EXPONENCIAL DISTRIBUTION COMPARE WITH THE CENTRAL LIMIT THEOREM

Course : Statistical Inference

Amarante, Geraldo Barbosa do

Mach 18,2018 Brazil

PROJECT DESCRIPTION

Part 1: Simulation Exercise Instructions

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. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

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

INTRODUCTION

The Central Limit Theorem(CLT) is one of the most important theorems in statistics. It states that the distribution of averages of iid (independent and identically distributed) variables becomes that of a standard normal as the sample size increases. So, the result of:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\text{Estimate} - \text{Mean of estimate}}{\text{Std. Err. of estimate}}$$

has a distribution like that of a standard normal for large nn. This will be applied later in this analysis to the exponential distribution and see the normality of the calculated results.
Hence, in order to apply the CLT to the exponential distribution we will investigate the distribution of averages of 40 exponentials. The average of 40 exponentials will be simulated 1000 times to the benefit of better CLT application results.
The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of the exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

SIMULATIONS

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

The initial idea is to analyze the example and compare with the theorem. To generate the necessary data we will use a series of 1000 simulations, so that it can be compared with the theorem. Each of the simulations will contain 40 observations and the distribution function will be performed with the lambda of 0.2.

Below the information defined for the project:

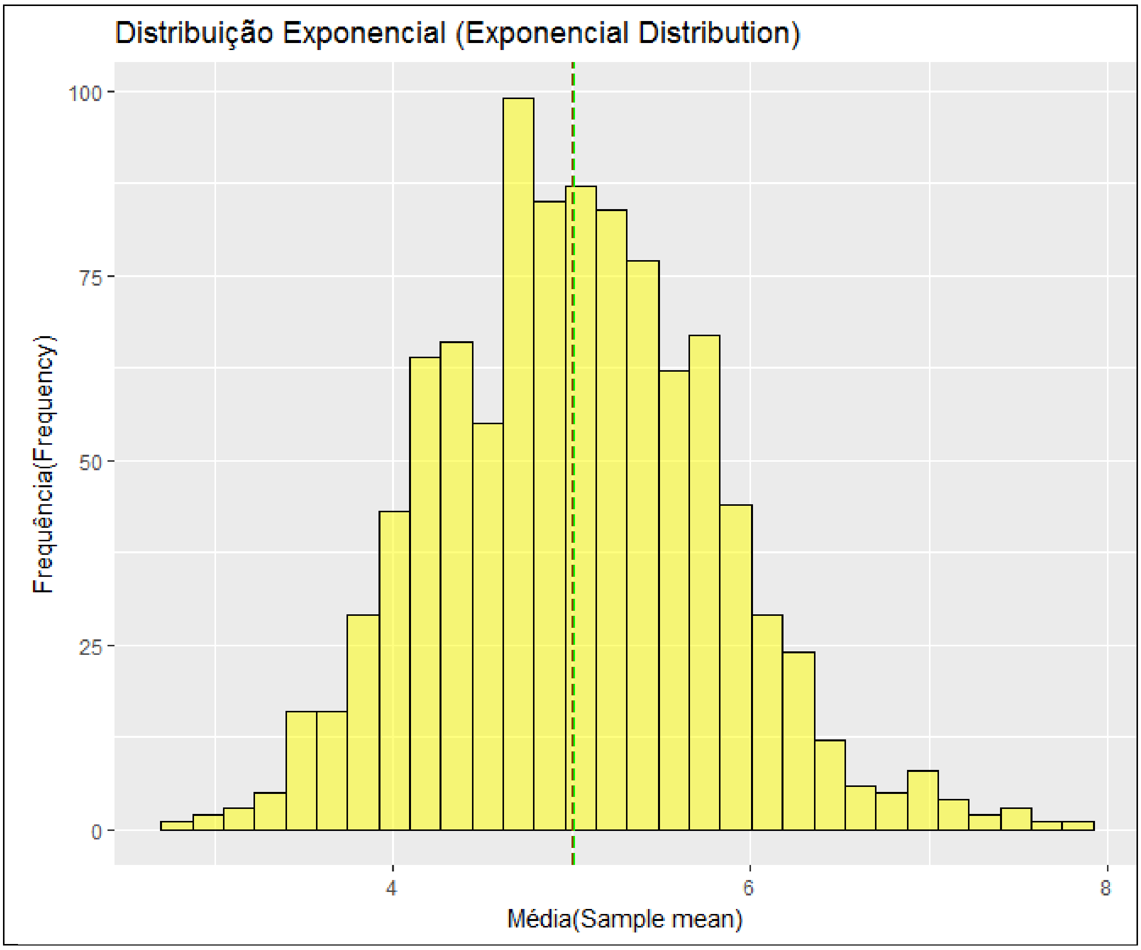
Lambda = 0.2
Number of distribuitions = 40
Number of simulations = 1000

