# Report: assignment 1

*C. Deramond*

*Sunday, December 07, 2014*

This document contains the report for assigment 1.

### Problem:

The exponential distribution can be simulated in R with rexp(n, $\lambda$) where $\lambda$ is the rate parameter. The mean of exponential distribution is $\mu = 1/\lambda$ and the standard deviation is also $\sigma = 1/\lambda$. Set $\lambda = 0.2$ for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should:

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

### Processing

First we load required packages and set the parameters for $\lambda$, theoretical values and n. Starting with 1000 simulations

```
library(ggplot2)
l <- 0.2
theo_mean <- 1/l
theo_sd <- 1/l
theo_var <- 1/l^2
n <- 40
nosim <- 1000
```

Remark on Theoretical mean with value *5*.

Now, let's review CLT for exponential distribution:

- Standard error of the mean is $SE = \sigma/\sqrt{n} = 1/(\sqrt{n} * \lambda)$
- Then
$$\frac{\bar{X} - \lambda^{-1}}{SE}$$
will be approximately normally distributed.

Having set the parameters, we'll generate the simulations:

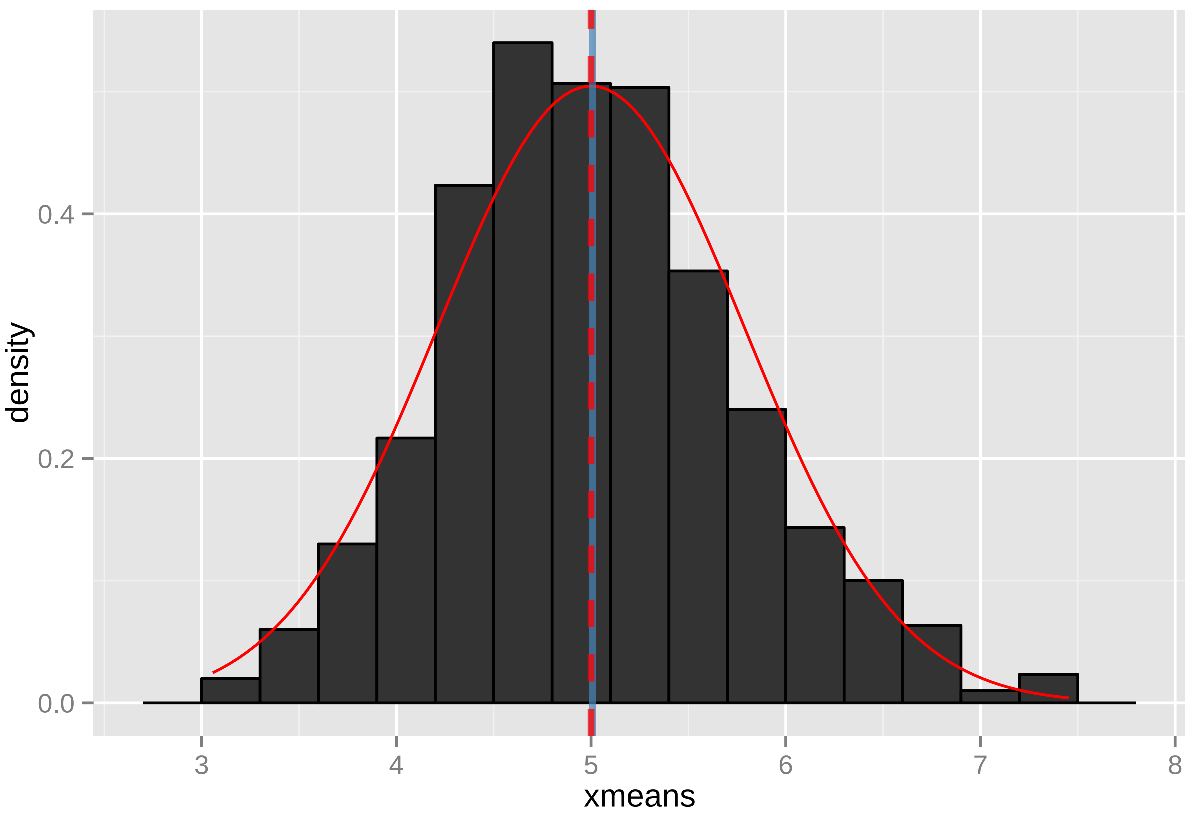
```
zConvert <- function(x, n) l * sqrt(n) * (mean(x) - theo_mean)
data <- matrix(rexp(n*nosim, l), nosim) #40 samples of 1000 iid exponential
xbar <- data.frame(xmeans = rowMeans(data)) # 1000 xbar (means of 40 observations)
xbar <- cbind(xbar, Zx = sapply(xbar$xmeans, zConvert, n)) #normalization
```

