

# Investigation of exponential distribution and comparison with Central Limit Theorem

*Mohammad Ullah*

*September 30, 2017*

## Synopsis:

The main goal of this project was to show that, whether the distribution of a large number (1000) of averages of 40 random exponential follow the Central Limit Theorem (CLT). It means, the distribution will follow an approximate normal distribution. Beside that, we have also showed that simulated sample mean and variance is very close to theoretical mean and variance.

## Given Directions and requirements of the project:

In this project we have to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The mean of exponential distribution is  $1/\lambda$  and the standard deviation is also  $1/\lambda$ . We have to investigate the distribution of averages of 40 exponentials. A thousand simulations is required.

We need to show

1. The sample mean and compare it to the theoretical mean of the distribution.
2. How variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. The distribution is approximately normal.

## Results:

### Simulated sample mean Vs. theoretical mean

```
## [1] "Simulated mean : 4.98661966880036"
```

```
## [1] "Theoretical mean : 5"
```

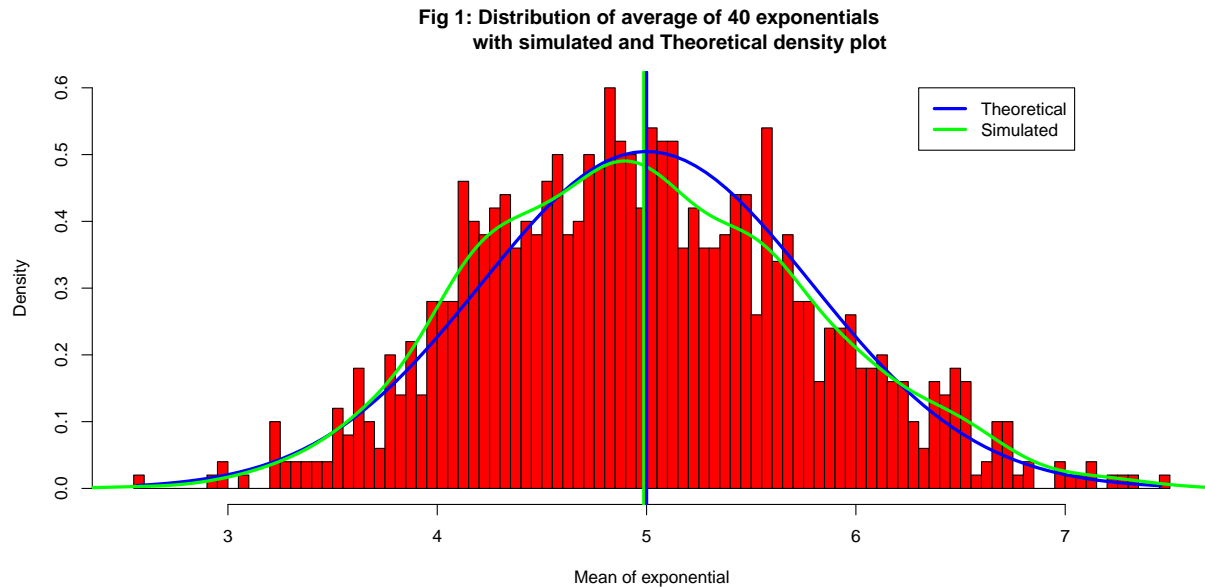
### Simulated sample variance Vs. theoretical variance

```
## [1] "Simulated variance : 0.631678852170819"
```

```
## [1] "Theoretical variance : 0.625"
```

So, simulated mean and variance are very close to theoretical values.

## Distribution of sample mean



In Figure 1, we have plotted the distribution of 1000 average of 40 random exponentials also the density from the simulated and theoretical values. The mean of the exponential follows the CLT and creates a normal distribution. The density plot from the simulated values overlap with the normal density curve from the theoretical values. It also shows the mean of the simulated and theoretical distributions in vertical lines.

