Statistical Inference Course Project - A simulation exercise (Part 1)

Andrey Fedyna

Synopsis

This is the first part of the project for the statistical inference course. In this project I investigate the exponential distribution in R and compare it with the Central Limit Theorem.

Problem Description

The exponential distribution can be simulated in R with **rexp(n, lambda)** where λ(lambda) is the rate parameter and **n** is the number of observations. 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.

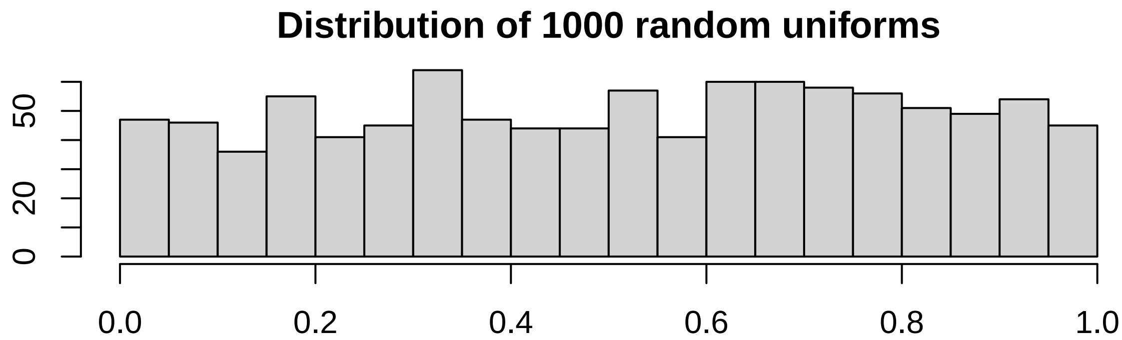
Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

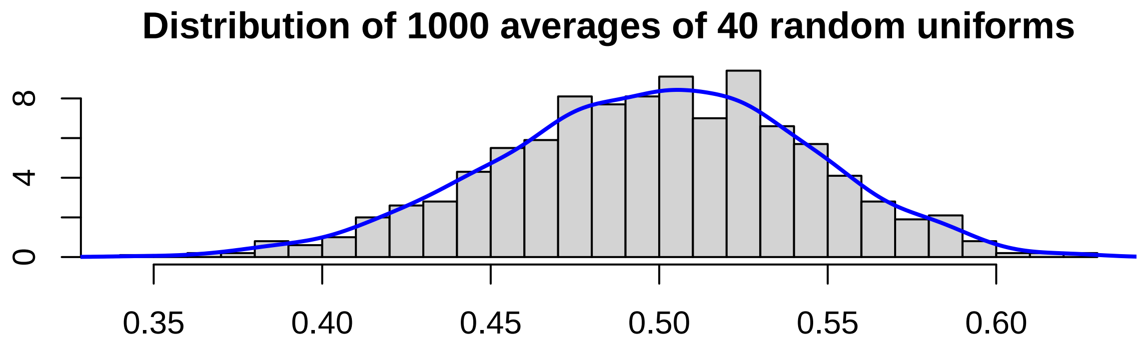
1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

