Inferencial Statistics Project

Irina White

18/04/2021

Part 1: Simulation Exercise

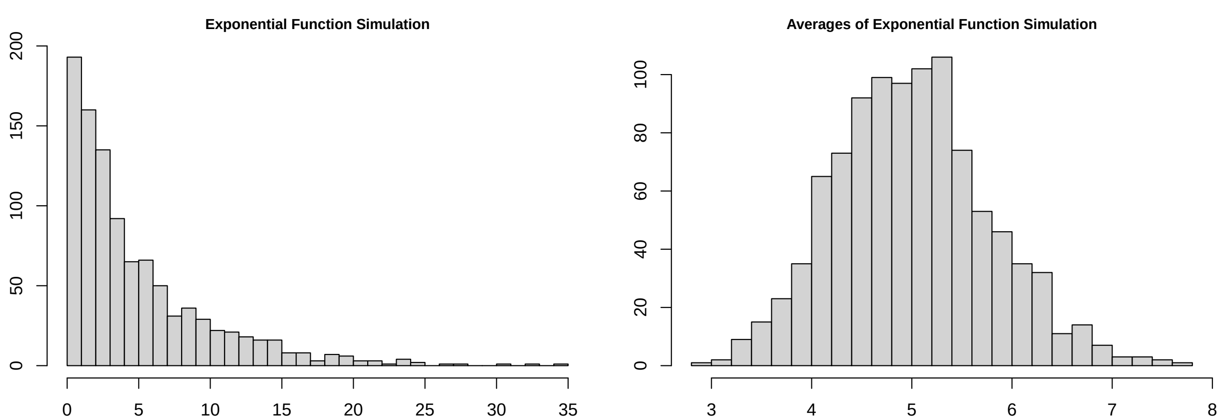
The goal is to investigate the exponential distribution and compare it with the distribution of averages, Central Limit Theorem.

```
#Set-up given parameters:
set.seed(4321)
options(digits=3)
n<-40
lambda <- 0.2
sim<-1:1000
```

Histograms

The exponential distribution vs. distribution of averages of the exponential distribution. Note that the form of the distribution of averages closely follows that of the normal distribution.

```
exp.distr<-rexp(1000, lambda)
av.exp.distr = NULL
for (i in sim) av.exp.distr = c(av.exp.distr, mean(rexp(n,lambda)))
par(mar=c(2,2,2,2), mfrow=c(1,2))
hist(exp.distr, breaks=30, main = "Exponential Function Simulation", xlab="", cex.main=0.8)
hist(av.exp.distr, breaks=30, main = "Averages of Exponential Function Simulation", xlab="", cex.main=0
```