This graph represents the distribution of the means. The vertical line represents therelative potion of the mean of distribution and theoretical mean.

R CODE

```
lambda <- 0.2
numdistribuicoes <- 40    #number of distribuitions
numsimulacoes <- 1000     #number of simulations

set.seed(349)
dado_simulado <- matrix(rexp(n=
numsimulacoes*numdistribuicoes,rate=lambda), numsimulacoes,
numdistribuicoes)
media_simulada <- rowMeans(dado_simulado)
media <- mean(media_simulada)
media_teorica <- 1/ lambda
result1 <-data.frame("Mean"=c(media,media_teorica),
          row.names = c("Média(Mean from the samples)
","Média Teórica(Theoretical mean)"))
kable(x = round(result1,3),align = 'c')
sampleMean_data <- as.data.frame (sample_mean)
ggplot(as.data.frame (media_simulada), aes(media_simulada))+
  geom_histogram(alpha=.5, position="identity", fill="yellow",
col="black")+
  geom_vline(xintercept = media_teorica, colour="darkorange4",
linetype = "longdash",show_guide=TRUE)+
  geom_vline(xintercept = media, colour="green", linetype =
"longdash", show_guide=TRUE)+
  ggtitle ("Distribuição Exponencial (Exponencial Distribution)")+
  xlab("Média(Sample mean)")+
  ylab("Frequência(Frequency")
```



NOTE:

```
lambda <- 0.2
numdistribuicoes <- 40     #number of distribuitions
numsimulacoes <- 1000     #number of simulations
dado_simulado <- matrix(rexp(numdistribuicoes =
numsimulacoes*numdistribuicoes,rate=lambda), numsimulacoes,
numdistribuicoes)
media_simulada <- rowMeans(dado_simulado)
mean(media_simulada)
```

This is the mean from the sample : **5.011672**

```
media_teorica <- 1/ lambda
media_teorica
```