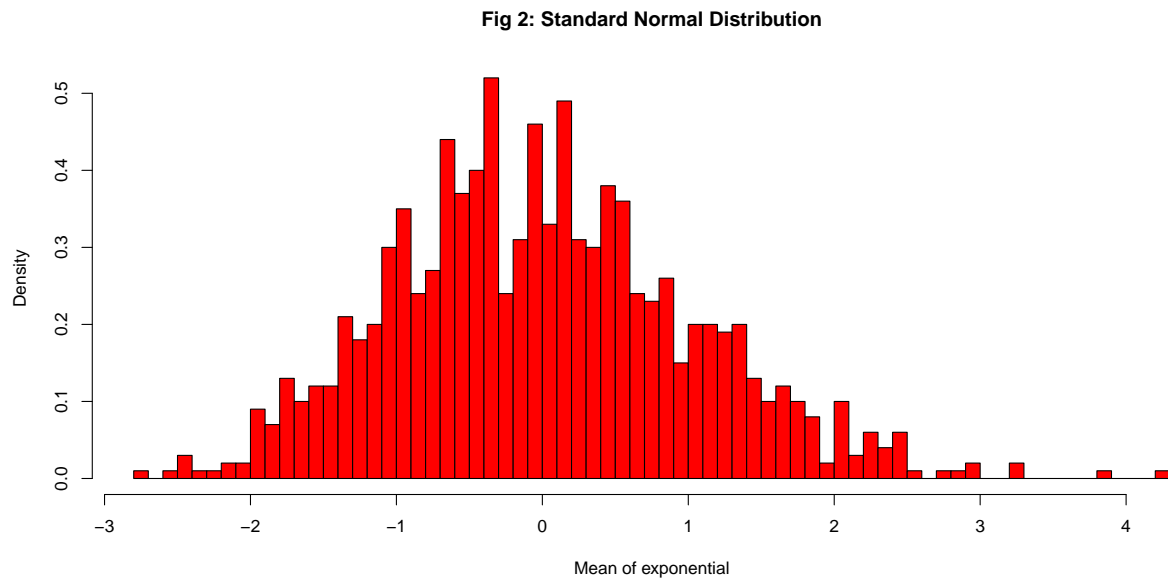
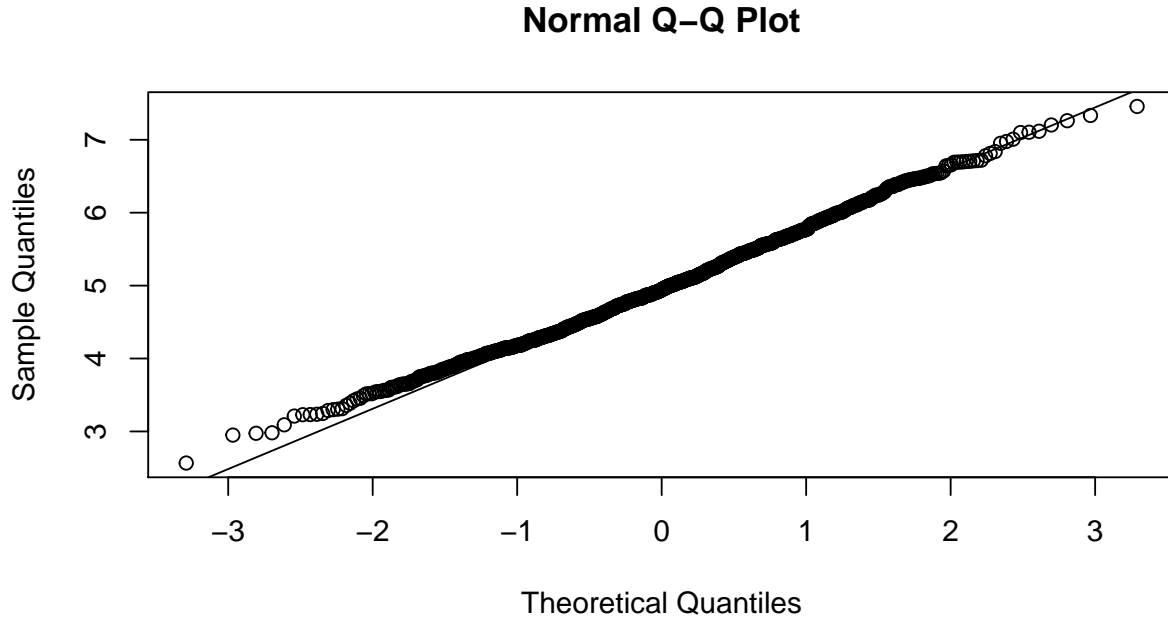


Figure 2 shows that by subtracting mean we can get approximately a standard normal distribution with mean zero.



From above qqplot, we see that the qqnorm function plotted the simulated sample against the normal distribution. The points in the figure overlap with the line added by the qqline function. This phenomenon suggest the normality of the distribution.

The project also asked to check the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

