

# Statistical Inference Course Project

Marta Regolo

## Part 1: Simulation

In this project we'll investigate the exponential distribution in R and compare it with the Central Limit Theorem. We will investigate the distribution of averages of 40 exponentials. I set  $\lambda=0.2$  for all of the simulations.

```
n<-40
lambda<-0.2
```

The theoretical mean of an exponential distribution is  $1/\lambda$  and the standard deviation is also  $1/\lambda$ .

```
mu<-1/lambda
mu
```

```
## [1] 5
```

```
sd<-1/lambda
sd
```

```
## [1] 5
```

## Theoretical and Sample Distributions

I simulate the exponential distribution 10000 times, taking 10000 random samples with replacement with 40 observations each.

```
B<-10000
x<-rexp(n,lambda)
resamples<-matrix(sample(x,n*B,replace=TRUE),B,n)
means<-apply(resamples,1,mean)
mean(means)
```

```
## [1] 5.878693
```

```
sd(means)
```

```
## [1] 0.821098
```

The *central limit theorem* states that if you have a population mean  $\mu$  and standard deviation  $\sigma$  and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed. This will hold true regardless of whether the source population is normal. For the random samples we take from the population, we can compute the mean of the sample means:

```
mean(means)=mu
```

and the standard deviation of the sample means:

