Statisticl Inference Project part1

Abdelrahman Elsehaily

Overview

This project from Statisticl Inference course on coursera, the goal is to investigate the exponential distribution in R and compare it with the Central Limit Theorem.

1.Simulation

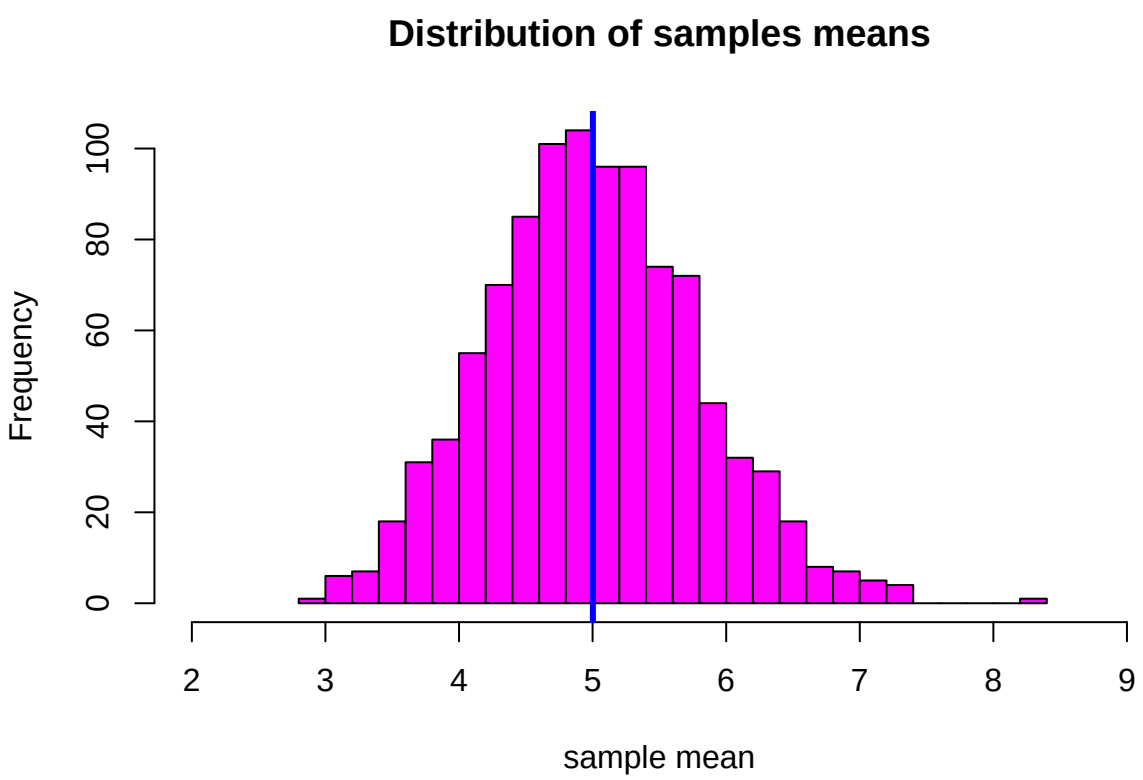
Created one thousand loops, in each iteration creating a new sample and getting the mean only as we only interested in the distribution of the mean.

```
lambda=0.2
sample_means=NULL
for (i in 1 : 1000) sample_means = c(sample_means, mean(rexp(40,rate =lambda)))
```

2.Sample Mean versus Theoretical Mean:

we wanna to check the mean of our samples means and see how it differs from the Theoretical Mean(Population Mean)

```
hist(sample_means,breaks = 20,xlim = c(2,9),col = "magenta",xlab = "sample mean",main="Distribution of s
abline(v=mean(sample_means),col="blue",lwd=3)
```




```
print(paste("Sample Mean: ",mean(sample_means)))
```

```
## [1] "Sample Mean:  5.00215802545185"
```

```
print(paste("Therortical Mean: ",1/lambda))
```

```
## [1] "Therortical Mean:  5"
```

The sample mean is approximately as the population mean as states in the Centeral Limit Theorm.

3.Sample Variance versus Theoretical Variance:

we wanna to check the sample varience given our simulation and the theoritcal varience and also check the variance of the sampe given the CLT.

```
print(paste("Theoretical Variance: ",1/lambda))
```

```
## [1] "Theoretical Variance:  5"
```

```
print(paste("sample Variance from our simulation: ",sd(sample_means)))
```

```
## [1] "sample Variance from our simulation:  0.785415282131449"
```

```
print (paste("sample varience from CLT: ",(1/lambda/sqrt(40))))
```

```
## [1] "sample varience from CLT:  0.790569415042095"
```