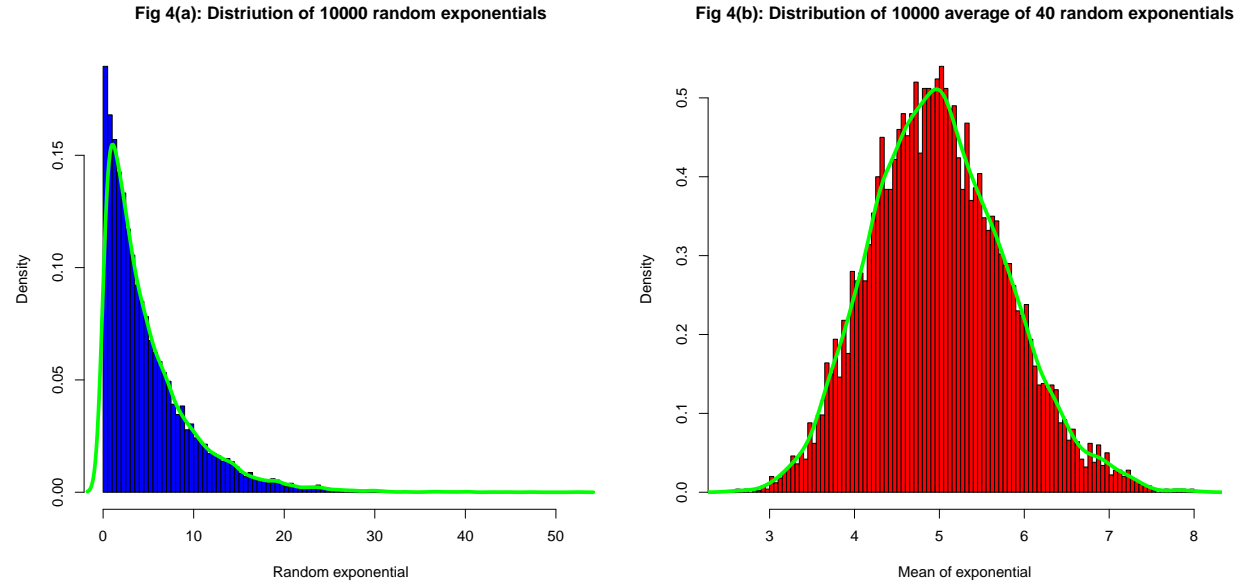


Figure 4 shows that random exponentials follow a exponential decay destrubtion, whereas the average of random exponentials follow normal distribution.

Finally, we can say that the distribution of means of 40 exponentials behave as predicted by the Central Limit Theorem.

## Appendix

Codes are given in this section:

### Simulated sample mean Vs. theoretical mean

Define variables:

```
set.seed(3)
expomean <- NULL
expomean2 <- NULL
normexpo <- NULL
lamda <- 0.2
num <- 40
```

Function:

```
expofunc <- function(num, lamda)
{
  mean(rexp(num, rate = lamda))
}
```

```
for (i in 1 : 1000) expomean = c(expomean, expofunc(num, lamda))

print(paste0("Simulated mean : ", mean(expomean)))
print(paste0("Theoretical mean : ", 1/lamda))
```

### Simulated sample variance Vs. theoretical variance

```
print(paste0("Simulated variance : ", var(expomean)))
print(paste0("Theoretical variance : ", ((1/lamda)^2)/40))
```

### Distribution of sample mean

```
h <- hist(expomean, breaks = 100, col = "red", prob = TRUE,
  main = "Fig 1: Distribution of average of 40 exponentials
  with simulated and Theoretical density plot",
  xlab = "Mean of exponential",
  ylab = "Density")
abline(v = mean(expomean), col = "green", lwd = 3)
xfit <- seq(min(expomean), max(expomean), length = 100)
yfit <- dnorm(xfit, mean = 1/lamda, sd = 1/lamda/sqrt(40))
lines(xfit, yfit, col = "blue", lwd = 3)
lines(density(expomean), col = "green", lwd = 3)
abline(v = 1/lamda, col = "blue", lwd = 3)
legend(6.3, 0.6, c("Theoretical", "Simulated"), lty = c(1,1),
  lwd = c(3,3), col = c("blue", "green"))
```

## Standard Normal Distribution

```
for (i in 1 : 1000) normexpo =  
  c(normexpo, (expofunc(num, lamda) - 1/lamda)*sqrt(num)*lamda)  
  
hist(normexpo, breaks = 100, col = "red", prob = TRUE,  
      main = "Fig 2: Standard Normal Distribution",  
      xlab = "Mean of exponential",  
      ylab = "Density")
```

## Q-Q plot

```
qqnorm(expomean)  
qqline(expomean)
```

Compare between the shape of original random exponential distribution and means of exponential distribution

```
par(mfrow = c(1,2))  
rexp <- rexp(10000, rate = 0.2)  
hist(rexp, probability = TRUE, col = "blue", breaks = 100,  
      main = "Fig 4(a): Distribution of 10000 random exponentials",  
      xlab = "Random exponential",  
      ylab = "Density")  
lines(density(rexp), lwd = 4, col = "green")  
  
for (i in 1 : 10000) expomean2 = c(expomean2, expofunc(num, lamda))  
h <- hist(expomean2, breaks = 100, col = "red", prob = TRUE,  
          main = "Fig 4(b): Distribution of 10000 average of 40 random exponentials",  
          xlab = "Mean of exponential",  
          ylab = "Density")  
lines(density(expomean2), lwd = 4, col = "green")
```