### Overview:

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem by taking many samples of size N and explore their distribution

### Simulations:

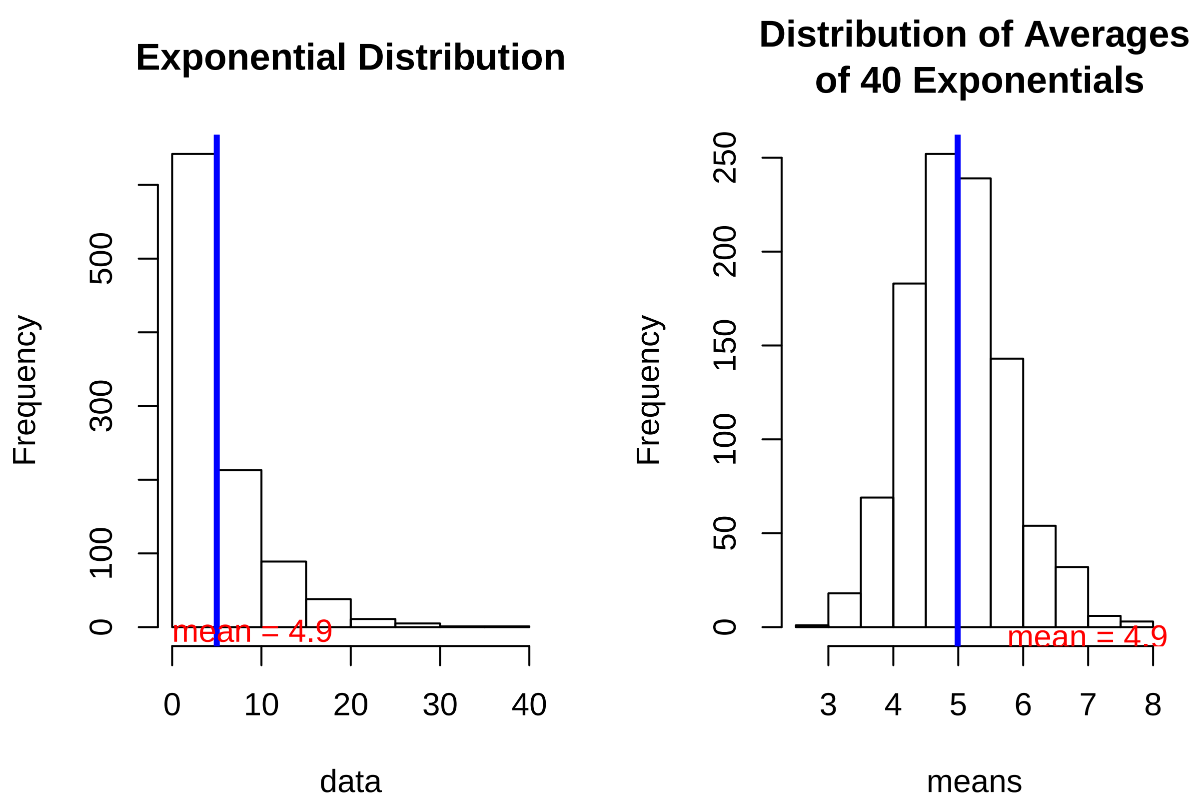
Let's take a sample of size 40 from our exponential distrubution

```
data=rexp(n=1000, rate=0.2)
means = NULL
for (i in 1 : 1000) means = c(means, mean(sample(data,40)))
```

### Sample Mean versus Theoretical Mean:

```
par(mfrow=c(1,2))
hist(data, main="Exponential Distribution")
abline(v=mean(data), col="blue", lwd=3)
text(x=mean(data)+4,y=-5, labels=c("mean = 4.9"),col="red")

hist(means, main="Distribution of Averages\n of 40 Exponentials")
abline(v=mean(means), col="blue", lwd=3)
text(x=mean(means)+2,y=-5, labels=c("mean = 4.9"),col="red")
```