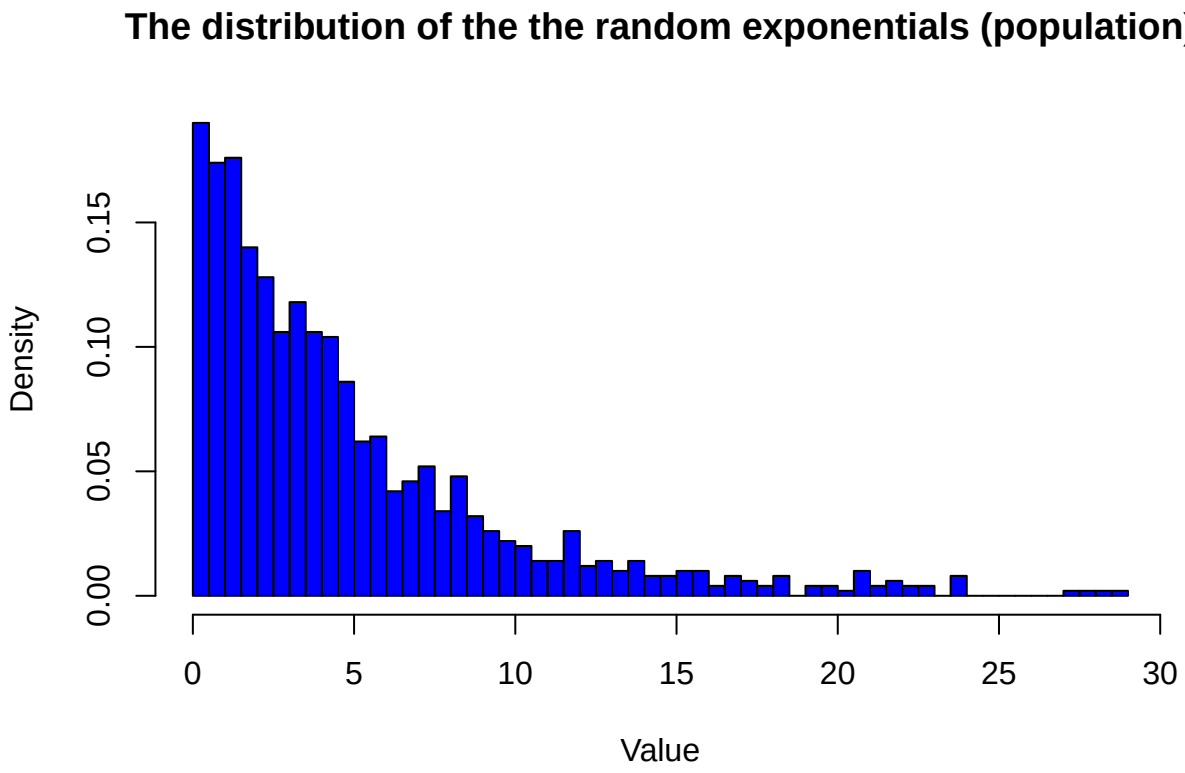
so the sample varience from the CLT and the sample varience from the simualtion is approximately the same.

4.Distribution

now we wanna see difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials we will plot the density instead of the count of the numbers.

The distribution of the random exponentials

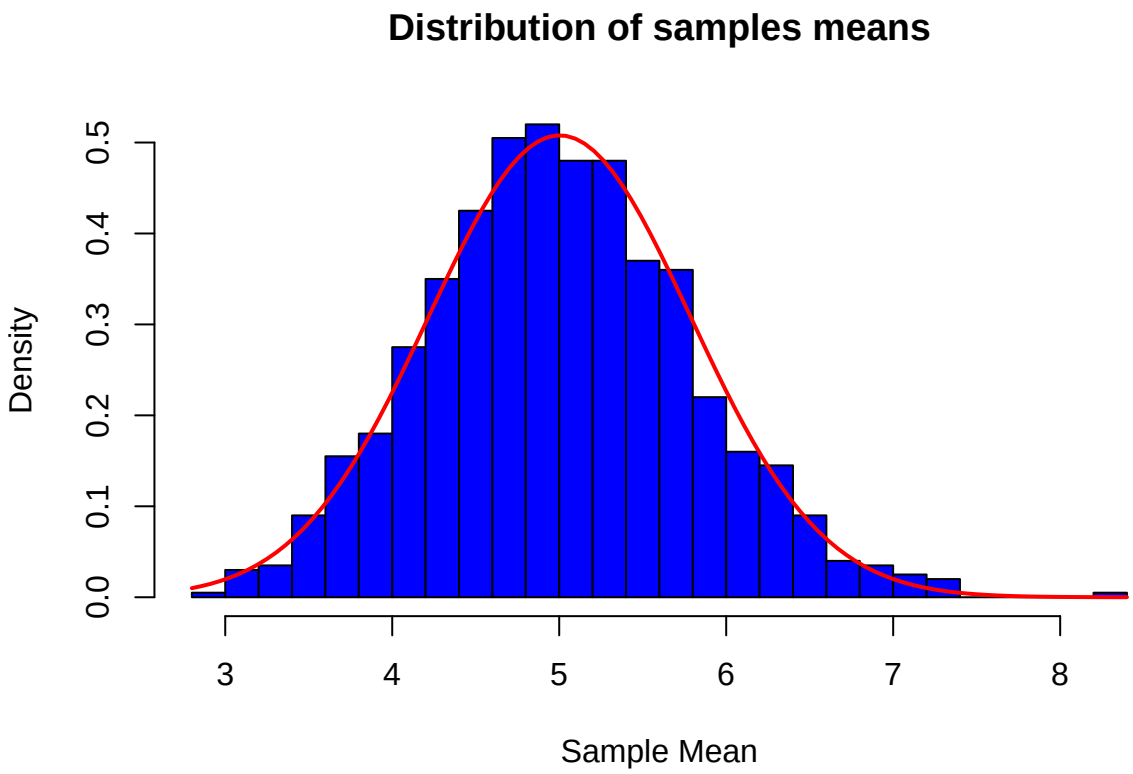
```
hist(rexp(1000,rate=lambda),breaks = 50,col = "blue",xlab = "Value",freq = F,main="The distribution of
```



The distribution of the samples means

To show if the samples mean distribution follow normal distribution or not I plotted the curve of normal distribution with mean and veriance equal to the mean and veriance of the samles means.

```
hist(sample_means,breaks = 20,col = "blue",xlab = "Sample Mean",main="Distribution of samples means",fr
curve(dnorm(x, mean=mean(sample_means), sd=sd(sample_means)), add=TRUE, col="red", lwd=2)
```



Conclusion

It's obvious that the distribution of the averages of the sampls follow normal distribution and it has no relationship with the distribution of the population. Also our simulation gives the excat mean and varience as derived from the Central Limit Theorm.