This is the theoretical mean from the sample : **5.00**

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

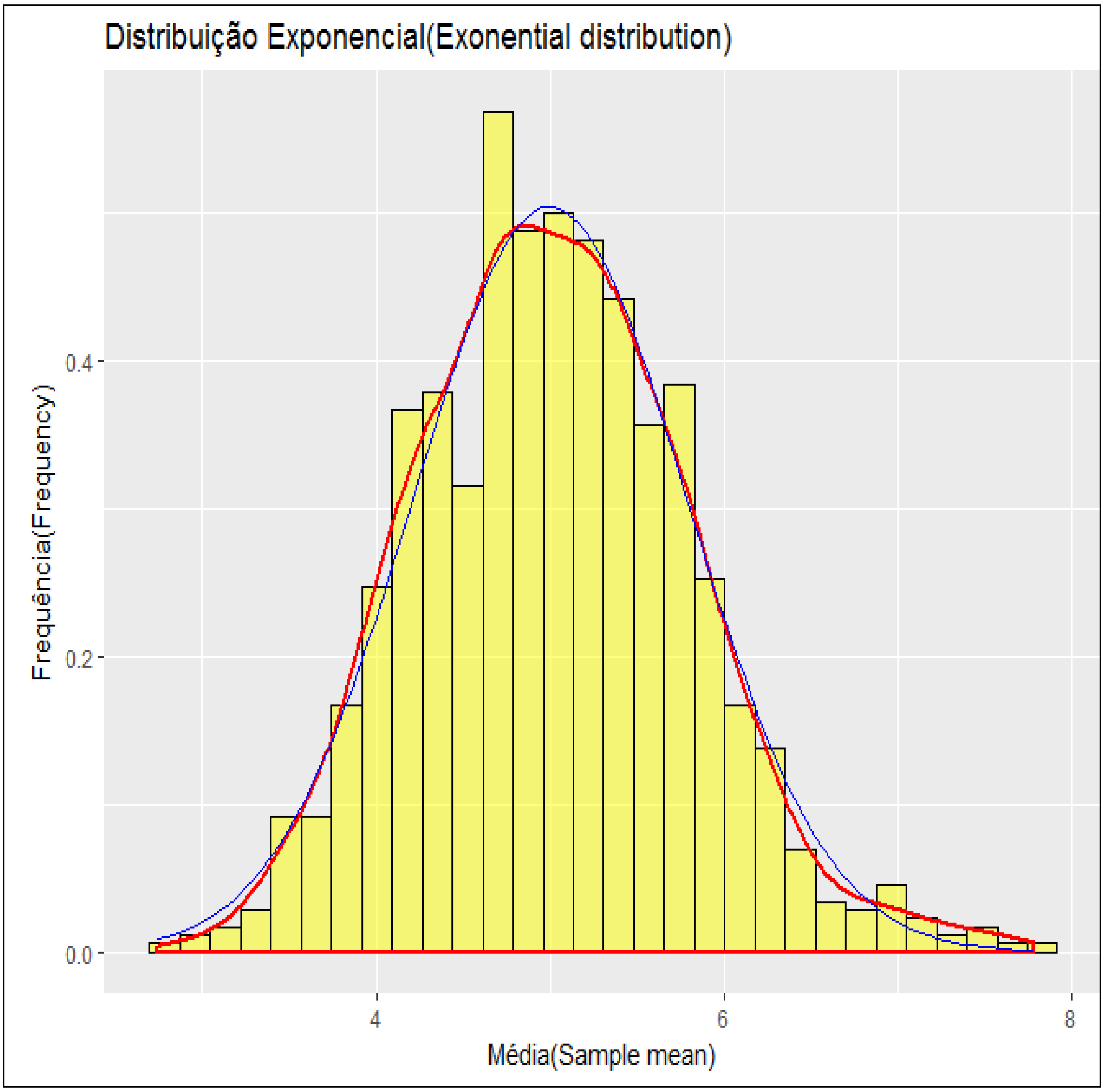
Theoretical variance of the distribution can be calculated as :

$$: \sigma_{\bar{x}}^2 = \frac{(1/\lambda)^2}{n} .$$

NOTE:
“n” is the number of distributions.

R CODE

```
lambda <- 0.2
numdistribuicoes <- 40
numsimulacoes <- 1000
set.seed(349)
dado_simulado <- matrix(rexp(numdistribuicoes=
numsimulacoes*numdistribuicoes,rate=lambda), numsimulacoes,
numdistribuicoes)
media <- rowMeans(dado_simulado)
media_exemplo <- as.data.frame (media)
variancia <- var(media)
variancia_teorica <- (1/ lambda)^2 / numdistribuicoes
dado <-data.frame("Variance"=c(variancia, variancia_teorica),
           row.names = c("Variância(Variance from the
sample)","Variância teórica(Theoretical variance)"))
kable(x = round(dado,3),align = 'c')
ggplot(media_exemplo, aes(media))+
  geom_histogram(aes(y=..density..), alpha=.5, position="identity",
fill="yellow", col="black")+
  geom_density(colour="red", size=1)+
  stat_function(fun = dnorm, colour = "blue", args = list(mean =
media_teorica, sd = sqrt(variancia_teorica)))+
  ggtitle ("Distribuição Exponencial(Exonential distribution) ")+
  xlab("Média(Sample mean)")+
  ylab("Frequência(Frequency)")
```



NOTE:

```
##Variance from the sample
dado_simulado <- matrix(rexp(numdistribuicoes= numsimulacoes*numdistribuicoes,rate=lambda), numsimulacoes, numdistribuicoes)
media <- rowMeans(dado_simulado)
variancia <- var(media)
variancia
```

Variance from the sample : **0.5938109**

```
variancia_teorica <- (1/ lambda)^2 /numdistribuicoes
variancia_teorica
```

