Statistical Inference Course Project. Simulation Exercise

Mariia Danilenko

Introduction

In this project we will make some research on the exponential distribution in R and compare it with the Central Limit Theorem. We will investigate the distribution of averages of 40 exponentials. We will need to do a thousand simulations.
Here and further fix $\lambda = 0.2$

Analysis

Let's simulate 1000 averages of 40 exponential iid values.
Also, calculate mean for each of them.

```
 n<-1000
for (i in 1:n) expDistSimul[i,]<-rexp(40,lambda)
for (j in 1:n) exp_means[j] <- mean(expDistSimul[j,])
```

Based on the properties of the mean theoretical distribusion mean would be:

$$E(E(\frac{1}{40}\sum_{i=1}^{40}\chi_{exp,i})) = E(\frac{1}{40}\sum_{i=1}^{40}E(\chi_{exp,i})) = E(\frac{1}{40} * 40 * \frac{1}{\lambda}) = \frac{1}{\lambda}$$

In this case mean equals 5. Compare simulation mean with theoretical:

```
t.test(exp_means, alternative = "two.sided", mu=1/lambda)
```

```
##
##  One Sample t-test
##
## data:  exp_means
## t = 0.32303, df = 999, p-value = 0.7467
## alternative hypothesis: true mean is not equal to 5
## 95 percent confidence interval:
##  4.957603 5.059106
## sample estimates:
## mean of x
##  5.008354
```

We failed to reject our hypothesis about mean.
Speaking about sd, we apply **Lindeberg–Lévy CLT** to our model:

$$\sqrt{40}((\frac{1}{40}\sum_{i=1}^{40}\chi_{exp,i}) - \frac{1}{\lambda}) \to N(0, \sigma^2)$$

$$\sqrt{40} * (mean(rexp(40, \lambda)) - \frac{1}{\lambda}) \to N(0, \frac{1}{\lambda}) \Rightarrow sd(mean(rexp(40, \lambda)) \to \frac{1}{\lambda\sqrt{40}} = 0.79$$

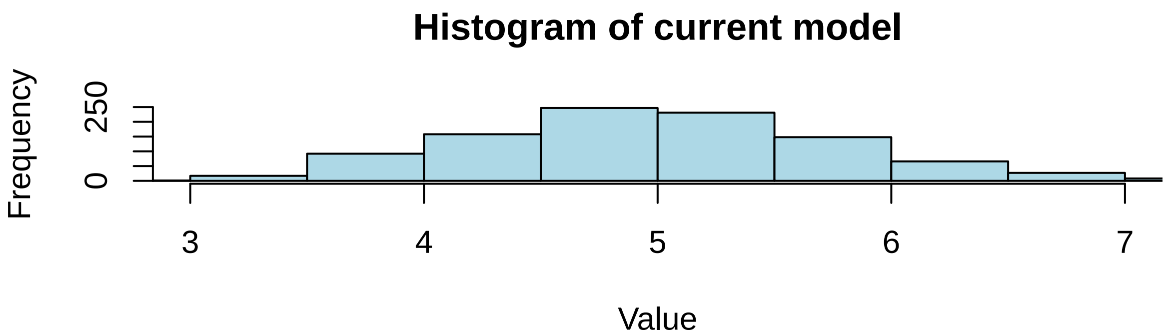
We want to check variances for hypothesis that final distribution almost $N(\frac{1}{\lambda}, \frac{1}{\lambda\sqrt{40}}) = N(5, 0.79)$. Let's check variance:

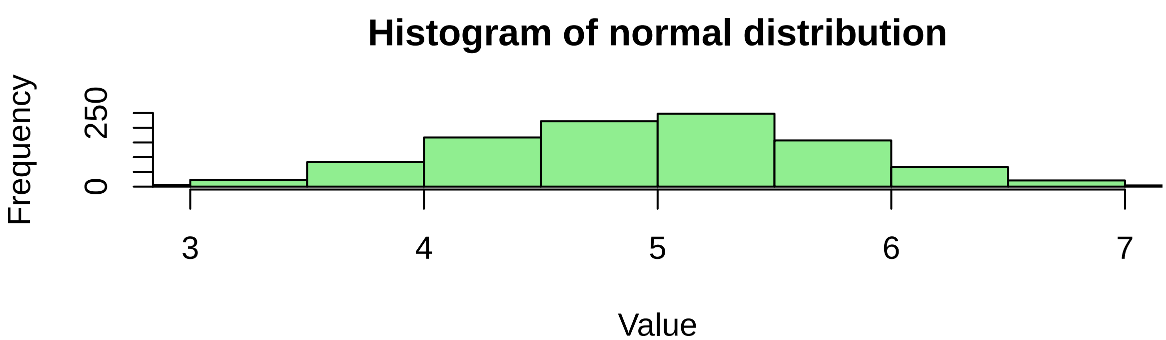
```r
var.test(exp_means,rnorm(1000,5,0.79))
```

```
##
##  F test to compare two variances
##
## data:  exp_means and rnorm(1000, 5, 0.79)
## F = 1.0986, num df = 999, denom df = 999, p-value = 0.1374
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.970387 1.243740
## sample estimates:
## ratio of variances
##           1.098594
```

We failed to reject out hypothesis.
Let's plot our distribution and 1000 instances $N(5, 0.79)$.

```r
par(mfcol=c(2,1))
hist(exp_means, col="lightblue",xlim = c(3,7),
     xlab="Value", main = "Histogram of current model")
hist(rnorm(1000,5,0.79),xlim = c(3,7), col="lightgreen",
     xlab="Value", main = "Histogram of normal distribution")
```





Conclusion

Our investigation coincides with CLT and final distribution is $N(5, 0.79)$