Central Limit Theorem and the Exponential Distribution

Riccardo Finotello

19 June 2020

Overview

We consider the exponential distribution and compare it with the Central Limit Theorem (CLT): we perform several simulations to extract samples and average a number of them to show that the distribution of the mean is approximately a normal distribution.

Simulations

In this simulation we will use exponential distributions with `lambda = 0.2` (the *rate* parameter):

```
lambda <- 0.2
```

We will then define the number of exponentials simulating draws from the distribution:

```
n.exp <- 40   # no. of exponentials
n.sim <- 1000 # no. of simulations
```

We then set the seed for reproducible results:

```
set.seed(42)
```

We then proceed to the simulation of the distributions:

```
means <- NULL
for (i in 1 : n.sim) {means = c(means, mean(rexp(n = n.exp, rate = lambda)))}
```

As a reference we also consider 1000 samples coming from one of the distributions. Instead of distributions of samples, we consider one distribution with samples:

```
exp.dist <- rexp(n = n.sim, rate = lambda)
```

We then plot the comparison as a preliminary graphic showing the key differences of the distributions in Figure 1.

Sample Mean and Theoretical Mean

We now consider the comparison between the population mean and the expected value of the estimator for an exponential distribution:

```
mean.pop <- mean(means)
mean.exp <- 1 / lambda

print(paste("Population mean:", round(mean.pop,3)))
```

```
## [1] "Population mean: 4.987"
```

```
print(paste("Expected mean:", mean.exp))
```

```
## [1] "Expected mean: 5"
```

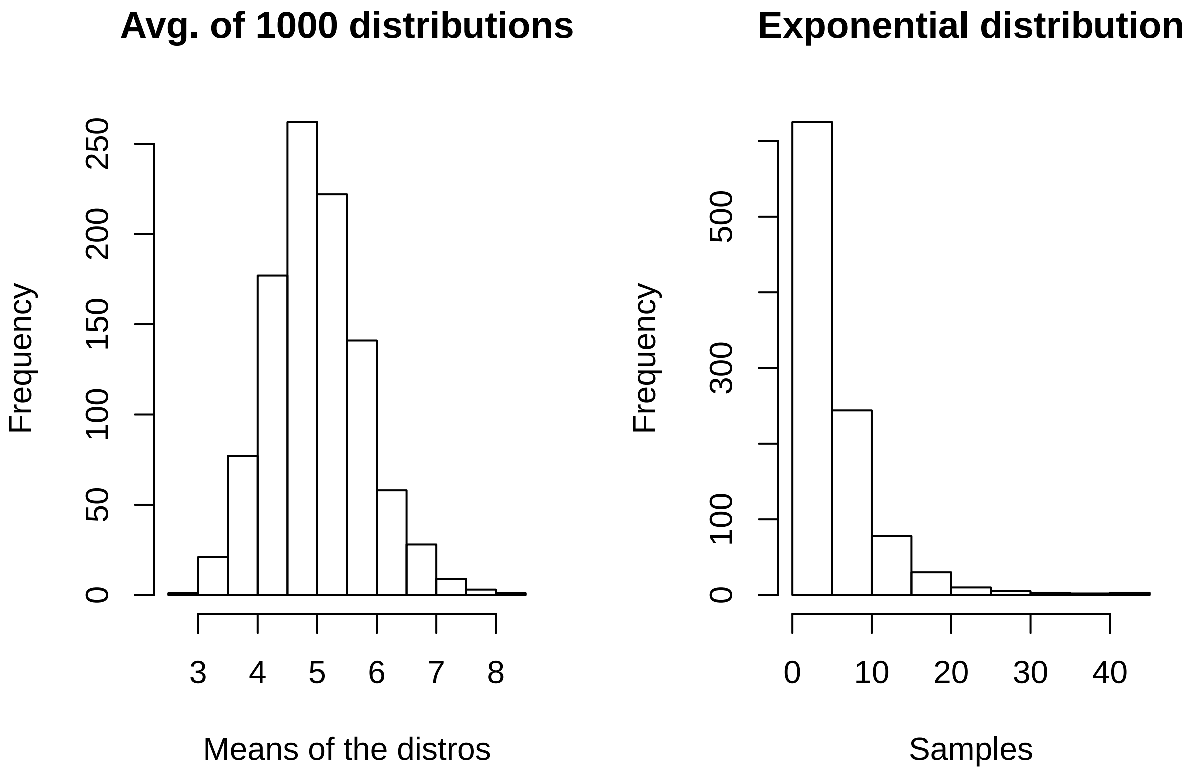

Figure 1: Difference between the distributions of 1000 exponential distribution holding 40 samples and one distribution with 1000 samples.

The results reflect the prediction of the CLT: the distribution of the averages will result in a normal distribution centred in the expected value of the mean.

