Program_assignment2 - Part 1

Illustration of the Central Limit Theorem

Here we're going to display and illustrate results of the Central Limit Theorem. Using simulation of random variables following exponential distribution. We will calculate the mean of 1000 random vairables and do it 40 times. And we will compare the distribution of these means with a normal distribution.

First we set the different parameters for the study: the rate parameter, the number of random variables per simulation and the number of simulations.

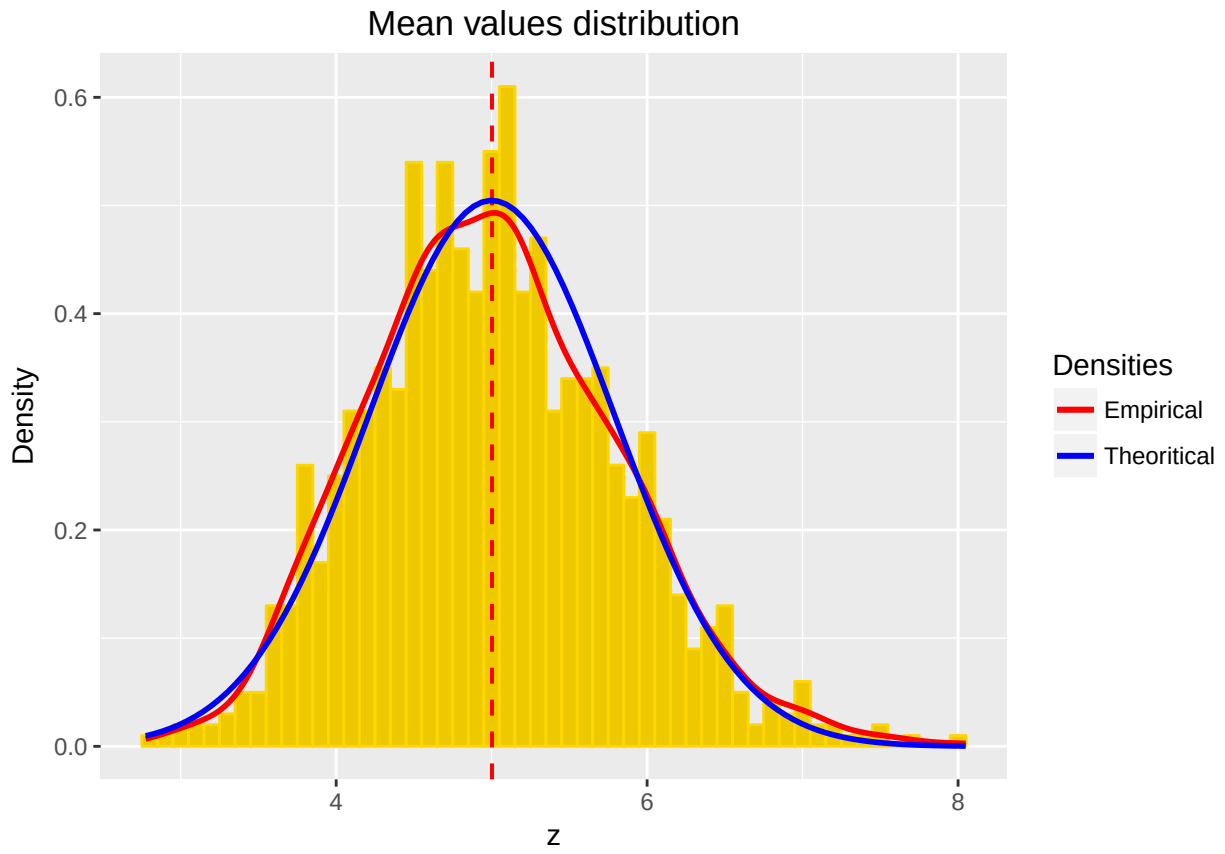
```
lambda <- 0.2
n <- 40
nb_simu <- 1000
```

We calculate now the mean of 1000 of 40 random variables simulated according to a exponential distribution with a rate parameter equals to 0.2 and display its distribution. We compare it to a normal distribution with a mean of $\frac{1}{\lambda}$ and a variance of $\frac{1}{\lambda * \sqrt{n}}$ to illustrate result of the Central Limit Theorem

```
means_simu <- data.frame(mean = replicate(nb_simu, mean(rexp(n, lambda))))
```

```
theoric_mean <- 1/lambda
theoric_sd <- (1/lambda)/sqrt(n)
```

```
ggplot(means_simu, aes(x=mean)) +
  geom_histogram(aes(y=..density..), binwidth=.1, colour="gold", fill="gold2") +
  geom_line(aes(y = ..density.., colour = 'Empirical'), stat = 'density', size = 1) +
  geom_vline(aes(xintercept=mean(mean, na.rm=TRUE)),
             color="red", linetype="dashed", size=0.7) +
  stat_function(fun = dnorm, aes(colour = 'Theoritical'),
                args = c(mean=theoric_mean,sd = theoric_sd), size = 1) +
  scale_colour_manual(values=c("red", "blue"), name="Densities") +
  ylab("Density") + xlab("z") + ggtitle("Mean values distribution") +
  theme(plot.title = element_text(hjust = 0.5))
```



The empirical mean, represented in red dotted line inthe graphic above, is almost equal to the theorical mean that is $\frac{1}{\lambda} = 5$.

