Statistical Inference Project - Simulation

Miriam Pavon

10/03/2018

Overview

This project is 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.

Click on links here to quickly view tasks completed in this assignment:

1. Simulation
2. Show the sample mean and compare it to the theoretical mean of the distribution.
3. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
4. Show that the distribution is approximately normal.

Data

The data for this assignment comes from this:

```
s <- 1000 #Number of simulations
lambda <- 0.2 #Rate for all simulations
n <- 40 #Number of samples
set.seed(240) #set seed for reproducability
```

Requirements

For this assignment you will need some specific tools

RStudio: You will need RStudio to publish your completed analysis document to RPubs. You can also use RStudio to edit/write your analysis.

knitr: You will need the knitr package in order to compile your R Markdown document and convert it to HTML

Before beginning the project, be sure to load the required R libraries and set any environmental variables. Note that setting messages in markdown to false suppresses messages from library loading such as version number and dependencies. Updating to latest versions of these libraries may improve ability to obtain results fairly similar to the steps outlined here.

```
# load libraries
library(ggplot2)
library(knitr)
```

1. Simulations

```
#### 1. Simulation

simulation <- replicate(s, mean(rexp(n,lambda)))
```

```r
simulation_mean <- mean(simulation) #Distribution Mean
simulation_mean
```

```
## [1] 5.026972
```

```r
simulation_median <- median(simulation) #Median
simulation_median
```

```
## [1] 4.997169
```

```r
simulation_sd     <- sd(simulation) #Standard Deviation
simulation_sd
```

```
## [1] 0.8111959
```

Back to top

2. Show the sample mean and compare it to the theoretical mean of the distribution.

The simulation and the theory means are the next ones:

```r
simulation_mean  <- mean(simulation) #Distribution Mean
simulation_mean
```

```
## [1] 5.026972
```

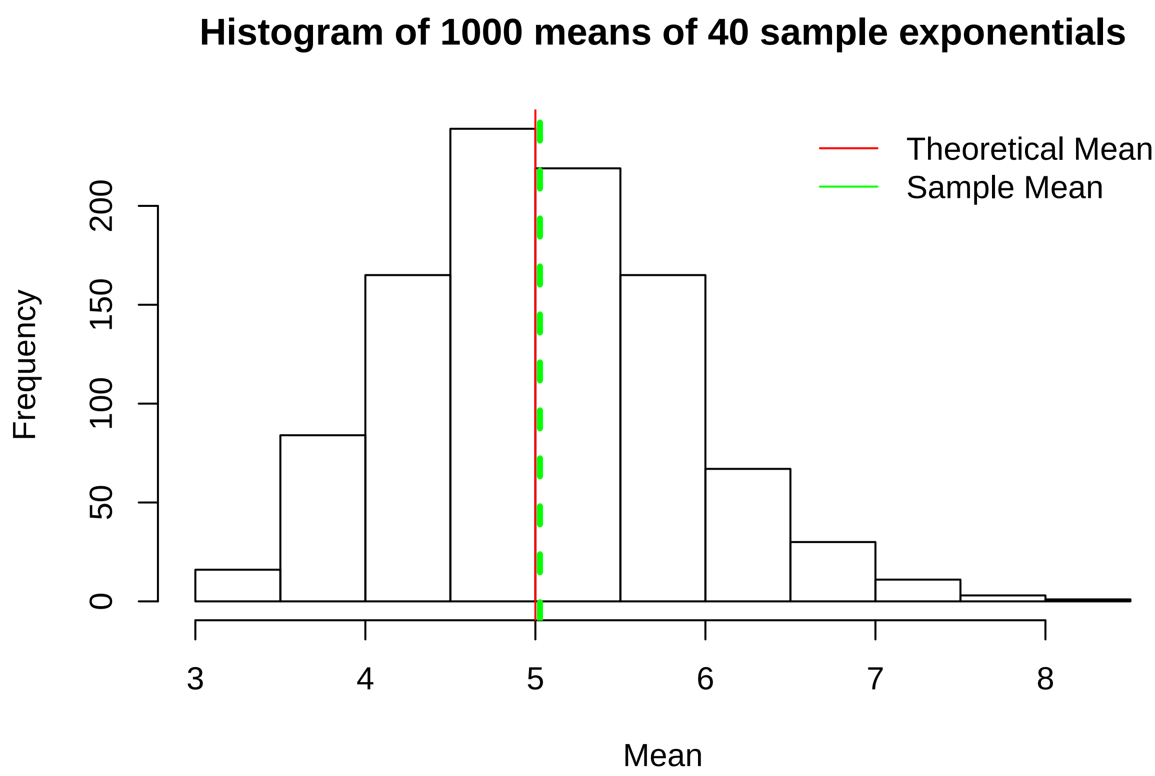
```r
theoretical_mean  <- 1/lambda #Analytical Mean
theoretical_mean
```

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

The actual center of the distribution of the average of 40 exponetialsis very close to its theoretical center of the distribution.

