Applying the Central Limit Theorem to the Exponential Distribution in R

Mykola Herasymovych

12.08.2015

Overview

The project concerns applying the Central Limit Theorem to the simulated sample of the exponential distribution, e.g. comparing computed and theoreticaly calculated mean, variance and density of the distribution. The aim of the project is to prove the fairness of applying the Central Limit Theorem to real cases.

1. Simulation of samples

First of all, we set the fixed values to lambda, sample size and number of simulations as noted in the instruction to the project.

```
# define fixed values for analysis
set.seed(1)
lambda <- 0.2
sample_size <- 40
simulations <- 1000
```

Next, we produce 1000 simulations of samples from the exponential distriburion 40 observation each and calculate averages of each simulation.

```
# do 1000 simulations
simulated_expos <- matrix(rexp(simulations * sample_size, rate = lambda), simulations, sample_size)
averages <- rowMeans(simulated_expos)
str(averages)
```

```
##  num [1:1000] 4.9 5.23 6.4 4.74 5.18 ...
```

2. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

To compare produced and theoretical centers of distribution we calculate the first using R "mean" function and the second using the formula. In order to get even more precise picture we draw a histogram of our averages samle and compare its density with the theoretical one.

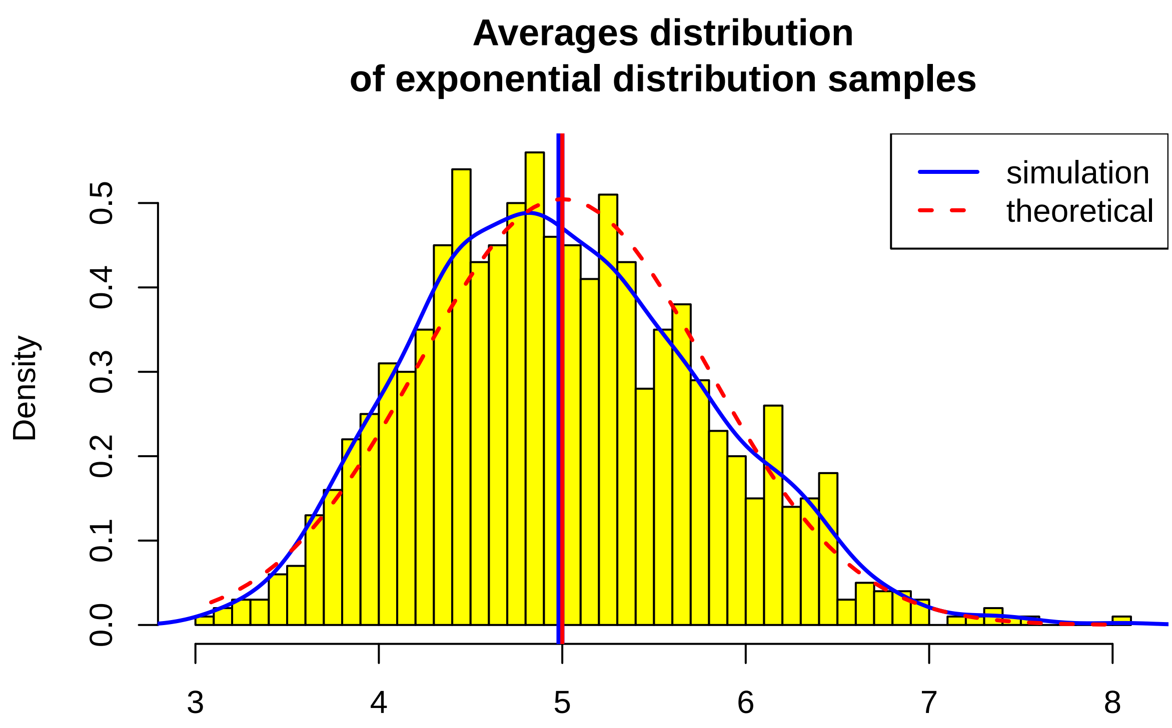
```
# calculate mean of distribution of averages of 40 exponentials
mean(averages)
```

```
## [1] 4.990025
```

```r
# calculate theoretical mean of distribution
1 / lambda
```

```
## [1] 5
```

```r
# draw a histogram of averages
hist(averages, breaks=50, prob=TRUE, col = "yellow",
     main="Averages distribution\nof exponential distribution samples",
     xlab="")
# density of the averages
lines(density(averages), col = "blue", lwd = 2)
# mean of averages of distribution
abline(v=mean(averages), col="blue", lwd = 4)
# theoretical center of distribution
abline(v=1/lambda, col="red", lwd = 2)
# theoretical density of the averages of samples
xfit <- seq(min(averages), max(averages), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sample_size)))
lines(xfit, yfit, pch=22, col="red", lty=2, lwd = 2)
# add legend
legend("topright", c("simulation", "theoretical"),
       lty=c(1,2), lwd = c(2, 2), col=c("blue", "red"))
```



Thus, produced and theoretical centers of distribution are almost equal: 4.99 and 5. Moreover, on the graph we can see that the simulated distribution's density almost fits the normal one.

