

# Statistical Inference Part1

Vamshi Krishna

12/9/2019

## Synopsis:

In this project I am going to investigate the exponential distribution in R and compare it with Central

## Instructions:

1.Show the sample mean and compare it to the theoretical mean of the distribution. 2.Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. 3.Show that the distribution is approximately normal.

## Preparing simulation data

Setting variables as defined in the problem and calculating means of 1000 simulations by running a for loop from 1 to 1000 and storing each time a mean value of rexp with n and lambda in in meanData variable.

```
set.seed(123)      #for reproducibility
n=40               #sample size
lambda=0.2         #given lambda for rexp
meanData=NULL
for(i in 1:1000) meanData = c(meanData, mean(rexp(n,lambda)))
```

## 1. Comparing Sample Mean and Theoretical Mean

### Sample Mean:

Lets calculate the mean of meanData to get the sample mean of simulation data

```
sampleMean = mean(meanData)
sampleMean
```

```
## [1] 5.011911
```

### Theoretical Mean:

Lets also calculate the theoretical mean. as we discussed earlier in the synopsis mean is the 1 over lambda,

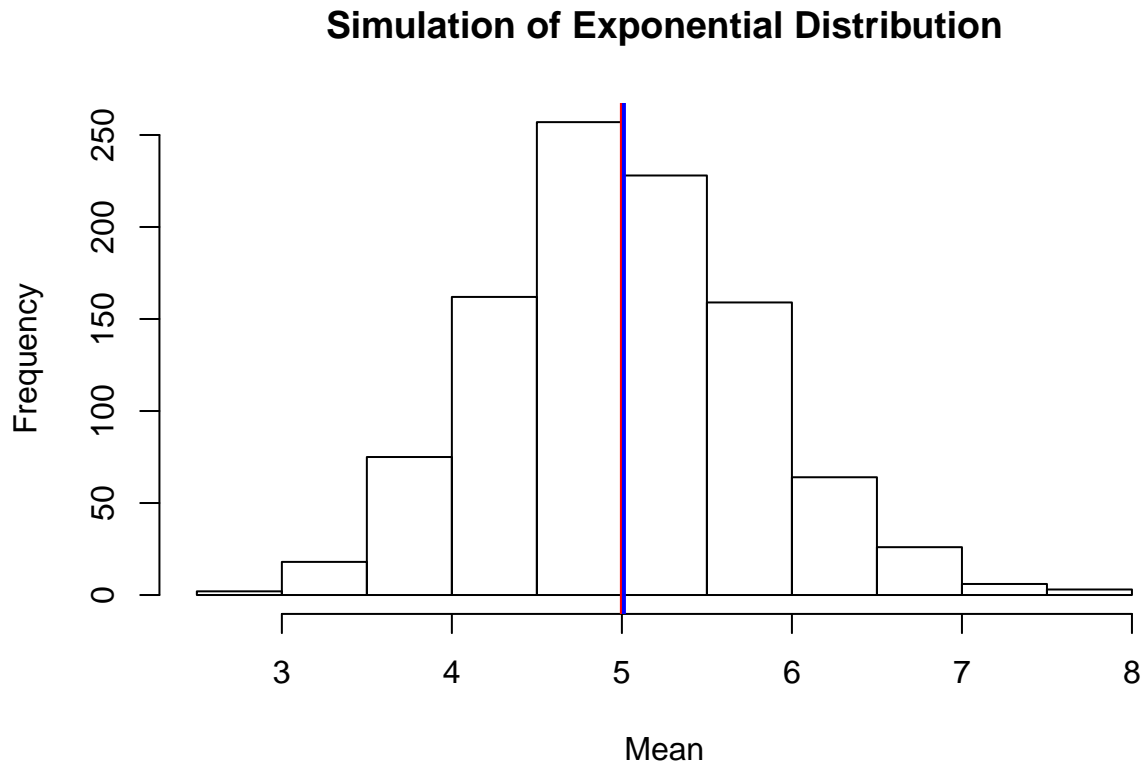
```
theoreticalMean=1/lambda
theoreticalMean
```

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

The distribution of mean of the sample mean is centered at 5.0119113 and theoretical mean is centered at 5. Both the means are very close.

