Statistical Inference Project - Part I

Exploration of the exponential distribution

Loïc BERTHOU

August 18, 2015

Overview

This document looks at the behaviour of the exponential distribution, specifically its mean distribution and variance distribution for a set of samples. The observations will help understand the Central Limit Theorem.

Introduction

For this study, we will investigate the distribution of averages of 40 exponentials that will be simulated 1000 times. The rate parameter λ is set to $\frac{1}{5}$ for all simulations.

Simulations

First, we generate the 1000 samples of 40 exponentials and store them in the variable *expMatrix*.

```
n <- 40
lambda <- 1/5
nbSim <- 1000
expMatrix <- matrix(rexp(n*nbSim, lambda), nbSim)
```

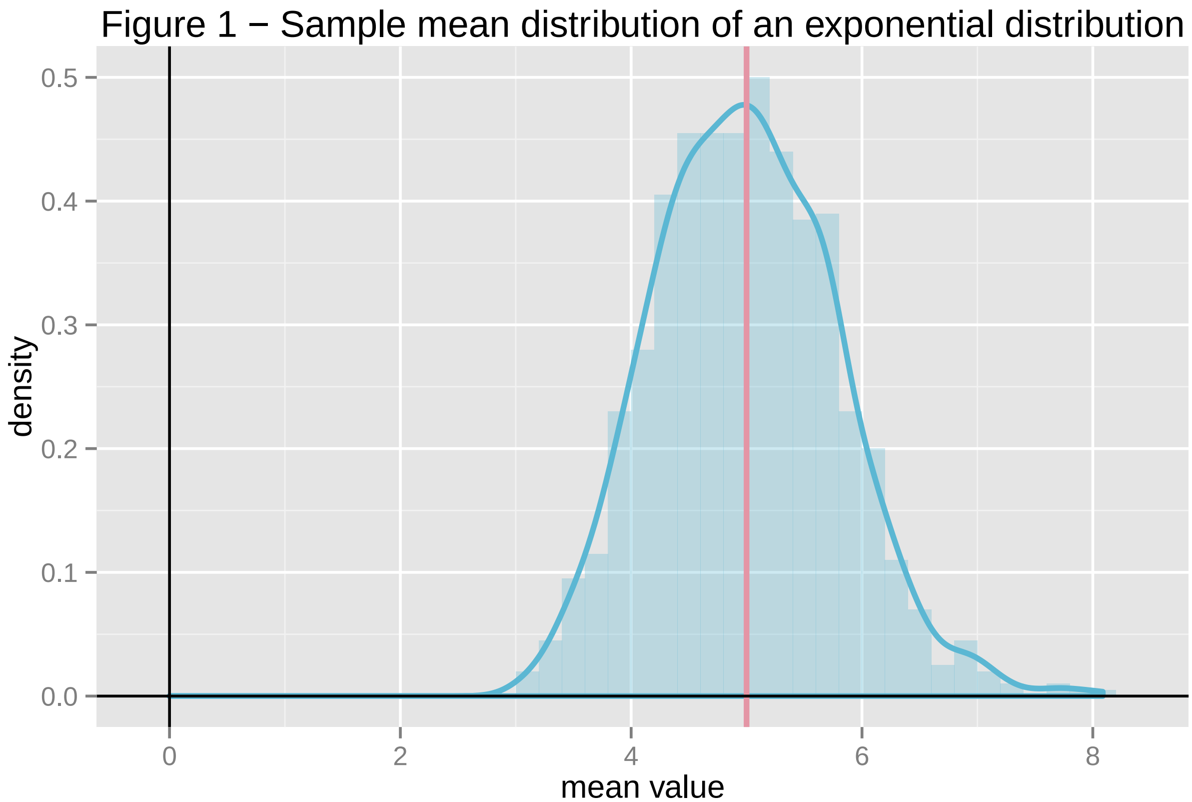
Sample Mean versus Theoretical Mean

The theoretical mean of the exponential distribution (our population) is $\mu = \frac{1}{\lambda}$. In our case, this means that $\mu = 5$.

We calculate the mean of each sample to obtain a distribution of 1000 means.

```
expMeans <- apply(expMatrix, 1, mean)
```

We can visualize the distribution of these means on a histogram in *Figure 1*. The theoretical mean has been represented on this figure by the vertical red line.



We observe that the sample mean is distributed around the theoretical mean of the population μ. Let's calculate the mean of this sample distribution and compare it to μ.

```
expMeansMean <- mean(expMeans)
```

The value $expMeansMean = 4.9934006$ is very close to the theoretical mean $\mu = 5$. We can then conclude that this demonstrates that the sample mean is approximately the popultation mean.

