Statistical Inference Project Part 1

Azmi Mohamed Ridwan

March 27, 2016

Overview

In this project, the exponential distribution is investigated and compared to 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. lambda will be set to 0.2 for all of the simulations. The target of study is the distribution of averages of 40 exponentials which will be simulated a thousand times.

The knitr to pdf is not working for R version installed on my machine. I have knitted to Word first before creating a PDF which resulted in some a spaces in the final document. The result is approximately 5 pages which is within the 3+3 pages allowed in the instructions

Simulations

Include English explanations of the simulations you ran, with the accompanying R code. Your explanations should make clear what the R code accomplishes.

The first code snippet sets up the necessary basic variables for the simultation.

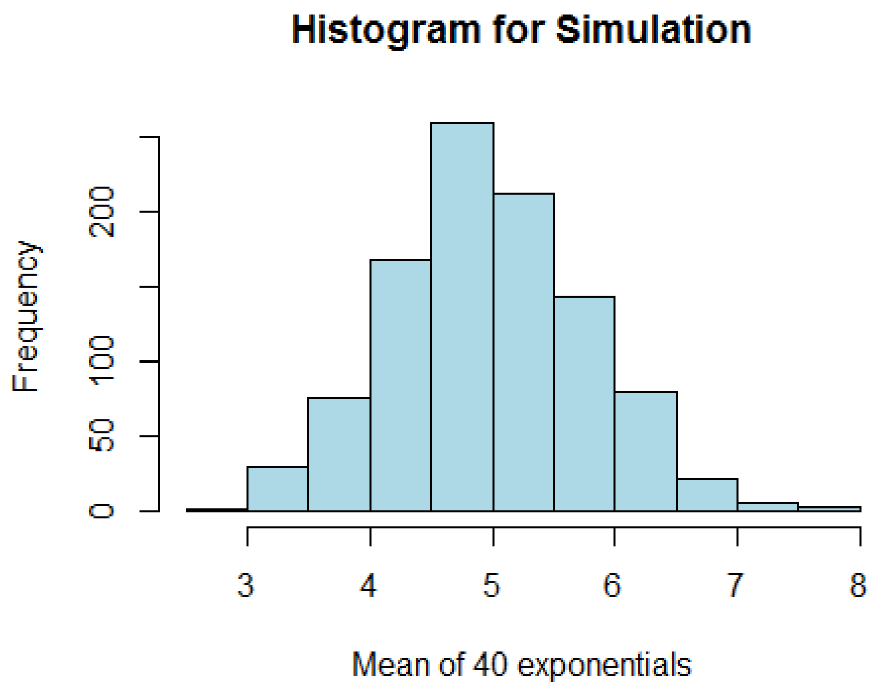
```
set.seed(2016)
lambda <- 0.2
n <- 40
num.sim <- 1000
```

Using the exponential distribution function *rexp*, a matrix is created with 1000 rows with the columns containing the 40 exponentials.

```
sim.matrix <- matrix(rexp(num.sim * n, rate = lambda), num.sim, n)
```

The mean for each row is calculated and a histogram is plotted.

```
sim.mean <- rowMeans(sim.matrix)
hist(sim.mean, xlab="Mean of 40 exponentials", ylab= "Frequency", col="light
blue", main="Histogram for Simulation")
```