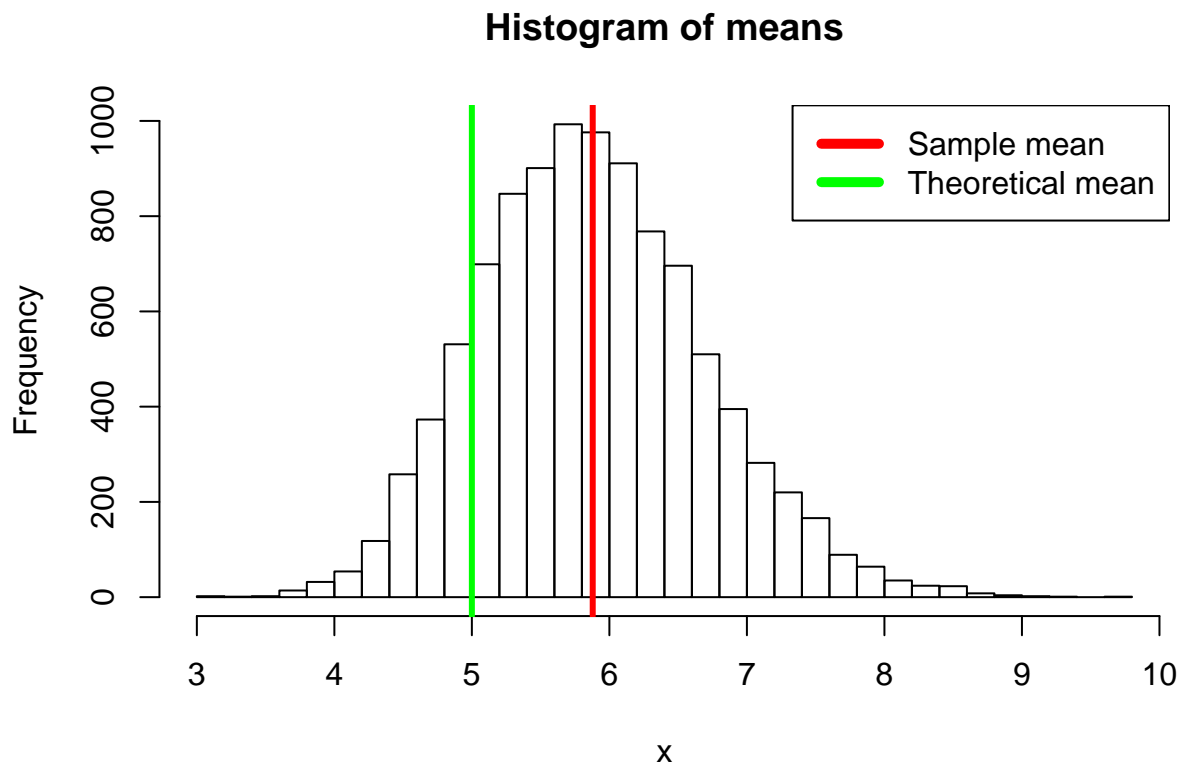
```
sd(means)=sd/sqrt(n)
```

So, in this case, the theoretical mean is:

```
## [1] 5
```

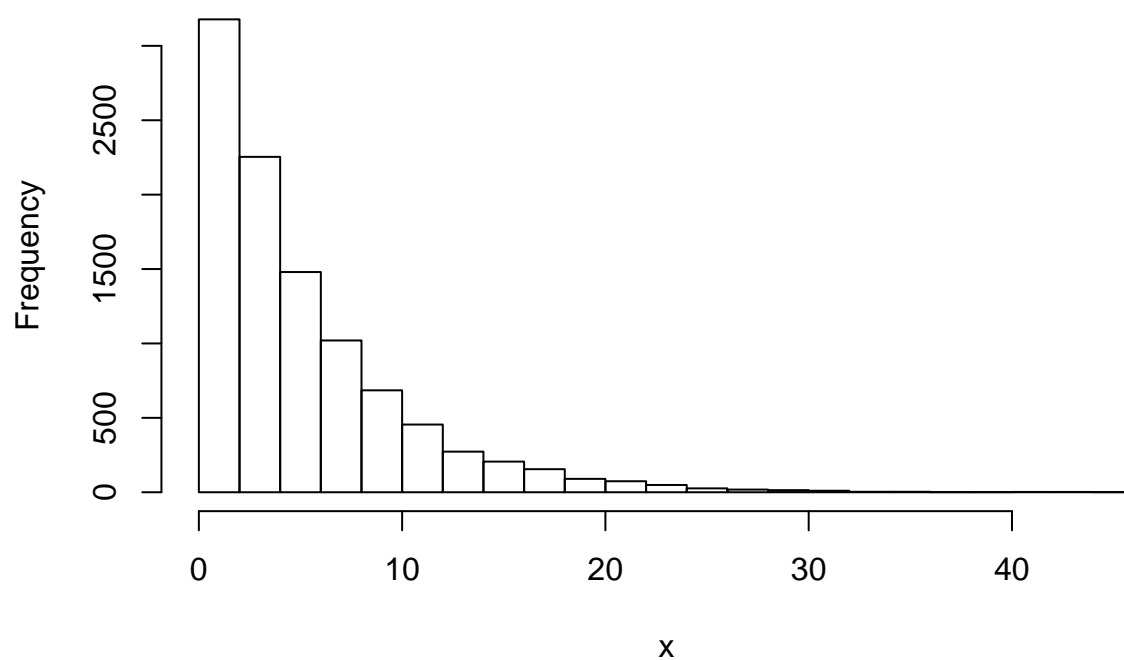
and the theoretical standard deviation is:

```
## [1] 0.7905694
```



If we compare the distribution of 1000 random exponential and the histogram above, we can see that the histogram above is far more Gaussian than the original distribution!

