

statistical Inference Course Project-part 1

In this project the exponential distribution in R is investigated and compared with the Central Limit Theorem. Then, the distribution of averages of 40 exponentials with a thousand simulations is investigated.

#Set the variables as defined in the problem

```
nosim <- 1000
n<-40
lambda<-0.2
ActualMean <- mean(apply(matrix(rexp(nosim*n, lambda), nosim), 1,mean ))
ActualMean
```

```
## [1] 4.990025
```

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

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

```
ActualSD <- sd(apply(matrix(rexp(nosim*n, lambda), nosim), 1,mean ))
ActualSD
```

```
## [1] 0.7853029
```

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

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

```
ActualVar <- var(apply(matrix(rexp(nosim*n, lambda), nosim), 1,mean))
ActualVar
```

```
## [1] 0.6362188
```

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

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

Answers to the following questions;

1. Showing where the distribution is centered at and compare it to the theoretical center of the distribution.

The actual(empirical) distribution is centered at 4.990025 compared to the theoretical center ie. 5

2. Show how variable it is and compare it to the theoretical variance of the distribution.

The actual(empirical) variance is 0.6362188 compared to the theoretical variance ie. 0.625

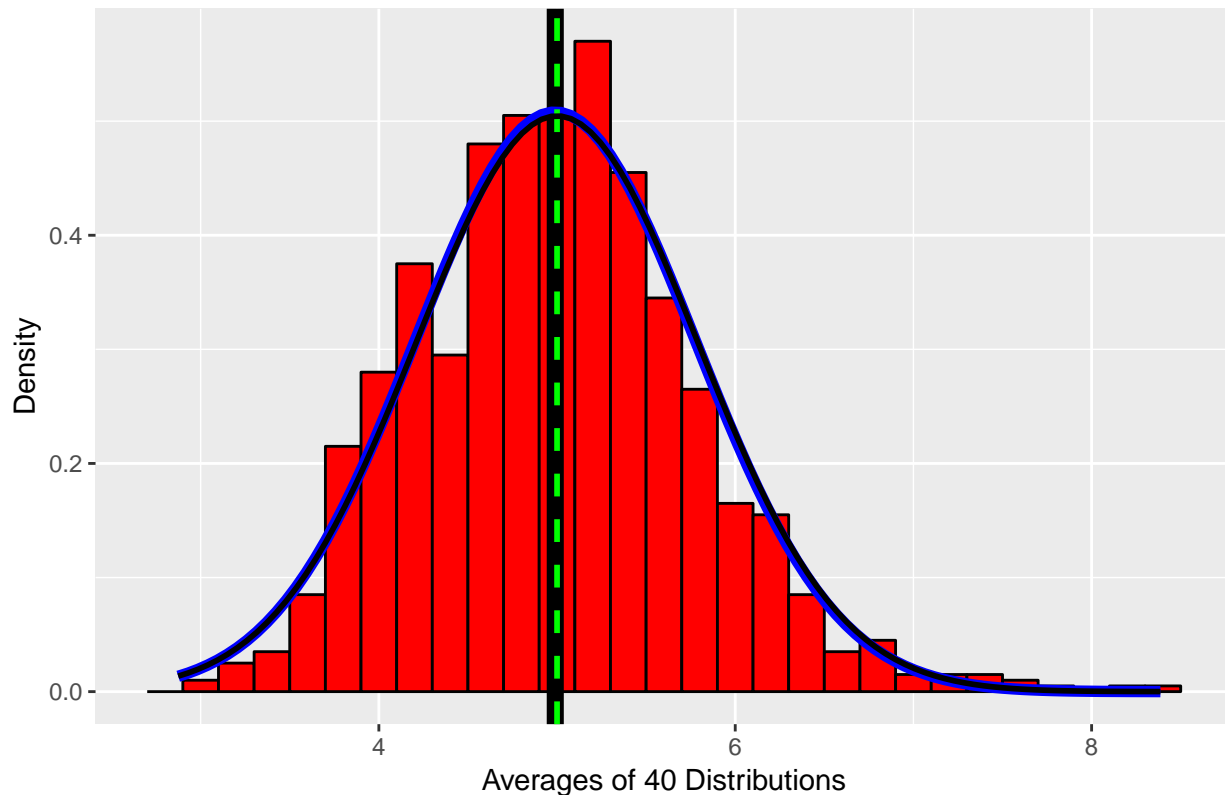
Actual(empirical) standard deviation is 0.7853029 compared to the theoretical std deviation ie. 0.7905694

The actual(empirical) distribution is centered at 4.990025 compared to the theoretical center ie. 5

3. Show that the distribution is approximately normal.

```
data <- data.frame(apply(matrix(rexp(nosim*n, lambda), nosim), 1,mean ))
row_means <- apply(matrix(rexp(nosim*n, lambda), nosim), 1,mean )
library(ggplot2)
ggplot(data ,aes(x=apply(matrix(rexp(nosim*n, lambda), nosim), 1,mean ))) +
  geom_histogram(binwidth = 0.2 ,fill="red",
                color="black",aes(y = ..density..)) +
  labs(title="Plot of 40 Exponential Distributions",
        x="Averages of 40 Distributions", y="Density") +
  geom_vline(xintercept=ActualMean,size=3.0, color="black") +
  stat_function(fun=dnorm,args=list(mean=ActualMean, sd= ActualSD),
                color = "blue", size = 2.0) +
  geom_vline(xintercept=TheoryMean,size=1.0
            ,color="green",linetype = "dashed", size=3.0) +
  stat_function(fun=dnorm,args=list(mean=TheoryMean, sd=TheorySD),
                color = "black", size = 1.0)
```

Plot of 40 Exponential Distributions



Graphical representation shows the comparison between the theoretical values and actual values for the exponential distribution. The plot clearly shows that we get a normal distribution (curve is shown in black (theoretical) and blue(actual)) due to Central Limit Theorem.

Theoretical Mean is shown in dashed Green line. Empirical (Actual) Mean is shown in thick black line. Empirical normal curve line(BLUE) from Actual Mean and Std Deviation. Theoretical normal curve line(BLACK) from Theoretical Mean and Std Deviation.