Decision

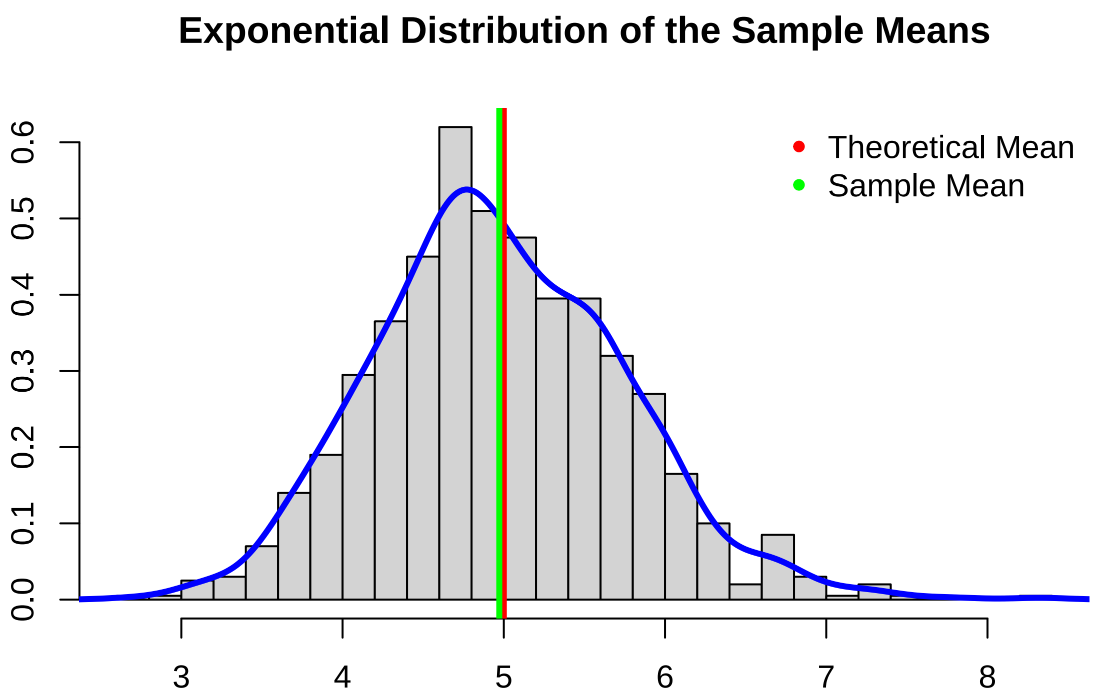
For example, compare the distribution of 1000 random uniforms and the distribution of 1000 averages of 40 random uniforms. Let's construct the histograms[(1)].





Build simulation

Next, we need to calculate and compare the theoretical and sample means, standart deviations and variances(2). Build the graphs "Exponential Distribution of the Sample Means"(3).



Answer for questions

value	Sample	Theory
mean	4.972	5
standard deviation	0.772	0.791
variance	0.595	0.625

Answer for Q1: Show the sample mean and compare it to the theoretical mean of the distribution.

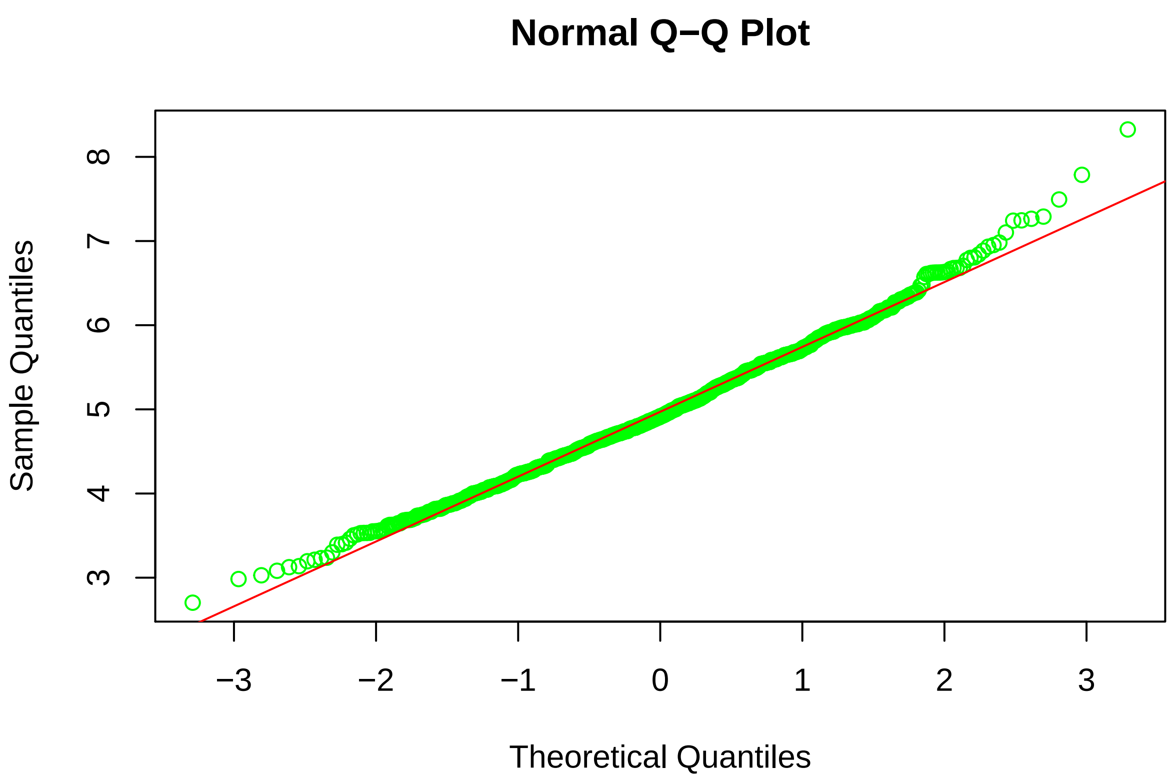
As seen from the graph and the computation above, the centre (mean) of the simulated sampling distribution of means, 4.972, almost the same as the calculated theoretical mean, 5.[(5)]

