Comparison of the exponential distribution and the central limit theorem

Rodrigo

September 21, 2015

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT). The exponential distribution can be simulated in R with rexp(n, lambda) where lambda λ is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. For this simulation, we set $\lambda = 0.2$. In this simulation, we investigate the distribution of averages of 40 numbers sampled from exponential distribution with $\lambda = 0.2$.

First, we are going to perform a thousand simulated averages of 40 exponentials.

```
set.seed(7)
simulation_number = 1000
size = 40
lambda = 0.2
mean.exp.distr = 1/lambda
sd.exp.distr = 1/lambda/sqrt(size)
dt = data.frame(sapply(1:simulation_number, function(x) mean(rexp(size, lambda))))
names(dt) = "means"

head(dt)
```

```
##      means
## 1 4.967949
## 2 5.185232
## 3 3.801033
## 4 3.991504
## 5 4.465423
## 6 4.693979
```

Now we compare the mean, the standard deviation and the variation of our data against the expected from CLT.

The mean of the theorical distribution is $\mu = \frac{1}{0.2} = 5$:

```
mean.exp.distr
```

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

The mean and the standard deviation of our data are:

```
mean(dt$means)
```

```
## [1] 4.983294
```

The mean of our data is very close to the expected mean of the theorical distribution. Moreover the standard deviation of the distribution is closer the expected standard deviation.

```r
sd(dt$means) #standard deviation of the distribution
```

```
## [1] 0.7610878
```

```r
sd.exp.distr #expected standard deviation
```

```
## [1] 0.7905694
```

According to the CLT we would expect that the variance of the sample of the 1000 means is approximately $\frac{\frac{1}{0.2^2}}{40} = 0.625$. The variance in our data is 0.58 which is close to the variance of the theoretical distribution.

The distribution is approximately normal.

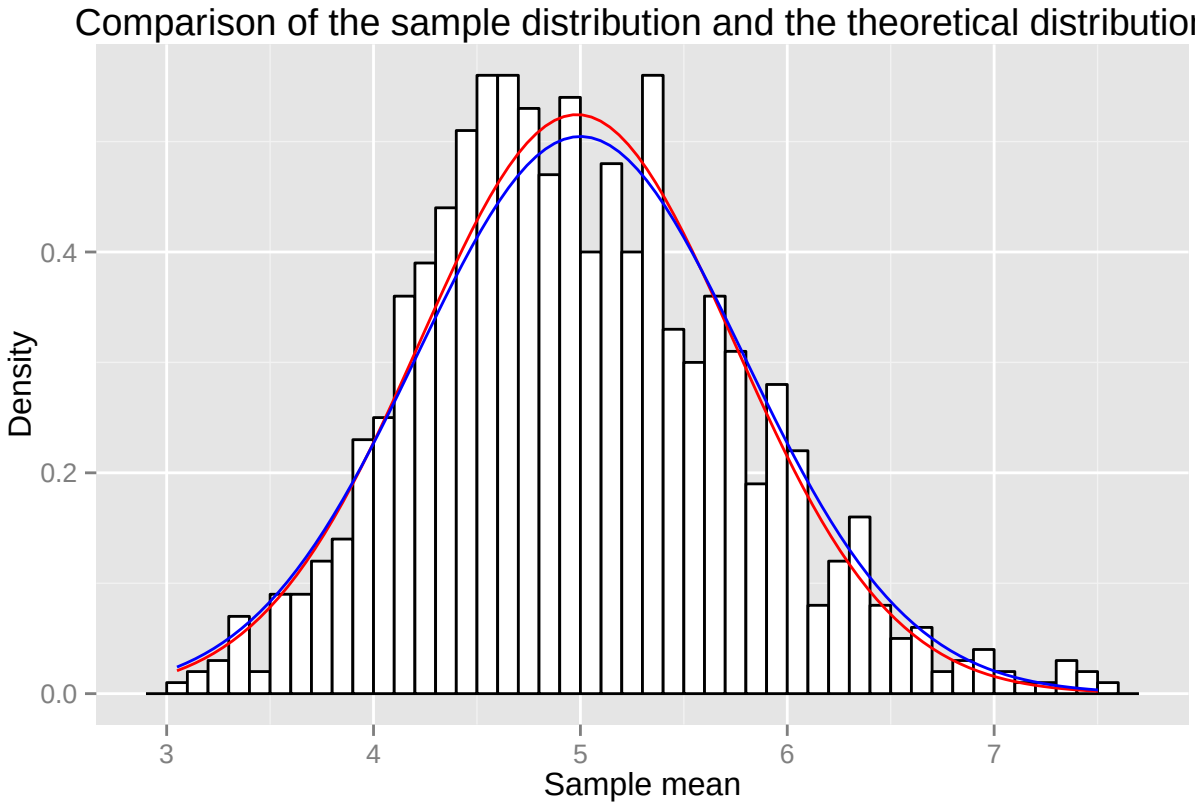
The histogram we show below confirm that the distribution is aproximately normal. The red line represent the density function from our distribution and the blue line represent the density function from the theoretical distribution.

```r
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```r
library(MASS)    # for fitsidtr(...)

ggplot(dt, aes(x = means)) +
    geom_histogram(aes(y=..density..),colour = "black", fill = "white", binwidth = 0.1)+
    stat_function(fun=dnorm, args=fitdistr(dt$means,"normal")$estimate, color = "red")+
    stat_function(aes(x = c(min(dt$means), max(dt$means))), fun = dnorm, color = 'blue',
                  args = list(mean = mean.exp.distr, sd = sd.exp.distr)) +
    xlab('Sample mean') +
    ylab('Density') +
    ggtitle('Comparison of the sample distribution and the theoretical distribution')
```


Comparison of the sample distribution and the theoretical distribution
Density
0.4
0.2
0.0
3
4
5
6
7
Sample mean