

# EXPONENTIAL DISTRIBUTION vs NORMAL DISTRIBUTION

*emmanuelq2*

*28 février, 2016*

```
Sys.setlocale("LC_TIME", "English")
```

The underlying idea of this project is to investigate the exponential distribution and compare it with the Central limit Theorem (CLT). The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda( $\lambda$ )` 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 observe the distribution of averages of 40 exponentials. We will perform a thousand simulations

## SIMULATION

```
# Set seed for reproducibility
set.seed(123)
lambda <- 0.2
# We perform 1000 simulations with 40 samples
sample <- 40
simulations <- 1:1000

# Let's do 1000 simulations
means <- data.frame(x = sapply(simulations, function(x) {mean(rexp(sample, lambda))}))
# Averages of 40 exponentials
avgfirst <- head(means)
knitr::kable(t(head(avgfirst)))
```

	1	2	3	4	5	6
x	4.811212	5.360077	4.592871	4.900051	5.51662	5.612834

## RESULTS

1. Show the sample mean and compare it to the theoretical mean of the distribution.

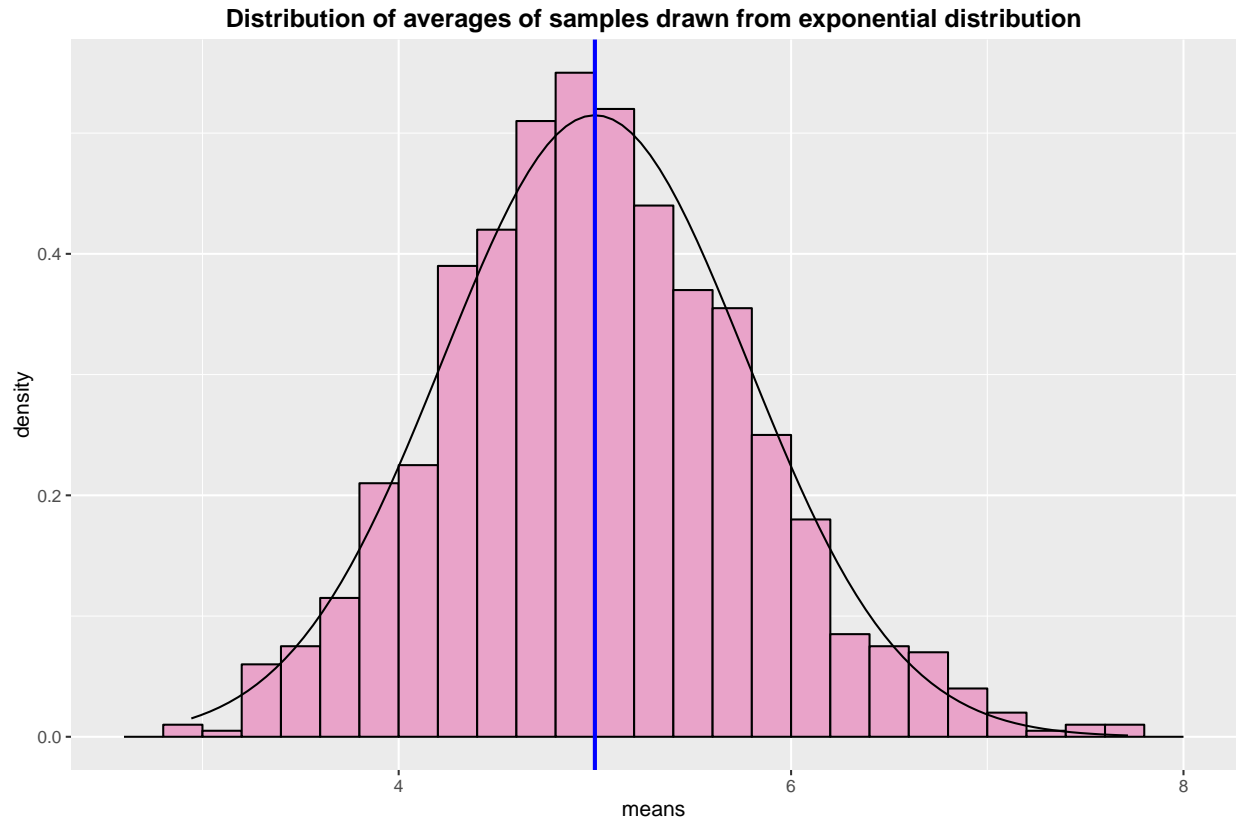
```
# mean of distribution of averages of 40 exponentials
sample_mean <- mean(means$x)
# expected mean from analytical expression
expected_mean <- 1/lambda
```