## Results

### 1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

as the graph shows, it centers at the blue line, being the mean of $\bar{X}$, with a value of 5.007 , very close to the theoretical $\lambda^{-1} = 5$

```
g <- ggplot(xbar, aes(x = xmeans)) + geom_histogram(binwidth=.3, colour = "black", aes(y = ..density..))
g <- g + stat_function(fun = dnorm, col = "red", args = list(mean = theo_mean, sd = theo_sd/sqrt(n))) +
        geom_vline(xintercept = mean(xbar$xmeans), col = "steelblue",alpha = .75, size = 1.2) +
        geom_vline(xintercept = theo_mean, col = "red",alpha = .75, size = 1.2, linetype = "dashed")
g
```




### 2. Show how variable it is and compare it to the theoretical variance of the distribution.

As defined variable $\bar{X}$ contains the means of `n=40` samples of 1000 simulations, therefore empirical mean, sd vs theoretical goes:

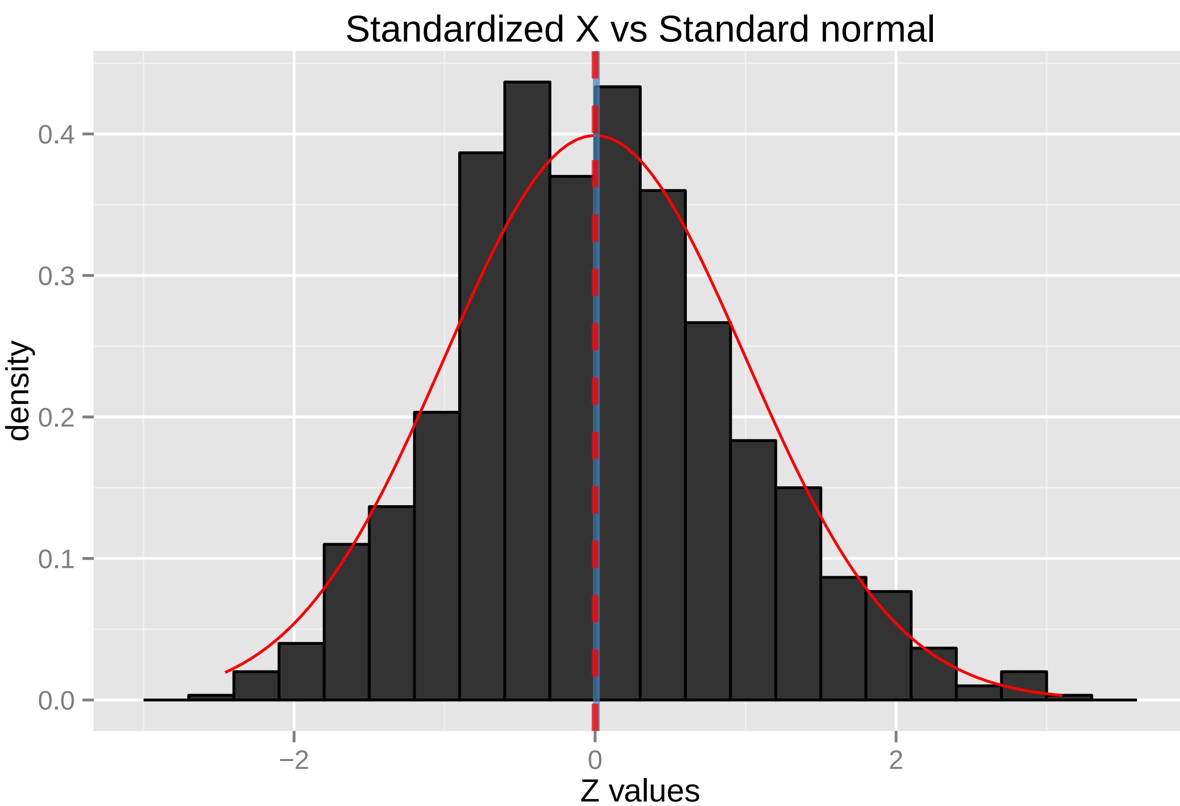
```
##        Variables Empirical Theoretical
## 1          Mean 5.0067403   5.0000000
## 2 Std. Deviation 0.7561972   0.7905694
## 3       Variance 0.5718342   0.6250000
```

**3. Show that the distribution is approximately normal.**

for this case, let's simplify by normalizing $\bar{X}$, which was already done, but repeat the code:

```
zConvert <- function(x, n) 1 * sqrt(n) * (mean(x) - theo_mean)
...
xbar <- cbind(xbar,
              Zx = sapply(xbar$xmeans, zConvert, n)) #normalization
```

so the normalized $Z_x$, will look like:




Which fits nicely with the normal distribution, having $Z_x$ centered at 0.009.