Statistical Reference Course Assignment, Part one

Xiaocheng Zeng

January 18, 2016

Overview

This assignment aims at investigating central limit theorem with a random exponential sample. We create a sample of averages with the method of bootstrapping and compare mean and variance of this sample with theoretical statistiscs of exponential distribution.

Create a random exponential sample and bootstrap from it

```
set.seed(3)
lambda=.2
b<-1000
n<-40
ex<-rexp(5000,rate=lambda)
resamples<-matrix(sample(ex,n*b,replace=TRUE),b,n)
```

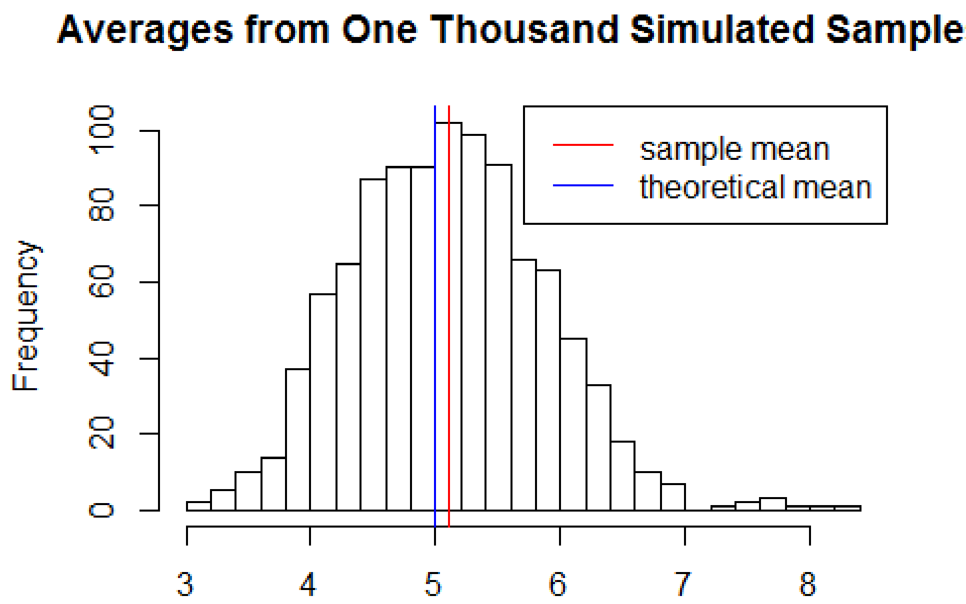
We create a sample of exponential distribution, ex, with 5000 observations and lambda=.2, which we will bootstrap from. Then we create a matrix of 1000 rows and 40 columns and fill it with observations from ex.

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

```
averages<-apply(resamples,1,mean)
mean_sample<-mean(averages)
mean_theoretical<-1/lambda
print(paste("sample mean=",round(mean_sample,4)," v.s. ","theoretical
mean=",mean_theoretical))

## [1] "sample mean= 5.1009  v.s.  theoretical mean= 5"

hist(averages,breaks=30,main="Averages from One Thousand Simulated
Samples",xlab="")
abline(v=mean_sample,col="red")
abline(v=mean_theoretical,col="blue")
legend("topright",c("sample mean","theoretical
mean"),lty=c(1,1),col=c("red","blue"))
```



As we can see the mean of simulated samples is very close to the theoretical mean of random exponentials.

