The Exponential Distribution and the CLT

Marco Totolo

Overview

This document will show how the distribution of averages of N exponential random variables follows the Central Limit Theorem. This means that as the number of sample means increases, the distribution asymptotically tends to a normal distribution with mean equal to the exponential mean and standard deviation equal to the exponential standard deviation divided by the square root of N.

Simulations

Let's first build an array of means of random exponential variables.

As specified in the assignment, I will take 1000 means of 40 random exponential variables with lambda 0.2.

```
library(ggplot2)
set.seed(18)
lambda<-0.2
N<-40
nosim<-1000
expMeans <- NULL
for (i in 1 : nosim) {
      expMeans = c(expMeans, mean(rexp(N,lambda)))
}
```

Sample Mean versus Theoretical Mean

The sample mean is

```
mean(expMeans)
```

```
## [1] 4.9928
```

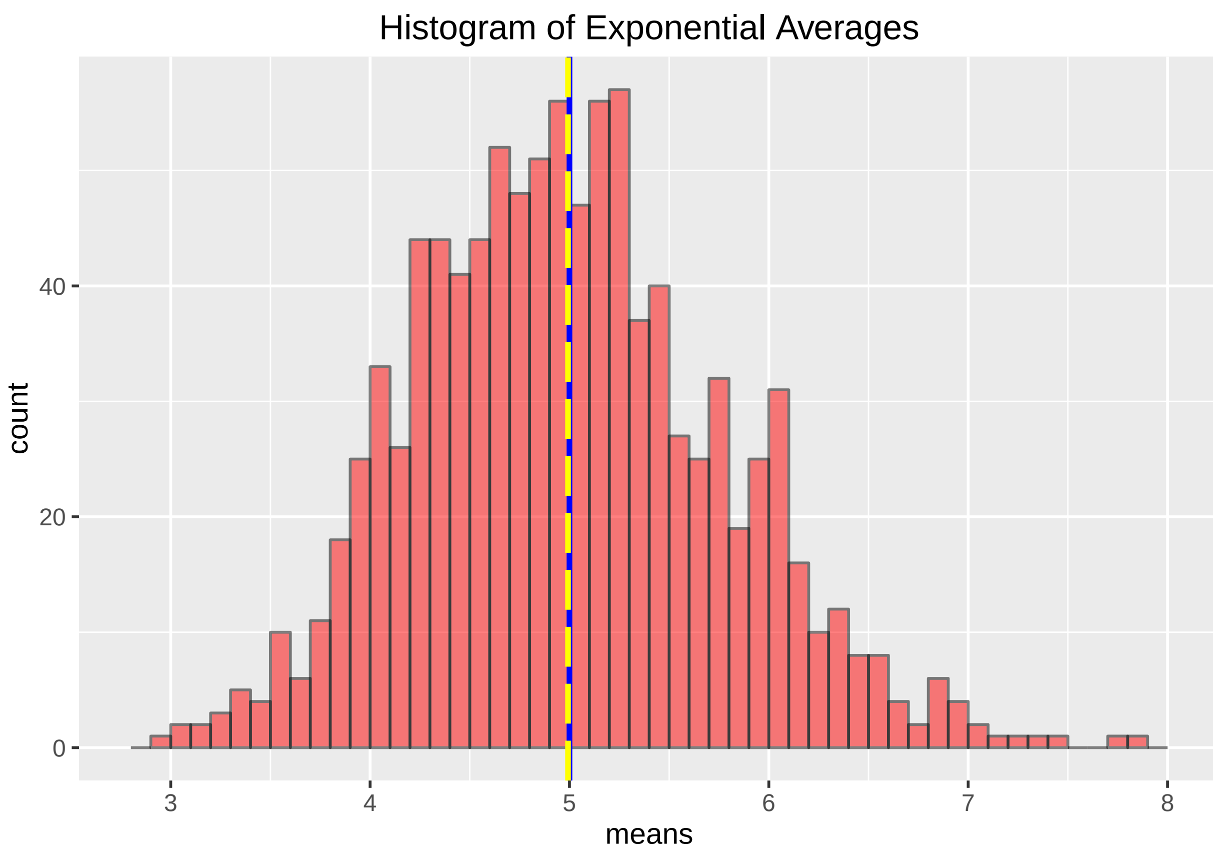
which is very close to the the CLT theoretical value, the mean of the distribution of the averages of exponential variables, 1/lambda.

```
1/lambda
```

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

To show how close the values are, I will plot them on top of the histogram of the simulated means.

```r
dfexpMeans <-as.data.frame(expMeans)
plot <- ggplot(dfexpMeans,aes(x=expMeans))
plot <- plot + geom_histogram(binwidth=.1,
                 fill="red", col="black",alpha=0.5)+
            geom_vline(xintercept=1/lambda,col="blue",size=1)+
            geom_vline(xintercept=mean(expMeans),col="yellow",size=1,
                       linetype="longdash") +
            xlab("means") +
            ggtitle("Histogram of Exponential Averages")
plot
```



The blue vertical line is the theoretical value and the yellow dashed one is the mean of the simulated data.

