Exponential Distribution Investigation and comparison with CLT

Donato Scarano

03 luglio 2018

Introduction

We will investigate the exponential distribution in R and 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.

We have set lambda = 0.2 for all of the simulations.

We investigate the distribution of averages of 40 exponentials and we will do 1000 simulations.

We illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

```
set.seed(1000)
lambda<-0.2
n<-40
simulation<-1000
simdata<-replicate(simulation,rexp(n,lambda))
str(simdata)
```

```
##  num [1:40, 1:1000] 5.02 2.59 12.19 10.82 2.39 ...
```

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

Let us calculate the sample mean of the simulations above:

```
meandata<-apply(simdata,2,mean)
str(meandata)
```

```
##  num [1:1000] 4.51 5.05 3.25 3.92 4.9 ...
```

```
meansim<-mean(meandata)
meansim
```

```
## [1] 4.986963
```

then we calculate both the theoretical and the simulation mean, standard deviation, variance:

```
print("THEORETICAL")
```

```
## [1] "THEORETICAL"
```

```
meantheory<-1/lambda
meantheory
```

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

```r
sdtheory <- (1/lambda)/sqrt(n)
sdtheory
```

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

```r
vartheory <- sdtheory^2
vartheory
```

```
## [1] 0.625
```

```r
print("SIMULATION")
```

```
## [1] "SIMULATION"
```

```r
meansim<-mean(meandata)
meansim
```

```
## [1] 4.986963
```

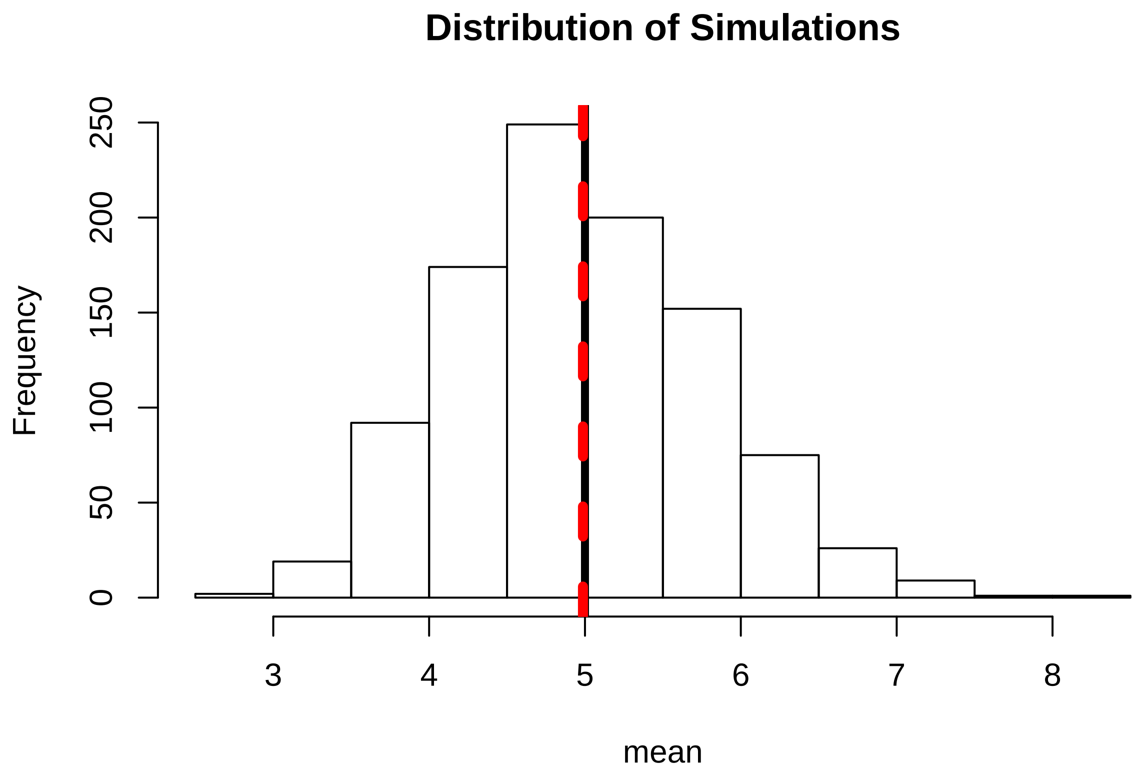
```r
sdsim<-sd(meandata)
sdsim
```

```
## [1] 0.8089147
```

```r
varsim<-sdsim^2
varsim
```

```
## [1] 0.654343
```

```r
hist(meandata,xlab="mean",main="Distribution of Simulations")
abline(v=meantheory,col="black",lwd=4, lty=1)
abline(v=meansim, col="red",lwd=5,lty=2)
```



