Investigation of exponential distribution in R

By: Mahesh Dave

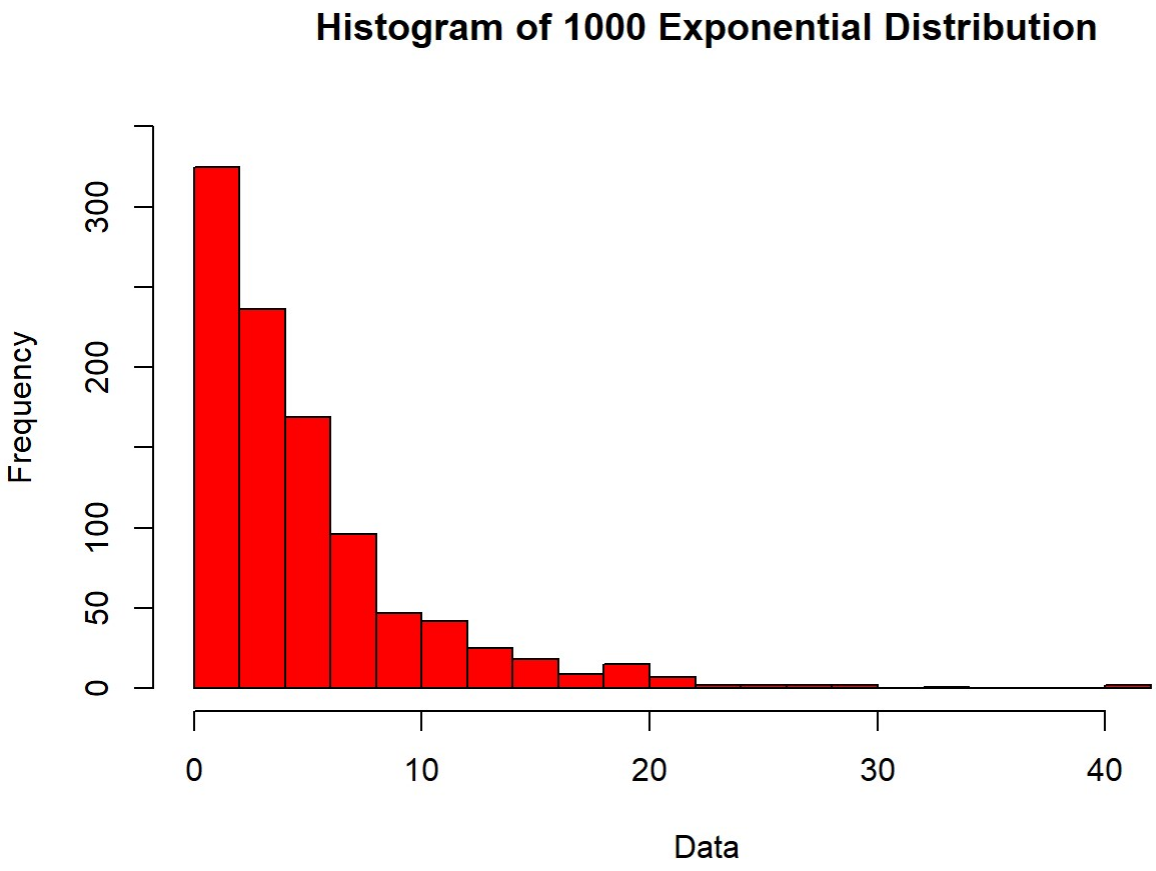
This project investigates the exponential distribution in R and compares it with the Central Limit Theorem. The lambda for the distribution was set to 0.2. The theoritical mean of this distribution was calculated by taking 1/lambda and the theoritical standard deviation was also calculated by taking 1/lambda. The central limit theorem was analyzed using means for 40 exponentials out of a 1000 data points. The simulated mean and variance were compared to the theoritcal mean and variance.

