Statistical Inference

Francois Chevallier

Overview:

In this study, we are going to analyse the exponential distribution and compare it to the Central Limit Theorem.

The Central Limit Theorem (or CLT) states that the sampling distribution of the mean of any independent random variable will be (nearly) normal.

Through this analysis, we will simulate the exponential distribution in R with the following parameters:

- n
- lambda

Study:

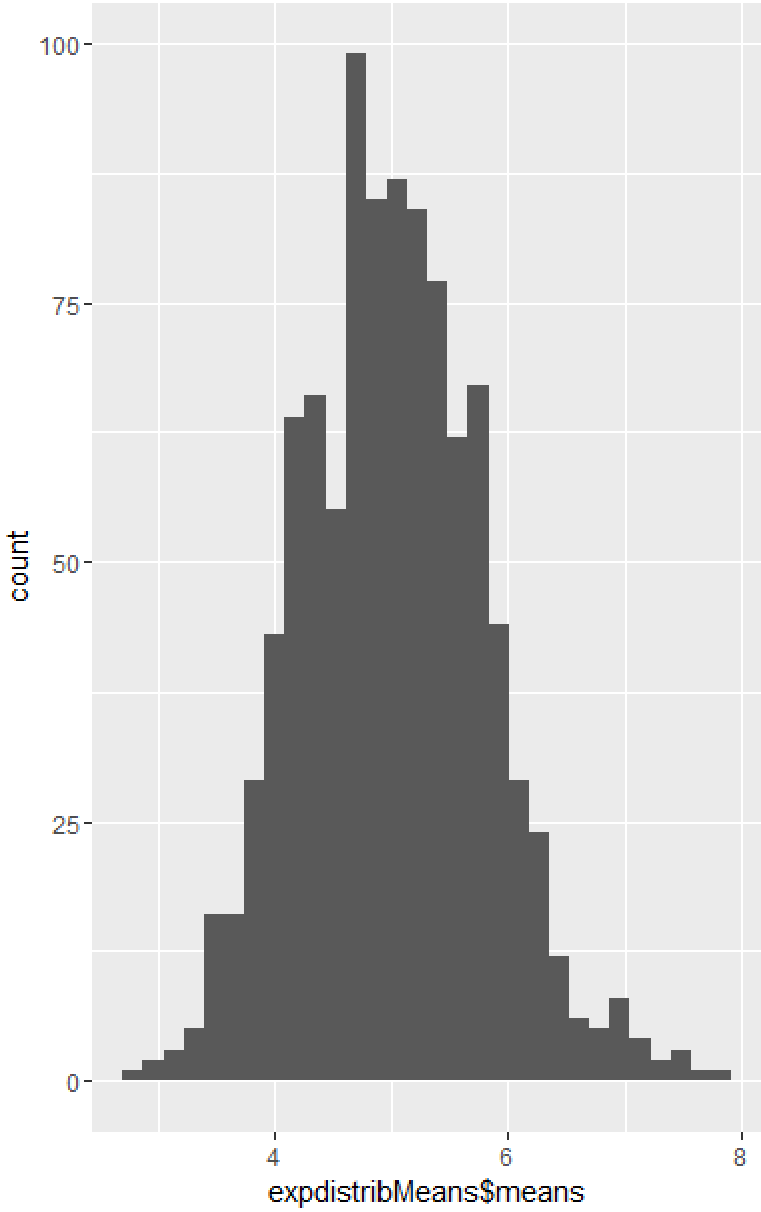
We set up the parameter:

```
library(ggplot2)
set.seed(123)

lambda <- 0.2
n <- 40
simulationNumber <- 1000
expdistrib <- matrix(data=rexp(n*simulationNumber, lambda), nrow=simulationNumber)
expdistribMeans <- data.frame(means=apply(expdistrib, 1, mean))
```

Let’s represent the exponential distribution:

```
qplot(expdistribMeans$means, geom="histogram")
```



The theoretical mean value is:
mu <- 1/lambda
Mu
5

The mean found is:
themean <- mean(expdistribMeans$means)
themean
5.011911

The theoretical standard deviation is:

```
sd <- mu/sqrt(n)
sd
0.7905694
```

The practical standard deviation is:

```
sdX <- sd(expdistribMeans$means)
sdX
0.7802751
```

The variance is:

```
variance <- sd^2
variance
0.625
```

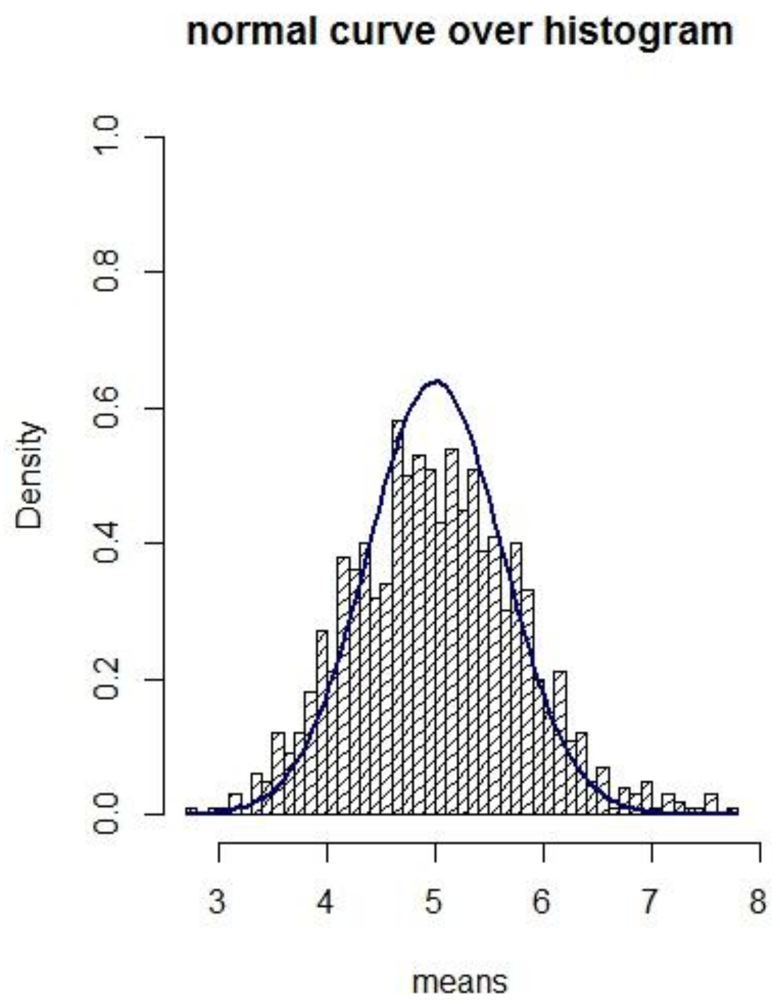
The practical variance is:

```
varX <- var(expdistribMeans$means)
varX
0.6088292
```

As the figures can tell, the CLT seems to be true for n=1000.

We could finally represent the normal and exponential distribution on the same plot:

```
g = expdistribMeans$means
m=5
std=0.625
hist(g, density=20, breaks=20, prob=TRUE,
+      xlab="means", ylim=c(0, 1),
+      main="normal curve over histogram")
curve(dnorm(x, mean=m, sd=std),
+       col="darkblue", lwd=2, add=TRUE, yaxt="n")
```


normal curve over histogram
Density
0.0
0.2
0.4
0.6
0.8
1.0
3
4
5
6
7
8
means