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. >> (Estimate - Mean of Estimates)/ Std Error of means –> Normal

Simulations

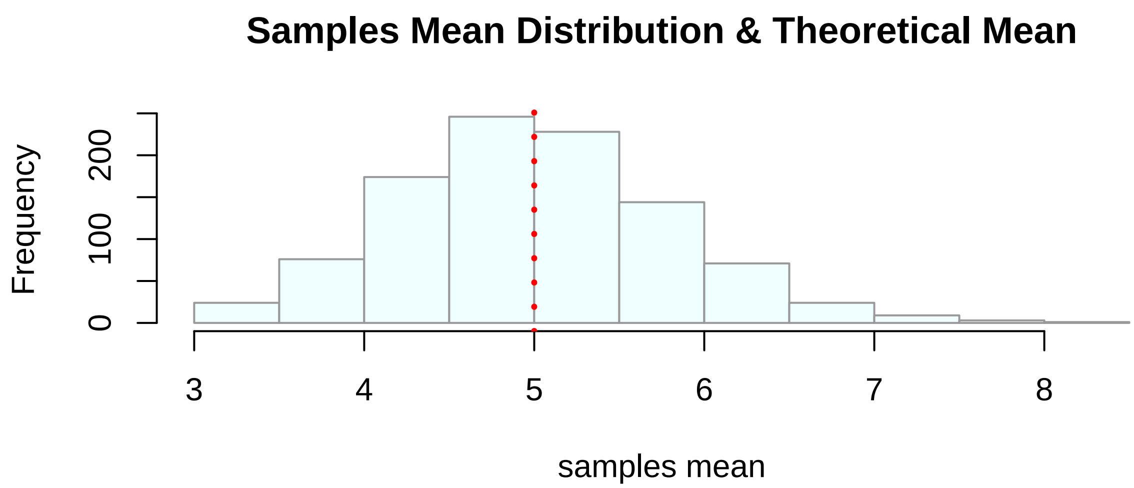
I will create 1000 random exponential distributions of iid variables, all with rate lambda = 0.2 and # of samples n=40. For each one of these I will be calculating and saving their mean.

```
#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, I am calculating each mean (mean(rexp(n, lambda)) and saving it to a table "mns"
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)
    #mean  & varianceof n samples simulation calculated and stored in the relevant table, aka mns and v
    mns <- c(mns, mean(dist)); vns <- c(vns, sd(dist)^2 ) }
# normalized distribution of means
d <- (ens - mns)*sqrt(n)/std
```

Sample Mean versus Theoretical Mean

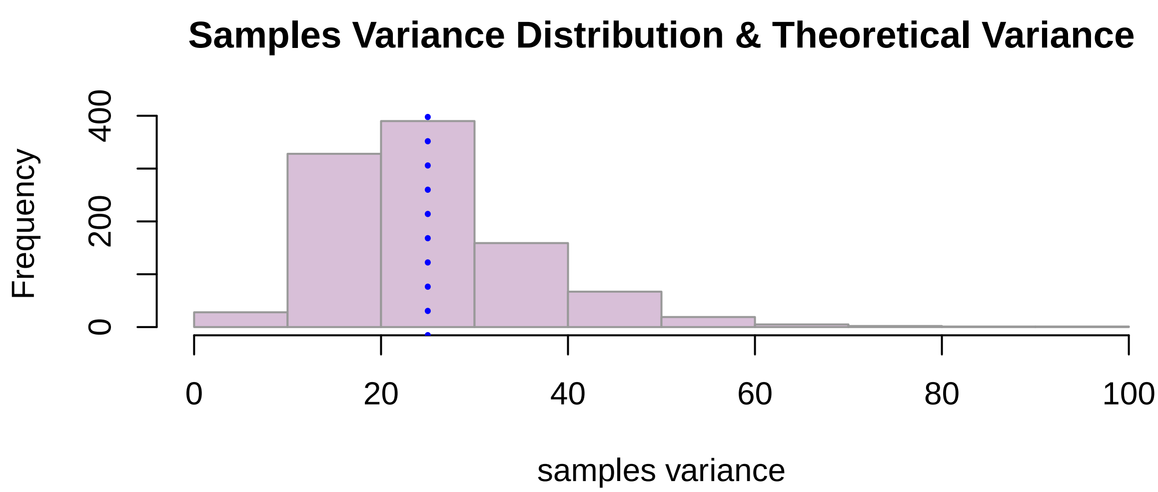
In the following diagram, I will be demonstrating the distribution of the means of the simulated exponentials vs the the theoretical mean, which only depends on the lambda. Since lambda does not change the theoretical mean is also fix = 1/lambda –> **Theoretical mean = 5**

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



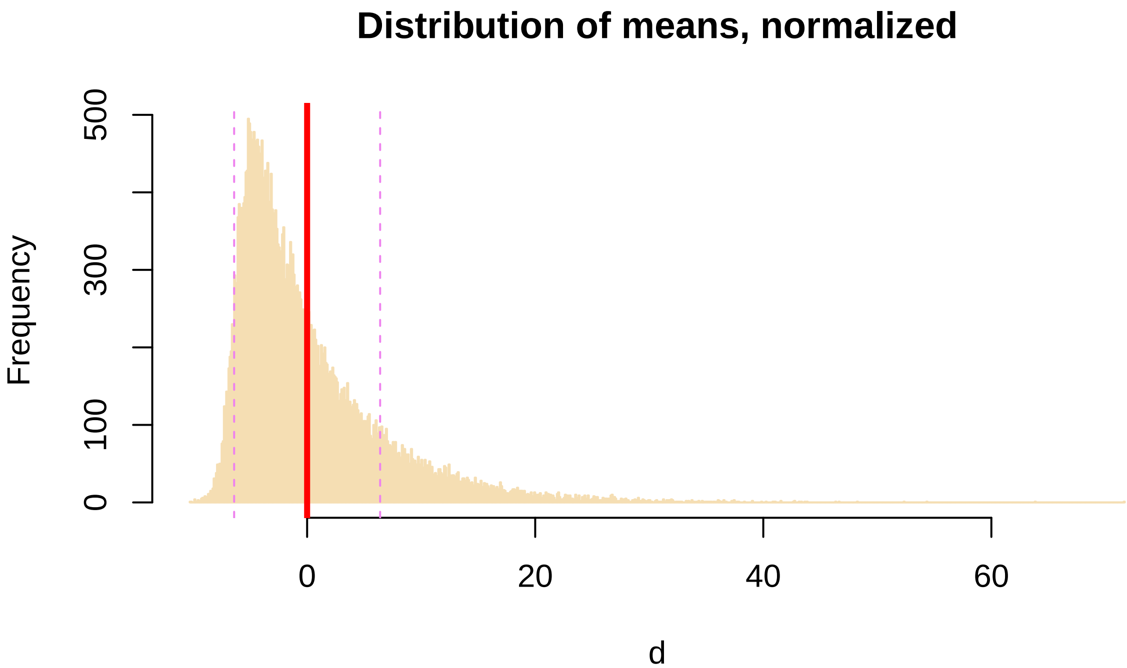
##Sample Variance versus Theoretical Variance In the following diagram, I will be demonstrating the distribution of the variances of the simulated exponentials vs the the theoretical variance, which only depends on the lambda. Since lambda does not change, the theoretical variance is also fixed = (1/lambda)^2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

```
#plot the distribution of sample variance
hist(vns, main="Samples Variance Distribution & Theoretical Variance", xlab="samples variance", col= "t
#plot the theoretical variance in blue
abline(v=var, col= "blue", lty=3, lwd= 3)
```



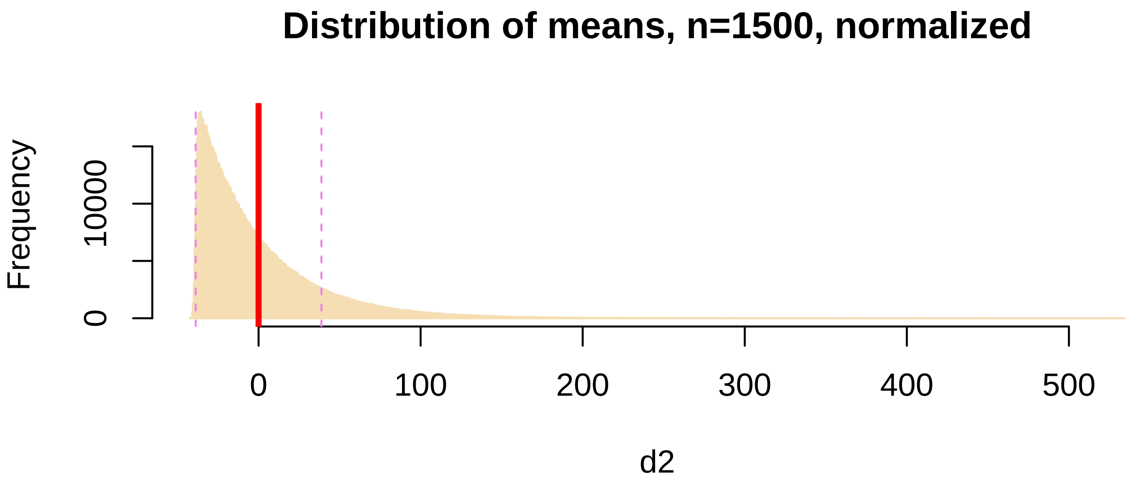
#Distribution Via figures and text, explain how one can tell the distribution is approximately normal.

```
hist(d, main="Distribution of means, normalized", col="wheat", border="wheat", breaks=1000)
abline(v=mean(d), col="red", lwd= 3) ; abline(v=sd(d), col="violet", lty=2, lwd= 1)
abline(v=-sd(d), col="violet", lty=2, lwd= 1)
```



Distribution **mean= 0** tends to equal the mean of standard normal distrubution, which is 0. Distribution **st. deviation= 6.4** tends to equal the st. deviation of standard normal distrubution, which is 1. As n= 40 is relatively small, I will rerun my simulations with a larger n, to see whether this round will result in a better approximation of a standard normal distribution for the simulation means.

```
#set a larger n
n2 <- 1500
#re-running 1000 simulations of exponential distributions, all with lambda=0.2, n=1500
ens2<- NULL; sens2<- NULL; mns2<-NULL ; vns2<- NULL
for (i in 1 : 1000)
  {
    #run simulation of exponential & add to the estimates table
    dist2<- rexp(n2, lambda) ; ens2<-c(ens2, dist2)
    #mean & variance of n samples simulation calculated and stored in the relevant table, aka mns2 and 
    mns2 <- c(mns2, mean(dist2)) ; vns2 <- c(vns2, sd(dist2)^2 ) }
# normalized distribution of means
d2 <- (ens2 - mns2)*sqrt(n2)/std
#plot new distribution with larger n
hist(d2, main="Distribution of means, n=1500, normalized", col="wheat3", border="wheat", breaks=1000)
abline(v=mean(d2), col="red", lwd= 3); abline(v=sd(d2), col="violet", lty=2, lwd= 1) ; abline(v=-sd(d2)
```



The **distribution mean 0** tends to equal the mean of standard normal distrubution, which is 0 The **distribution St deviation 38.81** tends to equal the St. Deviation of standard normal distrubution, which is 1