Answer for Q2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Selfsame, the standard deviation 0.772 and variance 0.595 are also very similar to the theoretically calculated standard deviation 0.791 and variance 0.625 respectively.[(5)]

Answer for Q3: Show that the distribution is approximately normal.

The distribution of the means of this simmulation follow a nearly perfect normal distribution as we can see on the graph above. Also we can confirme aproximate normality of this distribution by the normality test depicted in the Q-Q plot[(4)].



From the normal Q-Q plot, we can see that the diffrence between simulation and theoretical line is small, so we can say that the distribution is approximately normal.

Appendix: some calculations (i)

(1)

```
set.seed(12345)
par(mfrow=c(2,1), mar=c(4,4,2,1))

hist(runif(1000), breaks=30, col = "light grey",
     main = "Distribution of 1000 random uniforms", xlab = "", ylab = "")

mns = NULL
for (i in 1:1000) mns = c(mns, mean(runif(40)))

hist(mns, breaks=30, freq=FALSE, col = "light grey",
     main = "Distribution of 1000 averages of 40 random uniforms", xlab = "", ylab = "")
lines(density(mns), col="blue", lwd=2)
```

(2)

```
set.seed(12345)
lambda <- 0.2
nosim <- 1:1000 # Number of simulations
n <- 40

simulation = NULL
for (i in nosim) simulation = c(simulation, mean(rexp(n,lambda)))

sample_mean <- mean(simulation)
theory_mean <- 1/lambda
sample_sd <- sd(simulation)
theory_sd <- 1/lambda/sqrt(n)
sample_var <- sample_sd^2
theory_var <- theory_sd^2
```

(3)

```
hist(simulation, breaks=30, freq=FALSE, col = "light grey",
     main = "Exponential Distribution of the Sample Means", xlab = "", ylab = "")
lines(density(simulation), col="blue", lwd=3)
abline(v=theory_mean, col="red", lwd=3)
abline(v=sample_mean, col="green", lwd=3)
legend("topright", pch = 20, col = c("red","green"),
       legend = c("Theoretical Mean", "Sample Mean"), bty = "n")
```

(4)

```
qqnorm(simulation, col = "green")
qqline(simulation, col = "red")
```

(5)

```r
round(sample_mean,3)
```

```
## [1] 4.972
```

```r
round(theory_mean,3)
```

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

```r
round(sample_sd,3)
```

```
## [1] 0.772
```

```r
round(theory_sd,3)
```

```
## [1] 0.791
```

```r
round(sample_var,3)
```

```
## [1] 0.595
```

```r
round(theory_var,3)
```

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