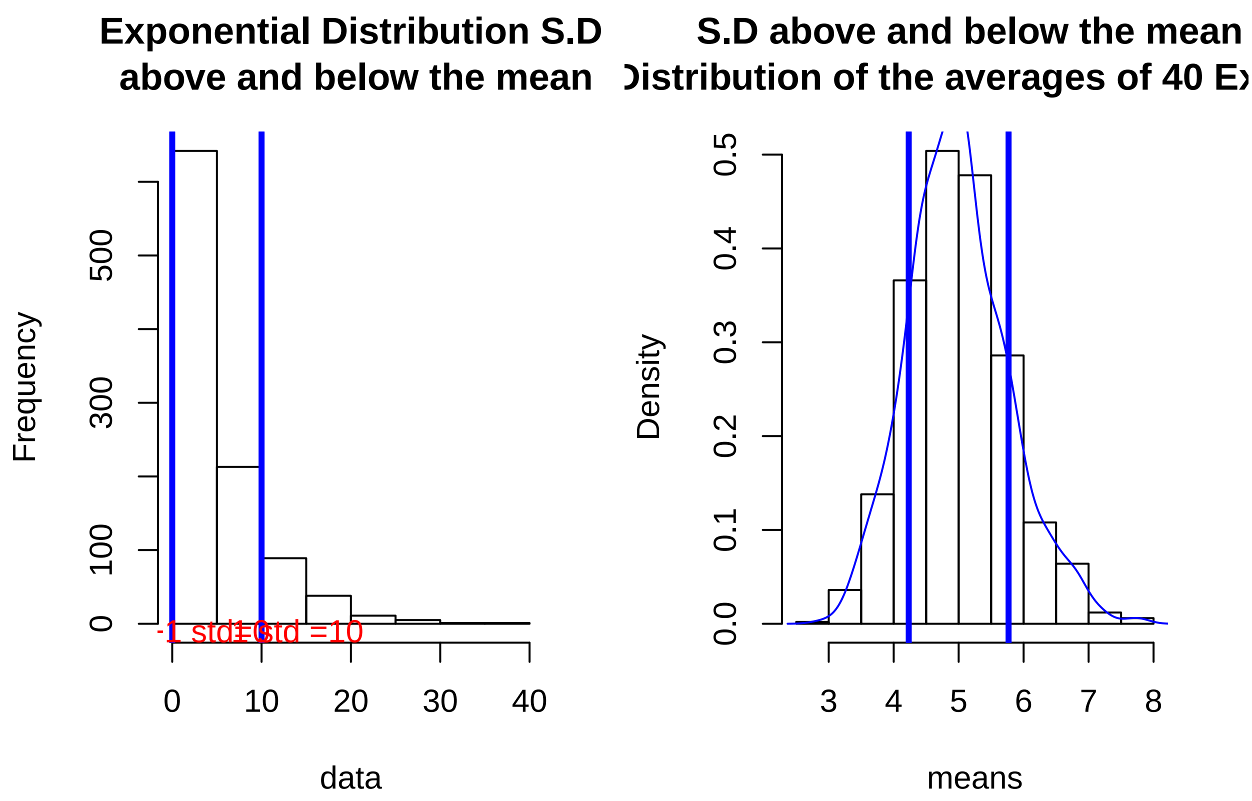


**the mean of the Distribution of the averages of 40 Exponentials is equal to the theoretical mean of the exponential distribution: which is 4.9**

## Sample Variance versus Theoretical Variance:

```
par(mfrow=c(1,2))
mu=5; std=5
hist(data, main="Exponential Distribution S.D\n above and below the mean")
abline(v=mu-std, col="blue", lwd=3)
abline(v=mu+std, col="blue", lwd=3)
text(x=mu-std+4,y=-10, labels=c("-1 std=0"),col="red")
text(x=mu+std+4,y=-10, labels=c("1 std =10"),col="red")

mu=5; std=sd(means)
hist(means,freq=FALSE, main="S.D above and below the mean \nin the Distribution of the averages of 40 E
lines(density(means),col="blue")
abline(v=mu-std, col="blue", lwd=3)
abline(v=mu+std, col="blue", lwd=3)
text(x=mu-std+0.3,y=-4, labels=c("-1 std=4.13"),col="red")
text(x=mu+std+0.3,y=-4, labels=c("1 std =5.67"),col="red")
```




**the distribution of the averages of 40 Exponentials have a mean of 0.8 as empirically shown below, which is equal to standard deviation of the original distribution/sqrt(sample size) = 4.9/sqrt(40)=0.77**

### Distribution Normality

we can tell the distribution of the averages of 40 Exponentials is approximately normal as the following properties are present:

**1-the distribution looks like a bell curve and symetric, having probability of first half=probability of second half = 0.5**

```
cdf=ecdf(means)##Compute an empirical cumulative distribution function
cdf(median(means))##cdf of the 50th percentile
```

```
## [1] 0.5
```

