StatInference_Project - Part A

Georgia P

September 14, 2015

Overview

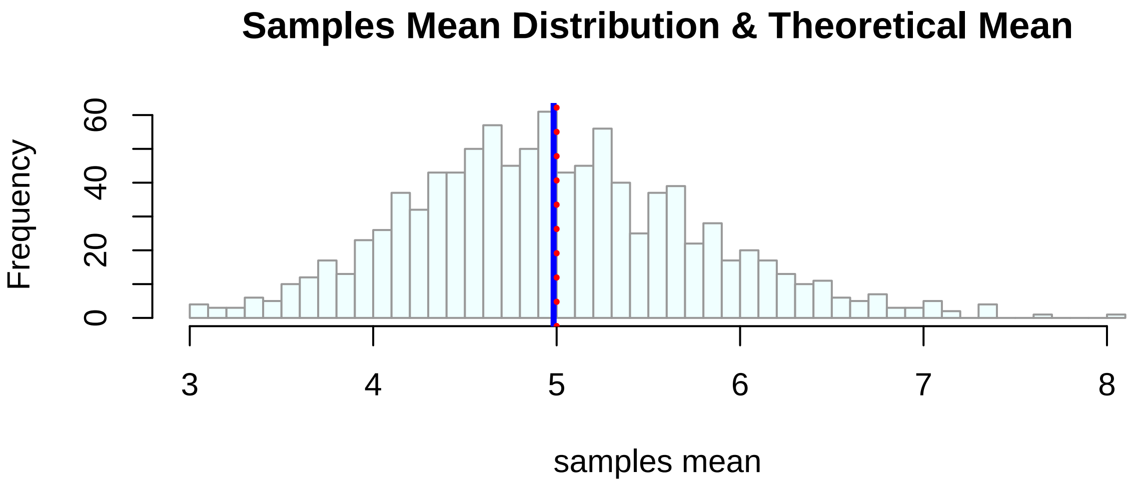
In this project I will investigate the **exponential distribution** in R and compare it with the **Central Limit Theorem**. The **Central Limit Theorem** in brief describes that the distribution of the means of iid variables tends to be a standard normal distribution. **Conclusions**: With a series of simulations of exponentials, and calculations of their means and st.deviations it is demonstrated that they present a normal distribution. ##Simulations First, I will create 1000 random exponential distributions of iid variables, all with rate **lambda = 0.2** and # of samples n=40. For each one of these I will be calculating and saving their mean.

```
#Setting lambda = 0.2 for all of the simulations
lambda <- 0.2
#calculate the mean, the standard deviation and the variance of each exponential distribution
mean<- 1/lambda ; std<- 1/lambda ; var= std^2 ; n<-40
#running 1000 simulations of exponential distributions, all with lambda=0.2, n=40 .For each one of thes
ens<- NULL; sens<- NULL; mns<-NULL ; vns<- NULL
for (i in 1 : 1000) {
#run simulation of exponential & add to the estimates table
dist<- rexp(n, lambda) ; ens<-c(ens, dist)
#means  & variances of 1000 simulated distributions, aka mns and vns
mns <- c(mns, mean(dist)); vns <- c(vns, sd(dist)^2 ) }
sample_mean <- mean(mns) ; sample_var<- mean(vns)
```

Sample Mean versus Theoretical Mean

In the following diagram, I will be demonstrating the distribution of the means of the simulated exponentials vs the the theoretical mean, which only depends on the lambda. As **lambda does not change across different distributions**, their **theoretical mean is also not changing** and equal to 1/lambda, therefore **Theoretical mean = 5**

