

Stat. Inf. Part 1: Simulation and Comparison between Exp. distribution and CLT

Jean Marie Cimula

5 March 2016

Overview

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. I have to investigate the distribution of averages of 40 exponentials. Note that I need to do a thousand simulations.

Below we illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials

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

Let's start the simulation

```
ECHO=TRUE
set.seed(3)
#Constant lambda
lambda <- 0.2
#Number of simulation or test
nbSim <- 1000
#Exponential
Sample <- 40

#Calculation
expDist <- matrix(data=rexp(Sample * nbSim, lambda), nrow = nbSim)
expDistMean <- data.frame(means=apply(expDist, 1, mean))

smu <- 1/lambda #Sample Mean
tmu <- mean(expDistMean$means) #Theoretical Mean
svar <- (1/lambda)/sqrt(Sample))^2 #Sample Variance
tvar <- var(expDistMean$means) #Theoretical Variance
dFM <- abs(smu-tmu) #Difference between Means
dFV <- round(abs(svar-tvar),3) #Difference between both Variances
```

Comparison between Sample Mean and Theoretical Mean

Sample-Mean : 5 and Theoretical-Mean : 4.9866197

Both Means are almost the same. We can compare them as follow : the abs of $5 - 4.9866197 =$

0.0133803. The difference is not significant. If we round 0.0133803 will get 0.

Comparison between Sample Variance and Theoretical Variance

Sample-Variance : 0.625 and Theoretical-Variance : 0.6257575

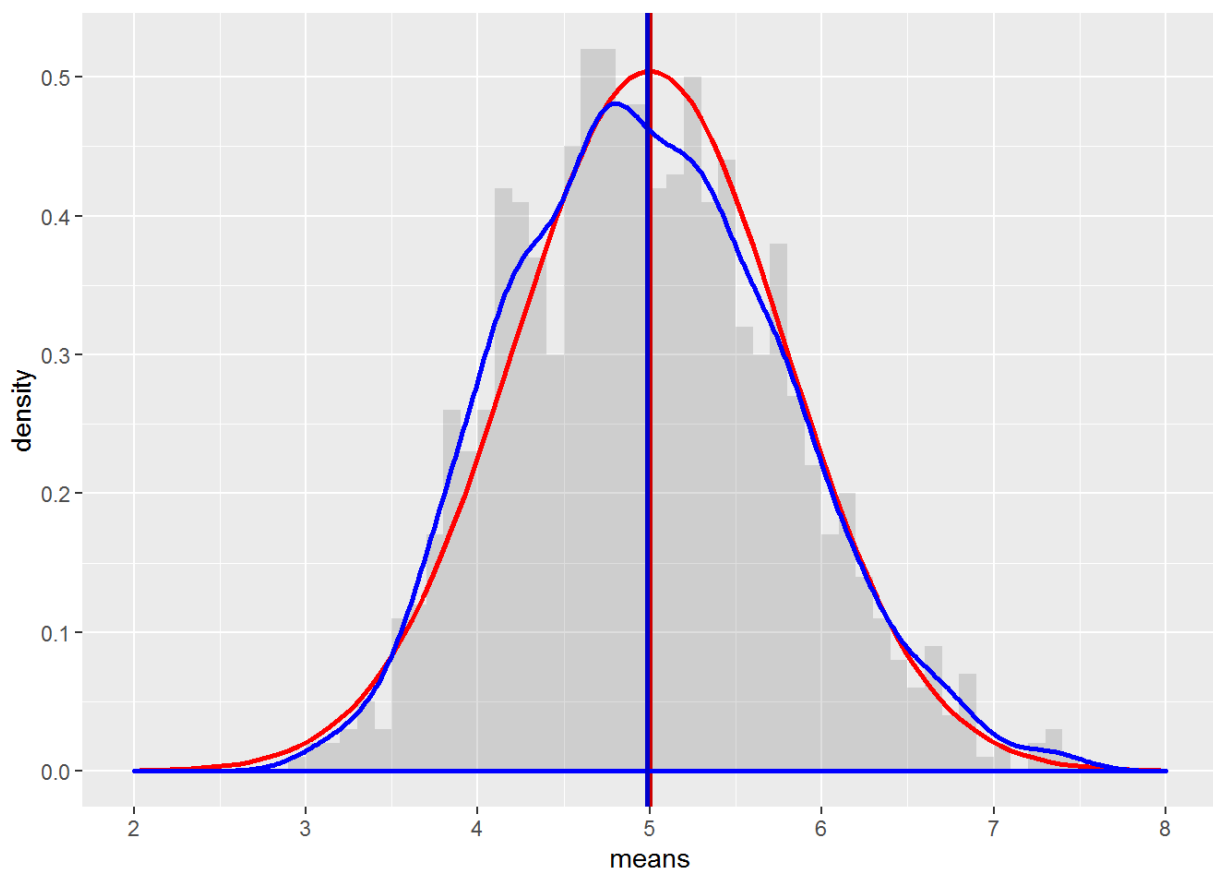
Both Variance are almost the same. We can compare them as follow : the abs of $0.625 - 0.6257575 = 0.001$. The difference is not significant. If we round 0.001 will get 0.

(3) Show that the distribution is approximately normal.

Below, please see how the distribution of our simulations appears normal.

```
#Load ggplot2
library(ggplot2)

#Plot to describe the mean
ggplot(data = expDistMean, aes(x = means)) + geom_histogram(binwidth=0.1, aes(y=..density..), alpha=0.2) +
  stat_function(fun = dnorm, arg = list(mean = 1/lambda , sd = 1/lambda/sqrt(Sample)), colour = "red", size=1) +
  geom_vline(xintercept = 1/lambda, size=1, colour="#CC0000") + geom_density(colour="blue", size=1) +
  geom_vline(xintercept = mean(expDistMean$means), size=1, colour="#0000CC") +
  scale_x_continuous(breaks=seq((1/lambda)-3, (1/lambda)+3, 1), limits=c((1/lambda)-3, (1/lambda)+3))
```



Above in the plot the the Bell Curve is the normal distribution.