Simulation Exercise

krrish16

18 August 2018

PART 1 - Overview

The purpose of this data analysis is to investigate the exponential distribution and compare it to the Central Limit Theorem. For this analysis, the lambda will be set to 0.2 for all of the simulations. This investigation will compare the distribution of averages of 40 exponentials over 1000 simulations.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.1
```

```
lam = 0.2
expos = 40
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(expos, lam)))
```

Sample mean and Theoretical mean

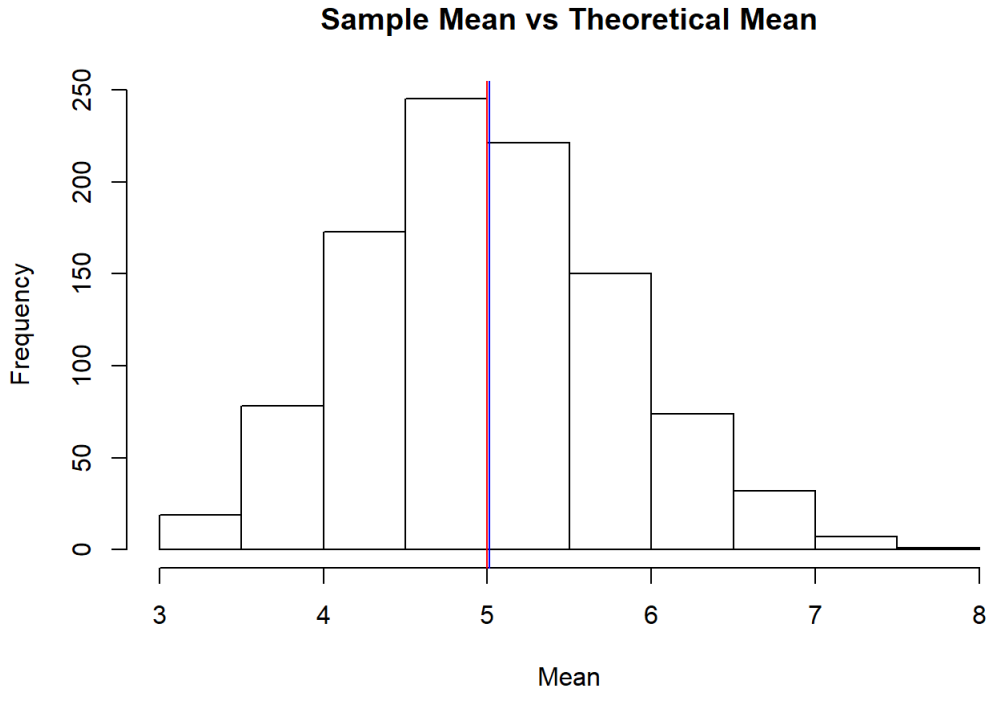
```
sample_mean <- mean(mns)
theo_mean <- mean(1/lam)
```

Comparing Sample and Theoretical mean

```
abs(sample_mean - theo_mean)
```

```
## [1] 0.0117172
```

```
hist(mns,main = "Sample Mean vs Theoretical Mean",xlab = "Mean")
abline(v=theo_mean, col="red")
abline(v=sample_mean , col ="blue")
```



There is only a slight difference between the simulations sample mean and the exponential distribution theoretical mean.

Sample Variance and Theoretical Variance

```
sample_var <- var(mns)
theo_var <- (1/lam)^2/expos
sample_sd <- sd(mns)
theo_sd <- 1/(lam * sqrt(expos))
```

Comparing Sample and Theoretical Variance

