Statistical Inference Project_Part 1

Moh Ahmed

04/02/2017

Synopsis

This project investigates the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. lambda = 0.2 is set for all of the simulations. In this investigation, a thousand simulations will be performed for the distribution of averages of 40 exponentials.

Simulations

In this part, a one thousand simulations will be perfomed on the distribution of means of 40 exponentials. Then the results are plotted. The following steps explain how the simulation was performed:

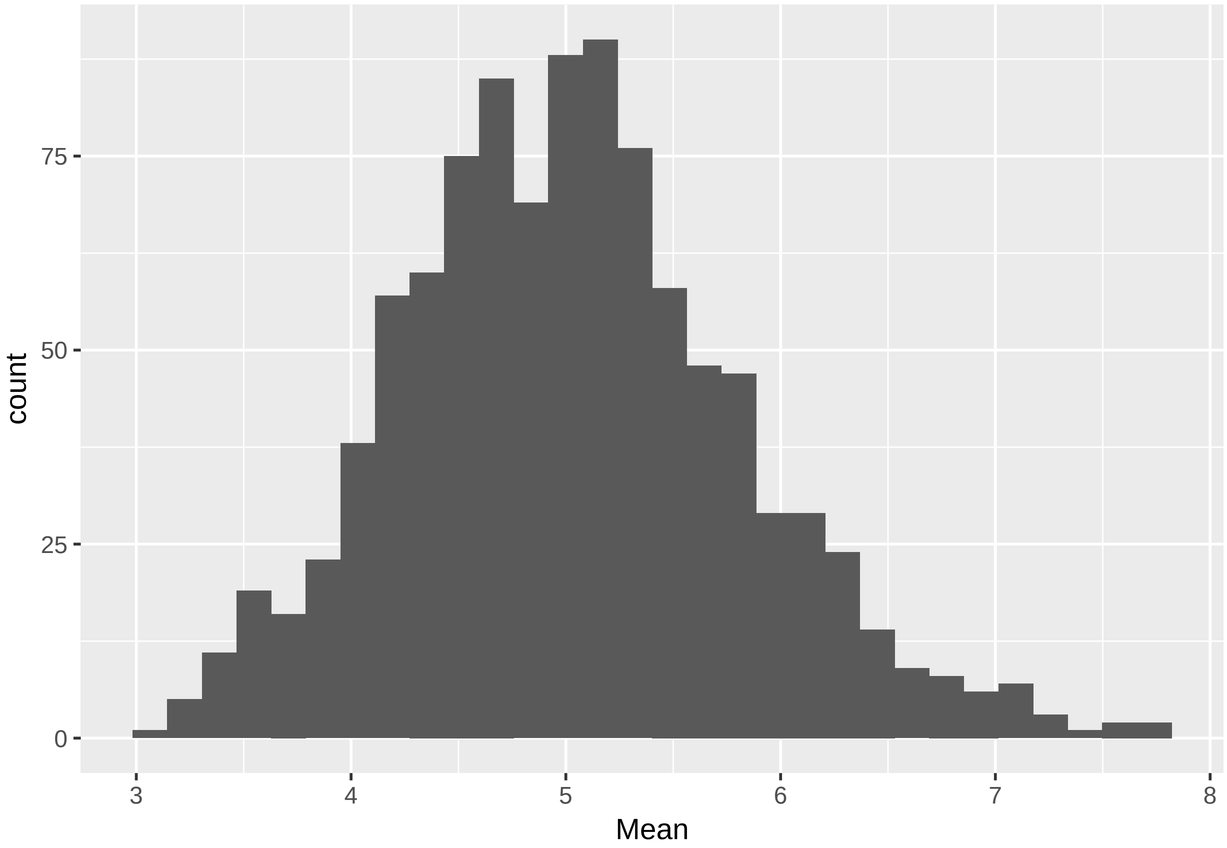
1. Set the seed to an arbitrary value for results reproducability.
2. Set the parameters for the exponential distribution function rexp(n, rate), where n= 40 observations and rate = 0.2.
3. Perform the 1000 simulations and store the results in matrix.
4. Calculate the mean distribution of 40 exponentials for the 1000 simulations.
5. Plot the results on a histogram.

```r
library(ggplot2) # load the plotting library

# set the parameter
set.seed(111)
n <- 40
lambda <- 0.2 # rate
num_simulation <- 1000

# running the simulations
exp_dist <- matrix(data = rexp(n * num_simulation, lambda), nrow = num_simulation)
exp_dist_mn <- data.frame(mn = apply(exp_dist, 1, mean)) #calculating the means

ggplot(exp_dist_mn, aes(mn)) + geom_histogram() + xlab('Mean') #ploting the results
```



