Statistical Interference Course Project 1:

analysis of normal distribution

by Henrik Gjerning, 27th September 2015

A simulation will be used to explore inference and do some simple inferential data analysis.

This project investiges the exponential distribution in R and compares it with the central Limit Theorem (CLT). The project consists of two parts:

```
1. Setting up simulation parameters and calc of statistics
2. Sample Mean vs Theoretical Mean
3. Sample Variance vs Theoretical Variance
4. Distribution, is it normal?
```

The analysis will compare samples to theorecital distributions to evaluate their similarities based on the parameteres below.

For the analysis we use the following Input:

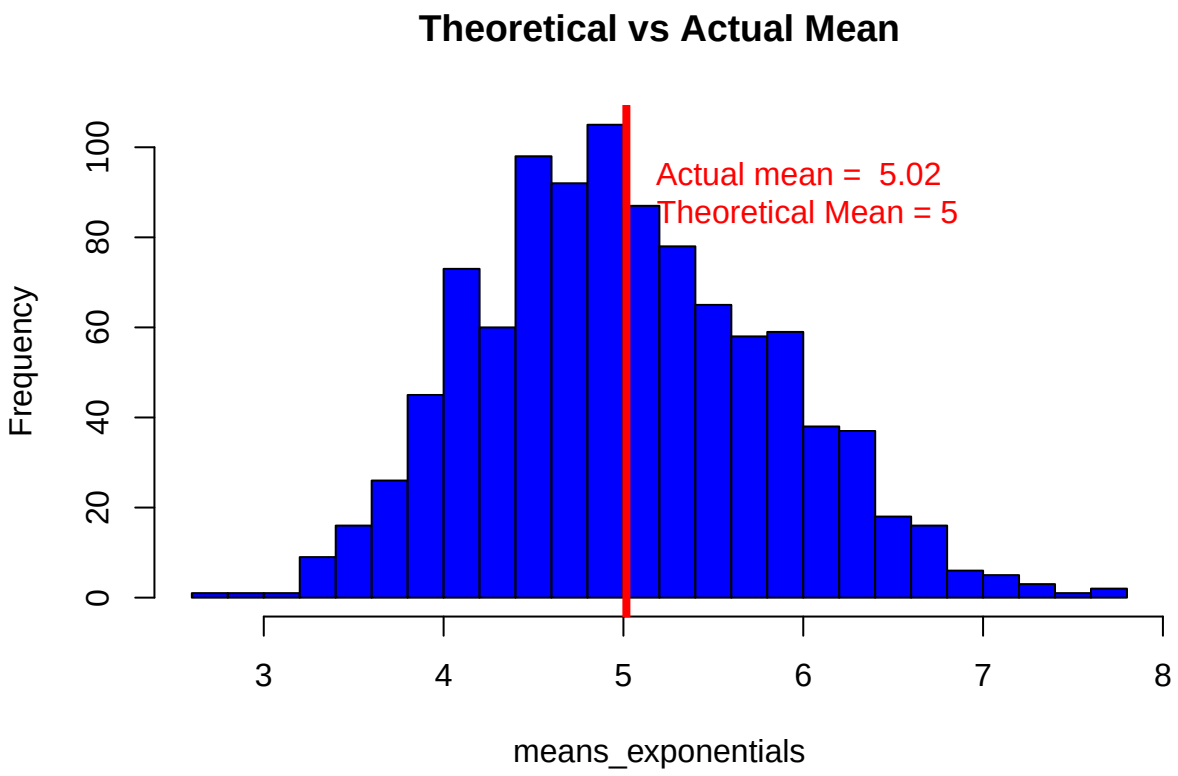
```
a) Lambda:          0.2
b) n:                40
c) simulations:    1000
```

1. Setting up simulation parameters and calc of statistics

```
set.seed(2)
lambda <- 0.2
n <- 40
simulations <- 1000

# simulate
simulated_exponentials <- replicate (simulations, rexp(n, lambda))

# calculate mean of exponentials
means_exponentials <- apply (simulated_exponentials, 2, mean)

actual_mean <- mean (means_exponentials)
theory_mean <- 1/lambda
theory_sd <- ((1/lambda) * (1/sqrt(n)))

actual_sd <- sd(means_exponentials)
theory_var <- theory_sd^2
actual_var <- var(means_exponentials)
```