Lets see the Sample Mean versus Theoretical Mean through a histogram:

```
hist(meanData, xlab = "Mean" , ylab = "Frequency", main = "Simulation of Exponential Distribution")
abline(v=theoreticalMean, col = "red", lw=2)
abline(v=sampleMean, col = "blue", lw=2)
```



## 2. Comparing Sample Variance and Theoretical Variance

**Sample Standard deviation:**

Lets calculate the sample standard deviation in order to calculate the sample variance

```
sampleSD = sd(meanData)
sampleSD
```

```
## [1] 0.7749147
```

**Sample Variance:**

and we get sample variance from sample standard deviation by squaring it or we can use `var()` function

```
sampleVariance = var(meanData)
sampleVariance
```

```
## [1] 0.6004928
```

### Theoretical Standard deviation:

Lets also calculate the theoretical standard deviation which is 1 over lambda

```
theoreticalSD = 1/(lambda*sqrt(n))
theoreticalSD
```

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

### Theoretical Variance:

and we get theoretical variance from theoretical standard deviation by squaring it

```
theoreticalVariance = 1/(lambda^2 * n)
theoreticalVariance
```

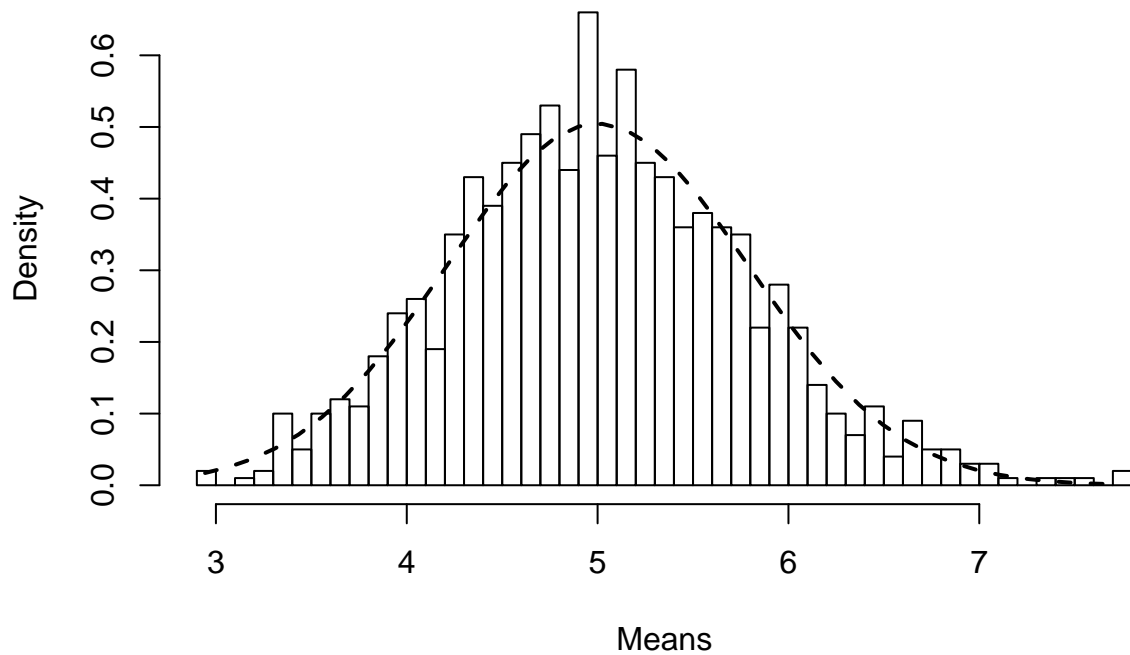
```
## [1] 0.625
```

we can see both sample standard deviation 0.7749147 and 0.7905694 are almost close and we can also see both sample variance 0.6004928 and 0.625 are also close.