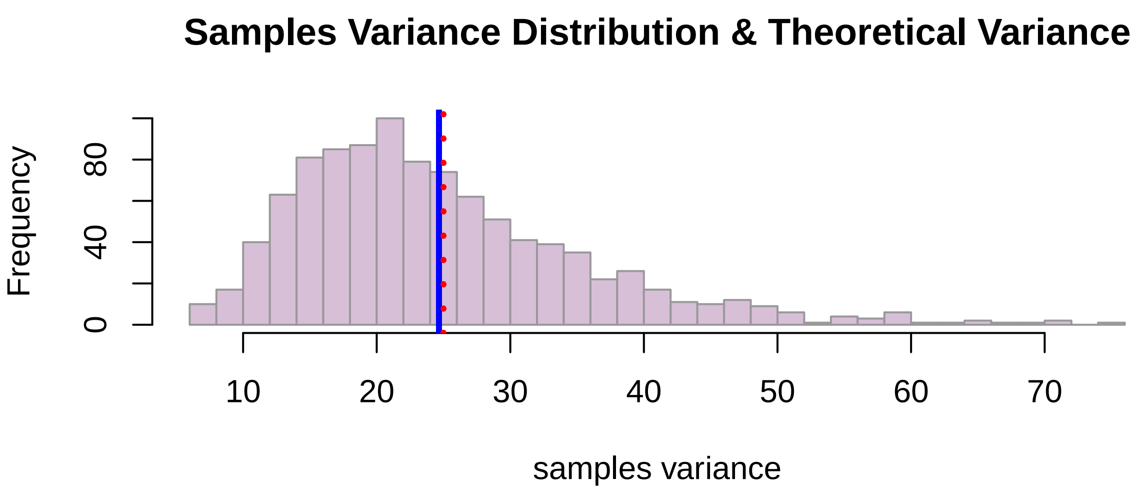
```
#plot the distribution of sample mean
hist(mns, breaks= 40, main="Samples Mean Distribution & Theoretical Mean", xlab="samples mean", col= "a
#plot the sample mean in blue
abline(v=sample_mean, col= "blue", lty=1, lwd= 3)
#plot the theoretical mean, in red
abline(v=mean, col= "red", lty=3, lwd= 3)
```



I observe that the samples' mean are distributed around the **sample mean 4.9835292** (in blue) that is very close to the **theoretical mean 5** (in dotted red).

##Sample Variance versus Theoretical Variance In the following diagram, I will be demonstrating the distribution of the variances of the simulated exponentials vs the the theoretical variance, which only depends on the lambda. Since lambda does not change, the **theoretical variance is also fixed = (1/lambda)^2,** therefore **Theoretical Variance = 25**

```
#distribution of sample variance is plotted in a histogram
hist(vns, breaks= 40,main="Samples Variance Distribution & Theoretical Variance", xlab="samples varianc
#the sample variance is added with a blue line
abline(v=sample_var, col= "blue", lty=1, lwd= 3)
#the theoretical variance is added with a red line
abline(v=var, col= "red", lty=3, lwd= 3)
```

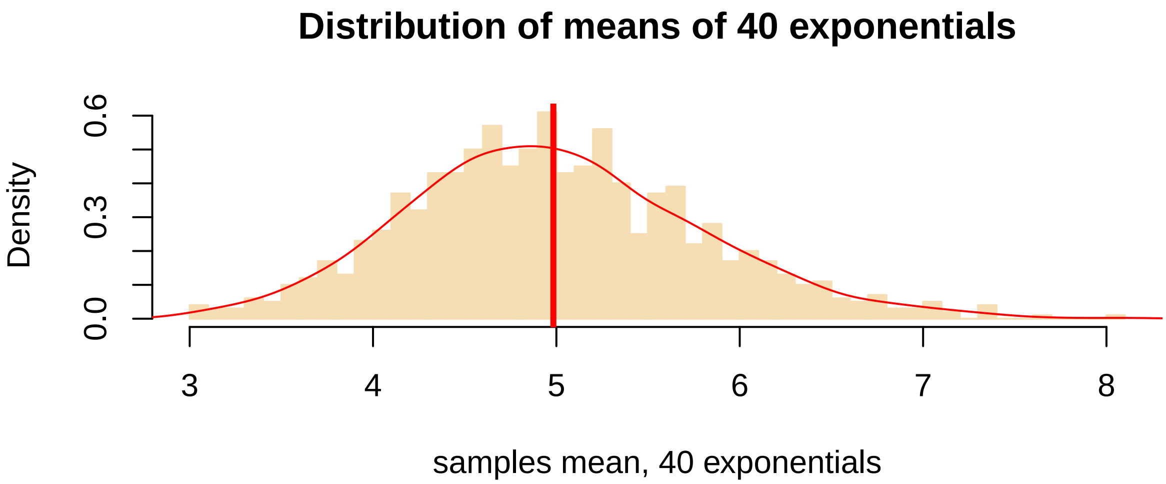


I observe that the samples' variances are distributed around the **sample mean variance 24.6606042** (in blue) that is very close to the **theoretical mean 25** (in dotted red).

