Exponential Distribution Simulation using R

Statistical Inference Course Project

Mouhammed Samer Houssien Hasn

July 20, 2019

Abstract. This project aims to simulate the exponential distribution, and make an application of the law of large numbers (LLN), and the central limit theory (CLT) using R programming language with basic plotting tools.

1. Exponential Distribution

The probability density function (pdf) of an exponential distribution is:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Where λ is the rate parameter.

Exponential distribution shape is determined by two parameters, number of observations n, and the rate λ. The mean of the distribution is λ^{-1}, and the variance is λ^{-2},and λ^{-2}/n for sample variance .

2. Theoretical and actual estimation of mean and variance

The theoretical mean of the distribution is λ^{-1}; however, the actual mean when simulating a random exponential distribution differs from the theoretical one.

LLN states that the sample average $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ converges to its expected value as n goes to infinity.

We simulate a $\lambda = 0.2$ and $n = 40$ exponential distribution thousand time. The theoretical mean of each of these distributions equals to 5, and the actual mean of each differs from 5.

To run the simulation in R, we first set the values of the parameters of the simulation. We create a 1000 rows 40 columns matrix contains each simulation values (which is assumed to be I.I.D random variables), and a matrix of 40 rows contains the mean of each simulation. as follow:

```r
#parameters of the simulation
n<-40              #number of exponentials
lambda<-.2         #rate parameter
exp.theo.mean<-1/lambda #The mean of exponential distribution
exp.val<- matrix(rexp(40000,lambda), 1000,n) #each simulation's values
exp.val.means <- rowMeans(exp.val) #mean of each simulation
```

We compare the theoretical mean and the actual mean graphically by plotting a histogram.

```
hist(exp.val.means,main = "Fig.1 \n Theoretical vs actual mean",
     xlab = "Means",col = "grey",probability = TRUE) #probability density of means
abline(v = exp.theo.mean,col="red",lwd=2,break=25) #to draw the theoretical mean
x<-exp.val.means
curve(dnorm(x,mean = mean(exp.val.means),sd=sqrt(var(exp.val.means)))
      ,add = T,lwd=2,col="blue") #distribution curve
```