This is code for the graphical representation:

```r
hist(simulation, xlab = "Mean", main = "Histogram of 1000 means of 40 sample exponentials")
abline(v = simulation_mean, col = "green",lwd=3, lty=2)
abline(v = theoretical_mean, col = "red",lwd=1, lty=1)
legend('topright', c("Theoretical Mean","Sample Mean"),
       col=c("red", "green"), lty=c(1,1), bty = "n")
```



Back to top

---

3. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
# Theoretical variance
theoretical_sd <- (1/lambda)/sqrt(n)
theoretical_variance <- theoretical_sd^2
theoretical_variance
```

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

```
# Simulated variance
simulated_variance  <- var(simulation)
simulated_variance
```

```
## [1] 0.6580389
```

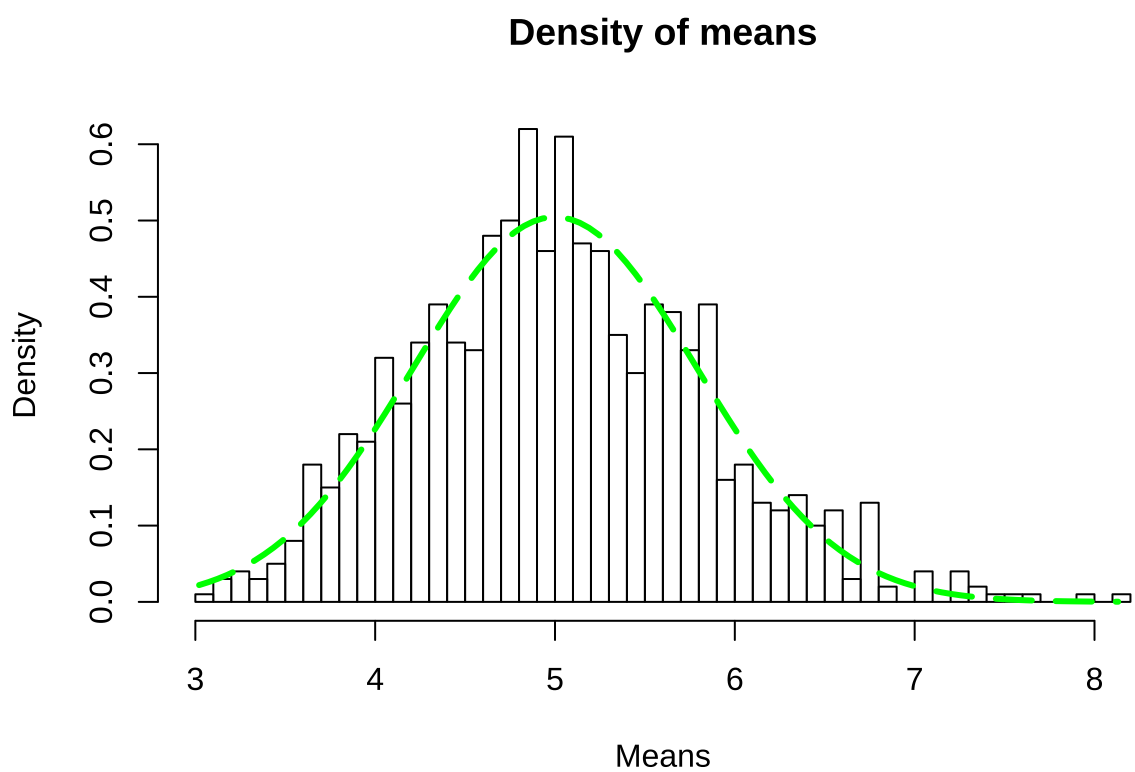
So, as we can see, both variances are quite close to each other.

Back to top

---

4. Show that the distribution is approximately normal.

```
xfit <- seq(min(simulation), max(simulation), length=100)
yfit <- dnorm(xfit, theoretical_mean , theoretical_sd)
hist(simulation,breaks=n,prob=T,xlab = "Means",main="Density of means",ylab="Density")
lines(xfit, yfit, pch=22, col="green", lwd=3,lty=5)
```



The distribution is approximately normal.

Back to top