As we can see the theoretical mean is equal to 5 (black line) and is very close to the mean from the simulation (red dotted line) which is equal to 4.99.

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

Theoretical Variance:

```
"THEORETICAL VARIANCE"
```

```
## [1] "THEORETICAL VARIANCE"
```

```
vartheory <- sdtheory^2
vartheory
```

```
## [1] 0.625
```

```
"SIMULATION VARIANCE"
```

```
## [1] "SIMULATION VARIANCE"
```

```
varsim<-sdsim^2
varsim
```

```
## [1] 0.654343
```

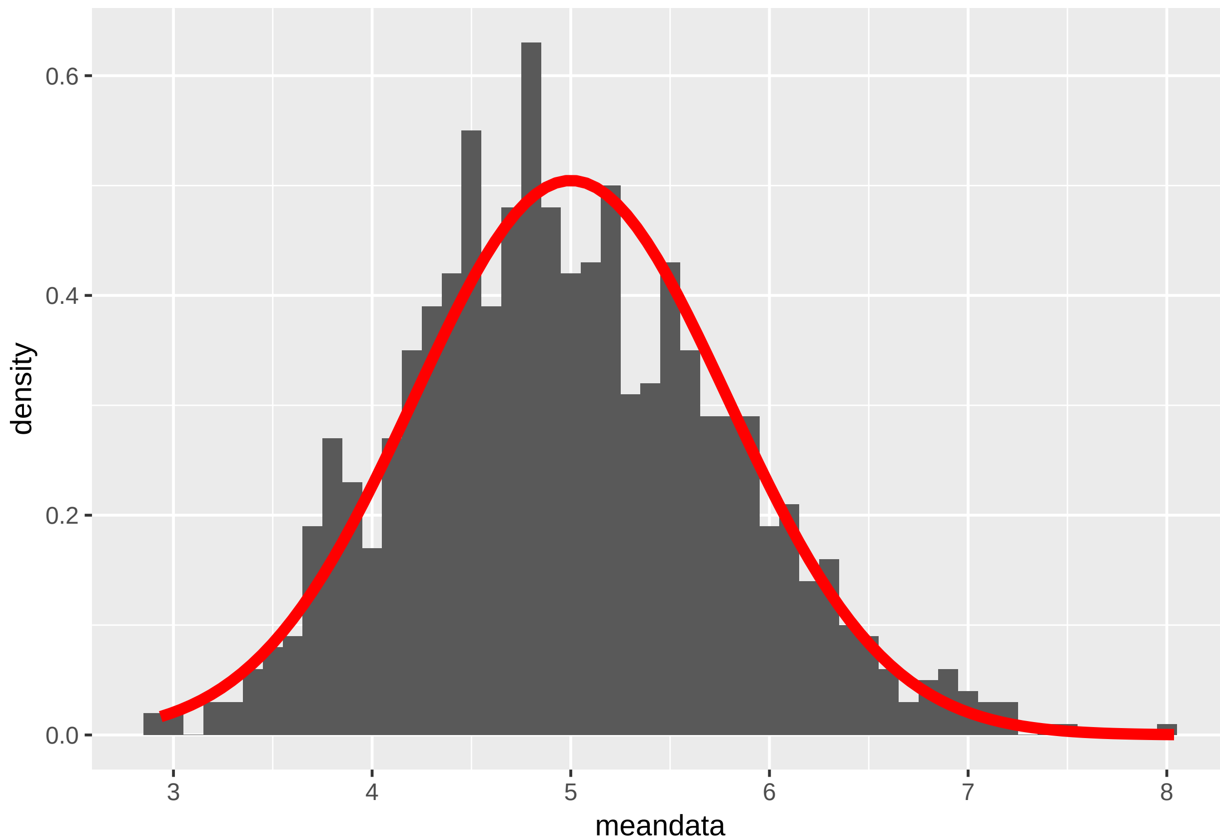
Show that the distribution is approximately normal.

We focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.1
```

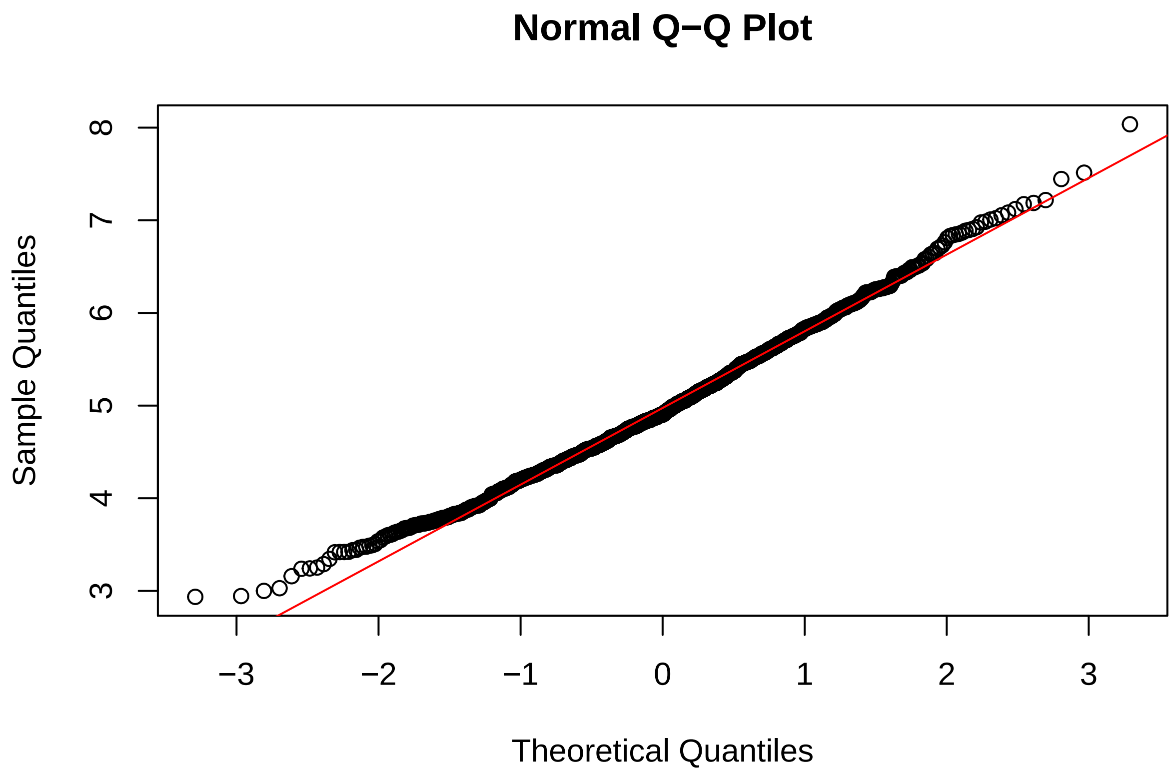
```
meanframe<-data.frame(meandata)
g<-ggplot(data=meanframe,aes(x=meandata))
g<-g+geom_histogram(aes(y=..density..),binwidth = 0.1)
g<-g+stat_function(fun=dnorm,args = list(mean=meantheory,sd=sdtheory),lwd=2, color="red")
g
```



As we can see on graph above the distribution is approximately normal.

As for the averages of 40 exponentials:

```
qqnorm(meandata)
qqline(meandata, col = 2)
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
3
4
5
6
7
8
−3
−2
−1
0
1
2
3