We can compare empirical and theorical means and variances.

```
result = data.frame(mean = c(mean(means_simu$mean), theoric_mean),
                    standard_deviatation = c(sd(means_simu$mean), theoric_sd),
                    variance = c(sd(means_simu$mean)^2, theoric_sd^2))
colnames(result) = c("Mean", "Standard Deviation", "Variance")
row.names(result) = c("Empirical", "Theoric")
```

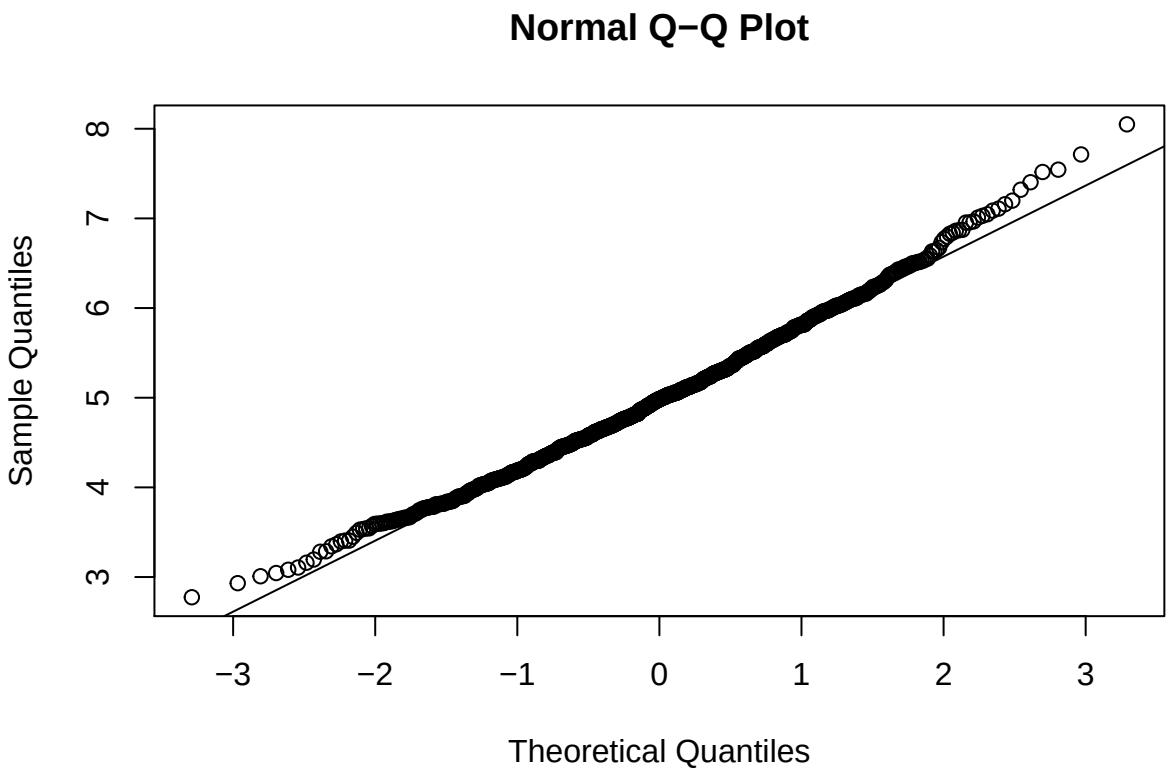
Table 1: Comparison Theoric vs. Empiric

	Mean	Standard Deviation	Variance
Empirical	5.002177	0.8036435	0.6458429
Theoric	5.000000	0.7905694	0.6250000

So theorical and empirical distributions are really close and this illustrates the Central Limit Theorem.

Finally to validate that we can approximate the means distribution by a Gaussian, we can draw a qqplot.

```
qqnorm(means_simu$mean)
qqline(means_simu$mean)
```



The theoretical quantiles also match closely with the actual quantiles.

Everything above shows that the distribution is approximately normal and illustrate the Central Limit Theorem.