Statistical Inference Course Project Part1 - Simulation Exercise

Toshiyuki Tachibana

2016-12-31

Overview

In this project you 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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

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.

SIMULATION

Setting wd and load necessary packages

```r
setwd("C:/Users/Tachibana/Documents/GitHub/Stastical Infrence")

require(knitr)
```

```
## Loading required package: knitr
```

```r
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```r
require(markdown)
```

```
## Loading required package: markdown
```

Compare the distribution of 1000 random uniforms and the distribution of 1000 averages of 40 random uniforms

Set Speed and variables that control the simulation

```r
set.seed=(1)

no_sim <- 1000
lambda <- 0.2
n <- 40
```

Create a matrix of 1000 rows with the columns corresponding to random simulation 40 times

```r
sim_matrix <- matrix(rexp(no_sim * n, rate=lambda), no_sim, n)
sim_mean <- rowMeans(sim_matrix)
```

Sample Mean versus Theoretical Mean

Calculation of the Sample Mean and Theoretical Mean

```r
sample_mean <- mean(sim_mean)
theory_mean <- 1/lambda
```

```r
sample_mean
```

```
## [1] 4.997276
```

```r
theory_mean
```

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

As the answer to the question, the sample mean is 5.019 while the teoretical mean is 5.

Calculation of the Sample Variance and Theoretical Variance

```r
sample_sd<-sd(sim_mean)
sample_var <- var(sim_mean)
theory_sd<-1/lambda*(1/sqrt(n))
theory_var <- theory_sd^2
```

```r
sample_sd
```

```
## [1] 0.7888932
```

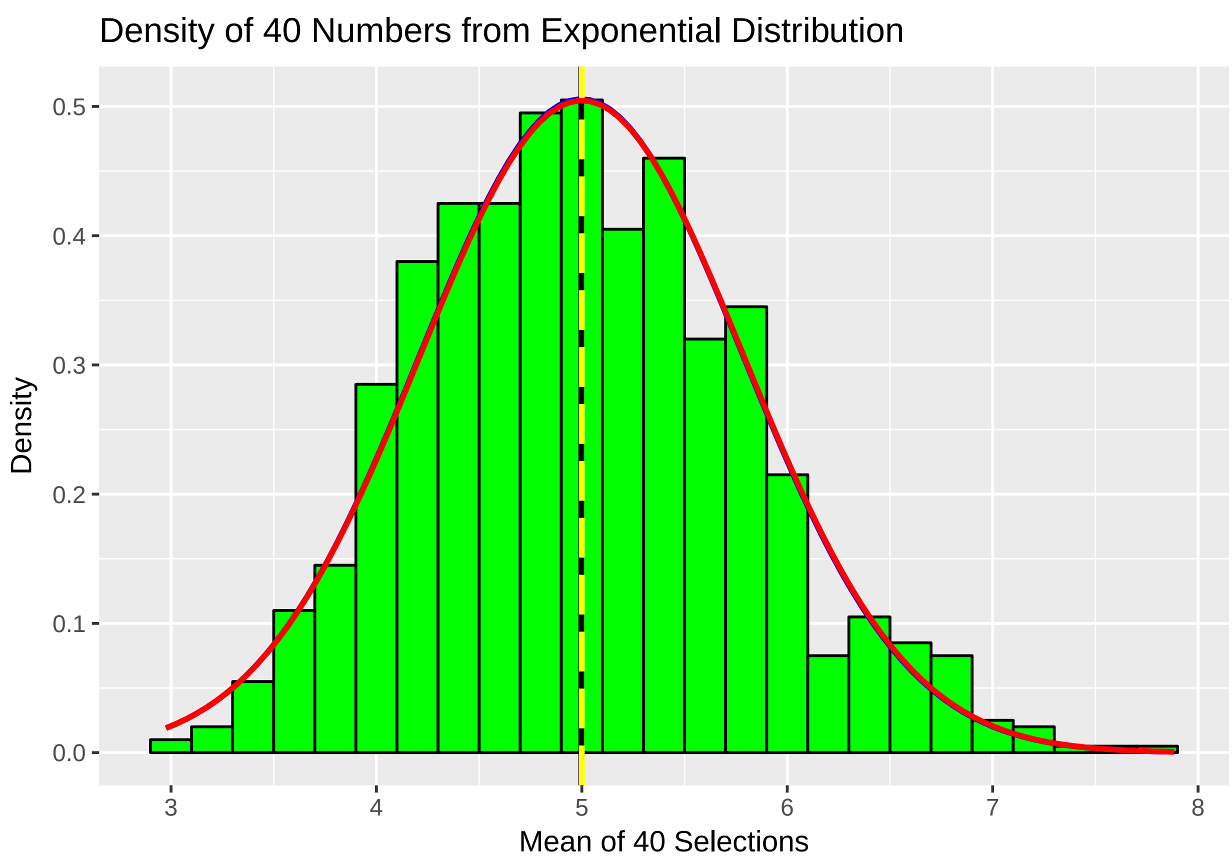
```
theory_sd
```

```
## [1] 0.7905694
```

Show that the distribution is approximately normal

```
plotdata <- data.frame(sim_mean)

g<-ggplot(plotdata, aes(x =sim_mean))
g<-g+geom_histogram(binwidth = lambda,fill="green",color="black",aes(y = ..density..))
g<-g+labs(title="Density of 40 Numbers from Exponential Distribution", x="Mean of 40 Selections", y="De
g<-g+geom_vline(xintercept=sample_mean,size=1.0, color="black")
g<-g+stat_function(fun=dnorm,args=list(mean=sample_mean, sd=sample_sd),color = "blue", size = 1.0)
g<-g+geom_vline(xintercept=theory_mean,size=1.0,color="yellow",linetype = "longdash")
g<-g+stat_function(fun=dnorm,args=list(mean=theory_mean, sd=theory_sd),color = "red", size = 1.0)
g
```



RESULT

The theoretical mean is shown in the graph as a dotted yellow line, while the actual mean is shown by the black line. The normal curve formed by the the theoretical mean and standard deviation is shown in red.

The curve formed by the actual mean and standard deviation is shown in blue. Finally, the density of the actual data is shown by the green bars. To answer question 3, we can observe the that the actual data follow a normal curve by observing the shape of the actual data on the graph shown in blue as compared to the normal curve shown in red.