Project 1: a study on the exponential distribution

0) Introduction

In this short note, we will try to study some properties of the average of 40 simulations of the expponential distribution with a parameter $\lambda = 0.2$. The first part will be devoted to the study of the mean of 10000 samples of such simulations. The second part will focus on its variance, while the third part will try to test the normality of the resulting distribution.

I) Sample Mean

We start by genrating our 10000 samples of averages of 40 simulations of the expponential distribution with a parameter $\lambda = 0.2$. The seed is set in order being able to reproduce the result.

```
set.seed(100)
lambda<-0.2
samplexp<-NULL
for (i in 1 : 10000) samplexp <- c(samplexp, mean(rexp(40,lambda)))
```

Since the mean is linear, we expect to find the mean of `samplexp` to be $1/\lambda$. Let us check this

```
meanexp<-mean(samplexp)
meanexp
```

```
## [1] 4.994001
```

This corresponds to a small deviation of

```
 library('scales')
percent(lambda*meanexp-1)
```

```
## [1] "-0.12%"
```

thanks to the Central Limit Theorem, more on this deviation will be said in part III

Sample Variance

the variance of our sample is given by

```
sdexp<-sd(samplexp)
sdexp
```

```
## [1] 0.7911362
```

The first question to ask is: whas was the expected result? The variance of the exponential distribution is equal to its mean: $1/\lambda$. But for the average of 40 simulations, we expect $1/(\lambda\sqrt{40})$. This is equal to

```
1/(lambda*sqrt(40))
```

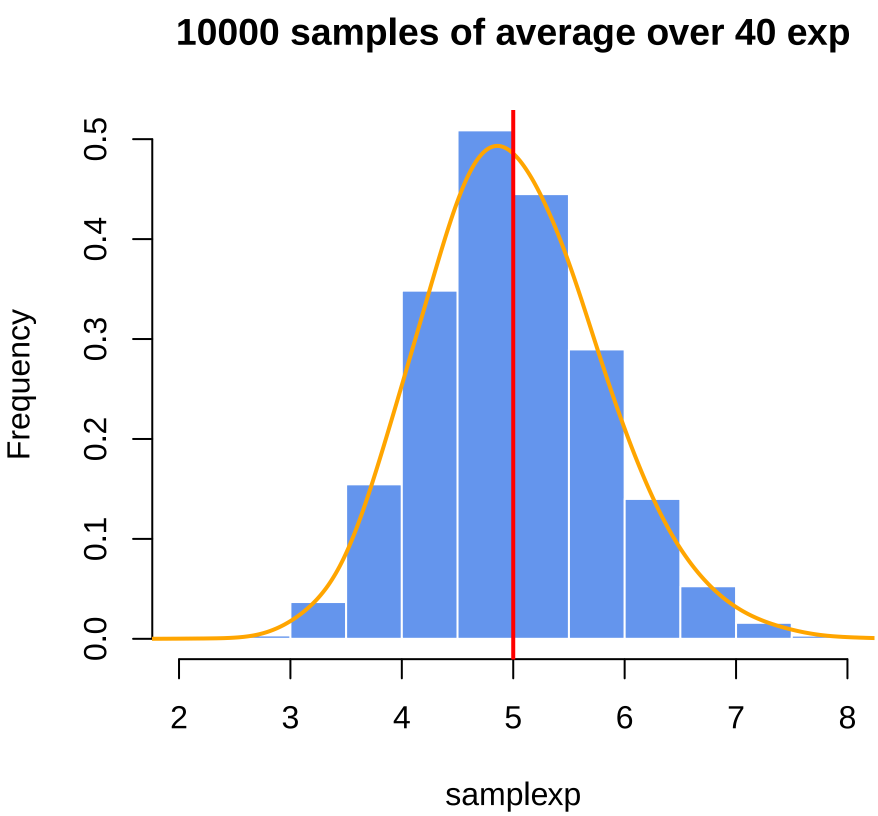
```
## [1] 0.7905694
```

This is again pretty close to what we found. More in this in the next part

Normality of the resulting distribution

Let us now look at the distribution of our 10000 samples (the orange line being a Gaussian fit to the histogram). The red color line indicate the value of `meanexp`