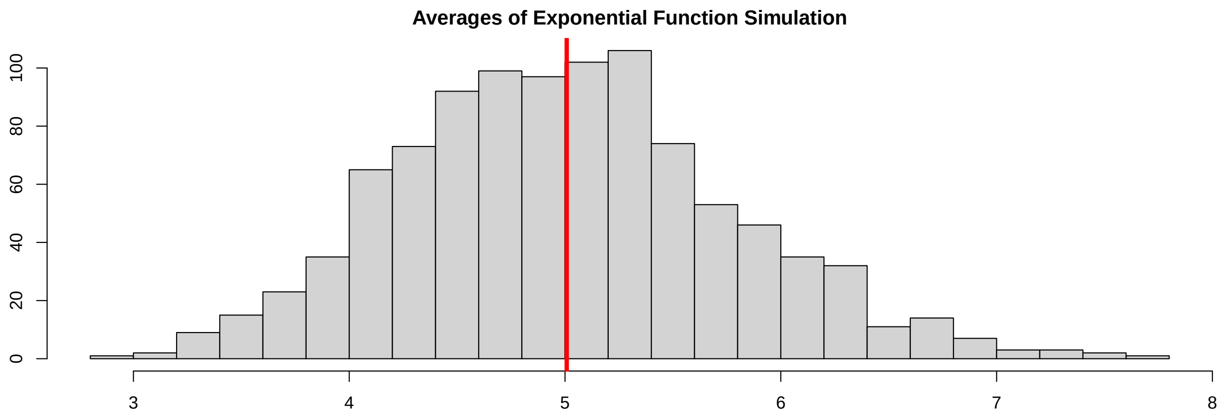


Compare the mean of the sample with that of the averages.

```r
mean_sample = 1/lambda
mean_averages = mean(av.exp.distr)
cbind(mean_sample, mean_averages)
```

```
##      mean_sample mean_averages
## [1,]           5          5.01
```

```r
par(mar=c(2,2,2,2))
hist(av.exp.distr, breaks=30, main = "Averages of Exponential Function Simulation", xlab="")
abline(v=mean_averages,col="red", lwd=4)
```




```r
t.test(av.exp.distr)[4]
```

```
## $conf.int
## [1] 4.96 5.06
## attr(,"conf.level")
## [1] 0.95
```

Results The mean of the averages is between 4.96 and 5.06 with 95% confidence interval (t-test).

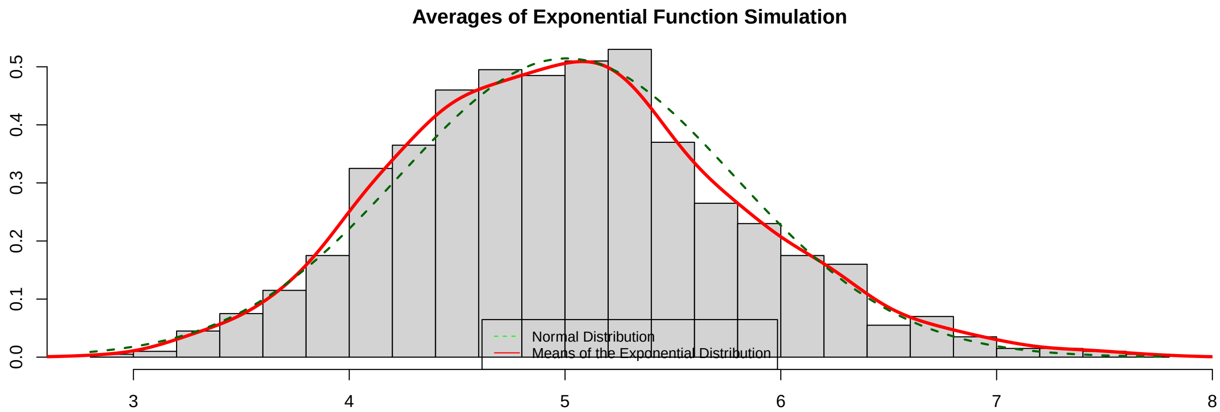
Compare the variance of the sample with that of the averages.

```r
var_sample = (1/lambda)^2/n
var_averages = var(av.exp.distr)
cbind(var_sample, var_averages)
```

```
##      var_sample var_averages
## [1,]      0.625        0.601
```

Outcome shows that variance of the averages data is very close to the variance of the population.

```r
par(mar=c(2,2,2,2))
hist(av.exp.distr, breaks=30, main = "Averages of Exponential Function Simulation", xlab="", freq=FALSE)
lines(density(av.exp.distr), lwd=3, col="red")
curve(dnorm(x,mean=mean_averages,sd=sd(av.exp.distr)), add=TRUE,col="darkgreen", lwd=2, lty=2)
legend("bottom", legend=c("Normal Distribution", "Means of the Exponential Distribution"),
       col=c("green", "red"), lty=2:1, cex=0.8)
```



Conclusion

Through the above simulation and analysis it can be clearly observed that the distribution of the mean of the exponential distribution (see red line on the graph) follows the Central LImit Theorem.In another words, it closely follows the normal distribution (see green dotted line on the graph. If we are to increase the the number of simulations dramatically, both lines will overlap.