Investigating Exponential Distribution

By: Syed Musharraf Ali

Overview

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. We will see how much theoritical **mean** and **standard deviation** of exponential distribution differ from the value obtain from **simulation**.
In this project we will assume Lambda = 0.2

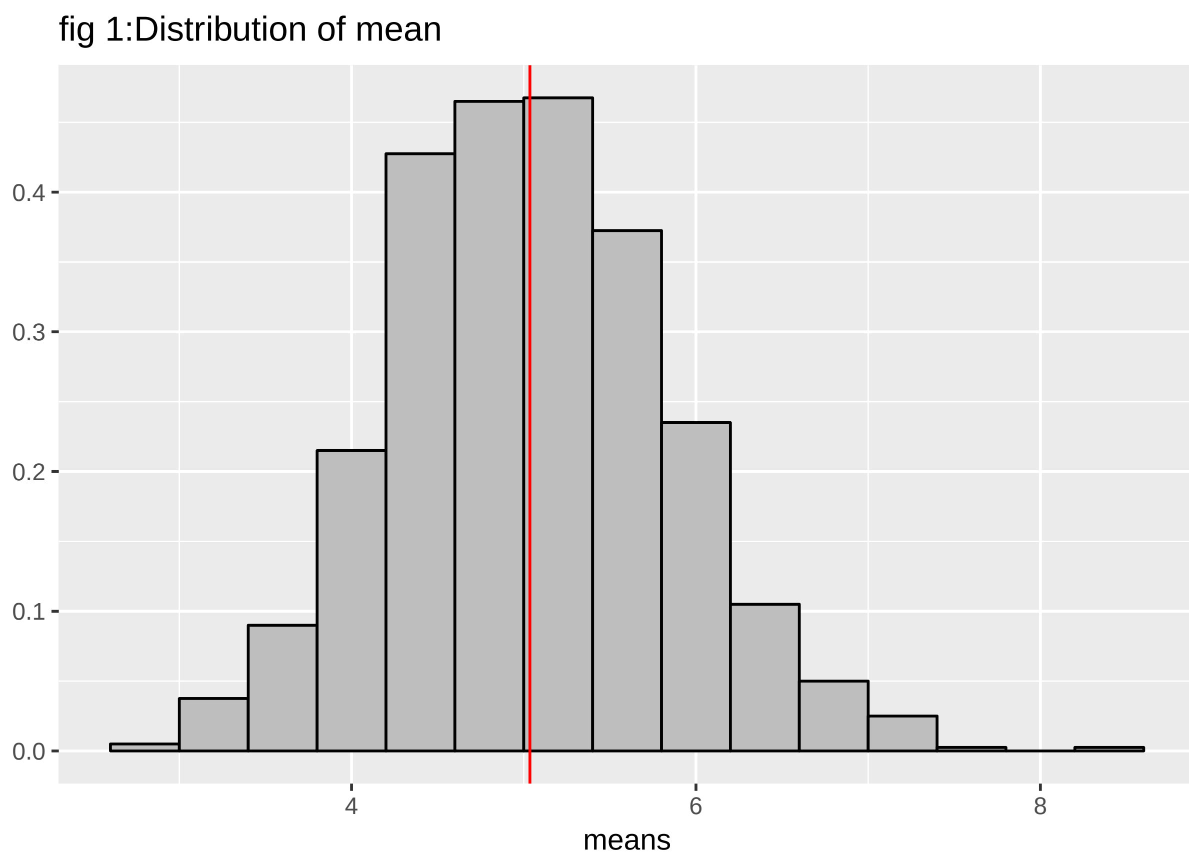
Compairing theoritical statistics with simulated statistics

The theoritical mean is 1/lambda which is 1/0.2=5. The simulation mean is found by **replicate()** command in which the second argument is repeated by the amount given in first argument.

```
mean_exp <- replicate(1000,{x<-rexp(n=40,rate=0.2)
                              mean(x)})
mean(mean_exp) # taking the mean of "mean_exp" to get the mean of distribution.
```

```
## [1] 5.035787
```

As the simulated mean is much closer to the theoritical mean of **5**. it shows that Central limit theorum has been correctly applied. From the below figure it can also be seen that the mean is the centre of distribution.



