Part 1: Simulation Exercise

Clément Brutti-Mairesse

26/04/2020

Overview

In this project we investigate the exponential distribution in R and compare it with the Central Limit Theorem. Here is the probability density function of an exponential distribution

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

The pdf of the exponential distribution only depends on λ, commonly called the *rate parameter*. The exponential distribution has a mean μ equals to $\frac{1}{\lambda}$ and a standard deviation σ of $\frac{1}{\lambda}$.

Simulations

Here, we simulate 1000 averages of random exponentials with the `rexp()` function given n = 40 and rate $\lambda =$ 0.2.

```
set.seed(43)
lambda = 0.2
sampleSize = 40
simulationCount = 1000

means = NULL
for (i in 1 : simulationCount) means = c(means, mean(rexp(sampleSize, rate = lambda)))
```

Sample Mean versus Theoretical Mean

Sample mean is noted $\bar{X}$

```
sampleMean = mean(means)
```

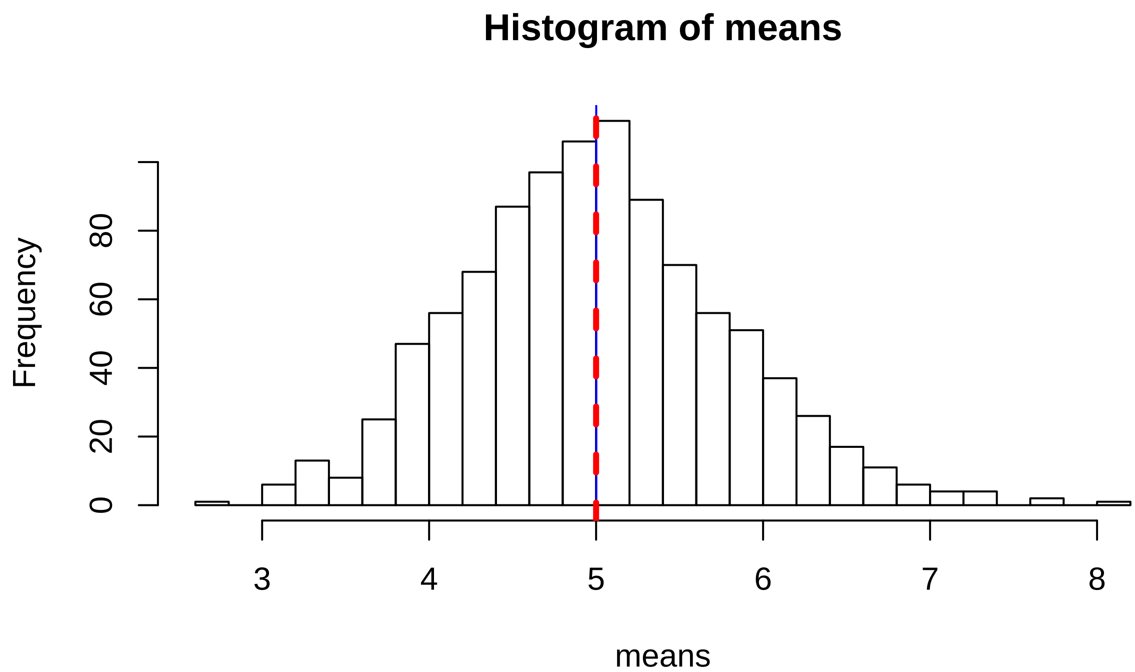
```
## [1] 5.000798
```

Theorical mean is $\frac{1}{\lambda}$

```
thoericalMean = 1 / lambda
```

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

In the following graph the theorical mean is painted in red, and the sample mean in blue.



Sample Variance versus Theoretical Variance

Sample variance is noted S^2

```
sampleVariance = var(means)
```

```
## [1] 0.6269514
```

Theorical variance is $\sigma^2 = \frac{1}{n\lambda^2}$

```
thoericalVariance = 1 / (sampleSize * lambda^2)
```

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

Distribution of the means

$\bar{X}$ is approximately normal with mean $\mu = \frac{1}{\lambda}$ and standard deviation $\frac{\sigma}{\sqrt{n}} = \frac{1}{\lambda\sqrt{n}}$.

The probability $\bar{X}$ being within $\frac{1}{\lambda} \pm \frac{2}{\lambda\sqrt{n}}$ is 95%.

```
confidenceInterval <- (1/lambda) + (c(-1, 1) * qnorm(0.975) * (1/lambda) / sqrt(sampleSize))
outOfInterval <- sum(means < confidenceInterval[1] | means > confidenceInterval[2])
print(confidenceInterval)
```

```
## [1] 3.450512 6.549488
```

```
paste(outOfInterval * 100.0 / simulationCount, "%")
```

```
## [1] "5.4 %"
```

As 5.4% of the averages are outside the 95% confidence interval, we consider that this distribution is approximately normal.

Here, we illustrate the sample distribution of the means with an histogram, the curve representing the theorical distribution, and the 95% confidence interval delimited by the two vertical dash lines.

```
h <- hist(means, breaks = 20)
xfit <- seq(min(means),max(means),length=simulationCount)
yfit <- dnorm(xfit, mean = 1 / lambda, sd = 1 / (lambda*sqrt(sampleSize)))
yfit <- yfit * diff(h$mids[1:2]) * simulationCount
lines(xfit, yfit, col="blue", lwd=2)
abline(v=c(confidenceInterval[1],
           confidenceInterval[2]),
       col=c("blue", "red"),
       lty=c(2,2), lwd=c(2, 2))
```