### 3. Distribution is approximately normal:

```
hist(meanData, breaks = n, prob = T, xlab = "Means" , ylab = "Density" , main = "Density curve of mean")
xPlot <- seq(min(meanData), max(meanData), length=40)
yPlot <- dnorm(xPlot, mean=1/lambda, sd=((1/lambda)/sqrt(n)))
lines(xPlot, yPlot, col="black", lty=2, lwd=2)
```

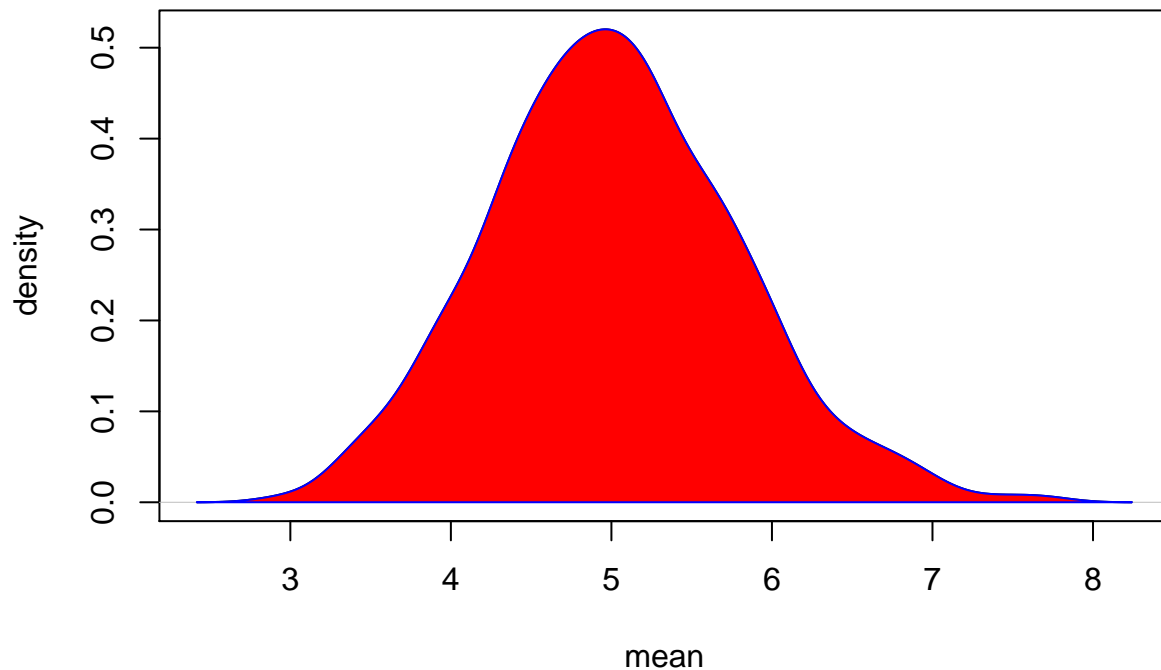
## Density curve of meanData



Lets plot a complete density plot

```
d=density(meanData)
plot(d, xlab = "mean" , ylab ="density", main = "Density plot of 1000 Exponential simulations")
polygon(d, col = "red" , border = "blue")
```

## Density plot of 1000 Exponential simulations



The above figures shows the distribution of sample means, it is approximately normal. The black density curve corresponds to  $N(5, 0.625)$

### 4. Confidence Interval Comparison:

Lets calculate both sample and theoretical Confidence Intervals:

```
theoreticalCI = theoreticalMean + c(-1,1)*qnorm(0.975)*theoreticalSD/sqrt(n)
round(theoreticalCI,3)
```

```
## [1] 4.755 5.245
```

```
sampleCI = sampleMean + c(-1,1)*qnorm(0.975)*sampleSD/sqrt(n)
round(sampleCI)
```

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

The sample confidence interval 4.7717671, 5.2520554 and theoretical confidence interval 4.7550045, 5.2449955 which are also close. Hence the distribution is proved to be approximately Normal.

### Conclusion:

It is determined that the distribution does indeed demonstrate the Central Limit Theorem, a bell curve. After graphing all the values and comparing means, variance and confidence intervals it is proved that the Exponential distribution is approximately Normal.