# Simulation Project - Statistical Inference JHU Data Science specialisation

Guynemer Cétoute

2025-03-20

## Simulation

In this project it is required to investigate the exponential distribution in R and to 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.

### Requirement

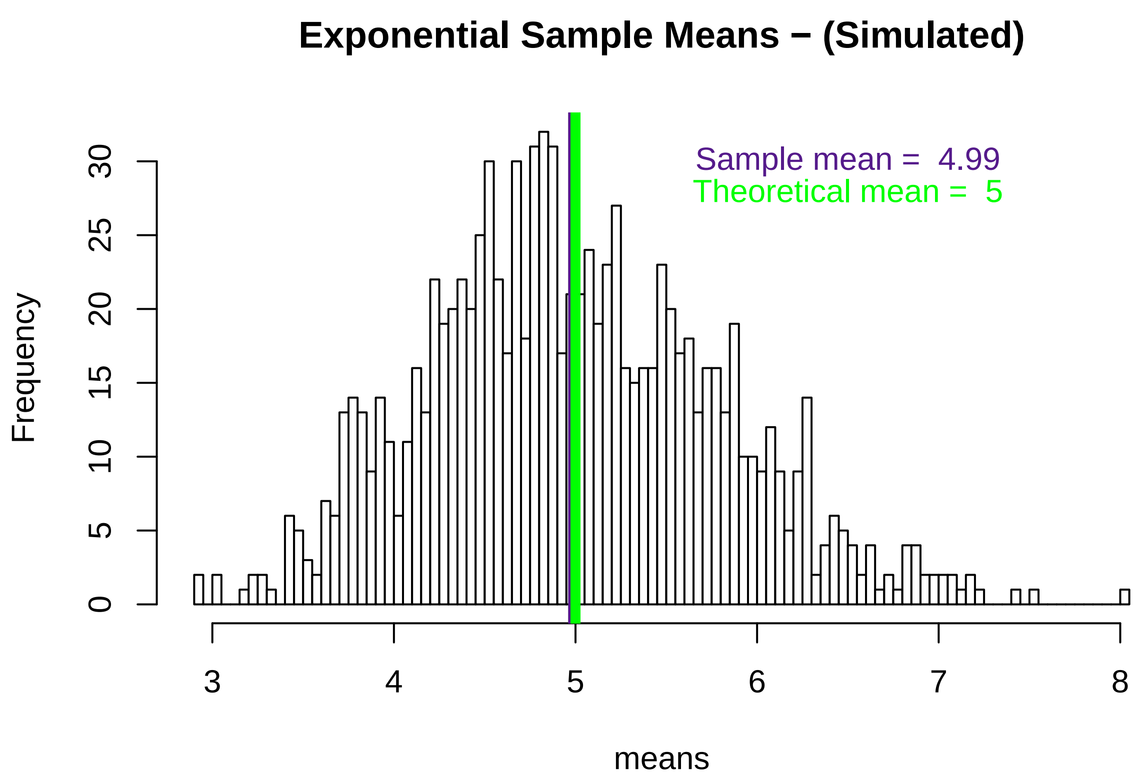
1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

```r
set.seed(1000) # random number generator before simulating random values
lambda <- .2 # value of lambda
nExpo <- 40 # value of exponential
nSimul <- 1000 # number of simulation
simulation <- replicate(nSimul, rexp(nExpo, lambda)) # simulation
means <- apply(simulation, 2, mean) # caluculation of the mean of the exponential simulation
```

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

```r
meanSample <- mean(means)
meanTheoretical <- 1/lambda
hist(means, main = "Exponential Sample Means - (Simulated)", col = "white", breaks = 150)
abline(v = meanSample, lwd = 5,  col = "purple4")
abline(v = meanTheoretical, lwd = 5,  col = "green")

# And show our numbers
text(6.5, 30, paste("Sample mean = ", round(meanSample, 2)), col = "purple4")
text(6.5, 28, paste("Theoretical mean = ", round(meanTheoretical, 2)), col = "green")
```




In this graph, we see that the Sample mean is lesser than the Theoretical mean.