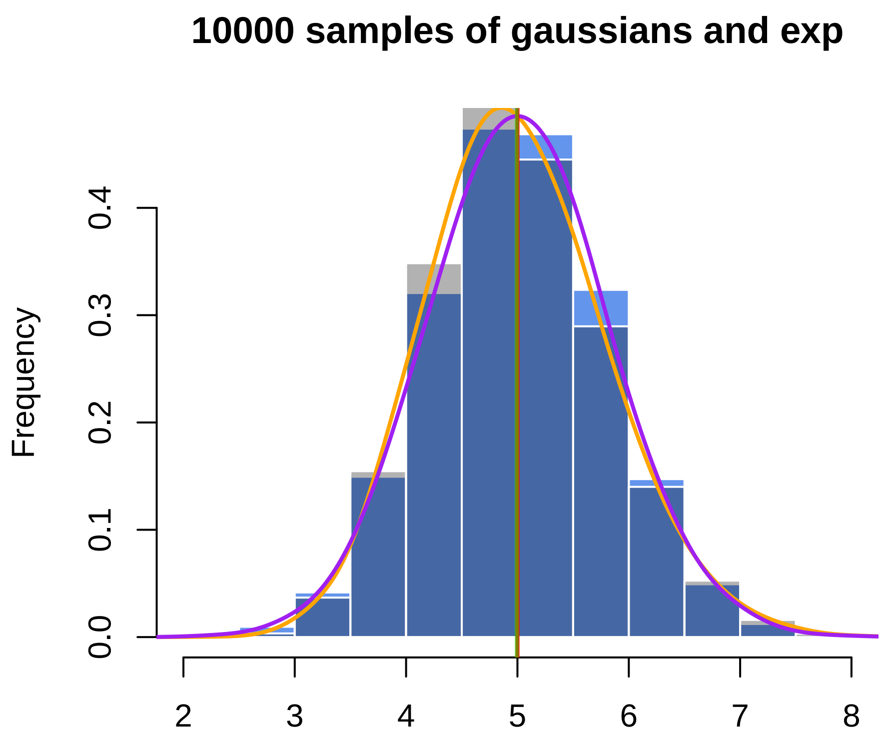
```
hist(samplexp, col="cornflowerblue", border="white",
    main="10000 samples of average over 40 exp",
    xlab="samplexp", ylab="Frequency",breaks=15,prob=TRUE,xlim = c(2,8))
lines(density(samplexp,adjust=2), lwd=2, col="orange")
abline(v=5,lwd=2,col="red")
```



And for comparison, let us do the same thing but put on top of it a normal distribution with a mean $\mu = 1/\lambda$ and a standard deviation $\sigma = 1/(\lambda\sqrt{40})$. The orange and purple lines being Gaussian fits to the histograms. The exp histogram is still in blue while the Gaussian one is black (and semi-transparant to ease readability). The color lines indicate the means of the distributions.

```
sampgauss<-rnorm(10000,mean=5,sd=5/sqrt(40))
hist(sampgauss, col="cornflowerblue", border="white",
    main="10000 samples of gaussians and exp",
```

```r
    xlab="", ylab="Frequency",breaks=15,prob=TRUE,
    xlim = c(2,8))
hist(samplexp, col=alpha("black",0.3), border="white",
    main="10000 samples of average over 40 exp",
    xlab="the two distributions", ylab="Frequency",breaks=15,prob=TRUE,xlim = c(2,8),add=T)
lines(density(samplexp,adjust=2), lwd=2, col="orange")
lines(density(sampgauss,adjust=2), lwd=2, col="purple")
abline(v=5,lwd=2,col="red")
abline(v=meanexp,lwd=2,col=alpha("green",0.5))
```



This is pretty similar! How is this? To understand that one just needs to recall the central limits theorem, that states that for X_i iid (with here i ranging from 1 to 10000), one has

$$\bar{X}_i \to \mathcal{N}(\mu, \sigma^2/40)$$

What we just found nicely illustrates this fact.