Demonstration of the Central Limit Theorem

Tanguy Levent

25 février 2017

Synopsis

The goal of this study is to demonstrate the Central Limit Theorem over an exponential distribution. First we plot the values of our dataset to validate that an exponential distribution is well simulate. Once exponential distribution validating, we compare the theoretical mean and variance with the sample mean and variance, which should be identical. Lastly we show that the mean tend toward a normal distribution.

Parameters

For the exponential distribution analysis the rate parameter lambda is equal to 0.2 with n = 40 and we will take 10 000 iterations for our simulation. To have a computation reproducible I declare the seed.set function.

```
set.seed(1000)
lambda <- 0.2
iteration <- 10000
n <- 40
```

Mean and variance : computation and comparaison

Step 1. The theoretical mean and variance

The computation is define as the original formulas.

```
theo_mean <- 1/lambda
theo_var <- 1/(lambda^2)/n
```

Step 2. The sample mean and variance

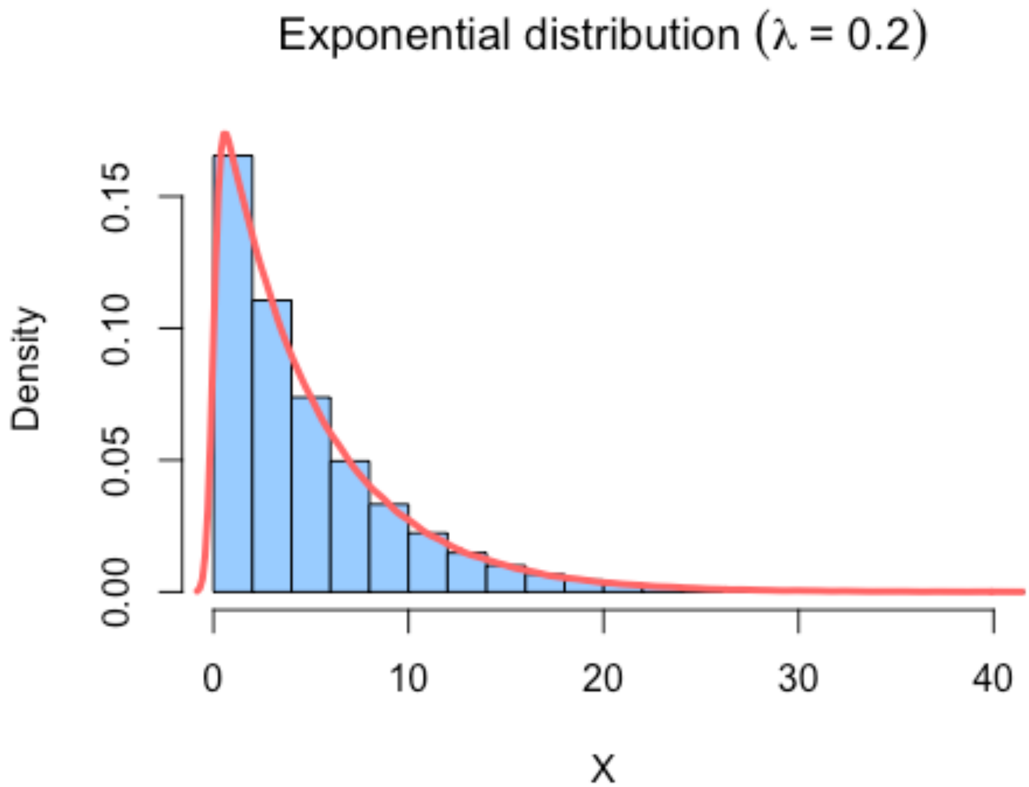
First we create an exponential distribution regarding our previous assumptions :

```
exp_distrib2 <- replicate(iteration,rexp(n, lambda))
dim (exp_distrib2)

## [1]    40 10000
```

