Statistical Inference Project 1

by,

Ravi Teja L

22-06-2020

Part-1

Comparing Exponential Distribution in R to Central Limit Theorem

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.

Loading Libraries

```
library("data.table")
library("ggplot2")
library("dplyr")
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##     between, first, last

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(ggpubr)
```

Production

```
# set seed
set.seed(31)

# set lambda
lambda <- 0.2
```

```r
# samples
n <- 40

# simulations
simulations <- 1000

samples <- matrix(sample(rexp(n, lambda), n*simulations, replace = TRUE), n, simulations)

sampleMeans <- apply(samples, 2, mean)
```

Question-1

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

```r
obtainedMean <- mean(sampleMeans)
obtainedMean
```

```
## [1] 5.683866
```

```r
basicMean <- 1/lambda
basicMean
```

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

```r
t.test(sampleMeans)$conf
```

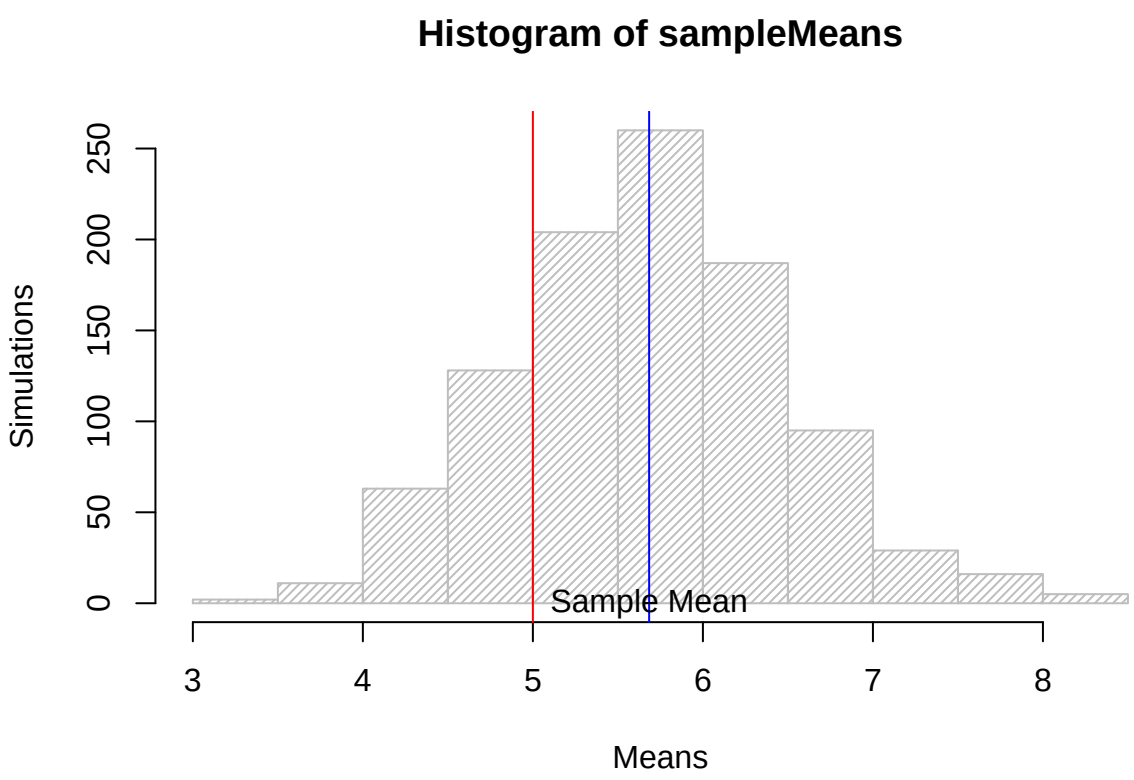
```
## [1] 5.633892 5.733841
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(sampleMeans)$p.value
```

```
## [1] 0
```

- Based on the comparision we can understand the fact that the true mean must be definitely greater than 5 i.e., Theoretical Mean.
- This is due to the fact that the theoretical mean is less than the 95% confidence interval of resampled data.
- Hence we reject the Null Hypothesis that True Mean is equal to 5.

```r
hist(sampleMeans, xlab = "Means", ylab = "Simulations", density = 30, col = "grey")
abline(v = basicMean, col = "red")
abline(v = obtainedMean, col = "blue", )
text(x = obtainedMean,y = 0,"Sample Mean")
```