Sample Variance versus Theoretical Variance

The sample variance is

```r
var(expMeans)
```

```
## [1] 0.5910052
```

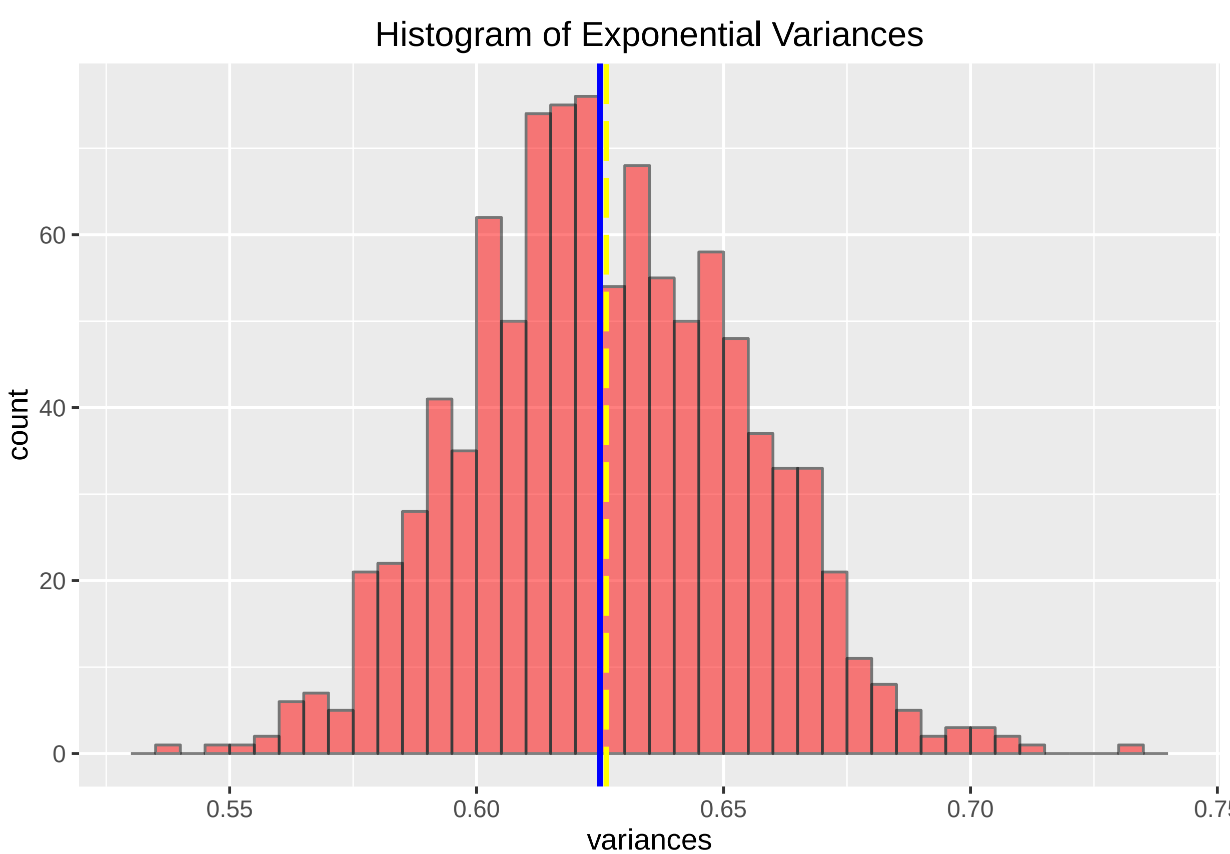
which is very close to the the CLT theoretical value, the variance of the distribution of the averages of exponential variables, 1/(lambda^2*N)

```
1/(lambda^2*N)
```