This is the theoretical mean from the sample : **0.625**

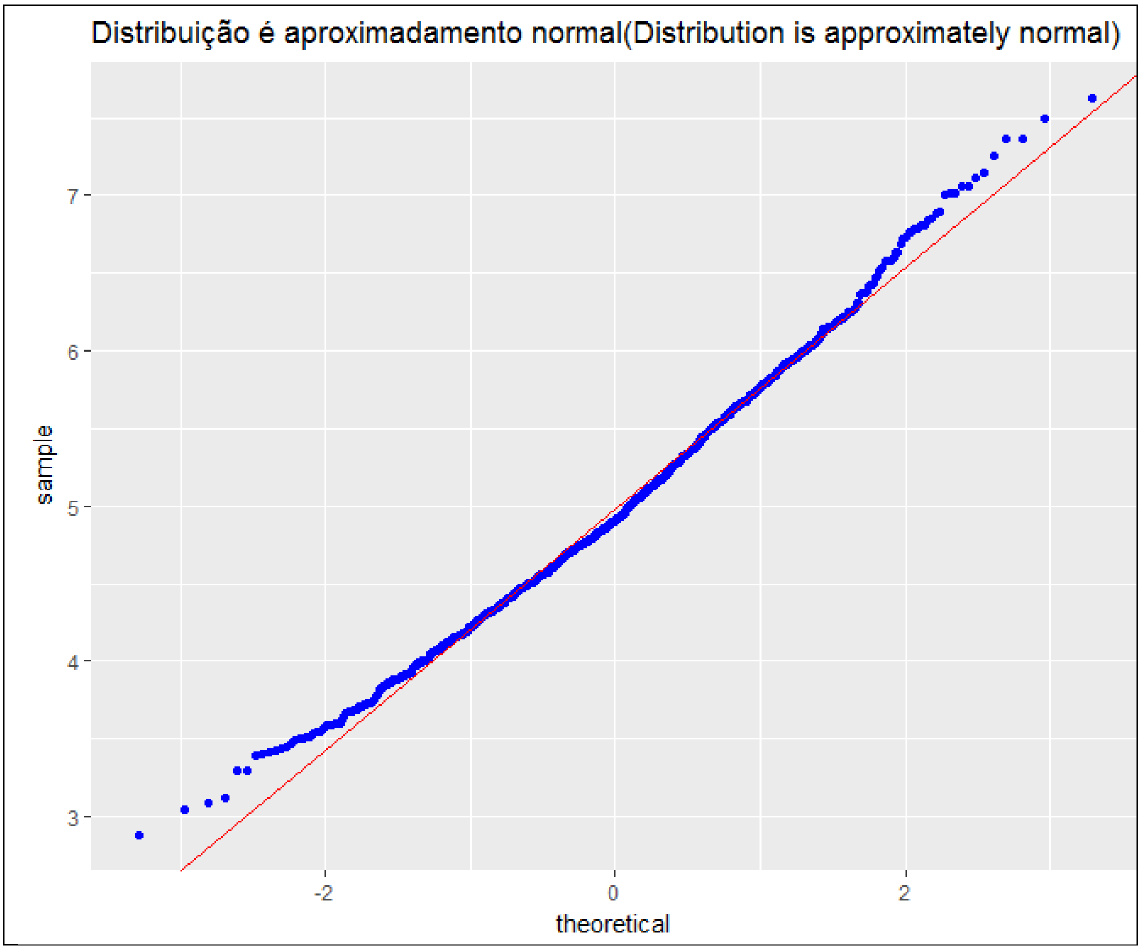
3- Show that distribution is approximately normal.

R CODE

4-

```
qqplot.data <- function (vetor)
{
  y <- quantile(vetor[!is.na(vetor)], c(0.25, 0.75))
  x <- qnorm(c(0.25, 0.75))
  slope <- diff(y)/diff(x)
  int <- y[1L] - slope * x[1L]
   d <- data.frame(dado = vetor)
   ggplot(d, aes(sample = dado)) + stat_qq(col="blue") +
geom_abline(slope = slope, intercept = int, col="Red")

}
qqplot.data (media) +ggtitle ("Distribuição é aproximadamento
normal(Distribution is approximately normal)")
```


Distribuição é aproximadamento normal(Distribution is approximately normal)
sample
theoretical