Simulation Exploration Excercise

Michael Lagunov

8/15/2020

Overview

In this project I would like to show how can the one use simulations to estimate theoretical mean and standard deviation. Also, I will demonstrate appliance of Central Limit Theorem

Data setting

```
lambda <- 0.2
num <- 40
simnum <- 1000
```

Make a matrix with "simnum" = 1000 rows and "num" = 40 columns, and then get an average across all rows, using "apply" function

```
expData <- matrix(rexp(num*simnum, lambda), simnum)
expDataMean <- apply(expData, 1, mean)
```

Point 1. Sample mean and comparing it to the theoretical mean of the distribution

```
simMean <- mean(expDataMean)
theoreticalMean <- 1/lambda
cat("Simulation mean: ", simMean, "\n", "Theoretical mean: ",theoreticalMean, sep = "")
```

```
## Simulation mean: 4.984655
## Theoretical mean: 5
```

The mean from the simulation is pretty close to the theoretical mean

Point 2. How variable the sample is (via variance) and compare it to the theoretical variance of the distribution?

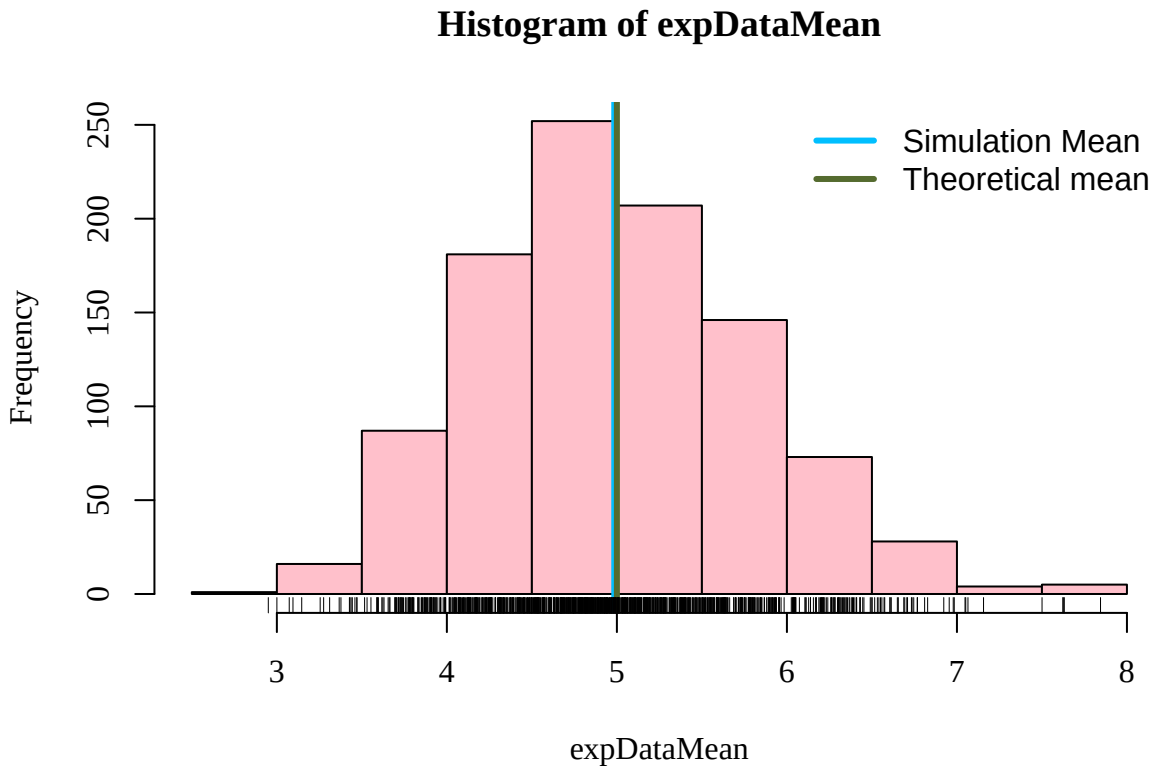
```
simSd <- sd(expDataMean)
theoreticalSd <- 1/lambda

cat("Simulation sd: ", simSd, "\n", "Theoretical sd:", theoreticalSd, sep = "")
```

```
## Simulation sd: 0.7975821
## Theoretical sd:5
```

The sd from the simulation is near to the theoretical sd, but not quite near. It happened mostly because of the low simulation number. The more of them is used, the closer it will be to the theoretical one.

Point 3. Show that the distribution is approximately normal

```
hist(expDataMean, col = "pink", family = "serif")
abline(v = simMean, lwd = 3, col = "deepskyblue")
abline(v = theoreticalMean, lwd = 3, col = "darkolivegreen")
legend("topright", legend = c("Simulation Mean", "Theoretical mean"),
       col = c("deepskyblue", "darkolivegreen"), lty = 1, lwd = 3,
       box.lty = 0)
rug(expDataMean)
```



The distribution really seems to be normal. It happens due to the CLT

Conclusion

In this project I demonstrated, that the mean and sd from the simulation are quite close to the theoretical. Also, there were provided a histogram of the means, that follows normal distribution