Statistical Inference Course Project2

Nadine.solov

16 01 2022

Part 1: Simulation Exercise.

We study the exponential distribution and compare it with the central limit theorem. Modeling the exponential distribution with parameters: lambda = 0.2, The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.Investigate the distribution of averages of 40 exponentials, 1000 simulations. Issues: 1.Show the sample mean and compare it to the theoretical mean of the distribution. 2.Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. 3.Show that the distribution is approximately normal.

```
n_sim <- 1000   # number of simulations
lambda <-  0.2
n <- 40             # sample size
```

1.Theoretical values: mean,variance, standart deviation:

```
TheoryMean<-1/lambda
TheoryMean
```

```
## [1] 5
```

```
TheorySD<-((1/lambda) * (1/sqrt(n)))
TheorySD
```

```
## [1] 0.7905694
```

```
TheoryVar<-TheorySD^2
TheoryVar
```

```
## [1] 0.625
```

We model the exponential distribution and calculate the actual values:

```
simulated_data <- matrix(rexp(n= n_sim*n,rate=lambda), n_sim, n)

RowMeans<-apply(simulated_data,1,mean)
ActualMean<-mean(RowMeans)
ActualMean
```

```
## [1] 5.011884
```

```
ActualSD<-sd(RowMeans)
ActualSD
```

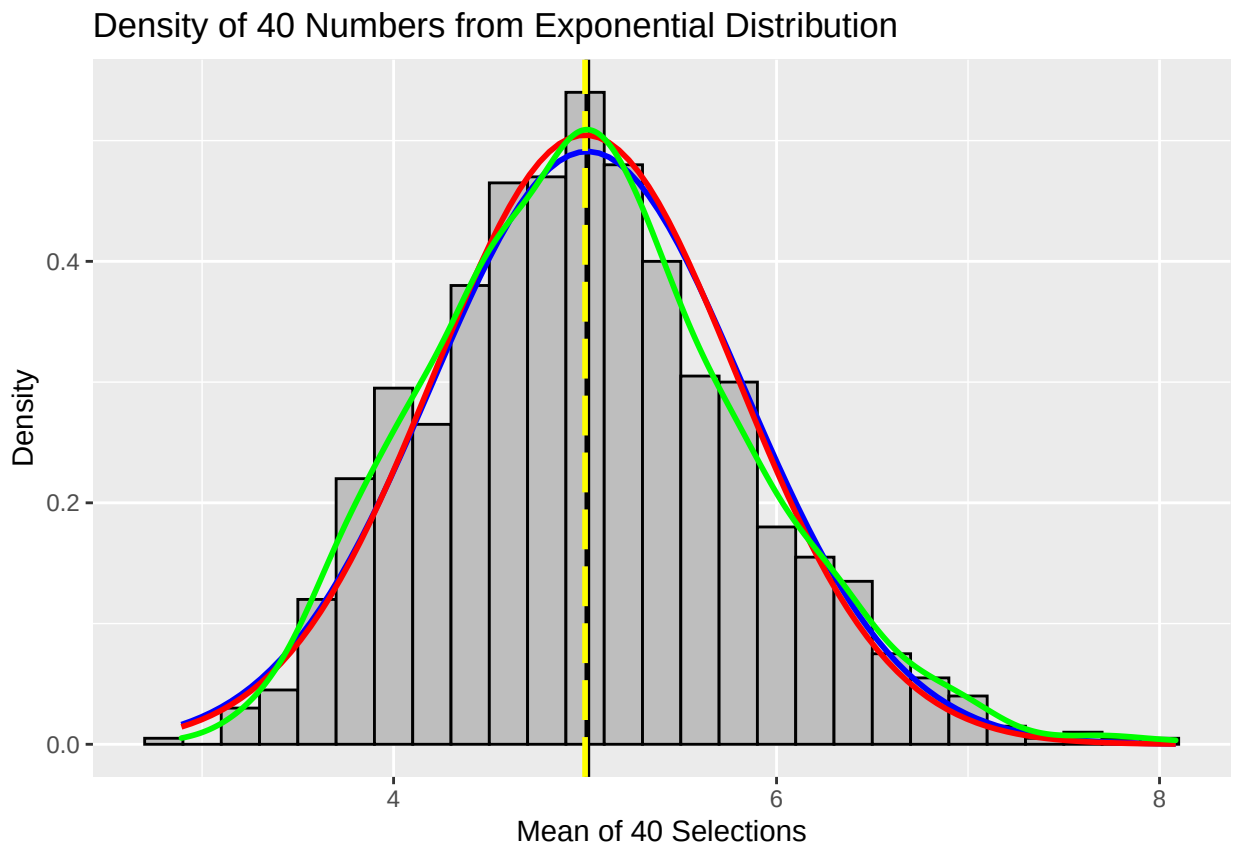
```
## [1] 0.8124412
```

```
ActualVar<-var(RowMeans)
ActualVar
```