Theoretical Mean compared to the Sample Mean

```
sample_mn <-  mean(exp_dist_mn$mn)
theoretical_mn <- 1/lambda
diff <-  theoretical_mn - sample_mn
diff #The difference between sample mean and theoretical mean
```

```
## [1] -0.02561954
```

As shown above, the theoretical mean = 1/ lambda = 5; while the sample mean = 5.026. This means that both theoretical and sample means have very close values.

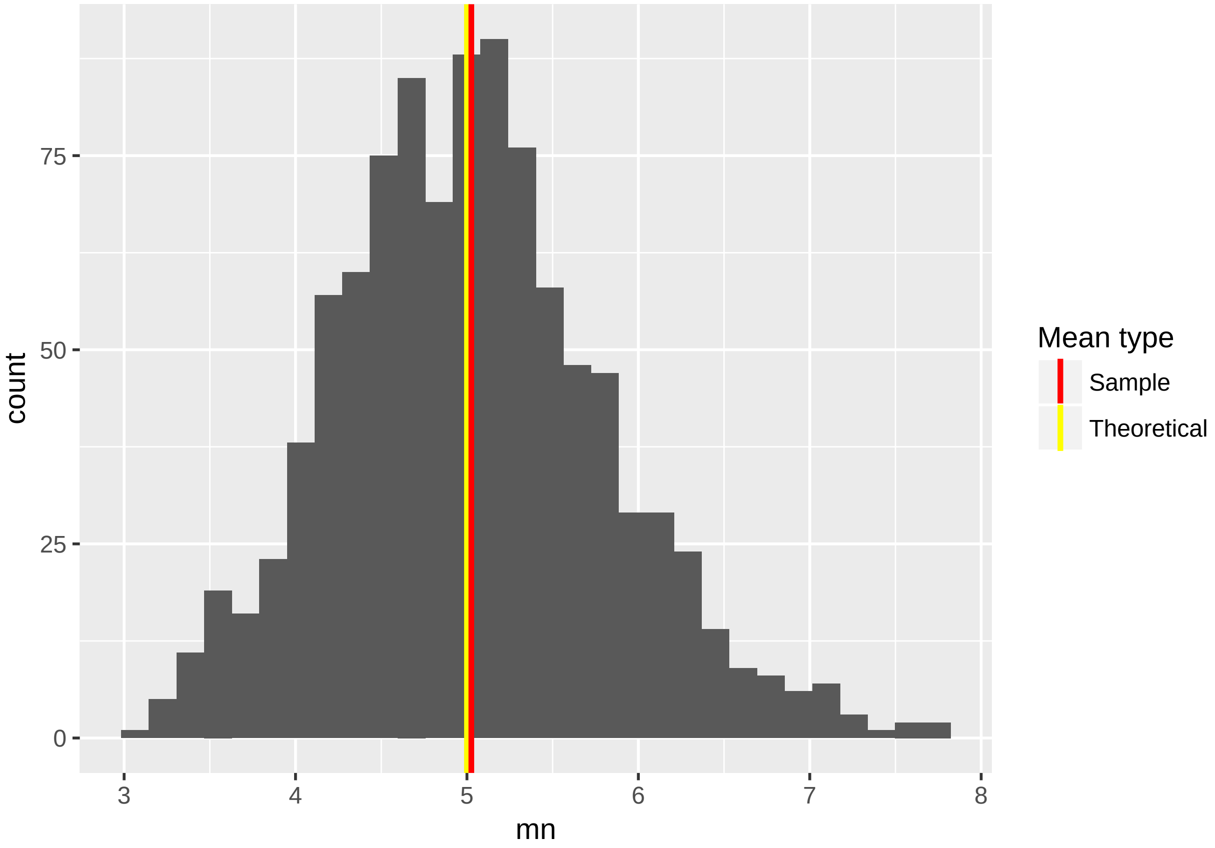
As can be seen in the following graph, the difference between the theoretical mean and the sample mean is very small.

```
sample_mn <-  mean(exp_dist_mn$mn)
theoretical_mn <- 1/lambda
diff_mn <-  theoretical_mn - sample_mn
diff_mn #The difference between sample mean and theoretical mean
```

```
## [1] -0.02561954
```

```
mns_mat <- data.frame(value = c(theoretical_mn, sample_mn), type= c('Theoretical', 'Sample') )

ggplot(exp_dist_mn, aes(mn)) + geom_histogram() +
    geom_vline(
    aes(xintercept = value, color = type),
    data = mns_mat,
    lwd = 1,
    show.legend = T
    ) +
    scale_color_manual("Mean type", values = c("Theoretical" = "yellow", "Sample" = "red"))
```