3. Show how variable it is and compare it to the theoretical variance of the distribution.

For this matter we will compute simulation sample's variance using R "var" command and compare it to the theoretical one, calculated using the formula.

```
# calculate variance of distribution of averages of 40 exponentials
var(averages)
```

```
## [1] 0.6177072
```

```
# calculate theoretical variance of distribution
(1 / lambda) ^ 2 / sample_size
```

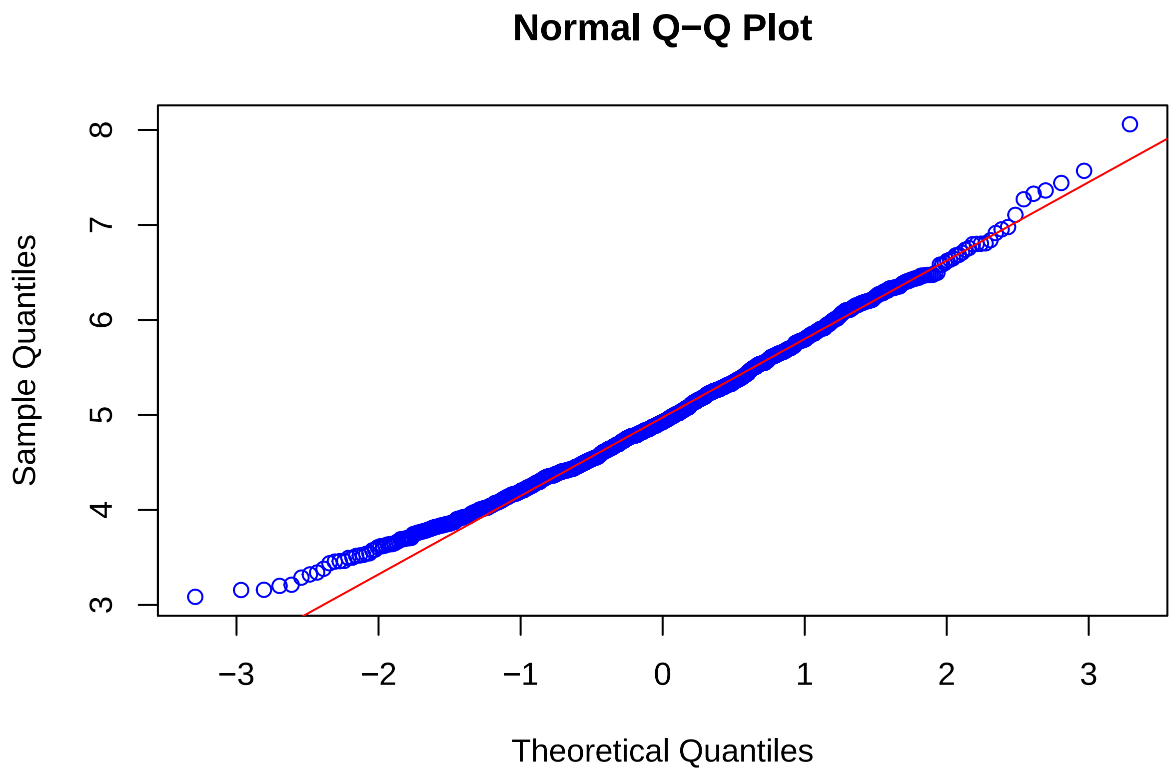
```
## [1] 0.625
```

Consequently, calculated variances are also quite similar: 0.6177 and 0.625.

4. Show that the distribution is approximately normal.

For this purpose we compare the density computed using the histogram and the normal density plotted with theoretical mean and variance values using Q-Q Plot.

```
# compare the computed distribution density and the normal density
qqnorm(averages, col = "blue"); qqline(averages, col = "red")
```



The plot shows that the sample quantiles fit the theoretical ones enough to make conclusion that the simulated distribution is approximately normal.