# Statistical Inference Project - Part 1

*Prashant Chand*

*2 Jan 2020*

**Part 1: Simulation Exercise**

**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. This project is to prove the Central LImit Theorem using distribuion of means generated using 40 random numbers simulated 1000 times

### 1. Show the sample mean and compare it to the theoretical mean of the distribution

Taking 1000 simulaions of average of 40 random numbers which has a rate parameter of lambda and storing it in a variable

```
set.seed(3)
lambda<-0.2
n<-40
simulated_means<-NULL
for(i in 1:1000){
  simulated_means<-c(simulated_means,sum(rexp(40,0.2))/n)
}
```

Theoretical Mean

```
theoretical_mean<-1/lambda
theoretical_mean
```

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

Sample Mean

```
simulated_mean<-mean(simulated_means)
simulated_mean
```

```
## [1] 4.98662
```

### Conclusion

The difference between the sample mean and the theoretical mean is very small i.e 0.01338 which means they are centred around 5.

### 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution

Theoretical Variance

```
theoretical_sd<-(1/lambda)/sqrt(n)
theoretical_variance<-(theoretical_sd)^2
theoretical_variance
```

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

Sample Variance

```
simulated_sd<-sd(simulated_means)
simulated_variance<-(simulated_sd)^2
simulated_variance
```

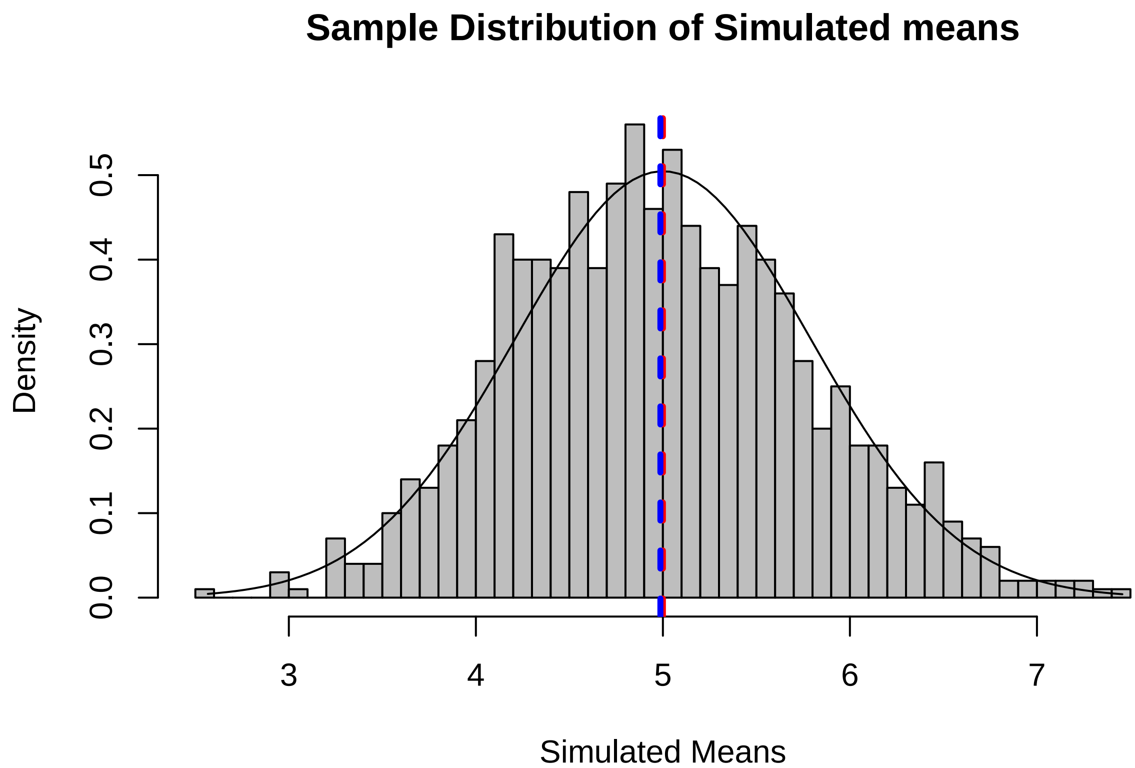
```
## [1] 0.6316789
```

#### Conclusion

The difference in variance is very small i.e 0.0066789.So both of the distriutions vary almost by the same number.

### 3. Show that the distribution is approximately normal.

```
x <- seq(min(simulated_means), max(simulated_means),length=100)
y <- dnorm(x, mean =1/lambda, sd =(1/lambda)/sqrt(n))
hist(simulated_means,breaks=n,prob=T,col="grey",xlab = "Simulated Means",main="Sample Distribution of Si
lines(x, y, pch=22, col="black", lty=1)
abline(v =theoretical_mean, col="red", lwd=3, lty=2)
abline(v =simulated_mean, col="blue", lwd=3, lty=2)
```




The blue line is sample mean line and the red line is the theoretical mean line.They almost coincide with each other that proves the distribution is centred around the same mean and have almost the same variance and it forms a normal or a gaussian distribution.

**Conclusion**

As we can see in the above plot that both the distributions are centred around the same mean i.e 5 and the have almost the same variance.So the distribution of 1000 simulations of 40 random numbers is almost a normal distribution which proves our Central Limit Theorem.