Final Project. Part 1: Simulation Exercise

Kadyrbek

11/18/2019

Synopsis

In this project we will run a thousand simulations with 40 exponential number each. We are going to investigate how all the simulated numbers are distributed and compare it with distribution of means of each simulations. Moreove, we will compare sample properties with theoretical ones and see the difference between them.

Simulation of the data

Here, the data will be simulated with mean=standard deviation=1/lambda, where lambda=0.2. The total number of simulations is equal to 1000 with 40 random exponential number each.

```
set.seed(12345)
lambda<-0.2
n<-40
sim<-1000
df<-matrix(rexp(n*sim,lambda),sim,n) #creates matrix sim by n (1000x40)
dfmean<-apply(df,1,mean) # Takes means of each simulation(row)
```

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

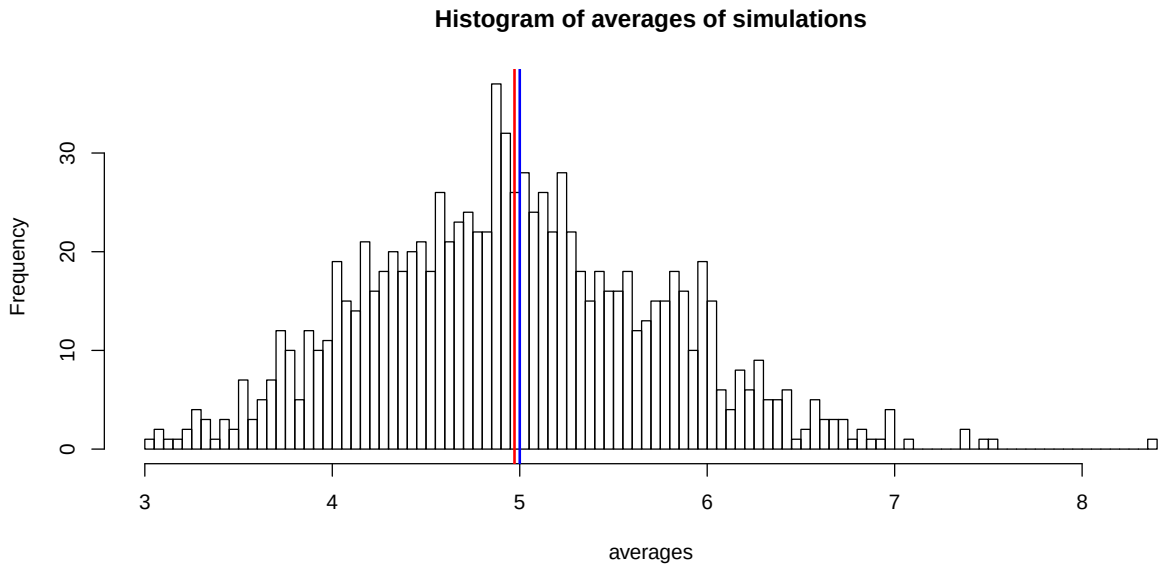
Sample mean =4.971972 and Theoretical mean = 5 Difference between theoretical and sample means is equal to 0.028028.

```
list("sample mean"=mean(dfmean),"theoretical mean"=(1/lambda),
     "Difference between theo and sample means"=(1/lambda)-mean(dfmean))
```

```
## $`sample mean`
## [1] 4.971972
##
## $`theoretical mean`
## [1] 5
##
## $`Difference between theo and sample means`
## [1] 0.02802804
```

Following graph shows us that the sample mean(red) and theoretical mean(blue) are very close to each other.

```
hist(dfmean,breaks = 100,main = "Histogram of averages of simulations",xlab = "averages")
abline(v=mean(dfmean),col="red",lwd=2)
abline(v=1/lambda,col="blue",lwd=2)
```