Sample Variance and Theoretical Variance

We can then proceed and perform the same analysis for the standard deviation of the population and the expected estimator:

```
sd.pop <- sd(means)
sd.exp <- 1 / lambda

print(paste("Population standard deviation:", round(sd.pop,3)))
```

```
## [1] "Population standard deviation: 0.797"
```

```
print(paste("Exponential standard deviation:", sd.exp))
```

```
## [1] "Exponential standard deviation: 5"
```

The difference in the values should not be regarded as worrisome: the CLT ensures that the distribution of the averages tends to a normal distribution centred in the expected value of the mean and a standard deviation equal to the standard error (i.e. $\sigma = \frac{s}{\sqrt{n}}$, where σ is the standard error, s is the population standard deviation and n is the number of degrees of freedom in the distribution):

```
sd.err <- sd.exp / sqrt(n.exp)

print(paste("Population standard deviation:", round(sd.pop,3)))
```

```
## [1] "Population standard deviation: 0.797"
```

```
print(paste("Standard error:", round(sd.err,3)))
```

```
## [1] "Standard error: 0.791"
```

Again we then have "experimental" evidence of the validity of the CLT, since the population mean and the standard error associated to the normal distribution of the averages are extremely close.

Distribution

In this last section we provide a visual reference of the compatibility of the distribution of the averages with a normal distribution. In fact such property can be immediately visualised in the histogram in Figure 2. We first create the intervals of reference values for the normal distribution.

```
x <- seq(0.75*min(means), 1.25*max(means), by = 0.01) #---- create the x axis
y <- dnorm(x, mean = mean.exp, sd = sd.err) #-------------- create PDF function
```

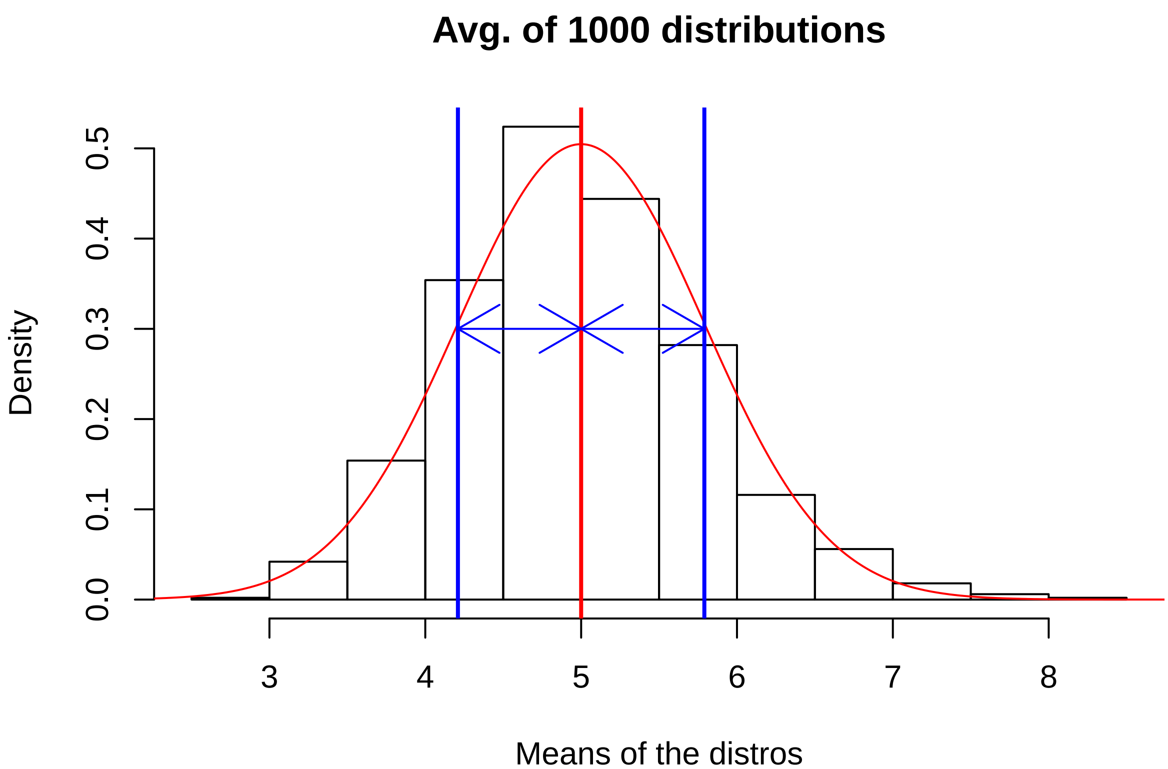

Figure 2: Comparison between the distribution of the averages and a normal distribution centered in the expected value of the mean estimator (shown in red) and a standard deviation equal to the standard error (shown in blue).