The theoritical variance of sampling distribution is given by **variance/n**. Its value is

```
((1/0.2)/sqrt(40))^2
```

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

the variance from the sampling distribution is

```
var(mean_exp)
```

```
## [1] 0.6455312
```

From the table we can easily see how close the **simulated variance** is to the **theoritical variance**

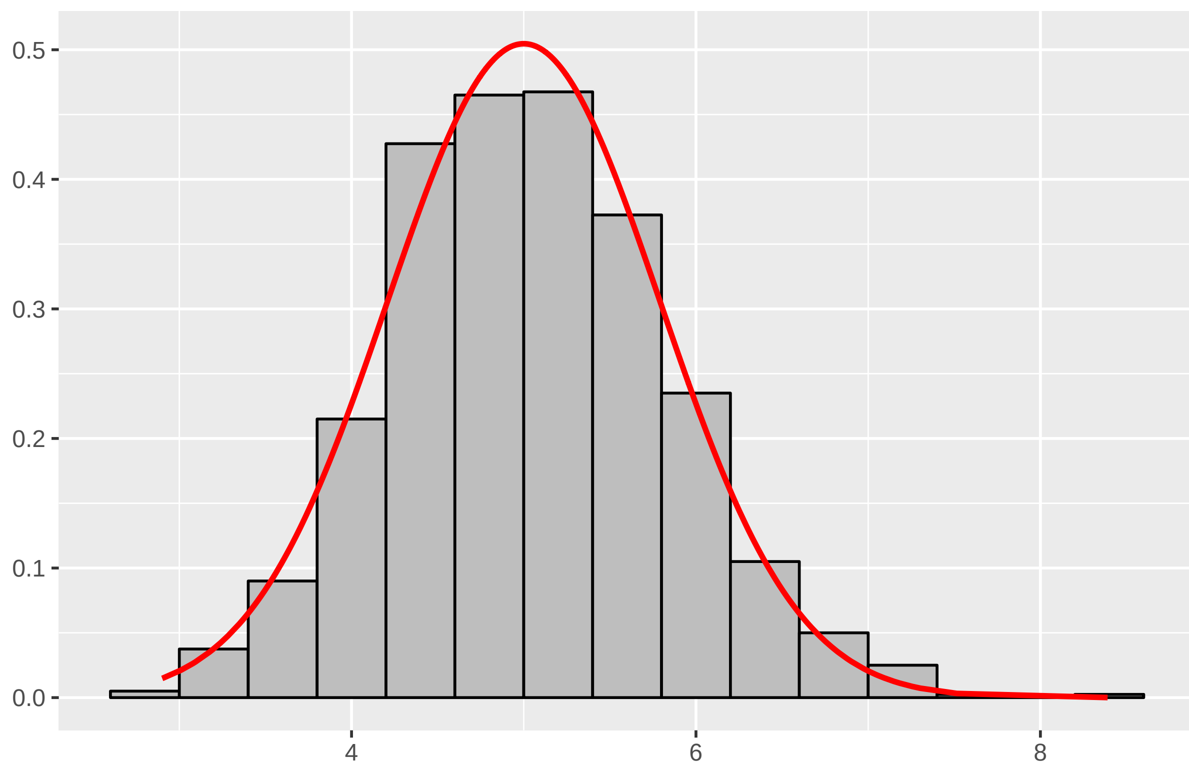
Table 1: Comparison of theoritical and simulated variance