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

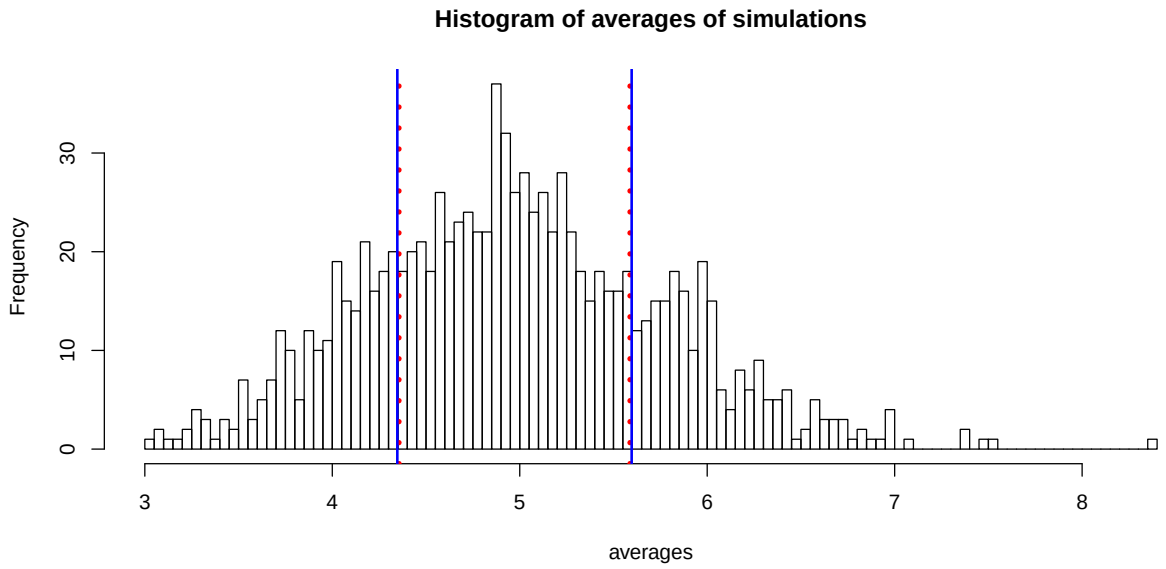
Sample variance =0.6157926 and Theoretical variance = 0.625 Difference between theoretical and sample variances is equal to 0.0092074.

```
list("sample variance"=var(dfmean),"theoretical variance"=(1/lambda)^2/n,
     "Difference between theo and sample variances"=(1/lambda)^2/n-var(dfmean))
```

```
## $`sample variance`
## [1] 0.6157926
##
## $`theoretical variance`
## [1] 0.625
##
## $`Difference between theo and sample variances`
## [1] 0.009207378
```

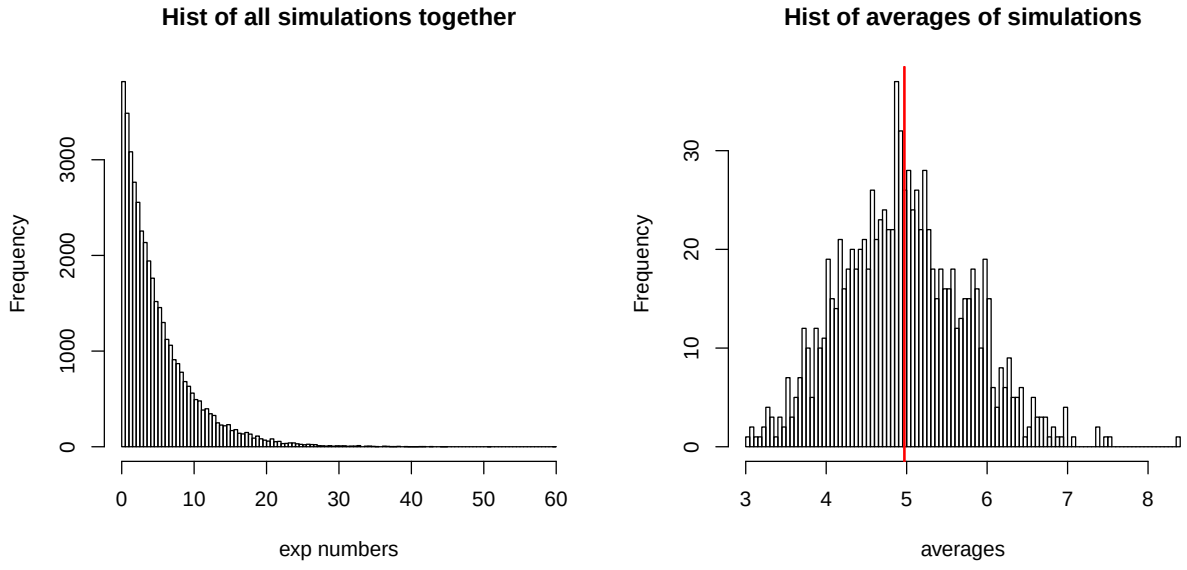
Following graph shows us that the sample variance(dashed red) and theoretical variance(blue line) are very close to each other.

```
hist(dfmean,breaks = 100,main = "Histogram of averages of simulations",xlab = "averages")
abline(v=c(mean(dfmean)+var(dfmean),mean(dfmean)-var(dfmean)),col="red",lwd=4,lty=3)
abline(v=c(mean(dfmean)+(1/lambda)^2/n,mean(dfmean)-(1/lambda)^2/n),col="blue",lwd=2)
```



#3.Show that the distribution is approximately normal.

```
par(mfrow=c(1,2))
hist(df,main="Hist of all simulations together",xlab = "exp numbers",breaks = 100)
hist(dfmean,main="Hist of averages of simulations",xlab = "averages",breaks = 100)
abline(v=mean(dfmean),col="red",lwd=2)
```



The above plot represents the histograms of all simulations numbers and averages of simulations with red lines indicating the mean of the distribution of the averages. We can observe that the averages is distributed symmetricly around mean(red line), has bell shape curve and peak is on the mean point.These properties are enought to claim that data is approximately normaly distributed.