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

```
theory_sd <- round((1/lambda)/sqrt(nExpo), 2)
sample_sd <- round(sd(means), 2)

theory_var <- round(theory_sd ** 2, 2)
sample_var <- round(sample_sd ** 2, 2)

paste('Theoretical variance = ', theory_var)
```

```
## [1] "Theoretical variance =  0.62"
```

```
paste('Sample variance = ', sample_var)
```

```
## [1] "Sample variance =  0.66"
```

```
paste('Theoretical standard deviation = ', theory_sd)
```

```
## [1] "Theoretical standard deviation =  0.79"
```

```
paste('Sample Standard deviation = ', sample_sd)
```

```
## [1] "Sample Standard deviation =  0.81"
```

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

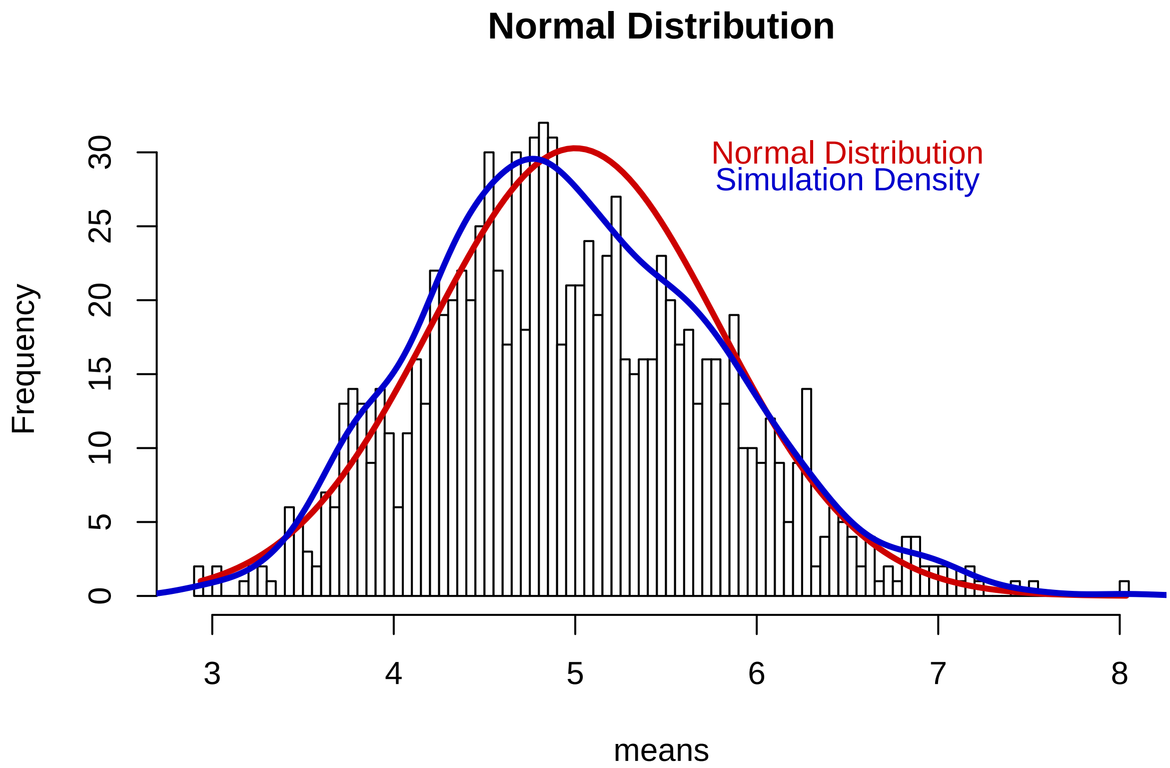
```
hist(means, main = "Normal Distribution", col = "white", breaks = 150) # Plotting the Histogram

# Overlap the normal distribution
```

```
xfit <- seq(min(means), max(means), length = 150)
yfit <- dnorm(xfit, mean = 1/lambda, sd = (1/lambda)/sqrt(nExpo))
lines(xfit, yfit*60, lwd=3, col="red3")

# And the simulation density
den <- density(means)
lines(den$x, den$y*60, lwd=3, col="blue3")

# Provide the legend
text(6.5, 30, "Normal Distribution", col="red3")
text(6.5, 28, "Simulation Density", col="blue3")
```




### Conclusion

In this simulation, we can see that the tail (the mean) of the Simulation density is larger than the tail (the mean) of the Normal distribution.