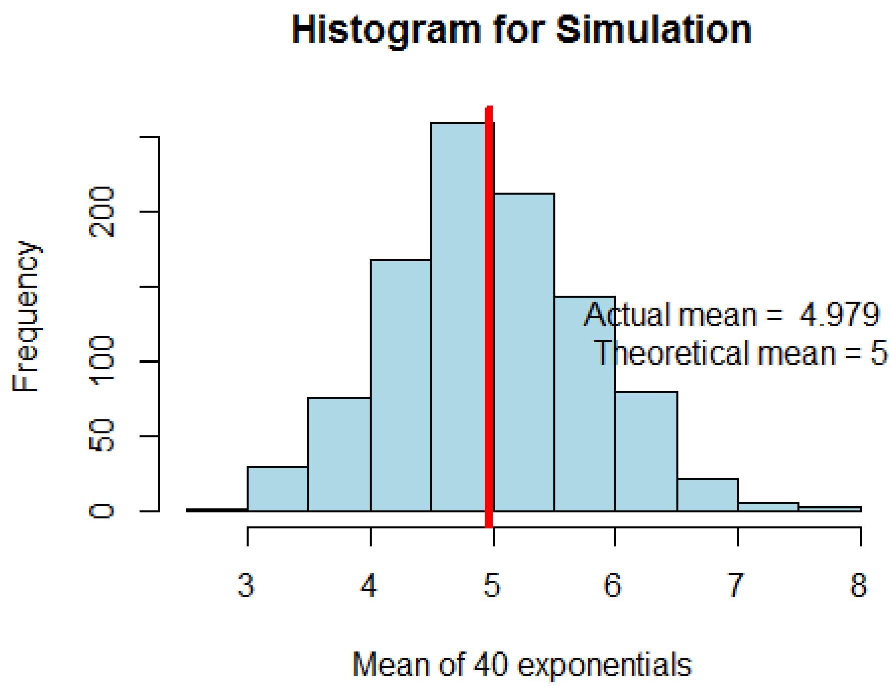
Sample Mean versus Theoretical Mean *Include figures with titles. In the figures, highlight the means you are comparing. Include text that explains the figures and what is shown on them, and provides appropriate numbers.*

The mean of the simulation and the theoretical mean are calculated here. The simulation mean values is then added to the histogram

```
actual.mean = mean(sim.mean)
theoretical.mean = 1/lambda
print(c(actual.mean, theoretical.mean))

## [1] 4.979186 5.000000

hist(sim.mean, xlab="Mean of 40 exponentials", ylab= "Frequency", col="light
blue", main="Histogram for Simulation")
abline(v=actual.mean, lwd="4", col="red")
text(7, 120, paste("Actual mean = ", round(actual.mean,3), "\n Theoretical
mean = 5" ), col="black")
```



The mean of the simulation is very close to the theoretical mean.

Sample Variance versus Theoretical Variance

Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

The actual and theoretical variance calculations are:

```
actual.var<- var(sim.mean)
theoretical.var <- (1/lambda)^2/n
```

The actual variance is **0.6379013** and the theoretical variance is **0.625**. Again both values are quite close.

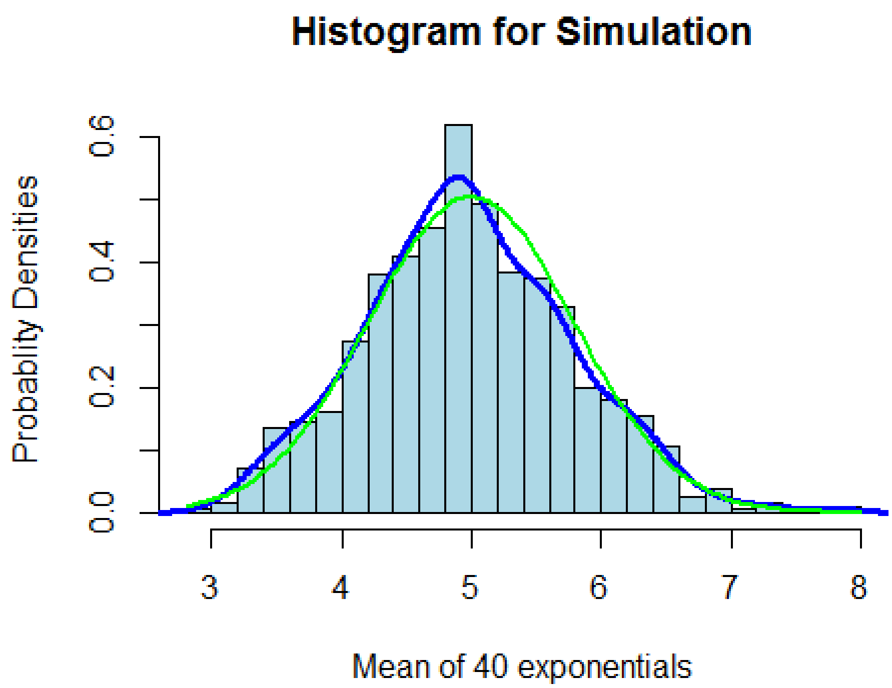
Distribution

Via figures and text, explain how one can tell the distribution is approximately normal.

