A simulation exercise

EvaLiang

2019.12.24

Overview:In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. I will investigate the distribution of averages of 40 exponentials.

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

```
# number of simulations
num <- 1000
# sample size
sampleSize <- 40
# rate parameter lambda
set.seed(19)
simMat <- matrix(rexp(n = num * sampleSize, rate = 0.2), num, sampleSize)
simMeans <- rowMeans(simMat)

sample_mean <- 1/0.2
theoratical_mean <- mean(simMeans)
c(sample_mean,theoratical_mean)
```

```
## [1] 5.000000 4.991311
```

The sample mean is 4.991311 The theoritical mean of the distribution is 5.000000 In this case, the sample mean and the theoritical mean is very close!

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

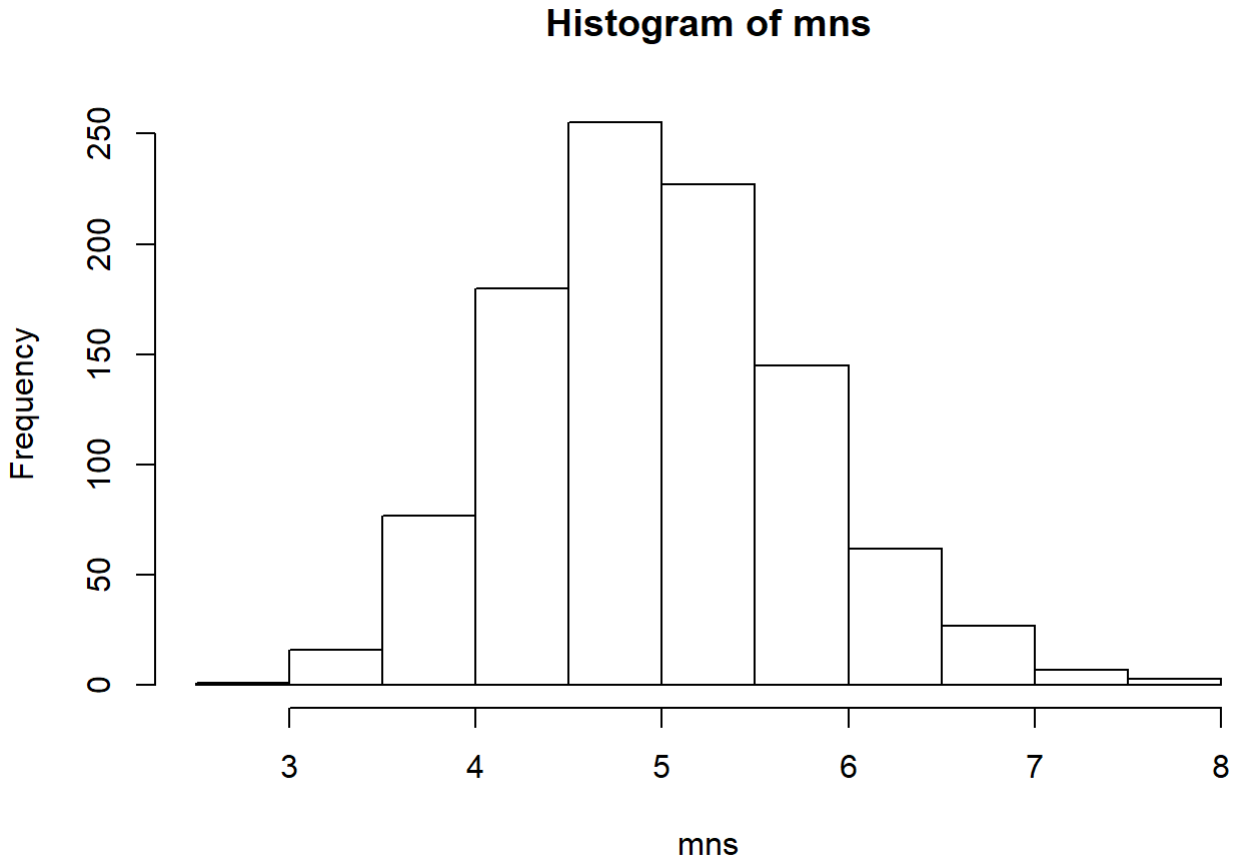
```
sample_var <- (1/0.2)^2/ sampleSize
theoratical_var <- var(simMeans)
c(sample_var,theoratical_var)
```

```
## [1] 0.6250000 0.6435493
```

The sample variance is 25.00000 The theoritical variance of the distribution is 13.38 In this case, the sample variance and the theoritical variance of the distribution is very close!

3. Show that the distribution is approximately normal.

```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(40, 0.2)))
hist(mns)
```



Judging from the plot, this distribution looks very Gaussian with the mean of approximately 5.