Means Distribution

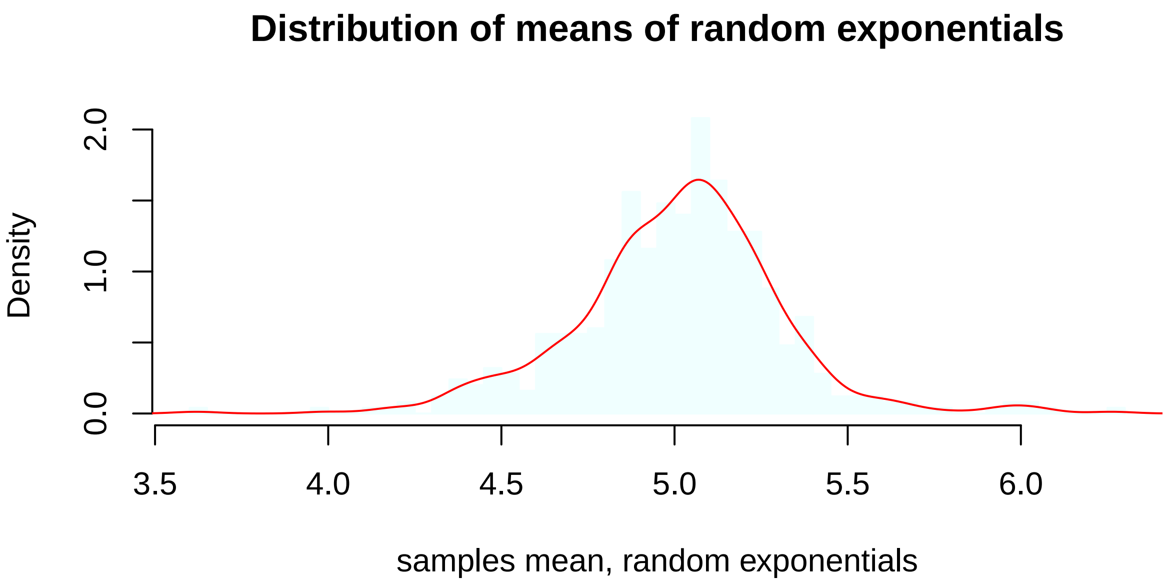
```
dh<- hist(mns, main="Distribution of means of 40 exponentials", col="wheat", border="wheat", breaks=40,
abline(v=mean(mns), col="red", lwd= 3) ; abline(v=sd(mns), col="violet", lty=2, lwd= 1)
abline(v=-sd(mns), col="violet", lty=2, lwd= 1)
```

```
#add the histogram linear envelope
lines (density(mns), col="red")
```



I observe that the **samples'mean= 4.98** is distributed at a Gaussian form that is very close to normal. Fitting a red line, helps me better visualize that. Then I will create a collection of **500 exponential distributions, with random exponentials**. Same as before I will collect their means and eval their distribution, to see whether it also tends to be a normal

```
rens<- NULL; rsens<- NULL; rmns<-NULL ; rvns<- NULL
#generate random number of exponentials
n_rdm<- sample(30:1000, 500)
for (i in 1 : 500){
rdist<- rexp(n_rdm[i], lambda) ; rens<-c(rens, rdist)
rmns <- c(rmns, mean(rdist)); rvns <- c(rvns, sd(rdist)^2 ) }
rdh<- hist(rmns, main="Distribution of means of random exponentials", col="azure",xlab="samples mean, ra
lines (density(rmns), col="red")
```



Same as above the distribution of means of random exponentials tends to a normal distribution as demonstrated by the red line.