The mean of the means of the exponential of 1000 simulations of 40 exponentials is 5.0119, which is very close to the expected mean of  $\lambda^{-1} = 5$ .

```
library(ggplot2)
```

```
ggplot(data = means, aes(x = means)) +
  geom_histogram(aes(y=..density..), fill = I('#E9A3C9'),
    binwidth = 0.20, color = I('black')) +
  stat_function(fun = dnorm, arg = list(mean = 5, sd = sd(means$x))) +
  ggtitle("Distribution of averages of samples drawn from exponential distribution") + theme(plot.title = element_text(hjust = 0.5))
```

## Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous



Above is displayed a histogram of the means of the 1000 simulations of  $\text{rexp}(n, \lambda)$ . It is overlaid with a normal distribution with mean 5 and standard deviation 0.7909. So, the distribution of our simulations appears to be normal.

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution

```
# Variance of our simulations:
sample_var <- var(means$x)
# Expected variance of the distribution
expected_var <- ((1/lambda)/sqrt(sample))^2

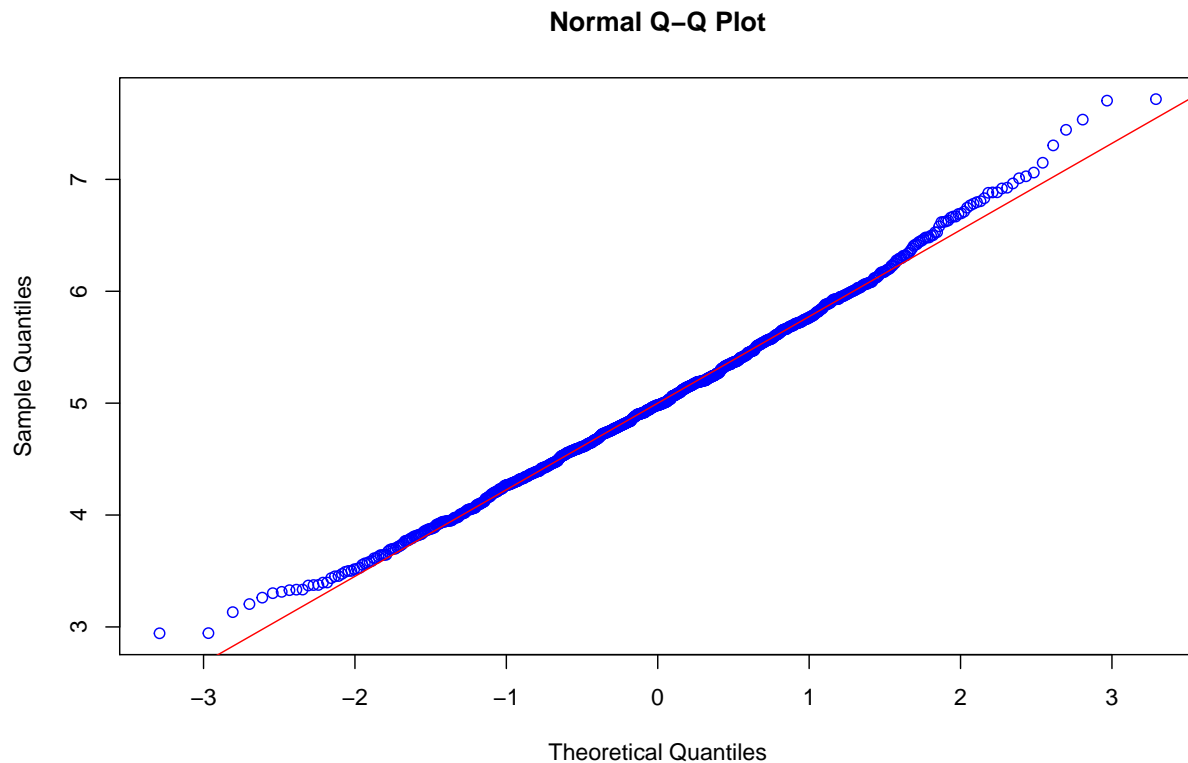
# standard deviation of distribution of averages of 40 exponentials
sample_sd <- sd(means$x)
# Expected standard deviation
expected_sd <- (1/lambda)/sqrt(sample)
```

The variance of sample means is 0.6005 whereas the theoretical variance of the distribution is  $\sigma^2/n = 1/(\lambda^2 n) = 1/(0.04 \times 40) = 0.625$ .

Similarly, the standard deviation in distribution of averages of 40 exponentials equates to 0.7749147 is close to the theoretical standard deviation of the distribution, 0.7906.

3. Show that the distribution is approximately normal

```
qqnorm(means$x, col="blue")  
qqline(means$x, col = "red")
```



Due to the Central Limit Theorem, the averages of samples follow normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. Also, the q-q plot suggests the distribution of averages of 40 exponentials is very close to a normal distribution