

# Exponential distribution Compare with the Central Limit Theorem

## Overview

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where `lambda` is the rate parameter. The mean of exponential distribution is  $1/\lambda$  and the standard deviation is also  $1/\lambda$ . Set `lambda = 0.2` for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

## Introduction

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

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

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

As a motivating example, compare the distribution of 1000 random uniforms

```
hist(runif(1000))
```

and the distribution of 1000 averages of 40 random uniforms

```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(runif(40)))
hist(mns)
```

This distribution looks far more Gaussian than the original uniform distribution!

## Output

Sample mean vs theoretical mean of the distribution.

We generate a series of 1000 simulation dataset for comparison to the theory. Each iteration contains 40 observation and the exponential distribution is set using 'rexp' function with `lambda = 0.2`

Below is the variables and its value :-

```
lambda = 0.2  
n = 40  
nosim = 1000
```

Now we perform the code below to simulate , collect the data and plot it.

rexp() - will generate 40 times(n) when the rate is 0.2 (lambda) and get the mean value from it on 1000 iteration(nosim).

```
exponen_simulation <- function(n, lambda)  
{  
  mean(rexp(n,lambda))  
}  
  
simData <- data.frame(ncol=2,nrow=1000)  
names(simData) <- c("Index","Mean")  
for (i in 1:nosim)  
{  
  simData[i,1] <- i  
  simData[i,2] <- exponen_simulation(n,lambda)  
}
```

The mean of the sample dataset is :-

```
sample_mean <- mean(simData$Mean)  
sample_mean  
  
## [1] 4.978711  
  
sample_mean # near to 5  
  
## [1] 4.978711  
  
theoretical_mean <- 1/lambda  
theoretical_mean  
  
## [1] 5
```

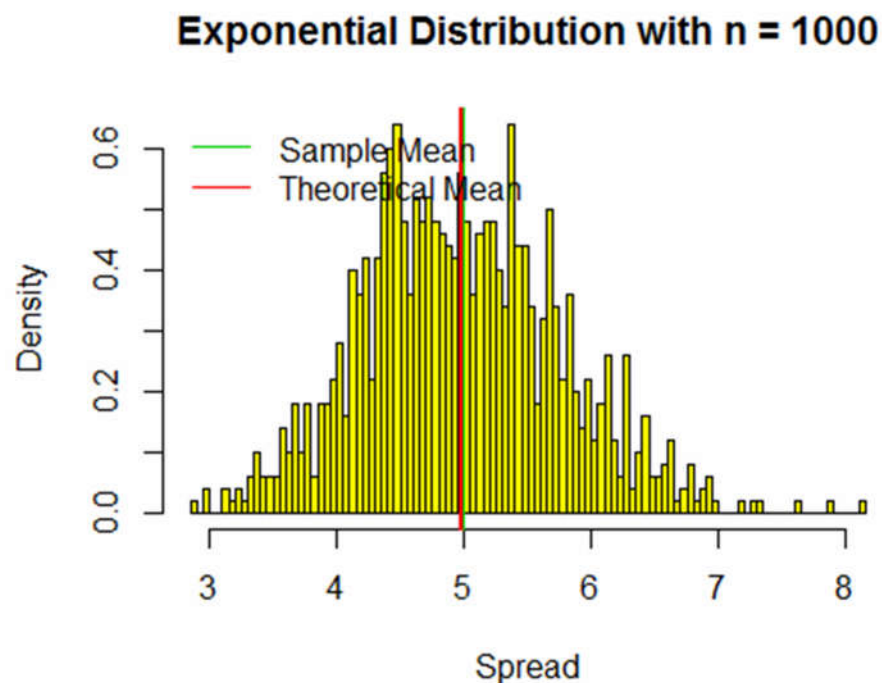
If simulation mean is near to the theoretical value of 5

Variance variable vs theoretical variance of the distribution.

Below is the historical plot of the exponential distribution of n = 1000 :-

```
hist(simData$Mean,  
breaks=100,  
prob=TRUE,  
col="yellow",  
main="Exponential Distribution with n = 1000",  
xlab="Spread")  
abline(v = theoretical_mean,
```

```
col= 3,
lwd = 2)
abline(v = sample_mean,
col = 2,
lwd = 2)
legend('topleft', c("Sample Mean", "Theoretical Mean"),
lty=c(1,1),
bty = "n",
col = c(col=3, col=2))
```



The following graph is to compare the variance in the sample means of the 1000 simulated value to the theoretical variance of the population

The sample means variance estimates the variance of the population using the value from the variance of the 1000 values in the means times the sample size which is 40 below :-

```
sample_var <- var(simData$Mean)
sample_var
```

```
## [1] 0.6170358
```