Sample Variance versus Theoretical Variance

The theoretical variance of the exponential distribution (our population) is $\sigma^2 = \frac{1}{\lambda^2}$. This means that the theoretical variance of the sample distribution of averages (or Standard Error) is $\frac{\sigma^2}{n}$. Let's calculate the variance of this sample distribution and compare it to the theoretical variance.

```
expMeansVar <- var(expMeans)
```

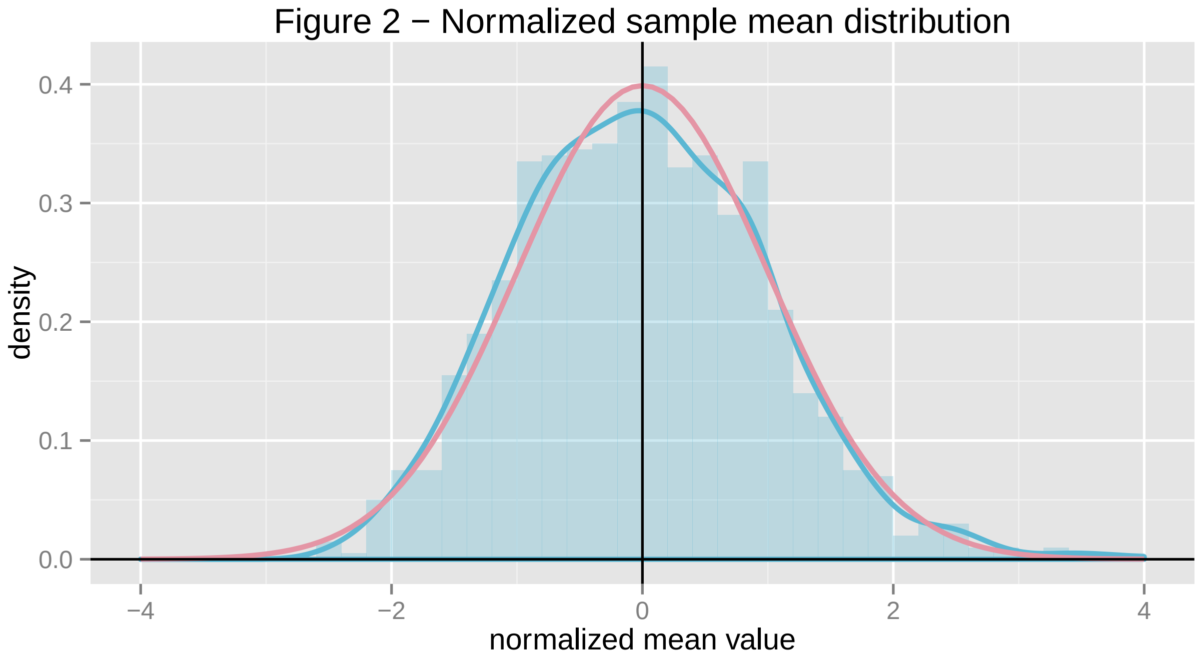
The value $expMeansVar = 0.6175667$ is very close to the theoretical variance $\frac{\sigma^2}{n} = 0.625$. We can then conclude that this demonstrates that the sample variance is approximately the Standard Error of the mean.

Distribution

As we could see in the figure 1, the distribution of means looks very much like a normal distribution. We will normalize the values obtained previously to be able to compare them with a standard normal curve. To do so we will use the formula: $\frac{estimate - meanofestimate}{StdErrofestimate}$

```
cfunc <- function(x) (mean(x) - 1/lambda) * sqrt(n) * lambda
expMeansNorm = apply(expMatrix, 1, cfunc)
```

We can visualize the distribution of these normalized means on a histogram in *Figure 3*. The standard normal distribution has been represented on this figure by the red curve.



We observe that the sample mean distribution is extremely close to the standard normal function. We can then conclude that the distribution is approximately normal.

Conclusion

We have demonstrated that the Central Limit Theorem does apply to the exponential distribution. The distribution of averages for an exponential distribution is approximately a normal distribution with:

- mean equals to the mean of the population.
- variance equals to the standard error of the mean.