The dimension of the variable is composed of 40 rows and 10 000 columns as we wanted. To validate the distribution as exponential, the best way is to plot our sample :

```
hist(exp_distrib2, breaks = 25, col = c('#99CCFF'), xlab ="X", main =
expression("Exponential distribution "(lambda *" = 0.2")), freq = FALSE ,
ylim = c(0,0.17), xlim = c(0,40))
lines(density(exp_distrib2), col = c('#FF6666'), lwd=3)
```



Once the sample validating as exponential, we want to calculate our sample mean and variance (before comparaison with the theoretical) :

```
colmean <- colMeans(exp_distrib2)

sample_mean <- mean(colmean)
sample_var <- var(colmean)
```

Step 3. The mean and variance : comparaison between the theoretical and the simulation

I used the print function to write our results :

```
print(paste("Theorical mean equal to", theo_mean, "and sample mean equal
to",round(sample_mean,3)))

## [1] "Theorical mean equal to 5 and sample mean equal to 4.98"
```

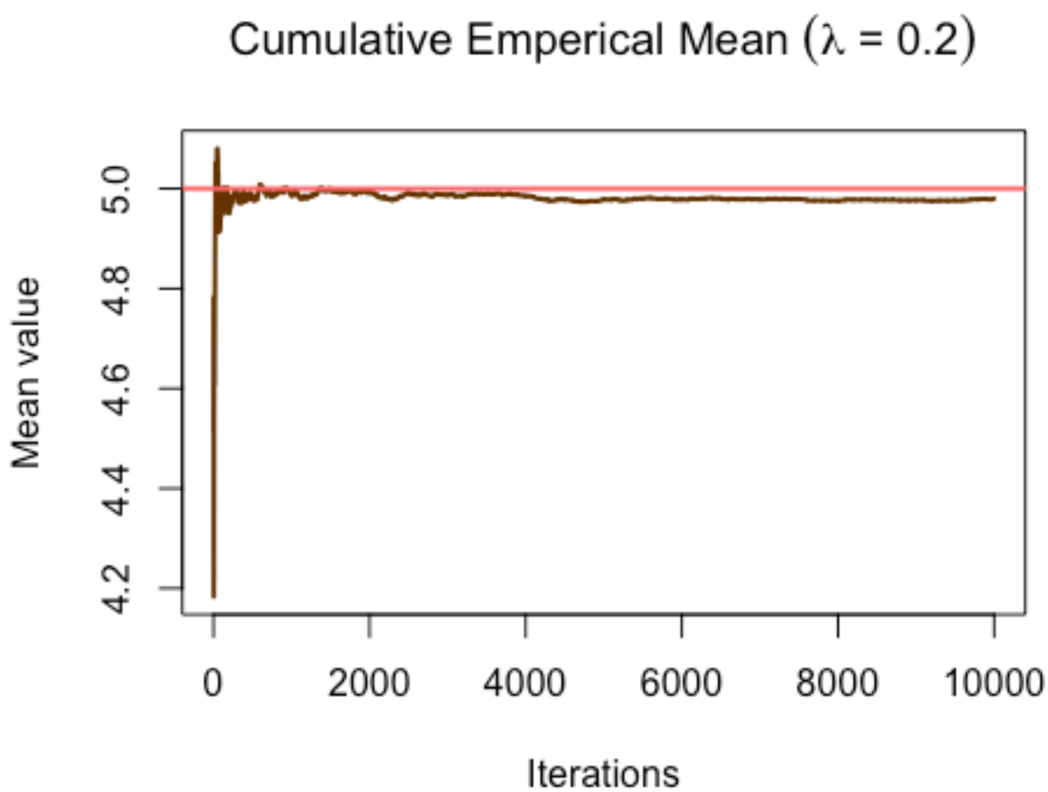
```r
print(paste("Theorical variance equal to", theo_var, "and sample variance
equal to",round(sample_var,3)))

## [1] "Theorical variance equal to 0.625 and sample variance equal to 0.639"
```