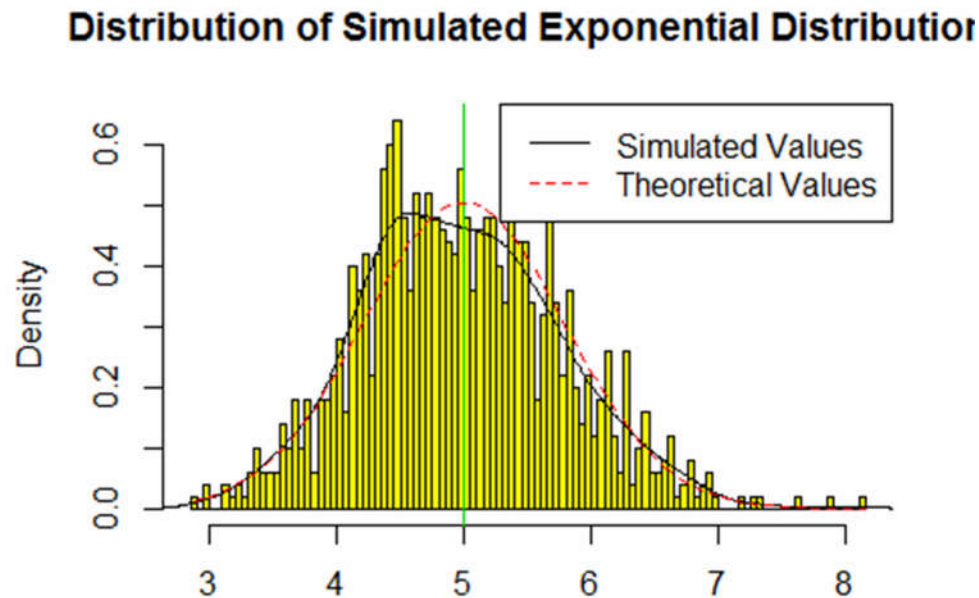
The theoretical variance of the population is :-

```
theor_var <- ((1/lambda)^2)/40
theor_var
```

```
## [1] 0.625
```

The distribution can be shown below :-

```
hist(simData$Mean,
breaks = 100,
prob = TRUE,
col="yellow",
main = "Distribution of Simulated Exponential Distribution", xlab="")
lines(density(simData$Mean))
abline(v = 1/lambda, col = 3)
xfit <- seq(min(simData$Mean), max(simData$Mean), length = 100)
yfit <- dnorm(xfit, mean = 1/lambda, sd = (1/lambda/sqrt(40)))
lines(xfit, yfit, pch=22, col="red", lty=2)
legend('topright', c("Simulated Values", "Theoretical Values"), lty=c(1,2),
col=c("black", "red"))
```

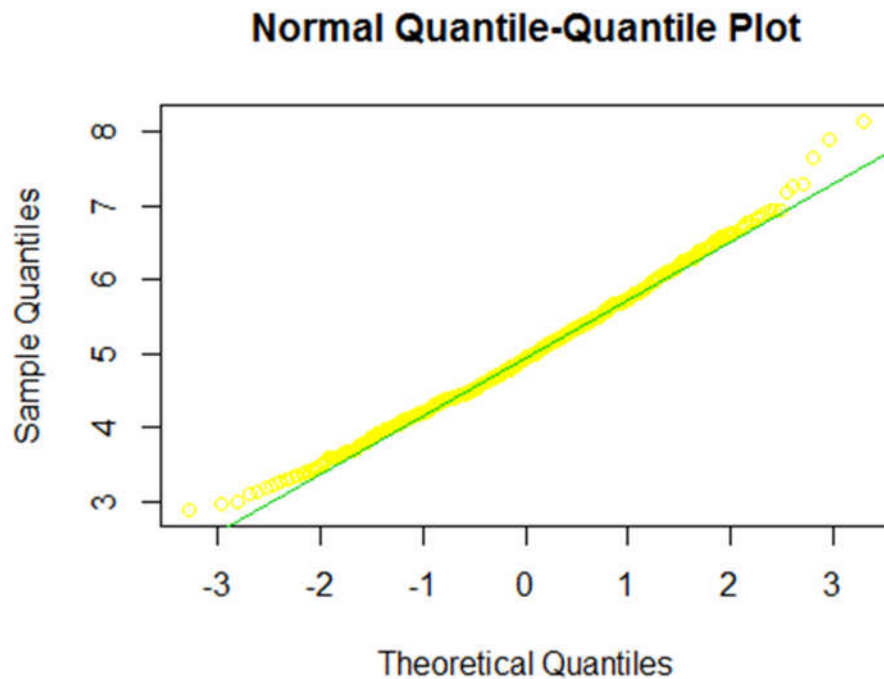


Distribution is approximately normal.

Because of the central limit theorem, the sample average will follow the normal distribution. The figure below shows the computed density using the histogram and the plotted normal density with additional values from theoretical means and variance values. The Quantile-Quantile plot describes the normality. The quantiles theoretical matches with the actual quantiles. The distribution shows normality and proves the methods.

```
qqnorm(simData$Mean,
col="yellow",
main="Normal Quantile-Quantile Plot")
```

```
qqline(simData$Mean,
col="3")
```



## Appendix

```
hist(simData$Mean,
breaks = 100,
prob = TRUE,
col="yellow",
main = "Exponential Distribution n = 1000",
xlab = "Spread")
xfit <- seq(min(simData$Mean), max(simData$Mean), length = 100)
yfit <- dnorm(xfit, mean = 1/lambda, sd = (1/lambda/sqrt(40)))
lines(xfit, yfit, pch=22, col = 3, lwd=2)
legend('topright', c("Theoretical Curve"), lty=1,lwd=2, col=3)
```

### Exponential Distribution $n = 1000$

