Coursera Statistical Inference Project

Seo-yeon Lim

2018/03/03

Part 1: Simulation Exercise - Exponential Distribution vs. Central Limit Theorem

Overview

In this first part of the project, our goal is to investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT). The exponential distribution in R is calculated using rexp(n,lambda) function. The mean of the exponential distirubtion is 1/lambda, and the standard deviation is 1/lambda as well. For the simulation, we set lambda = 0.2, n =40, number of simulation = 1000.

Simulations

Setting for the simulation analysis.

Set the variables according to the given values.

```
set.seed(12345678) #Set the seed for reproducibility
lambda <- 0.2 #Given value of lambda
n <- 40 #number of exponentials
sim_num <- 1000 #number of simuations
```

Create a matrix with 40 samples from the exponential distribution, with 1000 simulations. Then calculate the mean across the 40 values for each of the 1000 simulations.

```
#Conduct expoential distribution and check for its mean
exp_dist <- matrix(rexp(n*sim_num, lambda), sim_num)
exp_means <- rowMeans(exp_dist)
```

Conduct mean calculation for 1000 simulations of the exponential distribution

```
means <- NULL
for (i in 1: sim_num)means <- c(means, mean(rexp(n, lambda)))
```

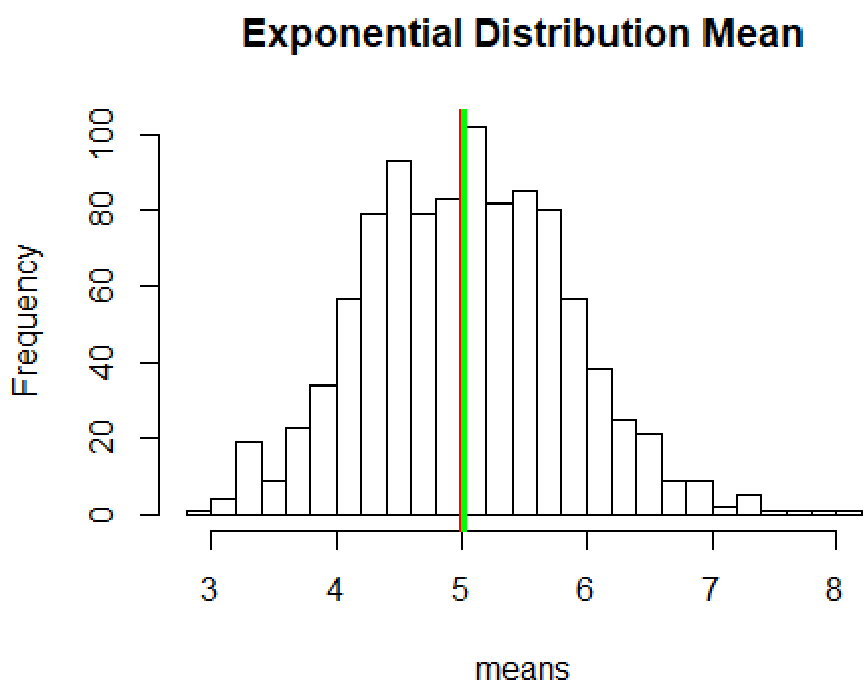
1. Show the sample mean and compare it to the theoretical mean of the distribution.

```
##Sample mean vs. theoretical mean
mu <- 1/lambda
mu #Theoretical mean

## [1] 5

sample_mn <- mean(exp_means)
sample_mn #Mean of sample means

## [1] 5.017885
```

The theoretical mean is 5, while the sample mean is 5.017885.

If we make a histogram, we can see that the sample mean (green) is slightly shifted toward right.

```
hist(means, main= "Exponential Distribution Mean", breaks = 30)
abline(v=mu, col="red",lwd=3)
abline(v=sample_mn, col="green", lwd=3)
```



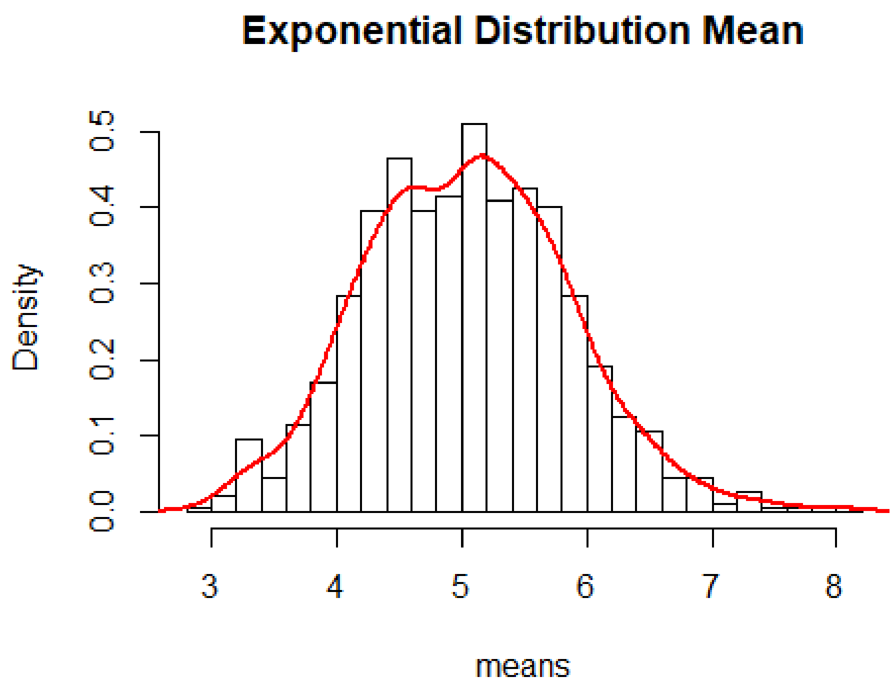
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
##Sample variance (and std. dev.) vs. theoretical variance (and std.dev.)
var <- (1/lambda)^2/n
var #Theoretical variance

## [1] 0.625

sigma <- 1/lambda/sqrt(n)
sigma #theoretical standard deviation

## [1] 0.7905694

sample_var <- var(exp_means)
sample_var #sample variance

## [1] 0.6020976

sample_sd <- sd(exp_means)
sample_sd #sample standard deviation

## [1] 0.7759495
```

The theoretical variance is 0.625 (sd = 0.7905694), while the sample variance is 0.6020979 (sd = 0.7759495).

3. Show that the distribution is approximately normal.

```
hist(means, prob=TRUE, main= "Exponential Distribution Mean", breaks = 30)
lines(density(means), lwd=2, col="red")
```



As we can see, the red line shows a smooth bell-curve, with centered mean. Hence we can conclude that the distribution is approximately normal.