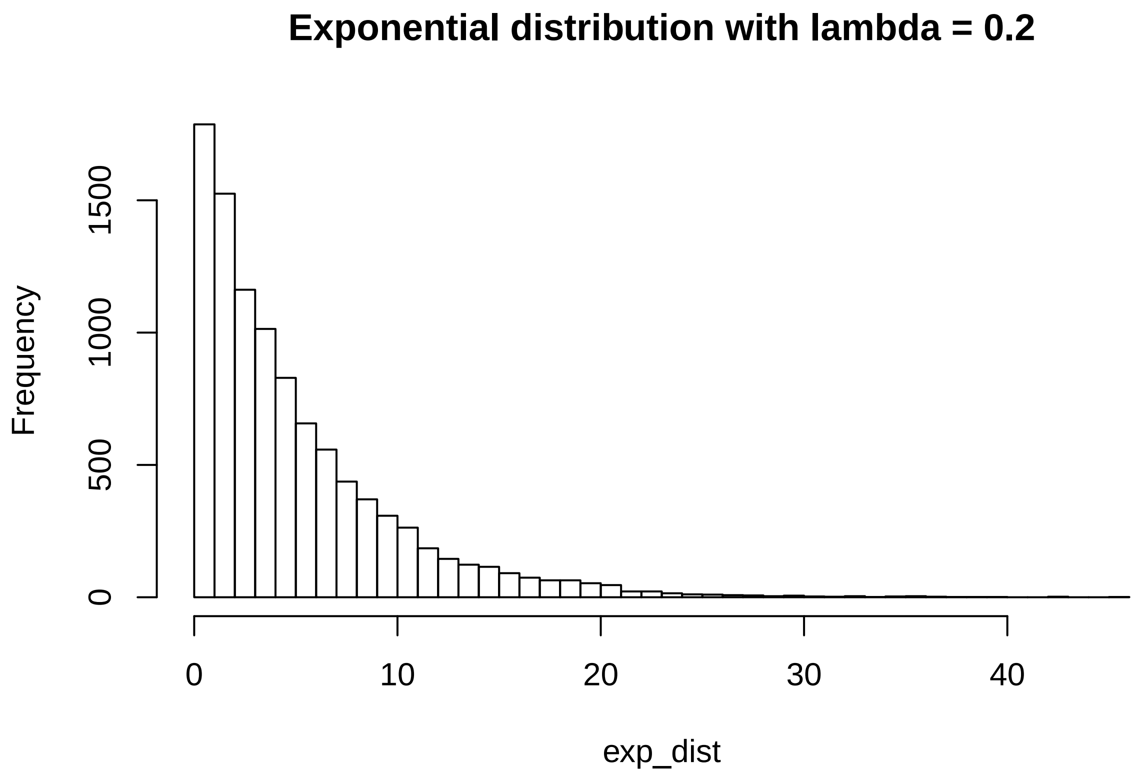
Statistical Inference - Project - Claudio Bogazzi

In this markdown file we are going to investigate the exponential distribution and compare it with the Central Limit Theorem. First, let us define our exponential distribution

```
lambda <- 0.2
size <- 40 #size of the sample
#mean and sd of the exponential distribution are defined as 1/lambda
mean <- 1/lambda
std <- 1/lambda
```

Now, we plot it:



Pretty, eh? Now we have to simulate the distribution of the mean of 40 exponentials.

```
#average of 40 exponentials
nsim <- 1000
Avexp2 <- data.frame(mean=numeric(nsim))

for (i in 1:nsim) {
  exp2<-rexp(size,lambda)     # simulation set of n exponential with l lambda
  Avexp2[i,1]<-mean(exp2) # mean of simulation set
}
```