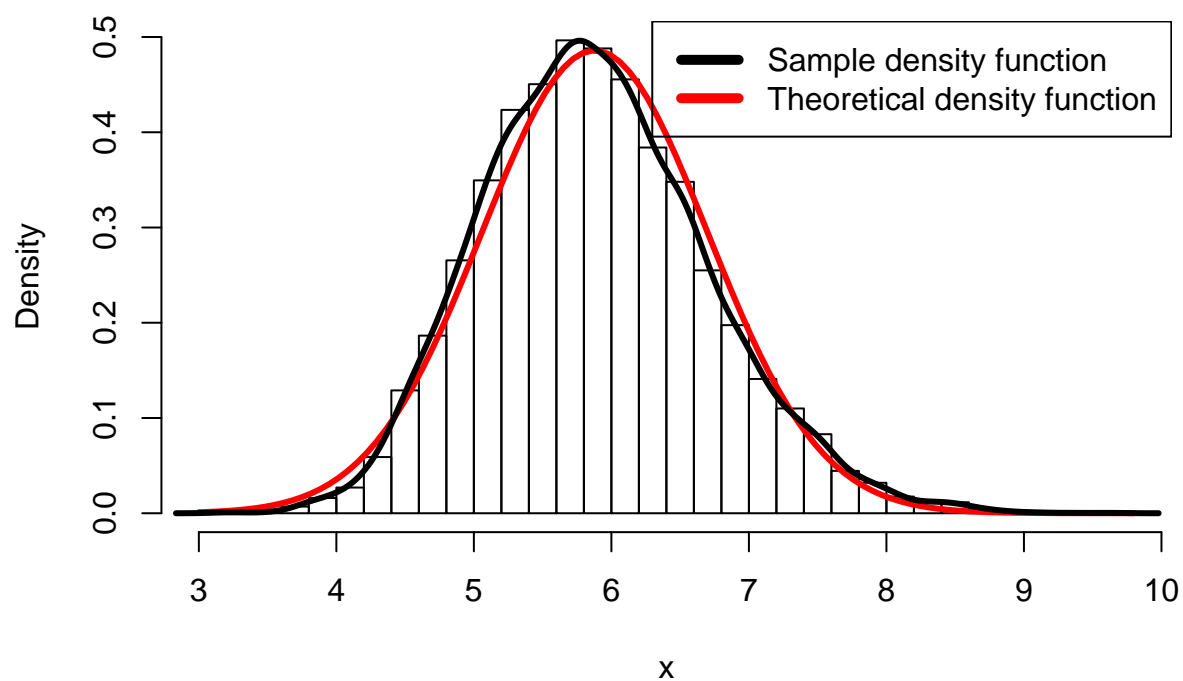
## Exponential distribution



## Normality test

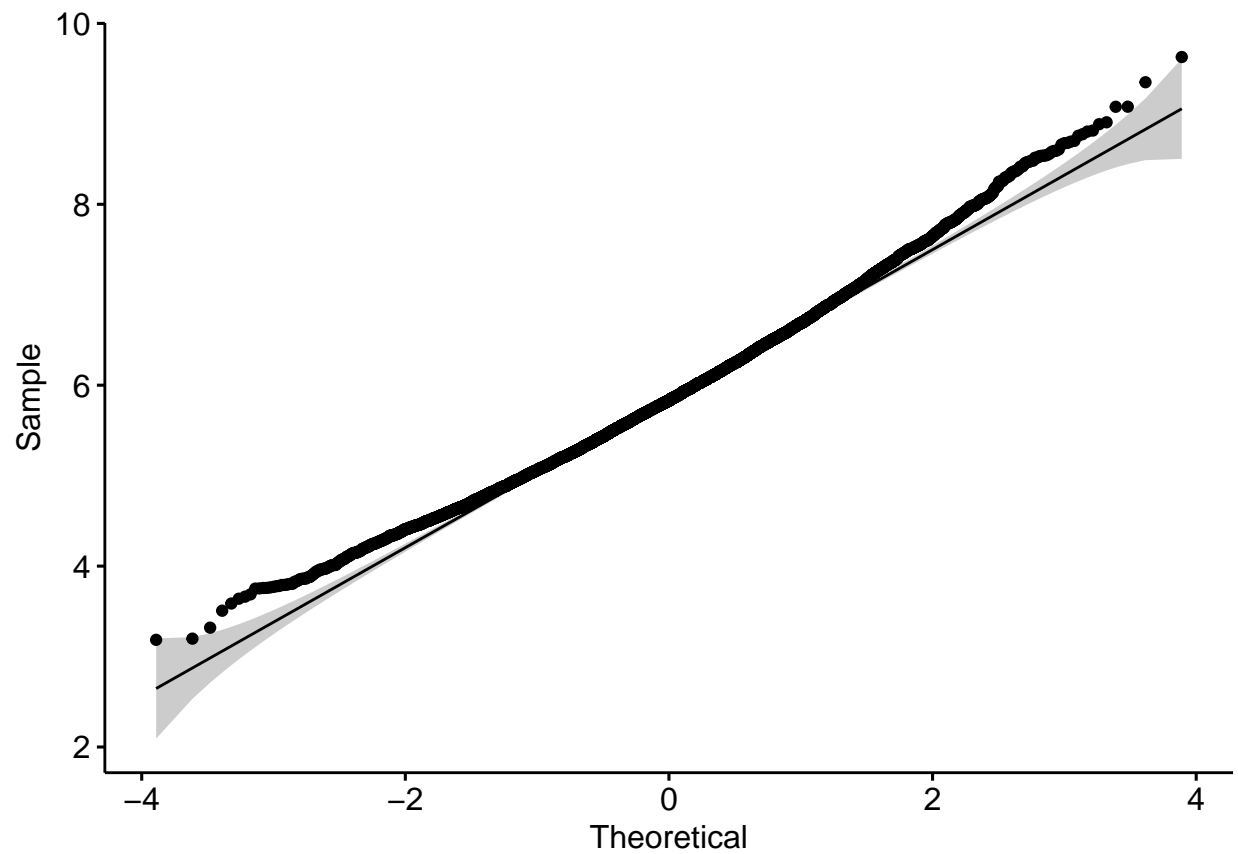
To show if the distribution of means is approximately normal we can use qualitative and visual methods. First, we plot a normal density function with mean `mean(means)` and standard deviation `sd(means)` over the previous histogram.

## Normal distributions comparison



We can see that the distribution is approximately normal.

Second, we use a Q-Q plot (or quantile-quantile plot) that draws the correlation between the given sample and the normal distribution.



Moreover, we can perform an Anderson-Darling test, for an accurate result.

```
ad.test(means,mean=mean(means), sd=sd(means), null="pnorm",estimated=TRUE)$p.value
```

```
## [1] 0.5124836
```

The p-value is greater than 0.05, so we can't reject the null hypothesis of normality.