Question-2

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

```
sampleSds <- sd(sampleMeans)
sampleSds
```

```
## [1] 0.805329
```

```
basicSd <- (1/lambda)/sqrt(n)
basicSd
```

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

```
sampleVariances <- sampleSds^2
sampleVariances
```

```
## [1] 0.6485548
```

```
basicVariance <- basicSd^2
basicVariance
```

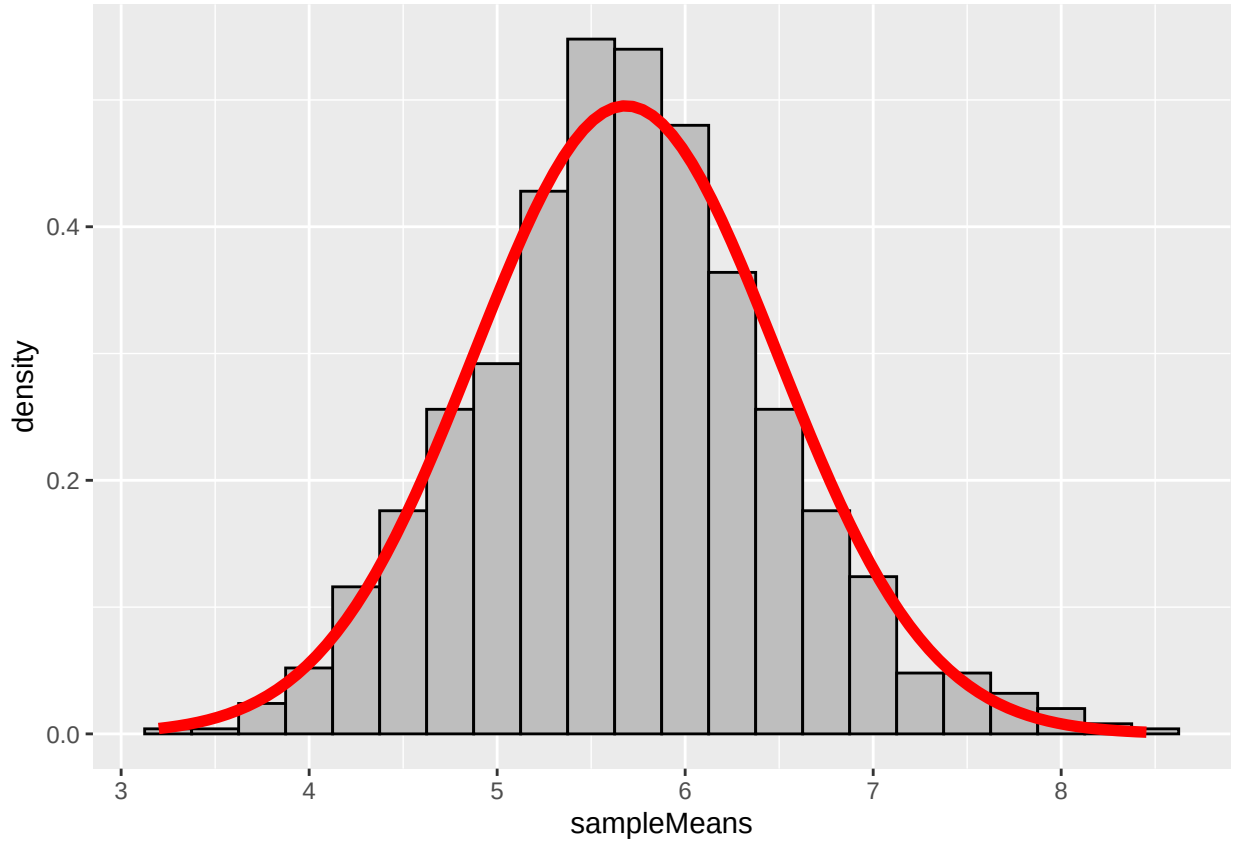
```
## [1] 0.625
```

- Standard Deviation of the distribution is 0.805329 with the theoretical SD calculated as 0.7905694. The Theoretical variance is 0.625.
- The actual variance of the distribution is 0.6485548