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

I will now plot a histogram of the variances of the averages of simulated data. For each simulated average distribution I only have one variance value, so I have to simulate much more data to plot a histogram for the variance.

```
expVars<-NULL
for (j in 1: nosim) {
      expMeans2 <- NULL
      for (i in 1 : nosim) {
            expMeans2 = c(expMeans2, mean(rexp(N,lambda)))
      }
      expVars<-c(expVars,var(expMeans2))
}
dfexpVars <-as.data.frame(expVars)
plot <- ggplot(dfexpVars,aes(x=expVars))
plot <- plot + geom_histogram(binwidth=.005,
                  fill="red", col="black",alpha=0.5)+
            geom_vline(xintercept=1/(lambda^2*N),col="blue",size=1)+
            geom_vline(xintercept=mean(expVars),col="yellow",size=1,
                       linetype="longdash") +
            xlab("variances") +
            ggtitle("Histogram of Exponential Variances")
plot
```



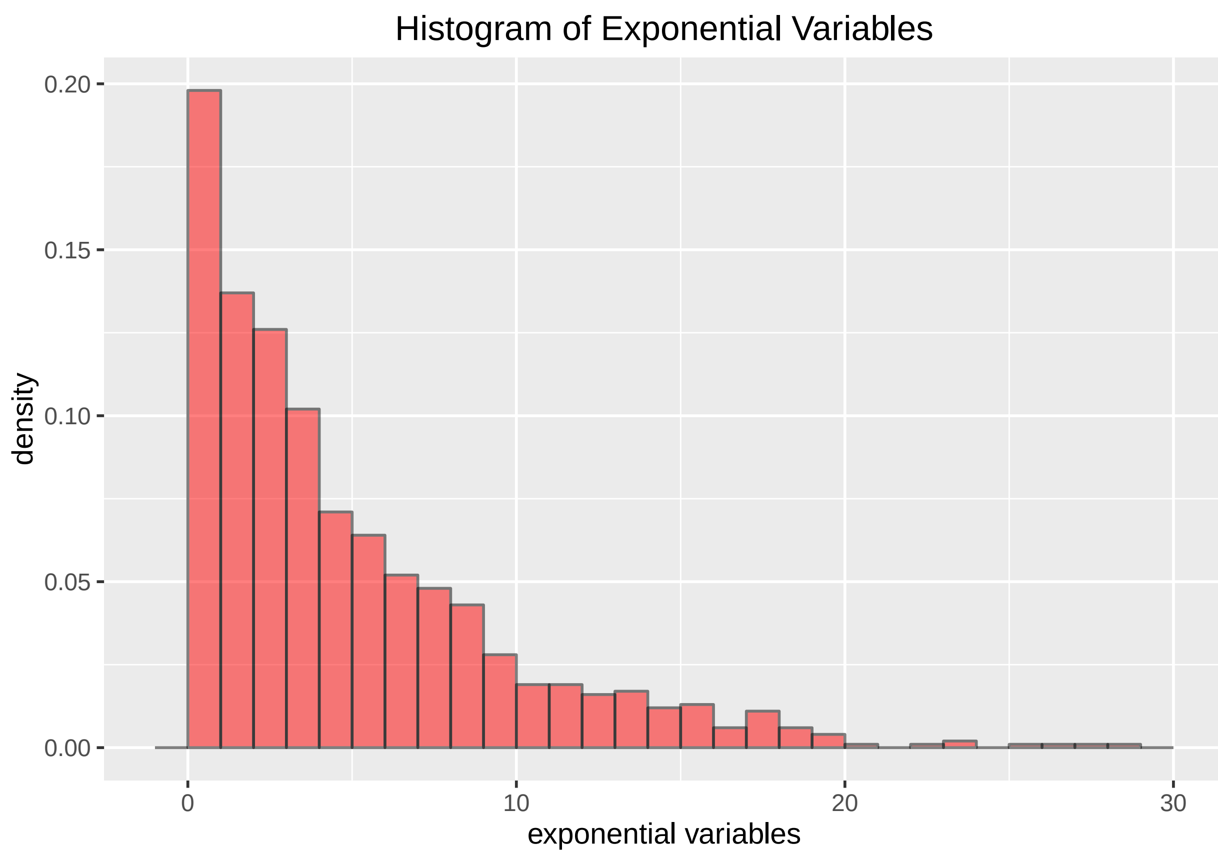
Again, the blue vertical line is the theoretical value of the variance and the yellow dashed one is the mean of the variance of the simulated data.

Note that since I've used a different set of simulations for this histogram, the variance will be different from the one previously calculated. More precisely, it will be much closer to the theoretical value.

Distribution

Let's take a look at the distribution of 1000 exponential variables.

