Part 1: Simulation Exercise

Mukesh Saxena

30/06/2018

Overview

In this first part of the project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The theoretical mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

We will simulate 1000 averages of 40 exponentials using a rate parameter lambda = 0.2. Then, we will compare the mean and the variance of the simulated distribution to the theoretical ones. And in the final step, we are going to show that the the exponencial distribution is approximately normal by comparing it to a similar sample simulated from normal distribution.

Simulation settings

We are going to investigate the exponential distribution, using a 1000 simulations of averages of 40 exponentials with a rate parameter lambda 0.2.

```
# number of simulations
nosim <- 1000
# sample size
n <- 40
# rate parameter
lambdaa <- 0.2

# simulate nosim averages of 40 exponentials
my_data <- apply(matrix(rexp(nosim * n, rate=lambdaa), nosim), 1, mean)
```

Sample Mean versus Theoretical Mean.

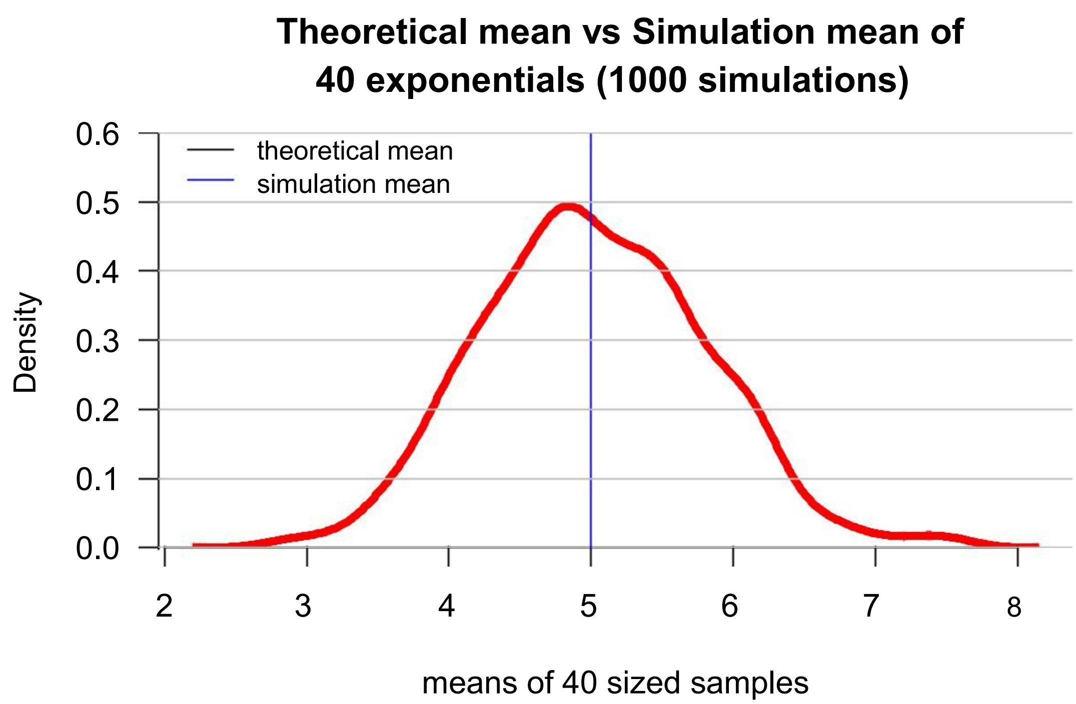
The theoretical mean of exponential distribution is 1/lambda. Then, according to the Central Limit Theorem the theoretical mean of a 40 exponentials must be 1/lambda. So, comparing it to the simulation mean :

```
# theoretical mean
theo_mean <- 1/lambdaa
# simulation mean
sim_mean <- mean(my_data)

m <- c(theo_mean, sim_mean); names(m) <- c("theo_mean", "sim_mean")
m
```

```
#   theo_mean  sim_mean
#    5.000000  5.020456
```

Looking then at the theoretecal and simulation mean in the figure below (The code for generation the figure is in the Appendix 1). It is clear so that the two values are two close.



Simulation Variance versus Theoretical Variance.