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

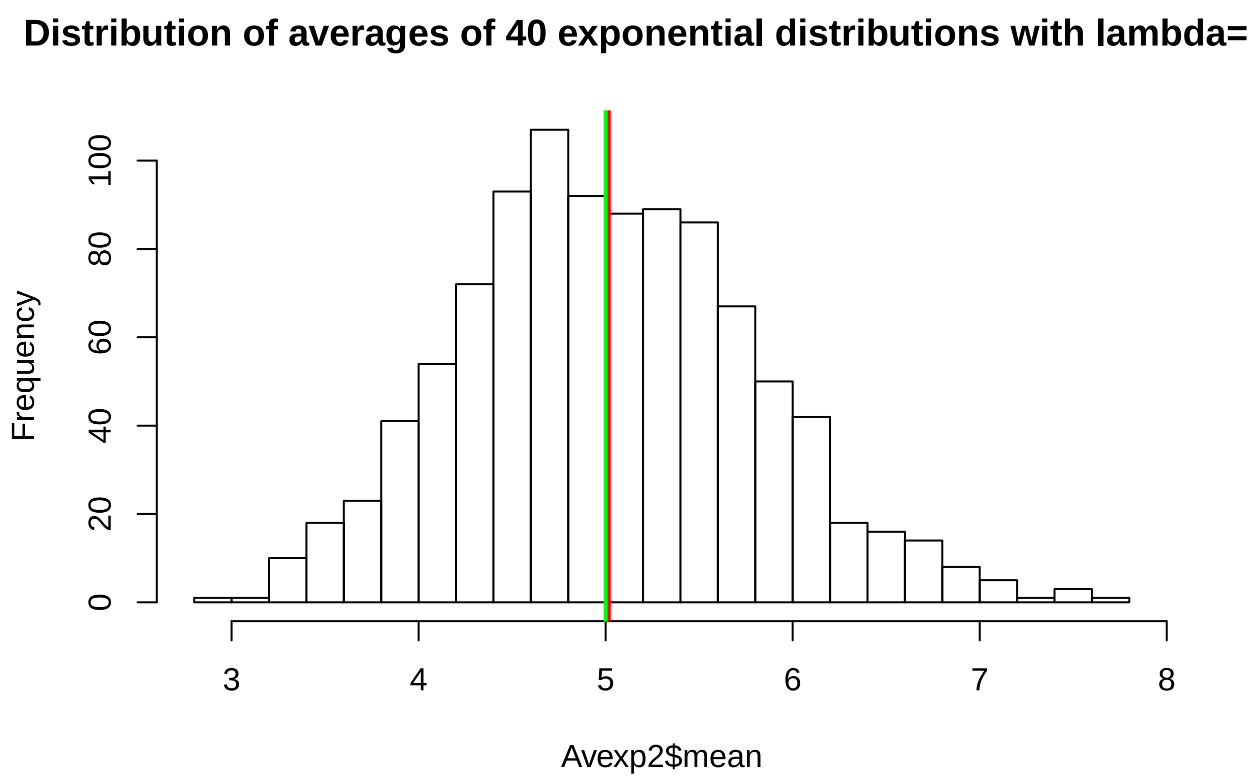
```r
#mean of new distribution and comparison with theoretical mean
smean = mean(Avexp2$mean)
cat("The sample mean is: ",smean)
```

```
## The sample mean is:  5.016402
```

```r
cat("The theoretical mean is: ", mean)
```

```
## The theoretical mean is:  5
```

As expected, the sample mean and the theoretical mean are very close. Let'see in a plot:



The two vertical lines represents the theoretical mean (red) and the sample mean (green).

Now we focus on the variance.