Theoretical Variance compared to the Sample Variance

```
theoretical_var <- ( (1/lambda) / sqrt(n) )^2
sample_var <- var(exp_dist_mn)
diff_var <- theoretical_var - sample_var
diff_var
```

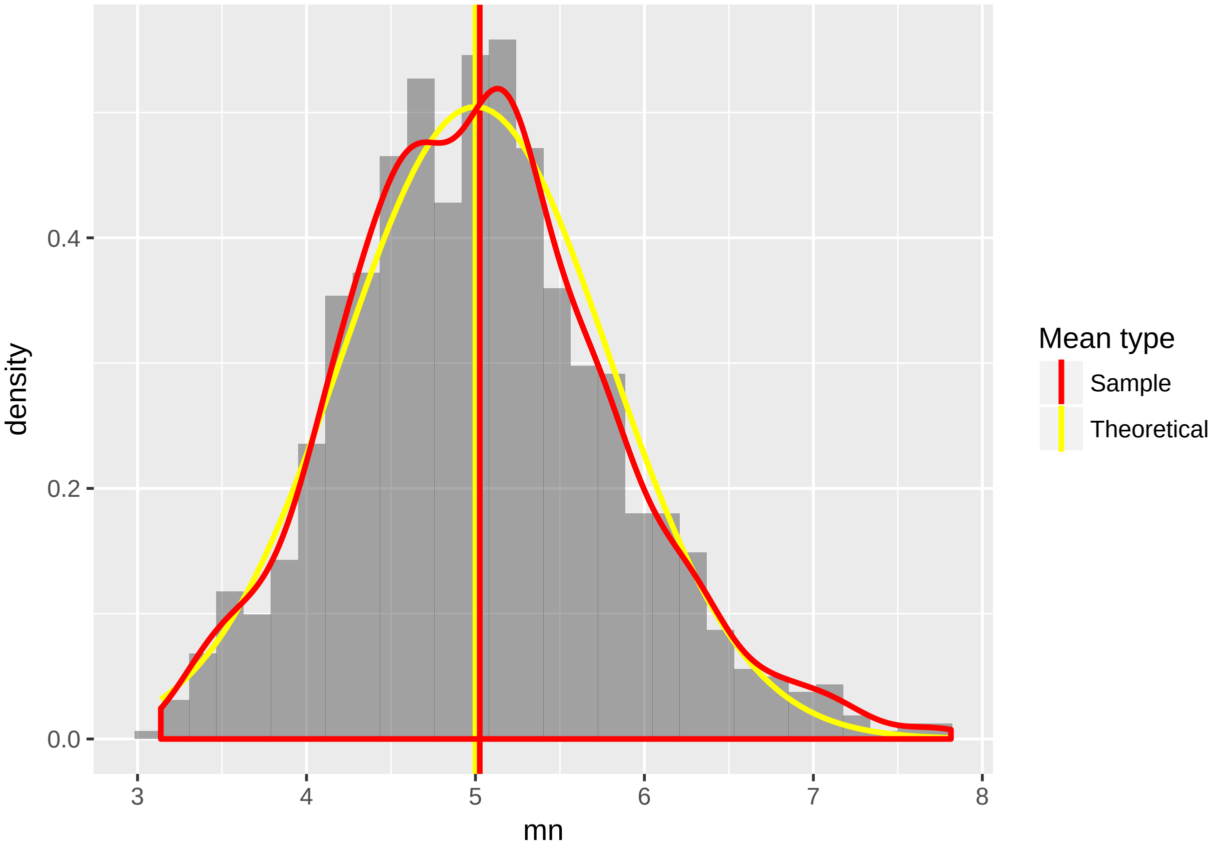
```
##              mn
## mn -0.006229641
```

As shown above, the theoretical variance = 0.625; while the sample variance = 0.631. Though variance equals standard deviation squared and hence differences will be enhanced, however both theoretical and sample var have very close values.