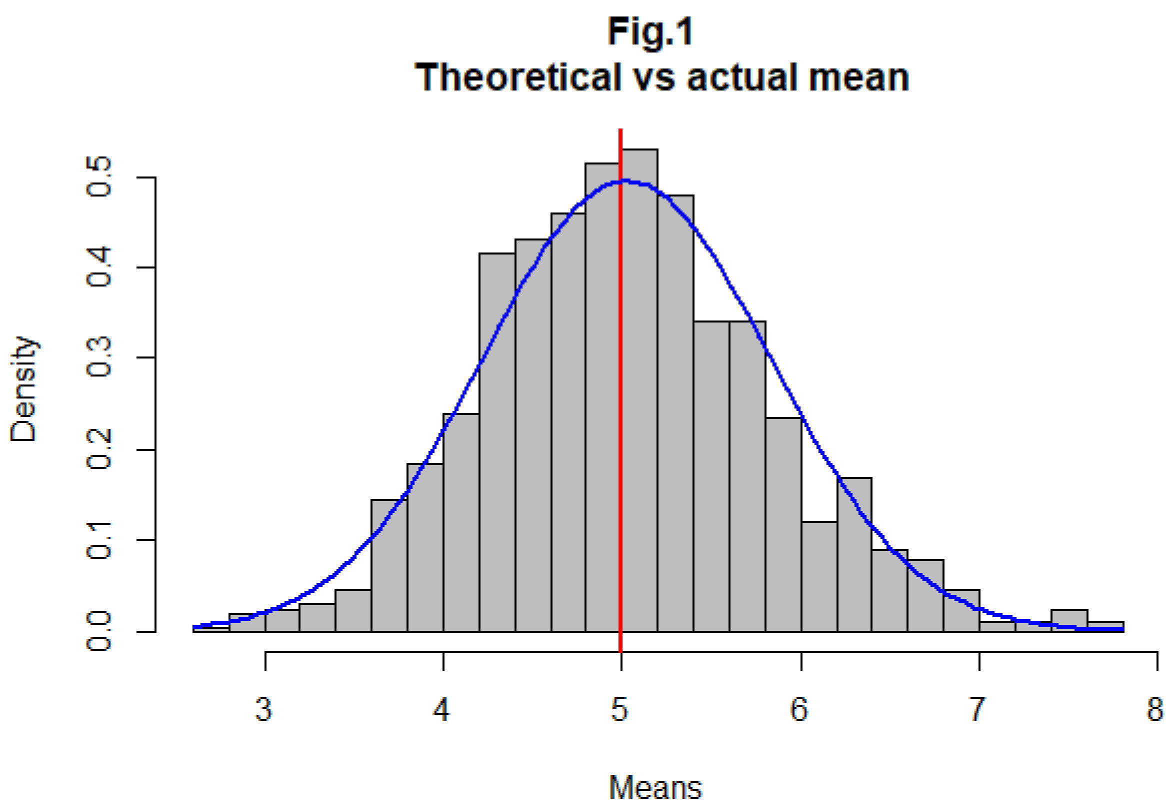

Fig.1 shows how the samples mean are distributed around the theoretical mean (the red line). The mean of these means sample will be always very close to the theoretical mean, which equals to 5.

CLT states that the values of $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}$ converge by distribution with the normal distribution as n goes to the infinity.

Fig.1 also shows how means sample distribution is asymptotic to the normal distribution. Also, we can test the normality of means sample by plotting the quantiles of the mean sample.

```
qqnorm(exp.val.means,main = "Fig.2 \n Normal Q-Q Plot");qqline(exp.val.means,col="red")
```

Fig.2
Normal Q-Q Plot

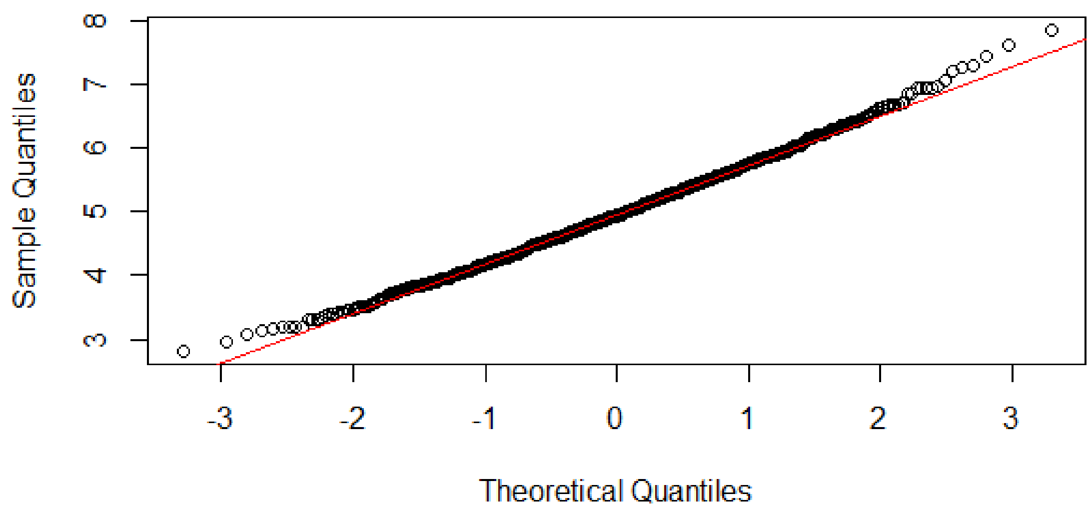

Fig.2 shows a robust convergence with the normal distribution as the quantiles consists with the normal quantiles (the red line).

The 95% confidence intervals of the theoretical and actual mean were calculated as follow:

```
conf.interval <- mean(exp.val.means) + c(-1,1)*1.96*(sd(exp.val.means)/sqrt(n))
conf.interval.theo <- exp.theo.mean + c(-1,1)*1.96*sqrt((1/lambda)^2/n)/sqrt(n)
```

By returning the values of each intervals we get [4.754261 , 5.238452] for actual and [4.755 , 5.245] for the theoretical, and they two do not differ significantly.

And the theoretical and actual variances were calculated as follow:

```
exp.val.var <- var(exp.val.means)
theo.var <- ((1/lambda)^2)/n
```

And the return 0.6499607 for actual and 0.625 for theoretical variance, and they two do not differ significantly.