Exponential Distribution and CLT

Stefano Masneri

14. April 2016

Synopsis

In this project we are going to investigate the exponential distribution and compare it with the Central Limit Theorem. The theory says that for an exponential distribution, both the mean and the standard deviation equals 1/lambda. For all our experiments we will use lambda = 0.2, consider 40 exponentials and will do 5000 simulations

```
library(ggplot2)
lambda <- 0.2
numexp <- 40
numsim <- 5000
```

Theory

We know from the theory that an exponential distribution has expected value and standard deviation equal to 1/lambda. This can be seasily seen by simulating 5000 random exponential, plotting its histogram and computing the sample mean and standard deviation.

```
myExp <- rexp(numsim, rate = lambda)
meanMyExp <- mean(myExp)
sdMyExp <- sd(myExp)
varMyExp <- sdMyExp^2
```

Both the mean (4.9848221) and the variance (24.1349717) are quite similar to the value we would expect from theory (that is, 5 for the mean and 25 for the variance).

Now, according to the central limit theorem, if we compute the distribution of the mean of 40 exponentials we should obtain a distribution which approximates a gaussian with the same mean and whose variance is the square root of the variance of the original distribution divided by the number of samples.

Simulations

To show an example of the CLT we will compute the mean of 40 exponentials as well as its mean and variance:

```
means <- NULL
for (i in 1 : numsim)
  means <- c(means, mean(rexp(numexp, rate = lambda)))

sampleMean <- mean(means)
sampleVar<- var(means)
```

And from the theory we should obtain those values for mean and variance:

```
meanCLT <- meanMyExp
varCLT <- sqrt(varMyExp / numexp)
sdCLT <- sqrt(varCLT)
```

As expected, the results are pretty similar

```
abs( meanCLT - sampleMean )
```

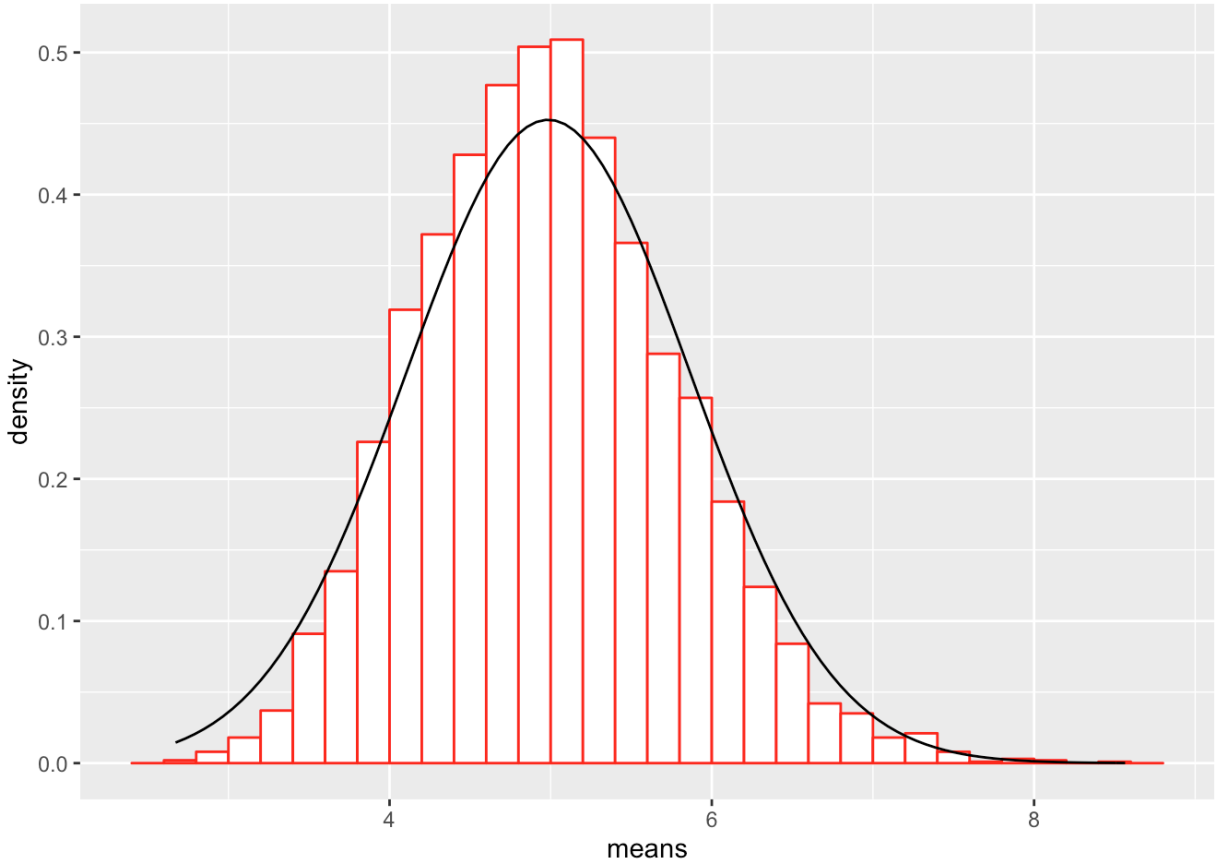
```
## [1] 0.002105386
```

```
abs( varCLT - sampleVar )
```

```
## [1] 0.1443416
```

Now we will plot the histogram of the means and compare it with the gaussian with the values of mean and variance defined above, to see graphically how similar they are.

```
g <- ggplot() + aes(means) + geom_histogram(binwidth=0.2, colour="red", fill
="white", aes(y = ..density..))
g  + stat_function(fun = dnorm, args = list(mean = meanCLT, sd = sdCLT))
```



The red bordered bars represents the density of the histogram, while the black line represents a standard normal centered at 5 and with variance equals to 0.7767717.