Step 4. Conclusion

The results of our computations give us a theoretical and practical values (mean and variance) almost even. To understand why, a good way is to draw the cumulative distribution, by the mean for example.

```r
cumul_mean <-cumsum(colmean)/seq_along(colmean)
cumum_var <- cumsum(var(colmean))
plot(cumul_mean, type="l", col = c('#663300'), lwd =2 ,xlab = "Iterations",
ylab = "Mean value", main = expression("Cumulative Emperical Mean " (lambda
*" = 0.2")))
abline(h=theo_mean, col = c('#FF6666'), lwd=2)
```

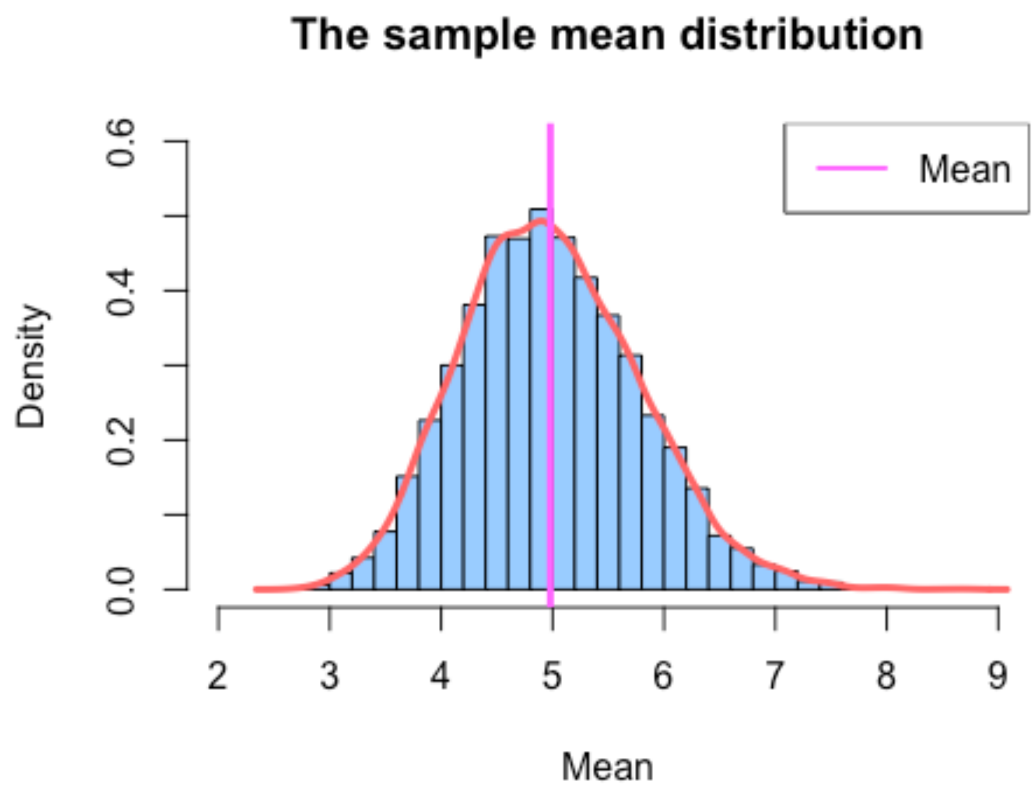


As you can see, our sample mean converge toward the theoretical mean (assuming a large # of iteration).

Does the scenarios tends toward a bell curve?

Below are draw the distribution of our sample mean :

```
hist(colmean, breaks = 25, freq = FALSE, xlim = c(2,9), ylim = c(0,0.6),col =
c('#99CCFF'), main = "The sample mean distribution", xlab = "Mean")
lines(density(colmean), col = c('#FF6666'), lwd=3)
abline(v=sample_mean, col = c('#FF66FF'), lwd=3)
legend("topright", legend = c("Mean"), col = c('#FF66FF'), lwd =2)
```



Conclusion

According to Wikipedia : *"In more precise terms, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined (finite) expected value and finite variance, will be approximately normally distributed, regardless of the underlying distribution. The theorem is a key concept in probability theory because it implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions."*

The more iterations we have, the more bell curve evidence is.