2. Sample Mean vs Theoretical Mean

```r
hist(means_exponentials, col="blue", main="Theoretical vs Actual Mean", breaks=20)
abline(v=mean(actual_mean), lwd="4", col="red")
text(6, 90, paste("Actual mean = ", round(mean(actual_mean),2), "\n Theoretical Mean = 5" ), col="red")
```




```r
### Results:
cat("The theoretical mean is defined as:\n 1/lambda, giving 1/0.2 = 5 \n whereas the simulated mean is:
cat("\n Giving a difference of: ", 5 - actual_mean)
```

```
## The theoretical mean is defined as:
##  1/lambda, giving 1/0.2 = 5
##  whereas the simulated mean is:  5.016356
##  Giving a difference of:  -0.01635615
```

```
## numeric(0)
```

Conclusion: So the number looks quite similar

3. Sample Variance vs Theoretical Variance

```r
cat(" the theoretical variance is defined as: \n 1/lambda / sqrt(n)^2, giving 1/0.2 / sqrt(40)^2 = 0.625
cat("\n Giving a difference of: ", 0.7906 - actual_var)
```

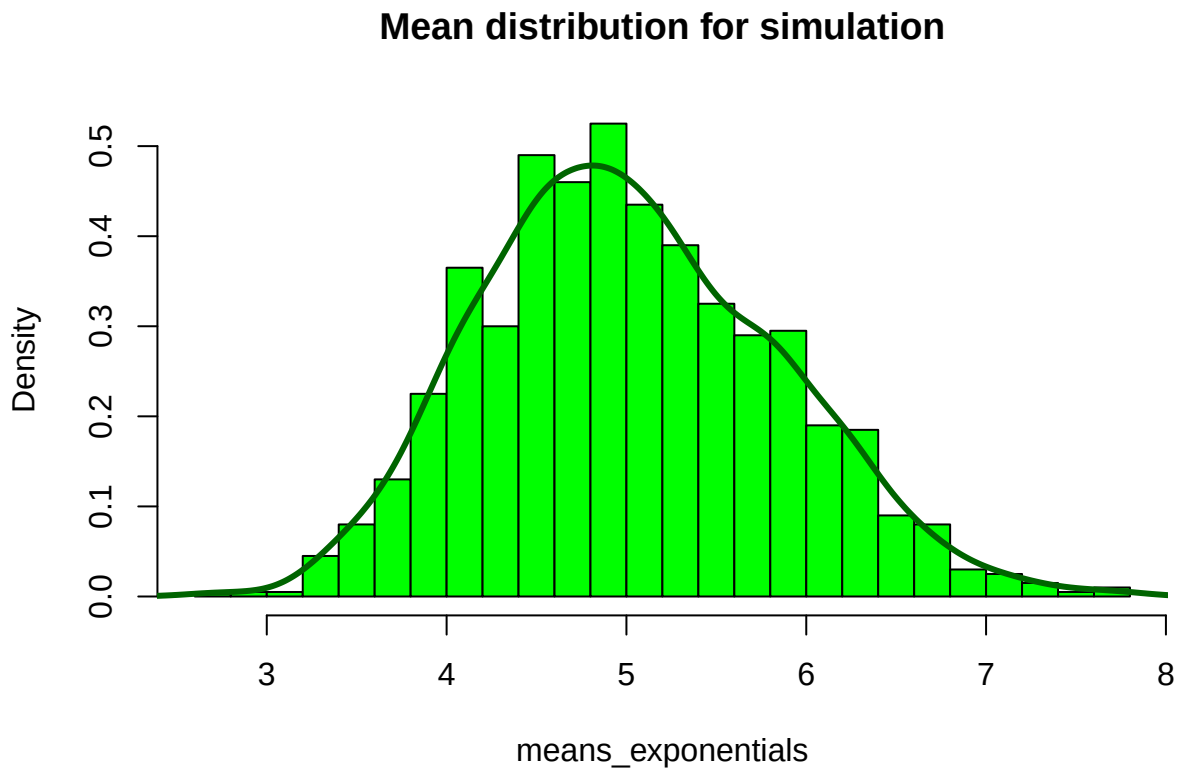
```
##  the theoretical variance is defined as:
##  1/lambda / sqrt(n)^2, giving 1/0.2 / sqrt(40)^2 = 0.625
##  whereas the simulated variance is:  0.6691305
##  Giving a difference of:  0.1214695
```

```
## numeric(0)
```

Conclusion: So the number looks quite similar

4. Distribution, is it normal?

```
hist(means_exponentials, prob=TRUE, col="green", main="Mean distribution for simulation", breaks=20)
lines(density(means_exponentials), lwd=3, col="darkgreen")
```



Conclusion: So the distribution looks normal (gaussian) distributed