Statistical Inferece Project_Part 1 Simulation Exercise

lnma

5/3/2020

1.Synopsis

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. 1000 simulations were performed to obtain reliable distribution. The following questions were answered.

1. The sample mean was shown and compared to the theoretical mean of the distribution.
2. The variable of sample was shown and compared to the theoretical variance of the distribution.
3. The distribution is shown to be approximately normal.

2.Simulation process

2.1 Data

The exponential distribution was simulated in R with rexp(n, lambda) where lambda is 0.2 and n is 40. The simulation was repeated 1000 times and the data was saved in the matrix of *exp*.Calculate the averages across the 40 values for each of the 1000 simulations and save the data in *mns*.

```
set.seed(100191)
n<-40
lambda<-0.2
nTest<-1000
exp<-matrix(rexp(n * nTest, lambda),nTest)
mns<-rowMeans(exp)
```

2.2 Mean comparison

Calculate the mean of the sample data, *sampleMean*, which is the average of the mean of the 1000 samples.

```
sampleMean<-mean(mns)
sampleMean
```

```
## [1] 5.010113
```

Calculate the theoretical mean, *theoMean*, which is simply 1/lambda.

```
theoMean<-1/lambda
theoMean
```

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

The sample mean is *5.010113*, which is close to the theoretical mean of *5*.

2.3 Variance Comparison

Calculate the variance of the sample data, *sampleVar*,and the standard deviation of sample, *sampleSd*, which are the variance and standard deviation of the mean of the 1000 samples, respectively.

```
sampleVar<-var(mns)
sampleVar
```

```
## [1] 0.6226745
```

```
sampleSd<-sd(mns)
sampleSd
```

```
## [1] 0.7890973
```

Calculate the theoretical mean, *theoVar*.

```
theoVar<-(1/lambda)^2/n
theoVar
```

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

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

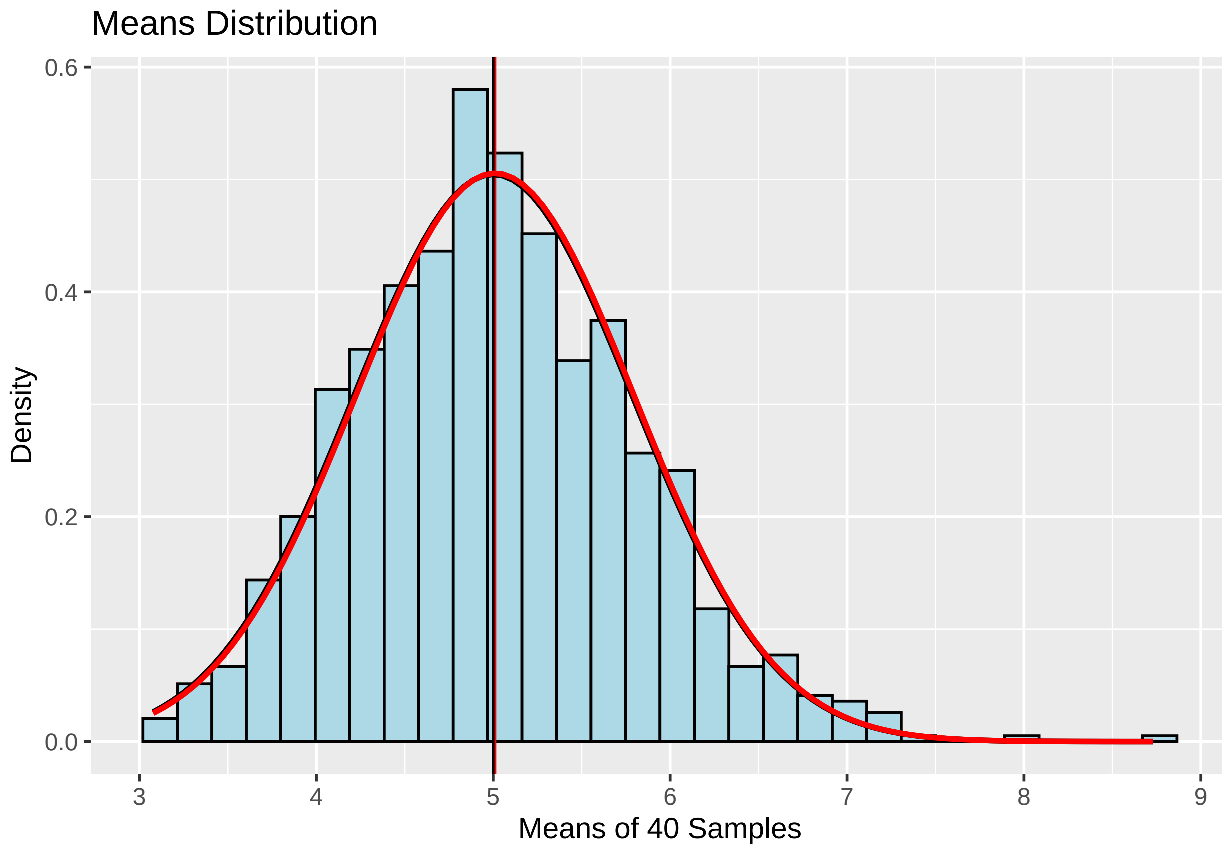
```
## [1] 0.7905694
```

The sample mean is *0.6226745*, which is close to the theoretical mean of *0.625*.
The sample mean is *0.7890973*, which is close to the theoretical mean of *0.7905694*.

3.Results

The distribution of the means were plotted using the histgram in *ggplot* package. The sample mean and theoretical mean were highlighted using red and black vertical lines. The theoretical normal distribution and the sample normal distribution (using the sample mean and sample standard deviation) were also plotted and compared with theoretical in black and sample in red.

```
library(ggplot2)
data<-data.frame(mns)
p<-ggplot(data,aes(x=mns))
p<-p+geom_histogram(aes(y=..density..),color='black',fill='lightblue')
p<-p+labs(title = 'Means Distribution',x = 'Means of 40 Samples', y = 'Density')
p<-p+geom_vline(aes(xintercept = sampleMean),color='red')
p<-p+geom_vline(aes(xintercept = theoMean),color='black')
p<-p+stat_function(fun = dnorm,args = list(mean=theoMean,sd=theoSd),color='black',size=1.0)
p<-p+stat_function(fun = dnorm,args = list(mean=sampleMean,sd=sampleSd),color='red',size=1.0)
p
```



The distribution of means is close to normal distribution (see the sample normal, red line). Moreover, the mean value of the mean distribution is close to that of theoretical distribution and sample normal overlaps well with the theoretical normal distribution.

4.Conclusions

From the example of exponential distribution simulation, the Central Limit Theorem has been proved to be validit that, in some situations, when independent random variables are added or averaged, their properly normalized sum or mean tends toward a normal distribution (informally a bell curve) even if the original variables themselves are not normally distributed