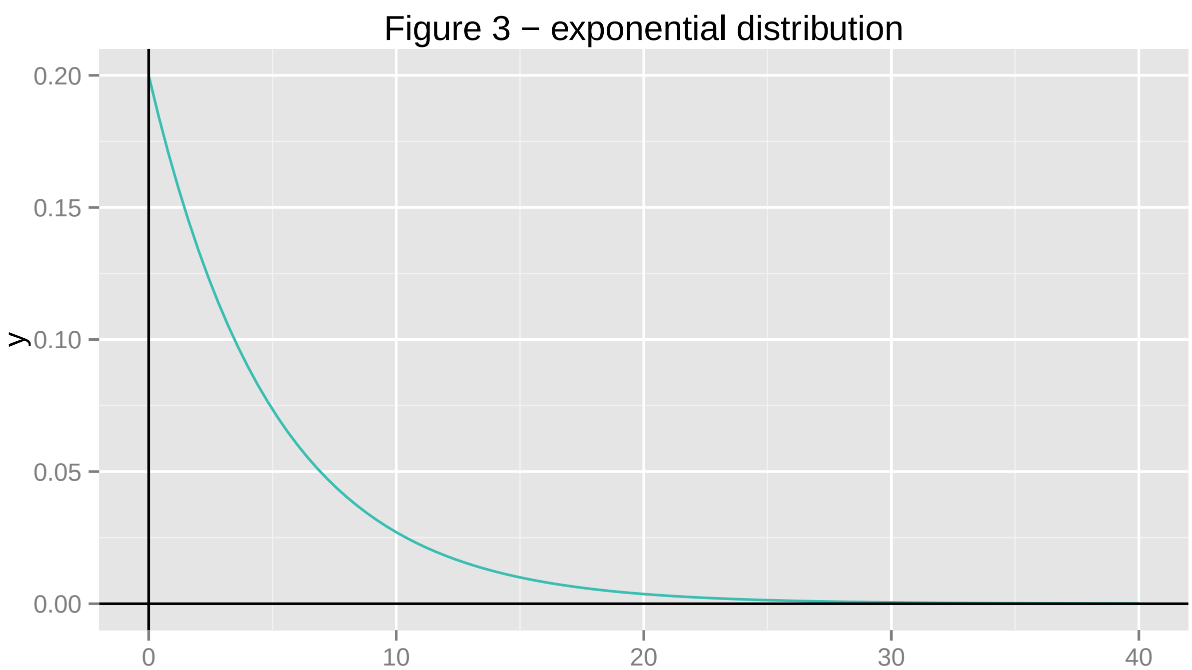
Appendix

Exponential Distribution

```
expValues <- expMatrix
dim(expValues) <- NULL

expValuesFrm <- data.frame(expValues=expValues)

g <- ggplot(data.frame(x = c(0, 40)), aes(x))
g <- g + stat_function(fun = dexp, args = list(rate = lambda), size = .5, color = hcl(180, 50, 70))
g <- g + geom_vline(xintercept = 0, size = .5, color = "black")
g <- g + geom_hline(yintercept = 0, size = .5, color = "black")
g <- g + ggtitle("Figure 3 - exponential distribution")
g <- g + labs(x=NULL)
g
```



Code for generating Figure 1

```
expMeansFrm <- data.frame(meanSim=expMeans)

g <- ggplot(expMeansFrm, aes(x=meanSim))
g <- g + geom_histogram(aes(y=..density..), alpha = .30, binwidth=0.2, linetype="blank", fill=hcl(220, 5
g <- g + geom_density(size = 1, color = hcl(220, 50, 70))
g <- g + geom_vline(xintercept = 1/lambda, size = 1, color = hcl(0, 50, 70))
g <- g + geom_vline(xintercept = 0, size = .5, color = "black")
g <- g + geom_hline(yintercept = 0, size = .5, color = "black")
g <- g + ggtitle("Figure 1 - Sample mean distribution of an exponential distribution")
g <- g + labs(x="mean value")
g
```

Code for generating Figure 2

```
expMatrixNormFrm <- data.frame(meanSimNorm=expMeansNorm)

g <- ggplot(expMatrixNormFrm, aes(x=meanSimNorm))
g <- g + geom_histogram(aes(y=..density..), alpha = .30, binwidth=0.2, linetype="blank", fill=hcl(220, 5
g <- g + geom_density(size = 1, color = hcl(220, 50, 70))
g <- g + stat_function(fun = dnorm, size = 1, color = hcl(0, 50, 70))
g <- g + geom_vline(xintercept = 0, size = .5, color = "black")
g <- g + geom_hline(yintercept = 0, size = .5, color = "black")
g <- g + ggtitle("Figure 2 - Normalized sample mean distribution")
g <- g + labs(x="normalized mean value")
g <- g + xlim(-4, 4)
g
```