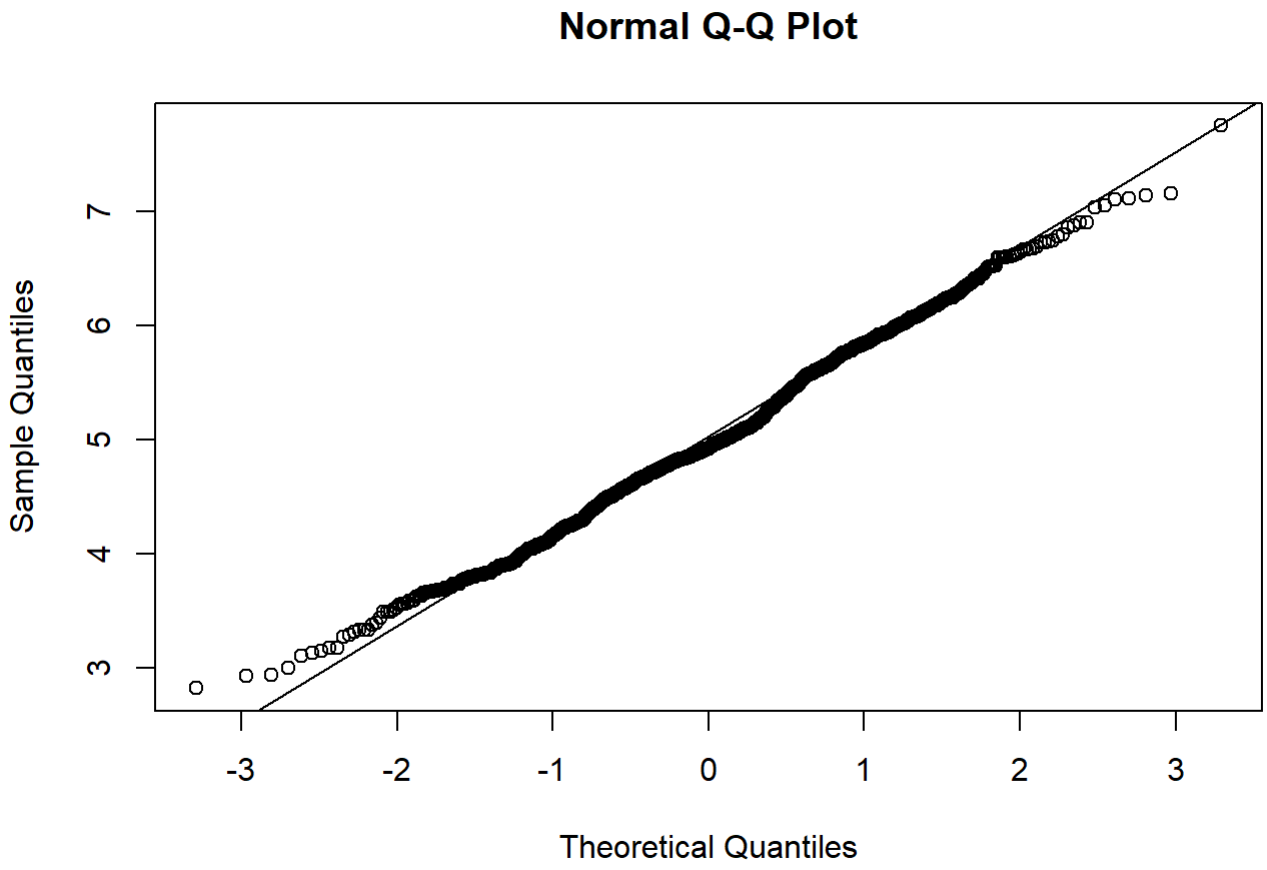
Compare the 95% con???dence intervals of the theoretical Normal distribution and simulated distribution

```
sampleCI <- simMeans + c(-1,1) * 1.96 * sqrt(sample_var) / sqrt(sampleSize)
theoCI <- theoratical_mean + c(-1,1) * 1.96 * sqrt(theoratical_var) / sqrt(sampleSize)
rbind(mean(sampleCI), theoCI)
```

```
##            [,1]     [,2]
##        4.991311 4.991311
## theoCI 4.742702 5.239920
```

The 95% con???dence interval of the simulated data [4.991311, 4.991311] is very close to the theoretical con??? dence interval [4.742702, 5.239920].

```
library(ggplot2)
qqnorm(simMeans)
qqline(simMeans)
```



According to the Q-Q plot the simulated quantiles follow a strong linear pattern and approximate the theoretical Normal quantiles.

Above, we can conclude that the simulated distribution approximately follows the Normal distribution.