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

We first compute the theoretical standard deviation of sample means:

```r
theo_sd <- std / sqrt(size)
```

and now the standard deviation of the sample means

```r
ssd <- sd(Avexp2$mean)
```

The results can now be compared:

```
cat("The theoretical standard deviation is ", theo_sd)
```

```
## The theoretical standard deviation is  0.7905694
```

```
cat("The standard deviation of the sample means is ", ssd)
```

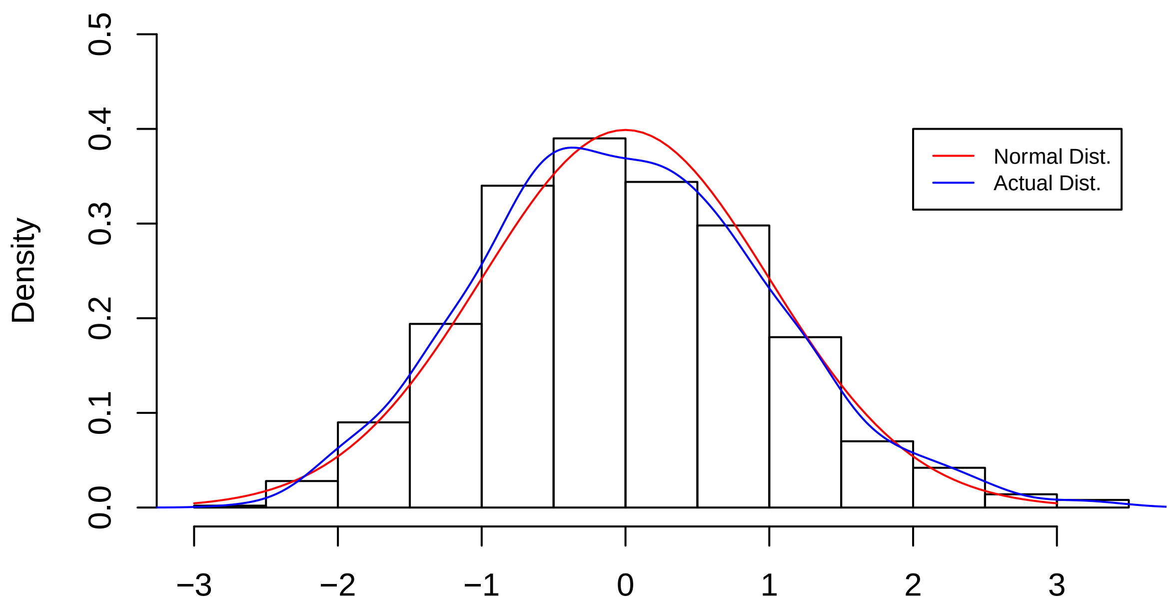
```
## The standard deviation of the sample means is  0.7845894
```

Now, the last part of the project.

3: Show that the distribution is approximately normal.