The theoretical standard deviation of exponential distribution is 1/lambda. Then, the theoretical vari-ance of exponential distribution is (1/lambda)ˆ2 and the variance of mean of a 40 exponentials must be ((1/lambda)ˆ2)/n according to the central limit Theorem. So, comparing it to the simulation variance :

```
# theoretical variance
theo_var <- ((1/lambdaa)^2)/n
# simulation variance
sim_var <- var(my_data)

v <- c(theo_var, sim_var); names(v) <- c("theo_var", "sim_var")
v
```

```
##   theo_var    sim_var
## 0.6250000 0.6296891
```

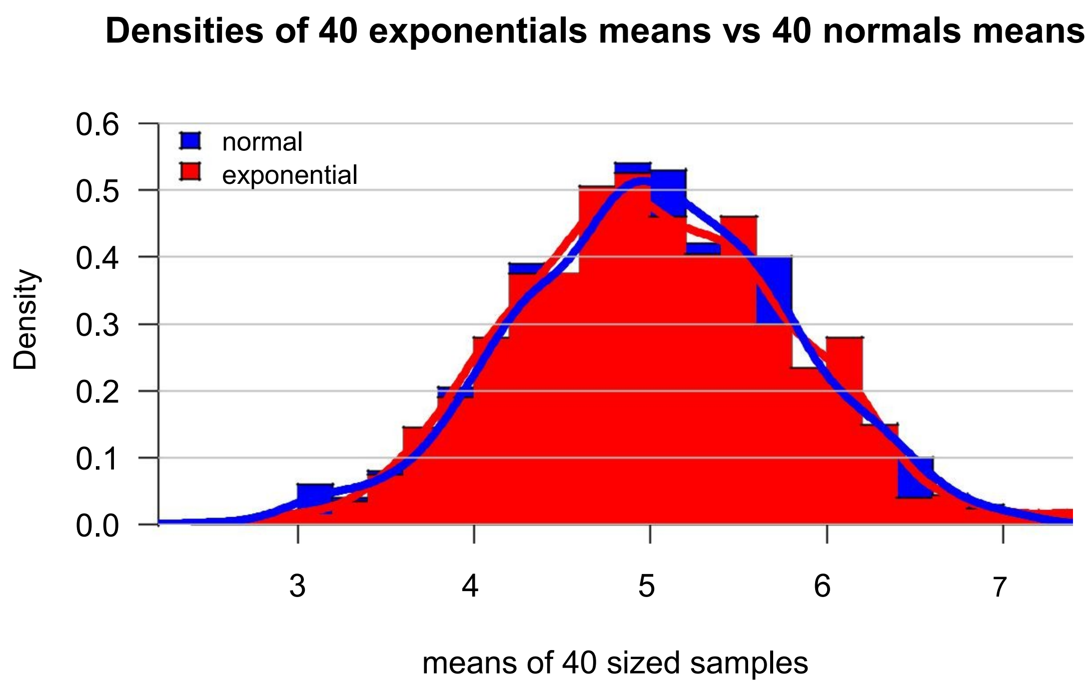
So the simulation variance is also very close to the thoretical variance.

Comparing with normal distribution.

In order to compare the exponential with the normal distribution, we are going to set a simimilar simulated sample using the normal distribution with a mean 1/lambda and stadard deviation 1/lambda.

```
# simulate nosim averages of 40 normals with mean = 1/lambda and sd = 1/lambda
my_norm <- apply(matrix(rnorm(nosim * n, mean=1/lambdaa, sd=1/lambdaa), nosim), 1, mean)
```

The figures above shows the density histograms of averages of 40 exponentials and to averages taken from normal distribution for a 1000 simulation.



It is clear from the figures of histograms that the exponential and the normal distributions (of 40 sample mean) have the the same shape of density. We can conclude so that the distribution is approximatly normal. This was in part predictable since we use the same mean and the same standard deviation.

Appendix 1

R code used to generate the plot in order to compare theoretical and simulation mean.

```
par(yaxs="i",las=1)
plot(density(my_data), col=rgb(1,0,0,1/2), lwd=4, type="l", ylim=c(0, 0.6),
        main="Theoretical mean vs Simulation mean of 40 exponentials
        (1000 simulations)",
        xlab="means of 40 sized samples", frame.plot=FALSE)
abline(v=sim_mean, lty=2)
abline(v=theo_mean, col="blue", lty=1)
grid(nx=NA,ny=NULL,lty=1,lwd=1,col="gray")
legend(x=2, y=0.62, legend=c("theoretical mean", "simulation mean"), lty=2:1,
         col=c("black", "blue"), cex=0.8, bty="n")
```

Appendix 2

R code used to generate the plot in order to compare exponential and normal distributions.

```
par(yaxs="i",las=1)
hist(my_norm, breaks=20, prob=TRUE, col=rgb(0,0,1,1/4),
       ylim=c(0, 0.62), main="Densities of 40 exponentials means vs 40 normals means",
       xlab="means of 40 sized samples")
hist(my_data, breaks=20, prob=TRUE, col=rgb(1,0,0,1/4), add=TRUE)
lines(density(my_data),col=rgb(1,0,0,1/2),lwd=4)
lines(density(my_norm),col=rgb(0,0,1,1/2),lwd=4)
grid(nx=NA,ny=NULL,lty=1,lwd=1,col="gray")
legend("topleft", legend=c("normal", "exponential"),
         fill=c(rgb(0,0,1,1/4), rgb(1,0,0,1/4)), cex=0.8, bty = "n")
```