**3-According to the Empirical Rule, in a normal distribution:**

68%, 95% and 99.7% of the data fall within one, two and three standard deviations of the mean. in the code shown below, we will check is this is the case in our distribution

```
p1=(cdf(mu+std)-cdf(mu-std))*100
p2=(cdf(mu+2*std)-cdf(mu-2*std))*100
p3=(cdf(mu+3*std)-cdf(mu-3*std))*100
print(paste0("probability of data within 1 standard deviation below and above the mean is: ",p1,"%"))
```

```
## [1] "probability of data within 1 standard deviation below and above the mean is: 70.9%"
```

```
print(paste0("probability of data within 2 standard deviations below and above the mean is: ",p2,"%"))
```

```
## [1] "probability of data within 2 standard deviations below and above the mean is: 94.8%"
```

```
print(paste0("probability of data within 3 standard deviations below and above the mean is: ",p3,"%"))
```

```
## [1] "probability of data within 3 standard deviations below and above the mean is: 99.7%"
```

*Thus, we can conclude that the distribution of the averages of 40 Exponentials is approximately normal.*

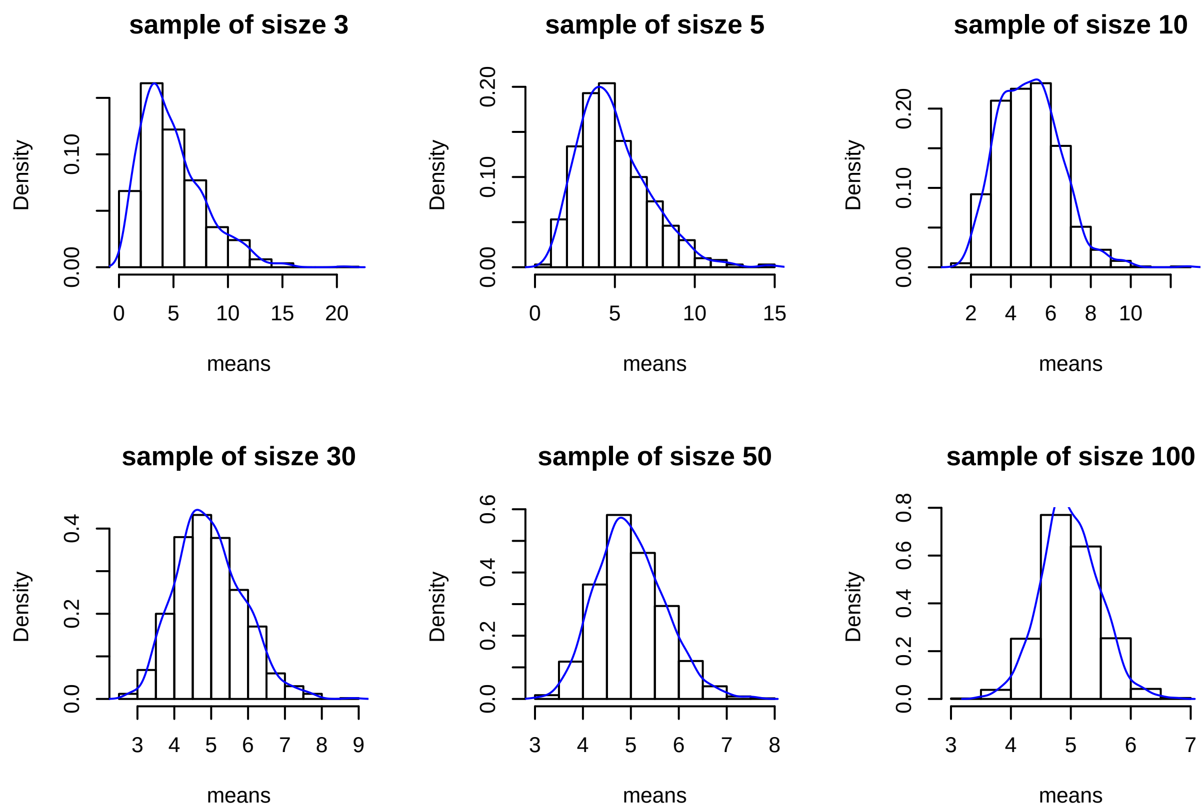
### Appendix

# Distribution of the averages of N Exponentials

Below are distribution of averages of different sample sizes: 3, 7, 15, 30, 50, 100

```
par(mfrow=c(2,3))

for (n in c(3, 5, 10, 30, 50, 100)){
  ##sample size n
   means = NULL
   for (i in 1 : 1000) means = c(means, mean(sample(data,n)))
   hist(means,freq=FALSE, main=paste0("sample of sisze ",n))
   lines(density(means),col="blue")
}
```




**As it's shown, as the sample size increases, the distribution becomes more normal, with a smaller standard deviation (gets narrower)**