Statistical inference: Assignment part1

Gurunath Gandikota

September 17, 2018

Synopsis

In this assignment, studies are carried out to understand the central limit theorem (CLT). Exponential distribution with a rate parameter 0.2 and sample size 40 are used to run 1000 simulations. The distribution of the sample means is looked at to see if it is a Gaussian distribution.Results indicate that the mean, standard deviation and the variance of the simulation are very close to the theoritical values. Further, a normality check carried out on the sample mean values indicates that the distribution of the sample means is very close to a Gaussian distribution.

Simulation for exponential distribution

Exponential distribution is used to carry out the study, to see if it follows CLT (central limit theorem). Samples of size 40 and rate parameter 0.2 are made 1000 times and mean of each sample is calculated as follows:

Parameters for the exponential distribution

```
#set the seed value
set.seed(1)

#Simulation parameters: lambda=rate parameter, n_sample=sample size,
simuls=number of simulations
lambda <- 0.2; n_sample <- 40; simuls <- 1000
mean_data<-NULL
```

Running 1000 simulations

```
#Generate the data
for (i in 1:simuls){
  data<-mean(rexp(n_sample,lambda))
  mean_data <- rbind(mean_data,data)
}
```

Calculating the mean SD and variance of the distribution of sample means

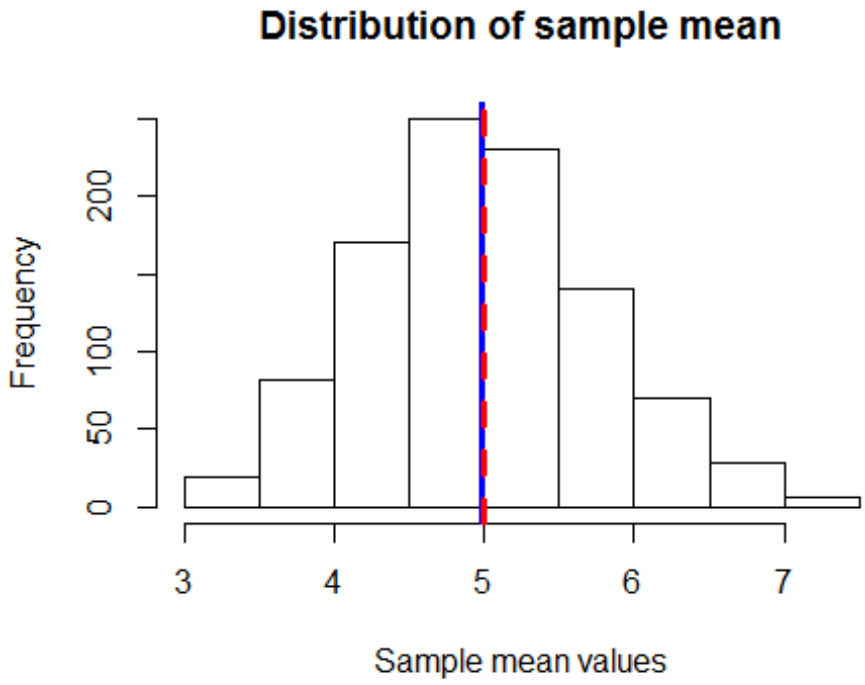
```
# Calculation of mean, SD and variance for the simulation
mean_simul<-mean(mean_data)
SD_simul<-sd(mean_data)
var_simul=var(mean_data)
```

Comparison of simulation with theory

Mean, SD and variance for an exponential distributions are available theoritically. A comparison of the theoritical values with those from simulation is carried out:

```
#Compare theoritical and simulation mean, SD and var
mean_theory <- 1/lambda
SD_theory <- (1/lambda)/sqrt(n_sample)
var_theory <- SD_theory^2

hist_mean<-hist(mean_data,xlab = "Sample mean values", main="Distribution of
sample mean")
abline(v=mean_simul,col="blue",lwd=3,lty=1)
abline(v=mean_theory,col="red",lwd=3,lty=2)
```



From the distribution, it can be seen that the theoritical and simulation mean values overlap. The following tables give a comparison of the SD and variance values for theory and simulations:

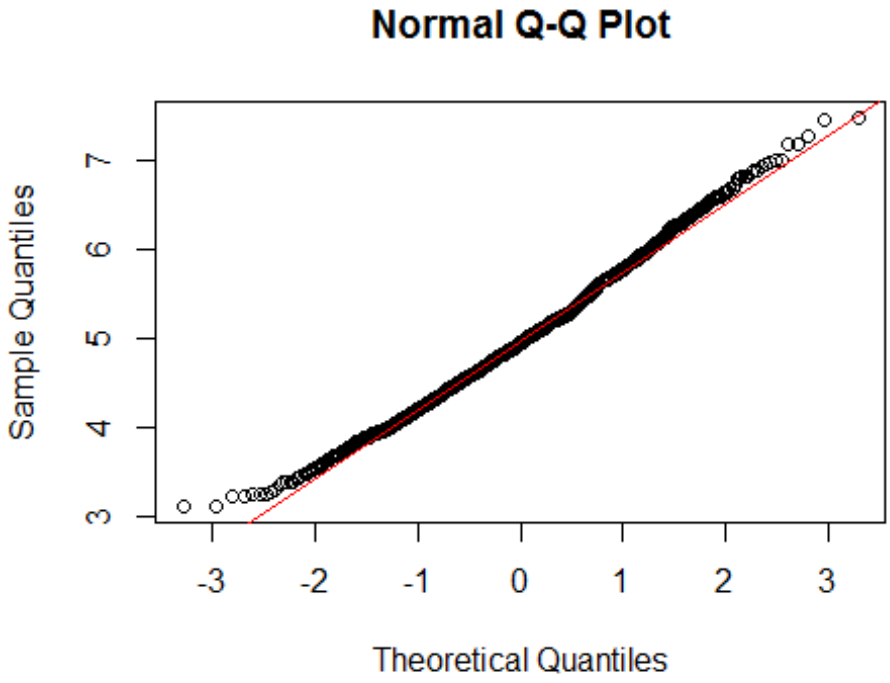
SD(Theory)	SD(simulation)
0.7905694	0.7817394
Var(Theory)	Var(simulation)
0.625	0.6111165

It can be seen that the SD and variance of the sample means is very close. They will get closer if we increase the sample size.

Gaussian nature of the distribution of the sample means

The distribution of the sample means is tested for normalcy:

```
# compare the distribution of averages of 40 exponentials to a normal
distribution
qqnorm(mean_data)
qqline(mean_data, col = 2)
```



The plot shows that the distribution is approximately Gaussian (Normal) in nature