance

```r
var_sample <- var(data_mean)
var_sample
```

```
## [1] 0.5319693
```

```r
var_th <- (1/lambda)^2/count
var_th
```

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

Again, we can see the sample variance **0.5319693** is close to the theoretical variance **0.625**.

Distribution

Let's check if the distribution can resemble the normal distribution.

```r
hist(data_mean,breaks=60,freq=FALSE,main="Distribution of 40 samples (lambda = 0.2)")
lines(density(data_mean),col="black")
abline(v=1/lambda,col="red",lwd=2)
curve(dnorm(x,mean_th,var_th),col="blue",add=TRUE)
legend('topright',c("simulation","theoretical"),lty=c(1,1),col=c("black","blue"))
```


Distribution of 40 samples (lambda = 0.2)
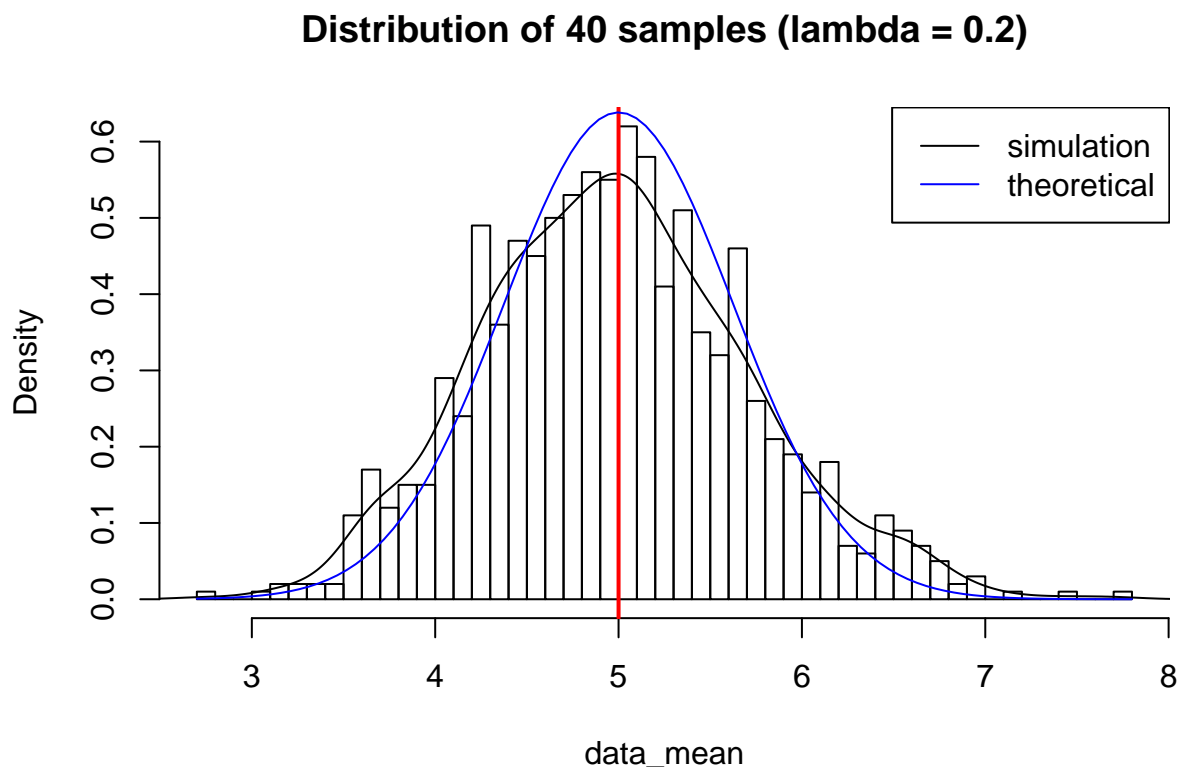
From the above histogram we can see the theoretical distribution overlapped with the histrogram. The distribution of sample data appears to approximate the normal distribution.

Also, we can find the below Q-Q plot suggests the normality.

```
qqnorm(data_mean); qqline(data_mean)
```

Normal Q–Q Plot