We compare the distribution of the mean with a theoretical normal distribution. The histogram is replotted as probability densities and increased breakpoints. The smoothed density line for the simulation mean (**blue**) is overlaid together with the theoretical normal distribution (**green**).

```
hist(sim.mean, freq = FALSE, breaks = 20, xlab="Mean of 40 exponentials",
ylab= "Probablity Densities", col="light blue", main="Histogram for
Simulation")
x <- c(0,1)
```

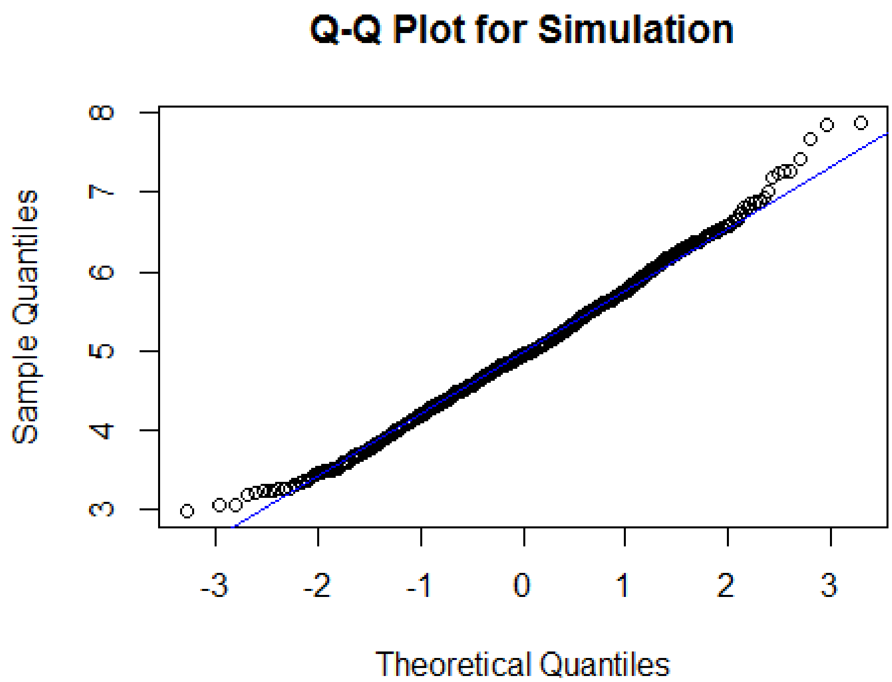
```
lines(density(sim.mean), lwd=3, col="blue")
curve(dnorm(x, mean=theoretical.mean, sd=sqrt(theoretical.var)), col="green",
lwd=2, add=TRUE, yaxt="n")
```



We can see that the simulation distribution is very close to the theoretical normal distribution.

We can do a Quantile-Quantile plot. We can see the values of the simulations fall close on the diagonal.

```
qqnorm(sim.mean, main="Q-Q Plot for Simulation", xlab="Theoretical
Quantiles", ylab="Sample Quantiles")
qqline(sim.mean, col="blue")
```



We can also compare the confidence interval. Again, we can see that the 2 values are very close.

```
actual.ci <- round (mean(sim.mean) + c(-1,1)*1.96*sd(sim.mean)/sqrt(n),3)
theoretical.ci <- theoretical.mean + c(-
1,1)*1.96*sqrt(theoretical.var)/sqrt(n)
```

95% confidence interval - **Actual=[4.732, 5.227]**, **Theoretical=[4.755, 5.245]**