Simulation of data

```
th.lambda <- 0.2
th.std <- 1/th.lambda
th.mean <- 1/th.lambda
th.var <- th.std^2
n <- 1000
set.seed(100)
thdata <- rexp(n,th.lambda)
```

Histogram of simulated 1000 points

```
hist(thdata,breaks = 25,col="red",main="Histogram of 1000 Exponential Distribution",x
lab="Data",ylim = c(0,350),xlim = c(0,45))
```



The mean of the above distribution is:

```
th.mean
```

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

The variation of the above distribution is:

```
th.var
```

```
## [1] 25
```

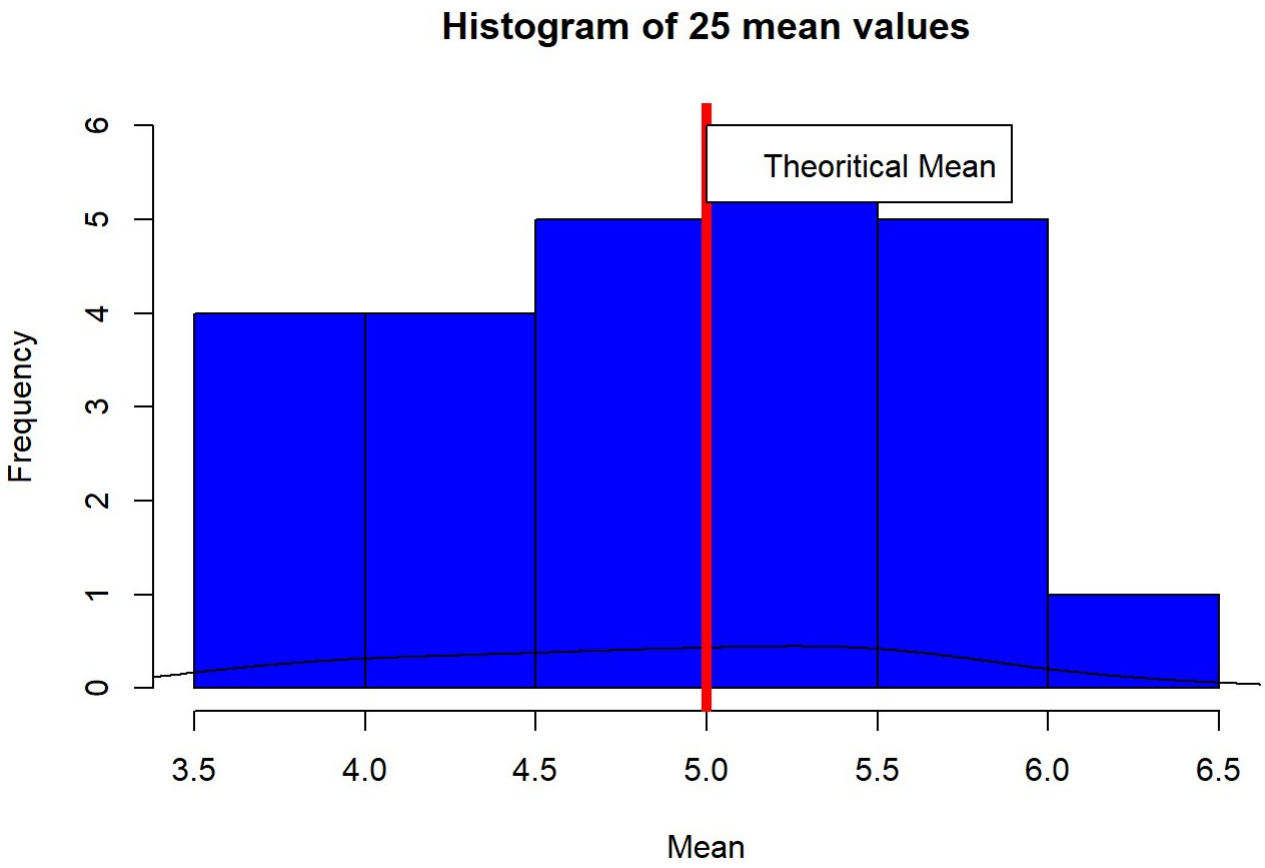
MEAN

Now we will plot 25 values of means of a sample size 40 drawn from the theorical data without replacement.

```
smatrix <- matrix(sample(thdata,40*25),40,25)
smeans <- apply(smatrix,2,mean)
svar <- apply(smatrix,2,var)
```

This is the histogram of 25 means collected from 1000 data points without replacement.

```
hist(smeans,breaks = 6,main="Histogram of 25 mean values",xlab="Mean",col = "blue")
lines(density(smeans,adjust = 1))
abline(v=th.mean,col="red",lwd=5)
legend(5,6,"Theoritical Mean",col = "red")
```