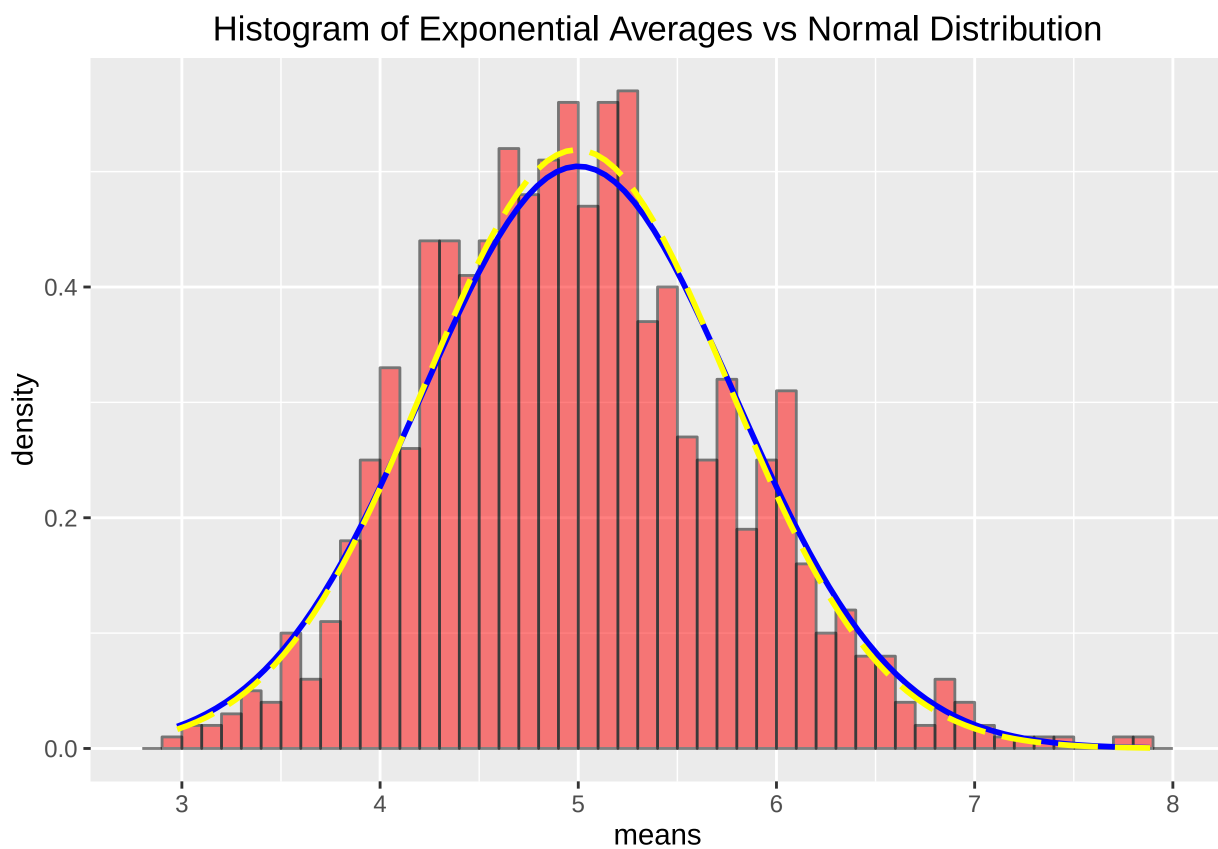
```
exp<-rexp(1000,lambda)
dfexp <-as.data.frame(exp)
plot <- ggplot(dfexp,aes(x=exp))
plot <- plot + geom_histogram(aes(y=..density..),binwidth=1,
            fill="red", col="black",alpha=0.5) +
            xlab("exponential variables") +
            ggtitle("Histogram of Exponential Variables")
plot
```



This of course doesn't look like a normal distribution at all. The CLT is valid for the distribution of the *avarages* of iid variables.

To show how the distribution of the avarages resembles a normal distribution I will plot a histogram of the simulated data and two normal distribution with mean and standard deviation as specified in the paragraphs above.

```r
library(ggplot2)
dfexpMeans <-as.data.frame(expMeans)
plot <- ggplot(dfexpMeans,aes(x=expMeans))
plot <- plot + geom_histogram(aes(y=..density..),binwidth=.1,
            fill="red", col="black",alpha=0.5)
plot <- plot + stat_function(fun=dnorm, size=1,
            col="blue", args=list(mean=1/lambda,sd=1/(lambda*sqrt(N))))
plot <- plot + stat_function(fun=dnorm, size=1, linetype="longdash",
            col="yellow", args=list(mean=mean(expMeans),sd=sd(expMeans)))+
      xlab("means") +
      ggtitle("Histogram of Exponential Averages vs Normal Distribution")
plot
```



The normal density in blue is drawn with the theoretical parameters, while the one in yellow is drawn with the values calculated from the sample data. We can see that the two curves are very similar.