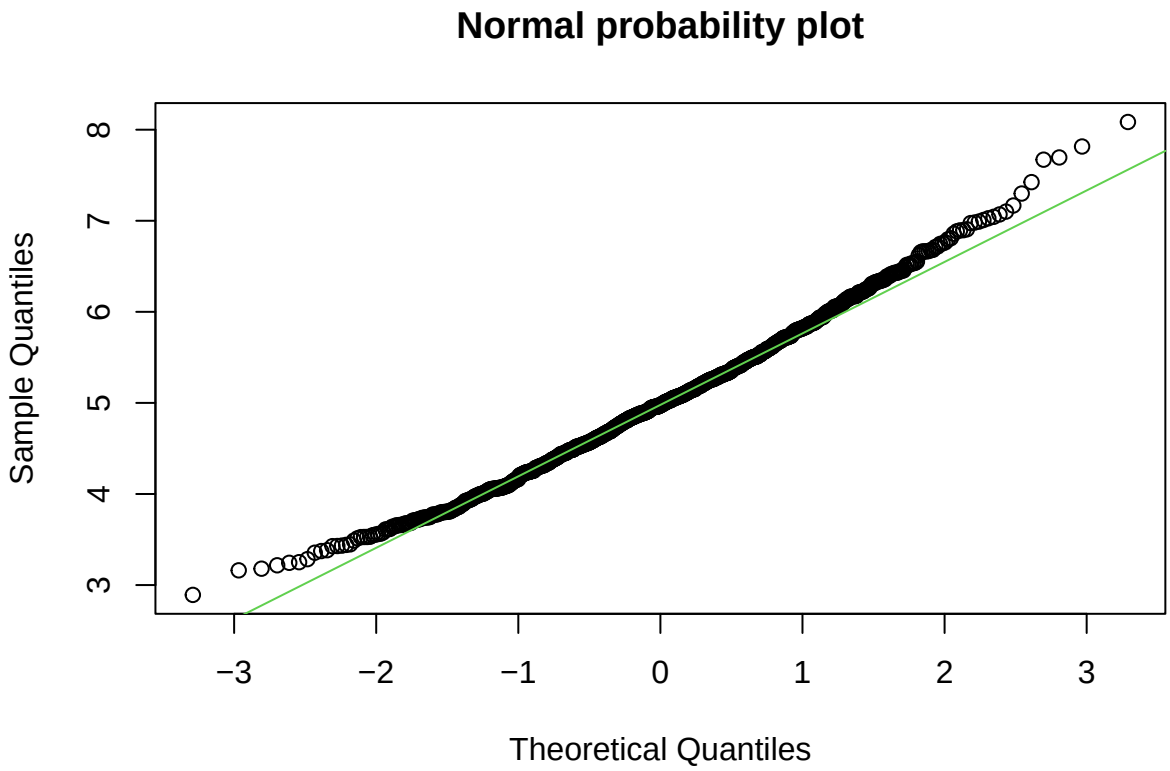
```
## [1] 0.6600607
```

1. The sample mean of 4.990836 is close to the theoretical value of 5 2.,3. converting data to data.frame for ggplot:

```
dfRowMeans<-data.frame(RowMeans)
g<-ggplot(dfRowMeans,aes(x=RowMeans))
g<-g+geom_histogram(binwidth = lambda,fill="gray",color="black",aes(y=..density..))
g<-g + labs(title="Density of 40 Numbers from Exponential Distribution",
            x="Mean of 40 Selections", y="Density")
g<-g + geom_vline(xintercept=ActualMean,size=1.0, color="black")
g<-g + stat_function(fun=dnorm,args=list(mean=ActualMean, sd=ActualSD),color = "blue", size = 1.0)
g<-g + geom_vline(xintercept=TheoryMean,size=1.0,color="yellow",linetype = "longdash")
g<-g + stat_function(fun=dnorm,args=list(mean=TheoryMean, sd=TheorySD),color = "red", size = 1.0)
g<-g+geom_density(colour="green", size=1)
g
```



```
qqnorm(RowMeans, main ="Normal probability plot")
qqline(RowMeans,col = "3")
```



In The graph: the theoretical mean is shown as a dotted yellow line; actual mean is shown by the black line; the normal curve formed by the theoretical mean and standard deviation is shown in red; the curve the actual mean and standard deviation is shown in blue;. the density of the actual data is shown by the gray bars and green line. Conclusion: Both histogram and the normal probability plot show that distribution of averages is approximately normal. The central limit theorem says that the distribution of averages is often normal, even if the distribution from which the data is taken is very abnormal. the distribution of average iid values of random variables becomes standard as the sample size increases.