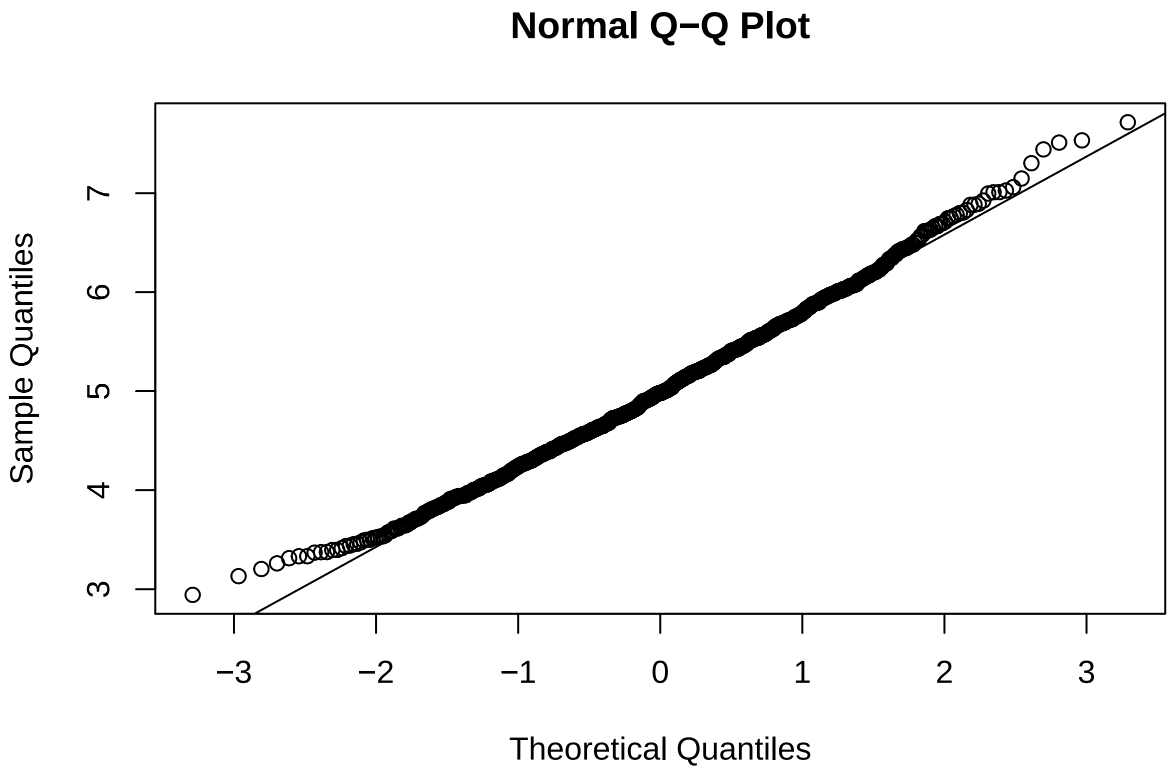
The best way to do this is to plot the distribution and compare it with a normal:

```
hist(scale(Avexp2$mean),probability="T",main='',ylim=c(0,0.5),xlab='')
curve(dnorm(x,0,1),-3,3, col='red',add=T) # normal distribution
lines(density(scale(Avexp2$mean)),col='blue') # actual distribution
legend(2,0.4,c('Normal Dist.','Actual Dist.'),cex=0.7,col=c('red','blue'),lty=1)
```



Another way to show how the distribution is approximately normal is with a quqnatile-quantile plot:

```
qqnorm(Avexp2$mean); qqline(Avexp2$mean)
```



Summary This simulation is performed with a population of 40,000 samples (1000 simulations x 40 samples). We conclude that the mean of all samples is approximately equal to the mean of the population. The normal distribution of the mean of 40 random exponentials is consistent with the Central Limit Theorem.

Appendix - R code

```
lambda <- 0.2
size <- 40 #size of the sample
#mean and sd of the exponential distribution are defined as 1/lambda
mean <- 1/lambda
std <- 1/lambda
set.seed(123)
exp_dist <- rexp(10000, lambda)
h <- hist(exp_dist, breaks=50,main="Exponential distribution with lambda = 0.2")
#average of 40 exponentials
nsim <- 1000
Avexp2 <- data.frame(mean=numeric(nsim))

for (i in 1:nsim) {
  exp2<-rexp(size,lambda)     # simulation set of n exponential with l lambda
  Avexp2[i,1]<-mean(exp2) # mean of simulation set
}
#mean of new distribution and comparison with theoretical mean
smean = mean(Avexp2$mean)
cat("The sample mean is: ",smean)
cat("The theoretical mean is: ", mean)
hist(Avexp2$mean, breaks=30,main="Distribution of averages of 40 exponential distributions with lambda=

abline(v=smean,col='red',lwd=2)
abline(v=mean,col='green',lwd=2)
```

```r
theo_sd <- std / sqrt(size)
ssd <- sd(Avexp2$mean)
cat("The theoretical standard deviation is ", theo_sd)
cat("The standard deviation of the sample means is ", ssd)


hist(scale(Avexp2$mean),probability="T",main='',ylim=c(0,0.5),xlab='')
curve(dnorm(x,0,1),-3,3, col='red',add=T) # normal distribution
lines(density(scale(Avexp2$mean)),col='blue') # actual distribution
legend(2,0.4,c('Normal Dist.','Actual Dist.'),cex=0.7,col=c('red','blue'),lty=1)

qqnorm(Avexp2$mean); qqline(Avexp2$mean)
```