```
abs(sample_var - theo_var)
```

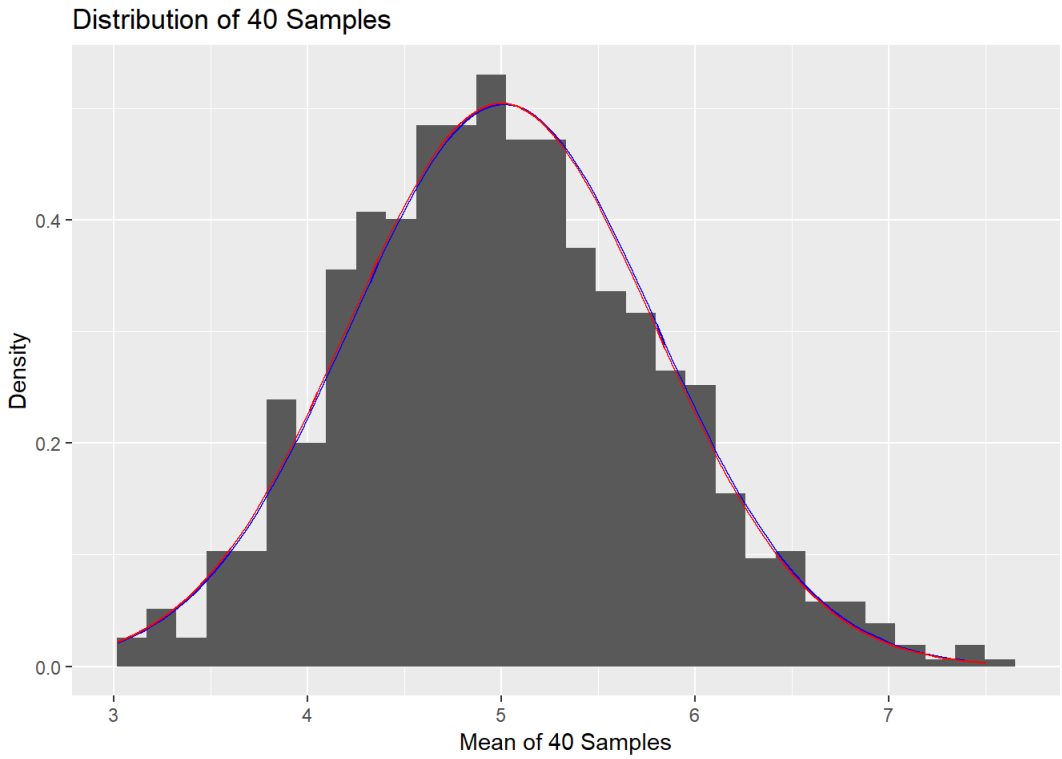
```
## [1] 0.001678311
```

The variance of this sample distribution and the theoretical variance are very close with only a little difference

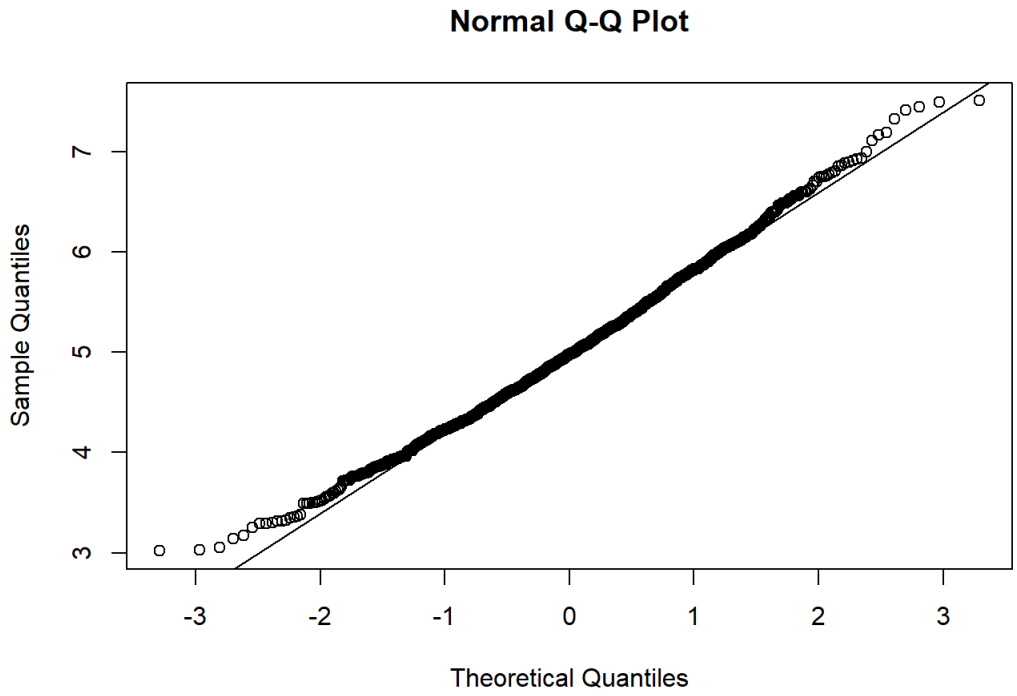
Plots

```
p <- data.frame(mns)
ggplot(p, aes(x =mns)) + geom_histogram(aes(y=..density..)) + labs(title = "Distribution of 40 Samples", x =
"Mean of 40 Samples", y = "Density") + stat_function(fun = dnorm, args = list(mean = sample_mean, sd = sampl
e_sd), color = "blue") + stat_function(fun = dnorm, args = list(mean = theo_mean, sd = theo_sd),color = "red
")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
qqnorm(mns)
qqline(mns)
```



Conclusion

From the plots of the 1000 samples taken from an exponential distribution, we see that the mean and variance of the sample means closely approximates the theoretical mean and variance. Additionally, the distribution of the sample means closely approximates a normal distribution.