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

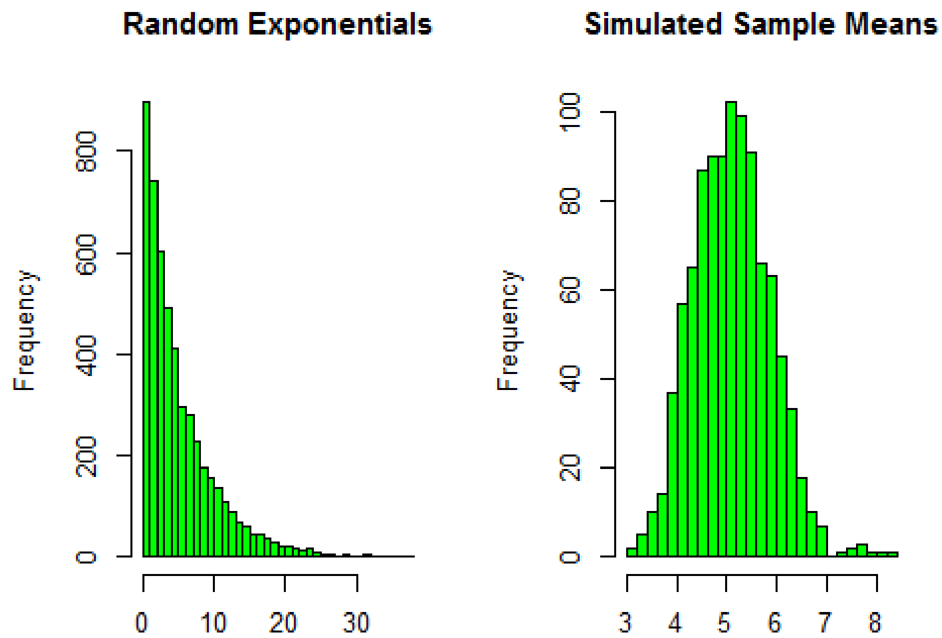
```
var_sample<-var(averages)
var_theoretical<-(1/(lambda*sqrt(n)))^2
print(paste("sample variance=",round(var_sample,4)," v.s. ","theoretical variance=",round(var_theoretical,4)))

## [1] "sample variance= 0.602  v.s.  theoretical variance= 0.625"
```

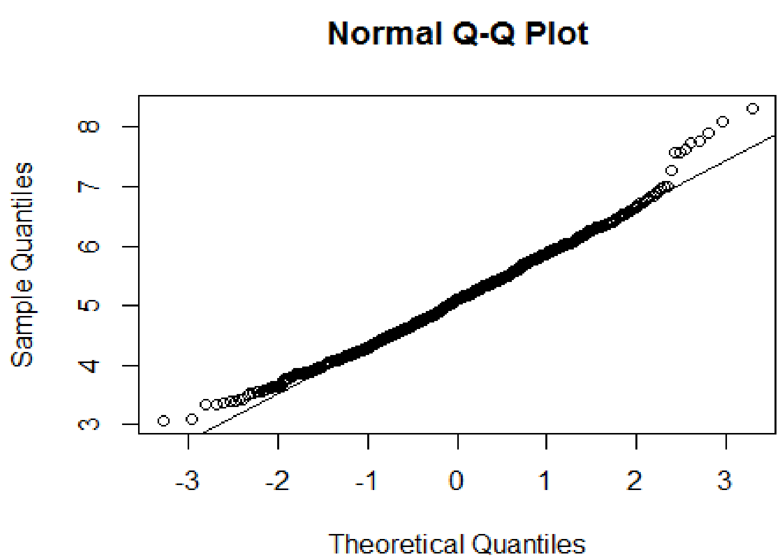
As we can see, sample variance is slightly different from theoretical variance. The reason is the sample that we bootstrap from, ex, with some randomness, still deviates from the theoretically perfect exponential distribution by a tiny bit.

Show that the distribution is approximately normal

```
par(mfrow=c(1,2),cex=.8)
hist(ex,breaks=30,col="green",main="Random Exponentials",xlab="")
hist(averages,breaks=30,col="green",main="Simulated Sample Means",xlab="")
```




```r
qqnorm(averages)
qqline(averages)
```



Clearly, with a bell shape,histogram from the bootsrapped data looks much more like a normal distribution. qqnorm and qqline also show that means from the simulated samples are very close to a normal distribution