theoritical_variance	simulated_variance
0.625	0.6455312

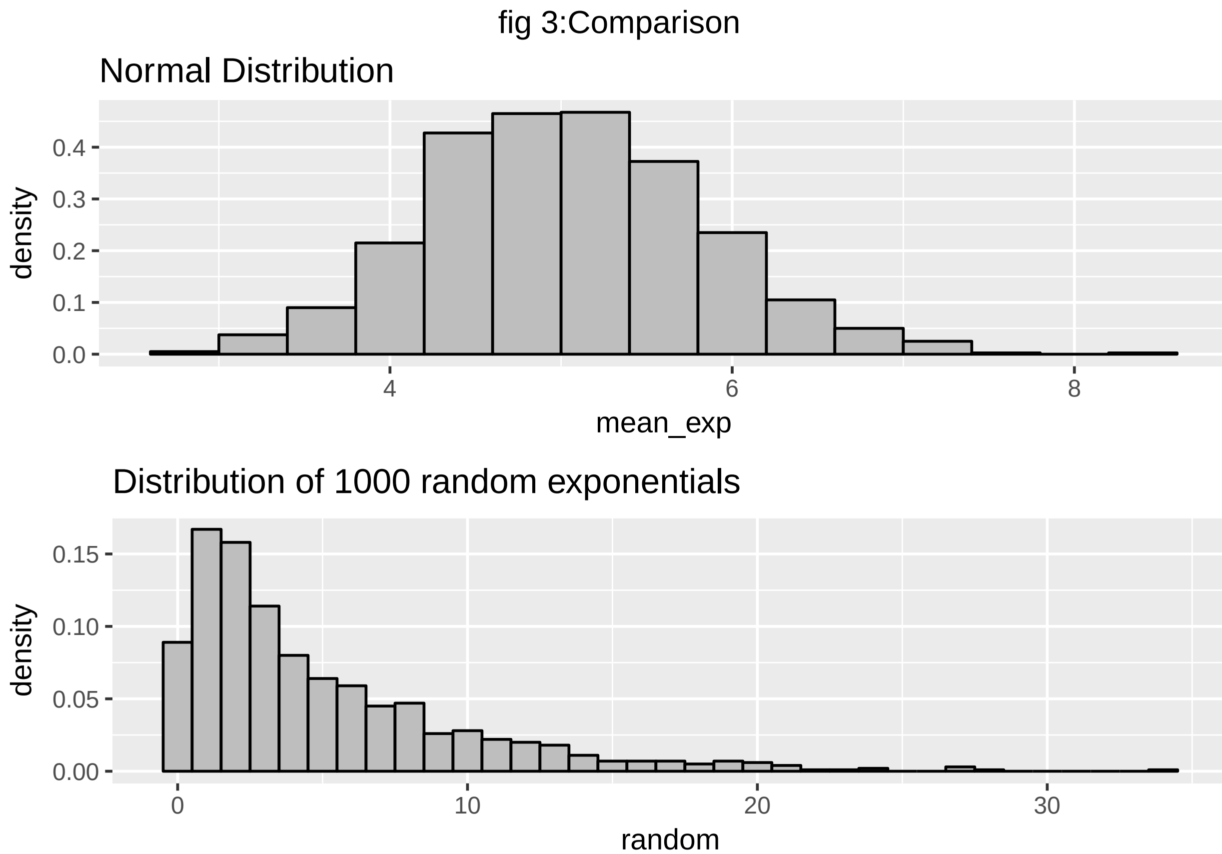
Simulated Exponential Distribution is Normal

As one can see from the graph below, since the distribution can be aproximated by the shape of a bell curve, it is safe to say that the distribution from simulation is normal.

fig 2:Resemblance with bell curve



Also if we compare the above distribution with the distribution of 1000 exponential variable , we will observe that the distribution of 1000 exponential values is right skewed.



APPENDIX

1-Library used are ggplot2,gridExtra,knitr

2-Code for the Table-1

dfvar<-data.frame(theoritical_variance=((1/0.2)/sqrt(40))^2,simulated_variance=var(mean_exp))
kable(dfvar,caption = "Comparison of theoritical and simulated variance")

3-Code for fig 1,2,3

fig 1
p<-ggplot(data=data.frame(mean=mean_exp),aes(mean_exp,..density..))+geom_histogram(fill="grey",color="black", binwidth = 0.4)+ ggtitle("Normal Distribution")
p+geom_vline(xintercept = mean(mean_exp),col="red")+ggtitle("fig 1:Distribution of mean")+xlab("means")+ylab("")

fig 2
theoritical_variance = ((1/0.2)/sqrt(40))^2
x<-dnorm(mean_exp,mean=5,sd=sqrt(theoritical_variance))
p+geom_line(aes(x=mean_exp,y=x),color="red",lwd=1)+ggtitle("fig 2:Resemblance with bell curve")+xlab("")+ylab("")

fig 3
y<-rexp(1000,0.2)

z<-ggplot(data=data.frame(random=y),aes(random,..density..))+geom_histogram(fill="grey",color="black",binwidth = 1)+ggtitle("Distribution of 1000 random exponentials")
grid.arrange(p,z,nrow=2,top="fig 3:Comparison")