Question-3

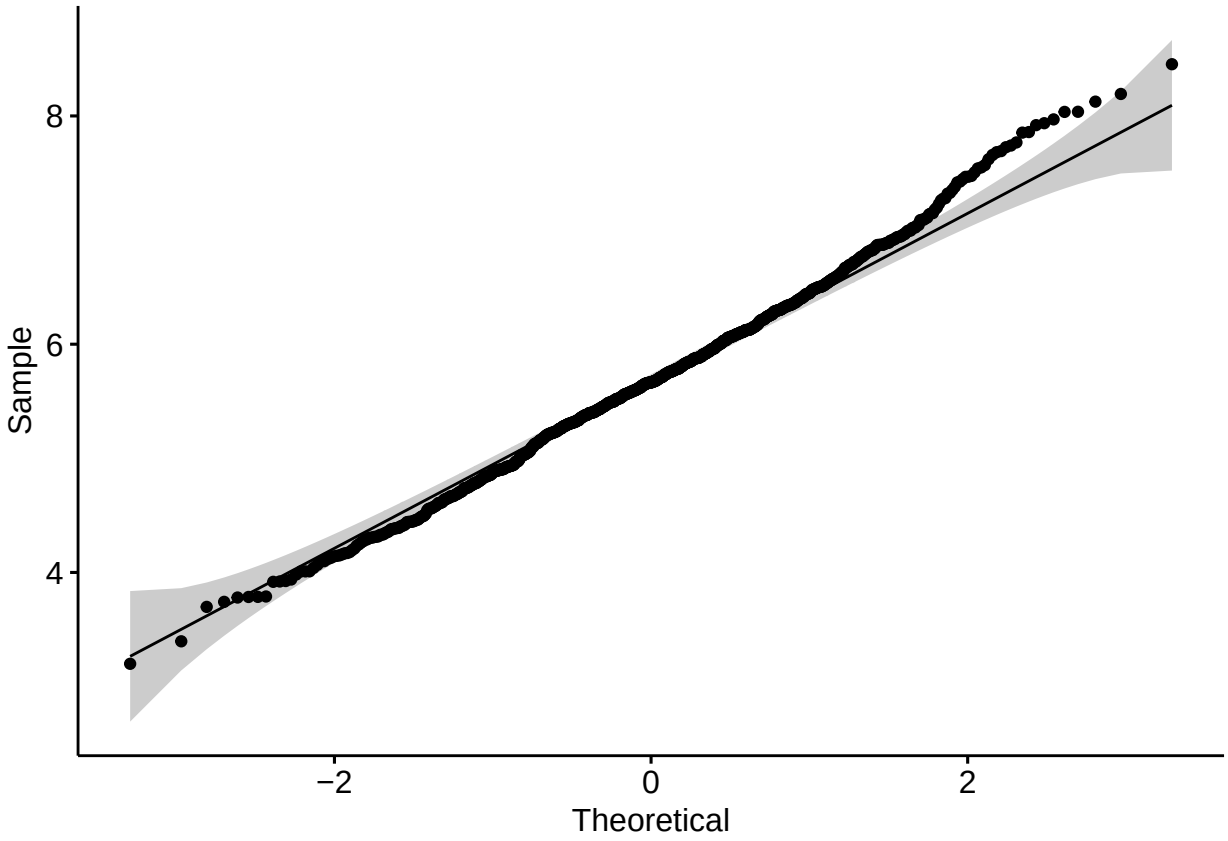
Show that the distribution is approximately normal

```
ggplot(data.frame(sampleMeans), aes(x=sampleMeans)) +
geom_histogram(aes(y = ..density..), binwidth=.25, colour="black", fill="grey")+
stat_function(fun = dnorm, lwd = 2, col = 'red', args = list(mean = mean(sampleMeans), sd = sd(sampleMea
```



From the above graph we can tell that the data is normally distributed.

```
ggqqplot(sampleMeans)
```



As all the points fall approximately along this reference line, Hence we can assume normality. Due to Due to the **Central limit theorem (CLT)**, the distribution of averages of 40 exponentials is very close to a normal distribution.