Distribution

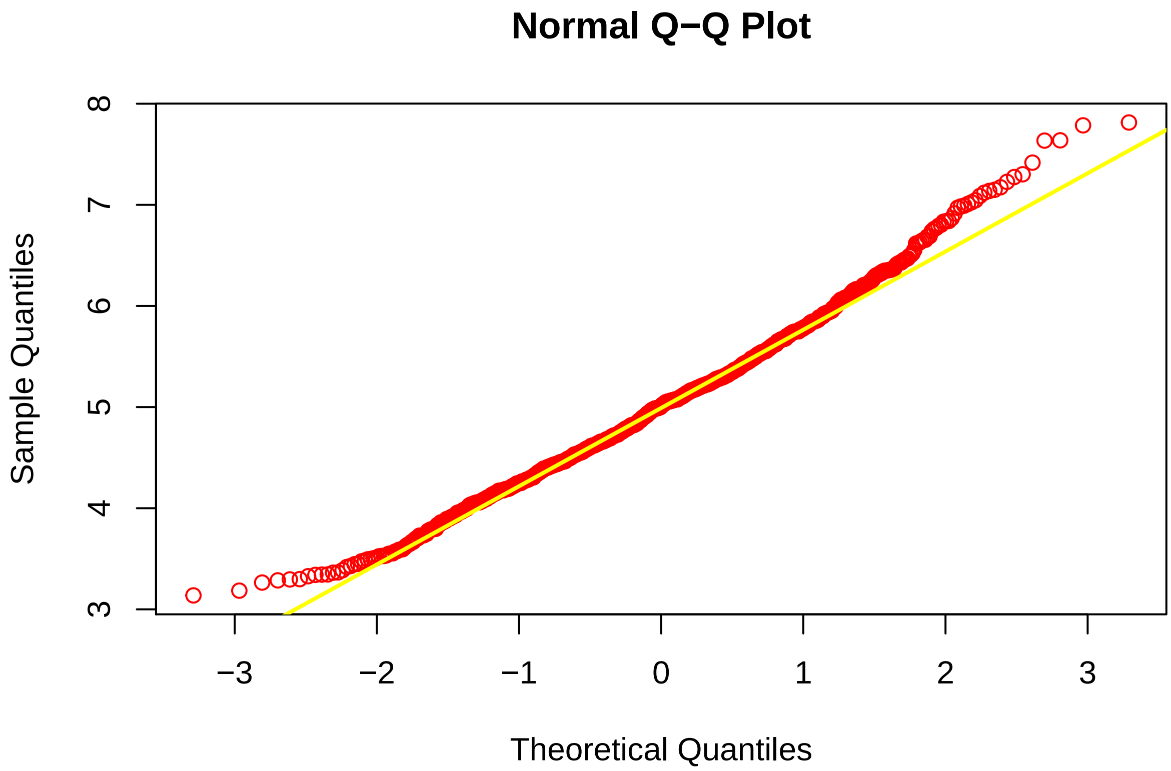
In this section, the sample distribution and the theoretical distribution are both plotted on the same scale. As shown in the figure below the sample means distribution closely match the theoretical means distribution. Hence, sample mean distribution follows a normal distribution.

```
ggplot(exp_dist_mn, aes(mn)) + geom_histogram(aes(y = ..density..), alpha = 0.5) +
    geom_vline(
    aes(xintercept = value, color = type),
    data = mns_mat,
    lwd = 1,
    show.legend = T
    ) +
    scale_color_manual("Mean type", values = c("Theoretical" = "yellow", "Sample" = "red")) +
    stat_function(
    fun = dnorm,
    args = list(mean = theoretical_mn, sd = sqrt(theoretical_var)),
    colour = 'yellow',
    size = 1
    ) +
    geom_density(color = 'red', size = 1)
```



Also shown below the Q-Q plot, or quantile-quantile plot, which is a graphical tool to help us assess if a set of data plausibly came from some theoretical distribution such as a Normal. The figure below supports the normality of the distribution.

```
qqnorm(exp_dist_mn$mn,  main ="Normal Q-Q Plot", col = 'red')
qqline(exp_dist_mn$mn,  col = "yellow", lwd = 2)
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles