Comparison between exponential distribution and the central limit theorem

Ricardo Arango Giraldo

20 de marzo de 2018

Overview

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution will be simulated in R with rexp(n, lambda) where lambda is the rate parameter. In theory the mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. We will set n = 40 and set lambda = 0.2 for all of the simulations and compare the resulting distribution of the smaple averages to the theory for 1000 simulations.

Our results will: Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. We shall:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show the variance of the sample and compare it to the theoretical population variance.
3. Show that the distribution is approximately normal.

Simulations

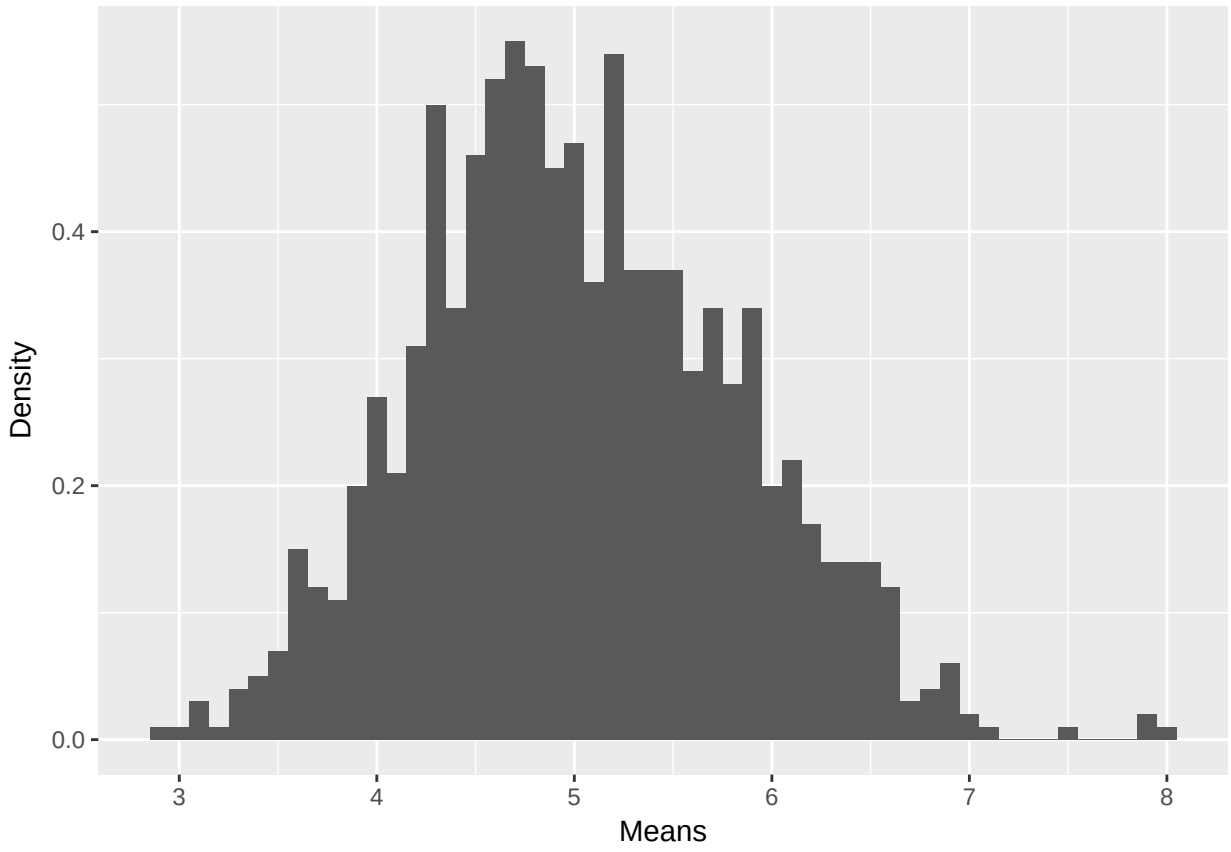
We will run a series of 1000 simulations to create a data set for comparison to theory. Each simulation will contain 40 observations and the expoential distribution function will be set to **rexp().**

```
#We define the parameters
n<-40
lambda<-0.2

#We set a seed for reproduce results
set.seed(207)

#We find the means
data<-data.frame(x=sapply(1:1000,function(x){mean(rexp(n,lambda))}))

#We show the histogram
library(ggplot2)
ggplot(data = data, aes(x = x)) +
  geom_histogram(binwidth=0.1, aes(y=..density..)) +
  labs(x="Means") +
  labs(y="Density")
```



Sample Mean vs. Theoretical Mean

```
#we found the theoretical mean
th_mean<-1/lambda

#we found the sample mean
sa_mean<-mean(data$x)
```

The theoretical mean of exponential distribution is 5 and the sample mean is 5.0263932.

Sample Variance vs Theoretical Variance

We need first to find the standard deviation:

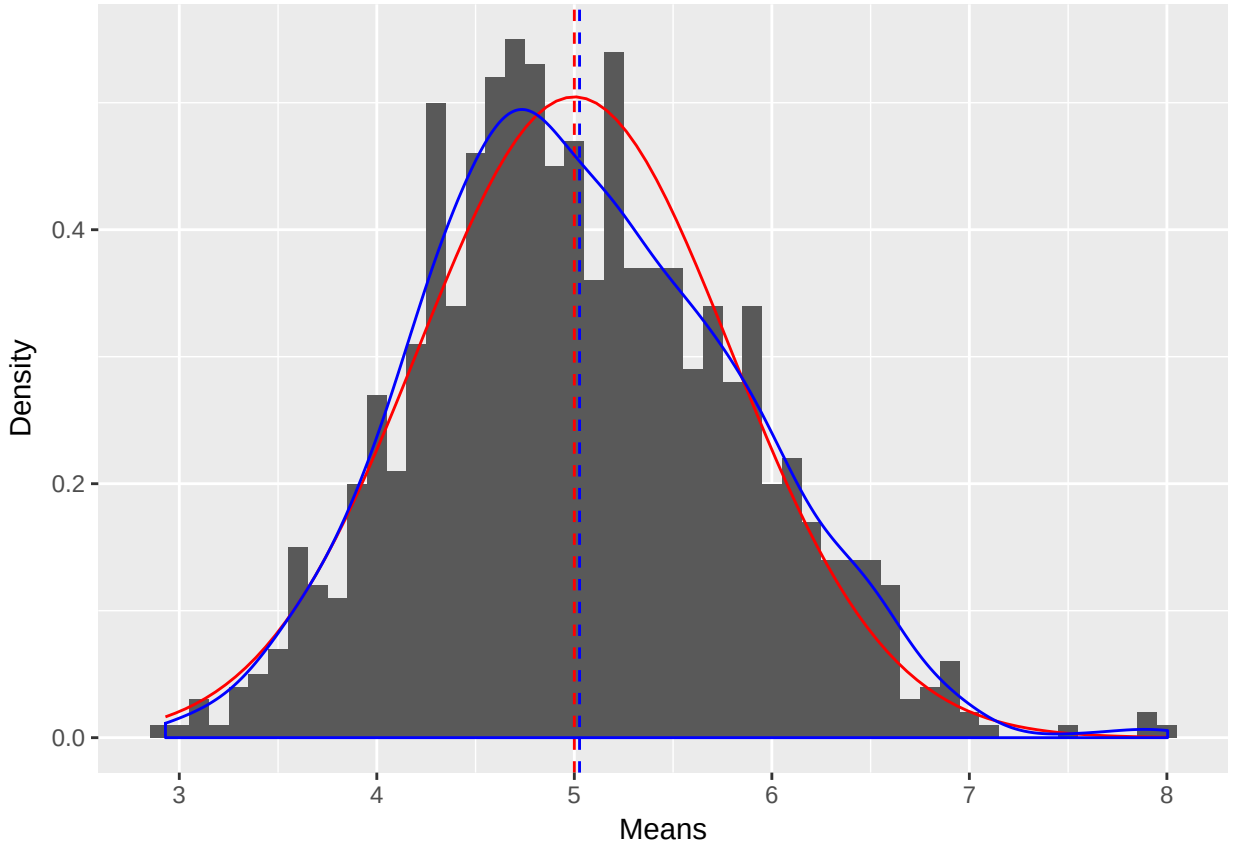
```
#we found the theoretical standard deviation
th_sd<-1/lambda/sqrt(n)

#we found the sample standard deviation
sa_sd<-sd(data$x)
```

You remember that variance is the standar deviation quare, so the theoretical variance of exponential distribution is 0.625 and the sample variance is 0.6449948.

```
ggplot(data = data, aes(x = x)) +
  geom_histogram(binwidth=0.1, aes(y=..density..)) +
  stat_function(fun = dnorm, args = list(mean = th_mean , sd = th_sd), colour = "red", size=.5) +
```

```
  geom_vline(xintercept = th_mean, size=.5,linetype=2, colour="red") +
  geom_density(colour="blue", size=.5) +
  geom_vline(xintercept = sa_mean, size=.5, colour="blue",linetype=2) +
  labs(x="Means") +
  labs(y="Density")
```



The red line is the normal distribution with lambda = 0.2 and the blue line is the sample exponential distribution. The figure shows that the 2 distribution lines, blue and red, are well aligned thus the distribution of simulated data is approximately normal.