Statistical Inference Course Project

Cesar Espitia

June 18, 2015

Purpose

This is the project for the statistical inference class. In it, you will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. A simulation exercise.
2. Basic inferential data analysis.

Problem 1

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.

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

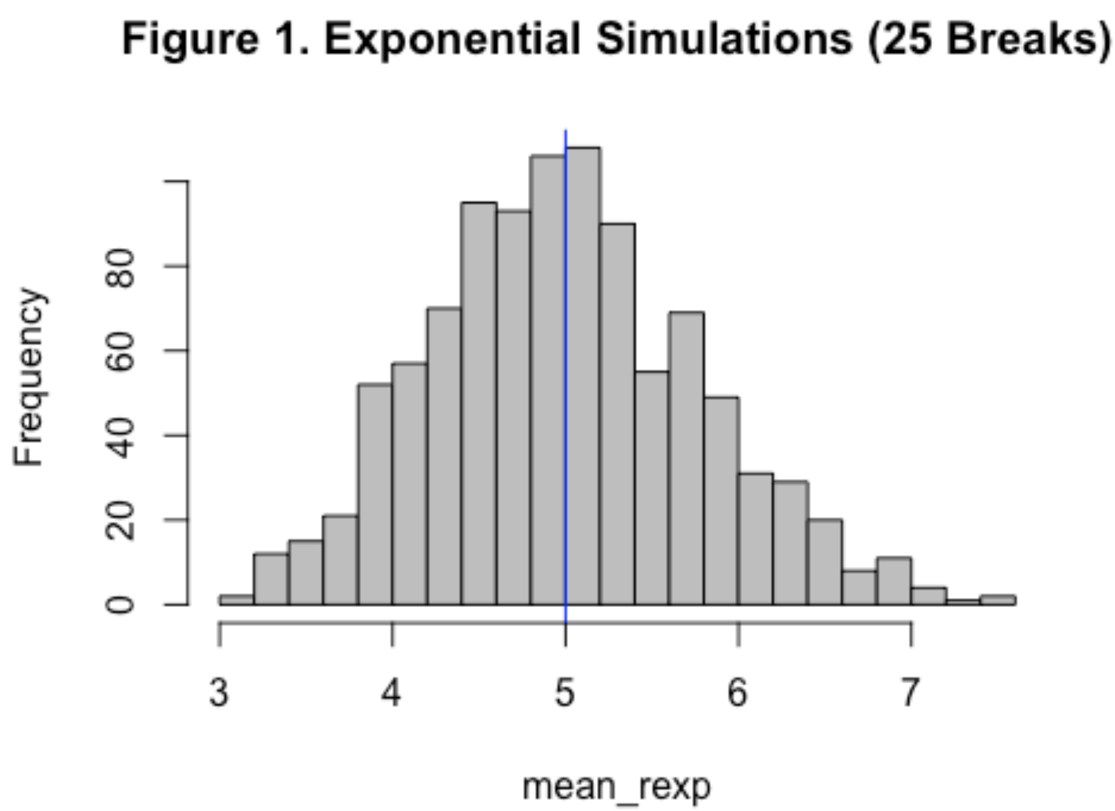
```
#PART 1
#Compare exponential distribution to CLT
#lambda = 0.2 for all simulations

set.seed(1)      #seed for reproducibility
l <- 0.2         #lambda
mn <- 1/l        #mean
n <- 40          #sample size
sim <- 1000      #simulations

#store simulations
store_rexp <- replicate(sim, rexp(n,l))

#calculate mean
mean_rexp <- apply(store_rexp, 2, mean)
```

Compare the simulation mean to the theoretical.

Note in Figure 1 that a histogram of the frequency of means is shown and overlayed is a blue line showing the theoretical mean.

```
#compare theory to simulations
sim_mean <- mean(mean_rexp)
```

```
theory_mean <- 1/l

print(sim_mean)

## [1] 4.990025

print(theory_mean)

## [1] 5

#plot results
hist(mean_rexp, col = "gray", main="Figure 1. Exponential Simulations (25
Breaks)", breaks=25)
abline(v=theory_mean, col = "blue")
```



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

Compare the simulation variance to the theoretical.

In theory, Variance = sigma^2; sigma = 1/lambda/sqrt(n) for an exponential distribution.

```
#compare variance of theory to simulations

sim_var <- var(mean_rexp)
```

```
theory_var <- (1/l/sqrt(n))^2

print(sim_var)

## [1] 0.6111165

print(theory_var)

## [1] 0.625
```

3. Show that the distribution is approximately normal.

Figure 2 below shows the simulations compared to a Normal Distribution of mean = 1/lambda and, sd = 1/lambda/sqrt(n).

The distribution of averages from our simulation, it is very close to a normal distribution which approached the CLT.

```
xfit <- seq(min(mean_rexp), max(mean_rexp), length = 100)
yfit <- dnorm(xfit, mean = 1/l, sd = (1/l/sqrt(n)))
hist(mean_rexp, col = "gray", main="Figure 2. Exponential Simulations (40
Breaks)", breaks=n, prob=T)
abline(v=sim_mean, col = "blue")
lines(xfit, yfit, col="red", lty=5, lwd=10)
```