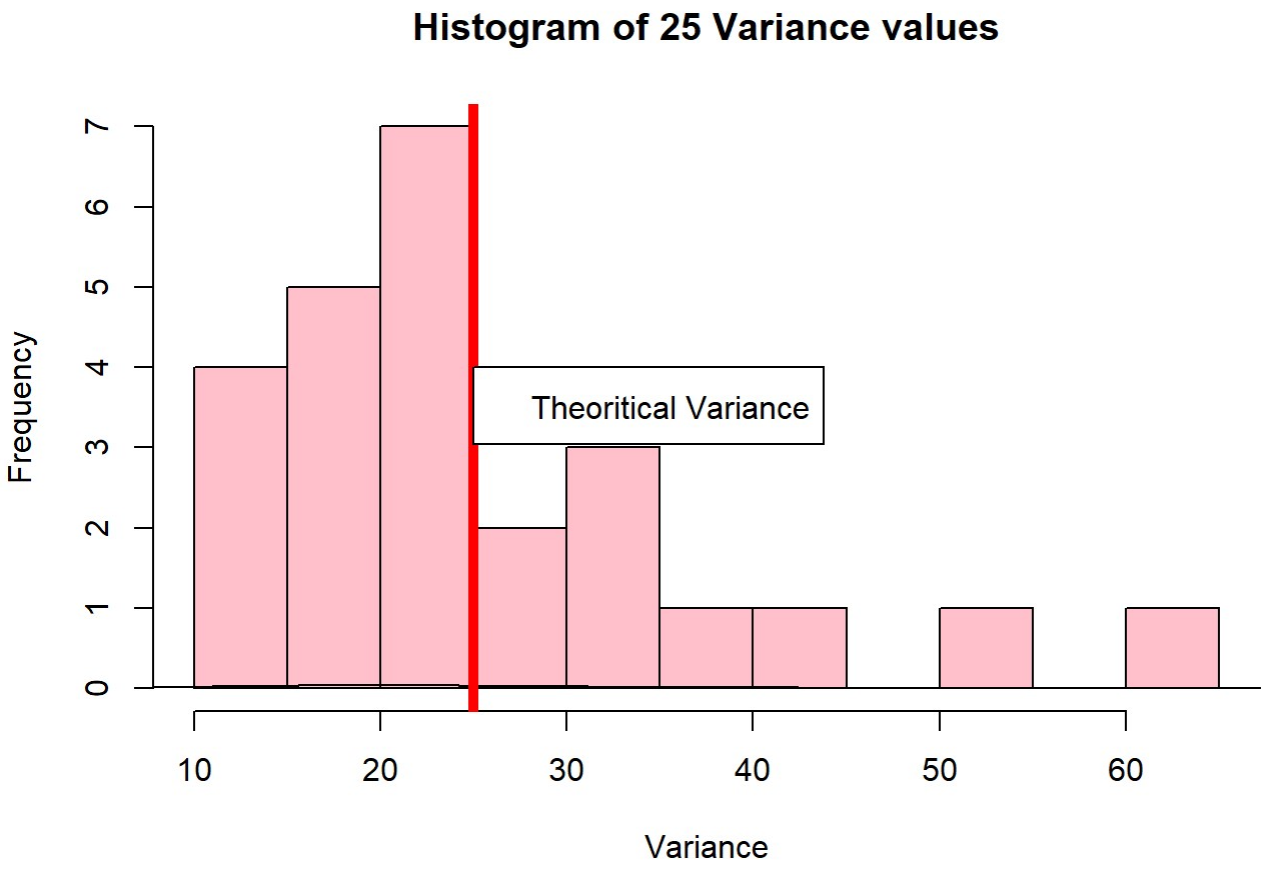


As you can see this mean is centered around the theoritical mean which is 5

VARIANCE

Now we will plot 25 values of variations of a sample size 40 drawn from the theoritical data without replacement.

```
hist(svar,breaks = 10,main="Histogram of 25 Variance values",xlab="Variance",col = "p
ink")
lines(density(svar,adjust = 1))
abline(v=th.var,col="red",lwd=5)
legend(25,4,"Theoritical Variance",col = "red")
```



As you can see this variance is centered around the theoritical variance which is 25

Conclusion:

As the sample size became larger and larger the histogram of the MEANS and VARIANCE would become more and more like NORMAL DISTRIBUTION even though the initial data is NOT normally distributed.