Stat Inference Project Part 1

Joe Logan

Statistical Inference Final Project

Part 1: Simulation Exercise

In this project I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution will 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. I will set lambda = 0.2 for all of the simulations. Finally, I will investigate the distribution of averages of 40 exponentials. Note that I will need to do a thousand simulations.

Initialization of required libraries and set Options

Setup the simulation

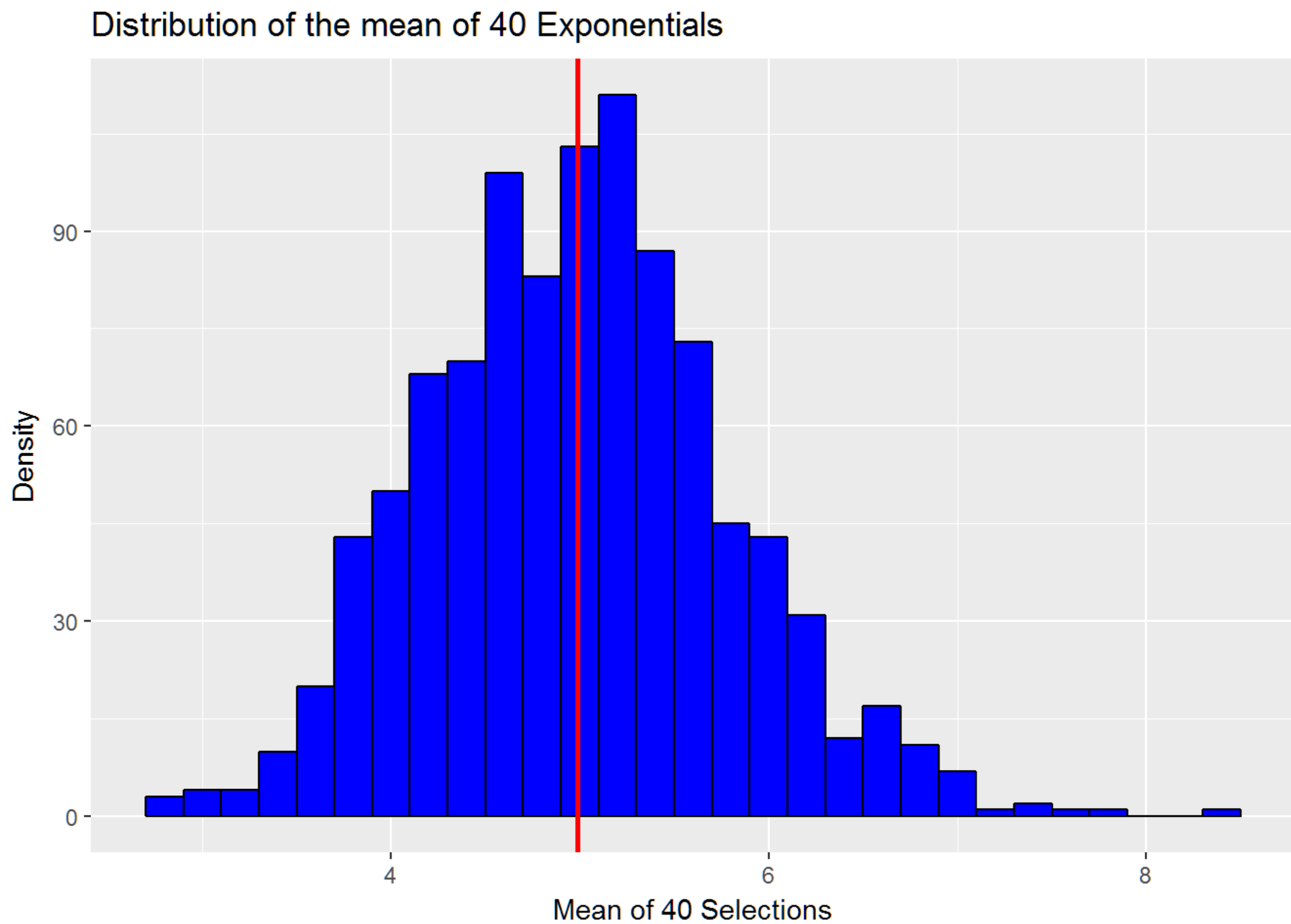
```
# Set the constants given in the problem statement
lambda <- 0.2 # given
nExp <- 40 # number of exponentials (given)
nSimulations <- 1000 # number of exponentials (given)

set.seed(567)

# run the test resulting in n x numberOfSimulations matrix
exponentialDistributions <- matrix(data=rexp(nExp * nSimulations, lambda), nrow=nSimulations)
exponentialDistributionMeans <- data.frame(means=apply(exponentialDistributions, 1, mean))
actualMean <- mean(exponentialDistributionMeans$means)
```

Plot the data

```
# plot the means
ggplot(data = exponentialDistributionMeans, aes(x = means)) +
        geom_histogram(binwidth=lambda, fill="blue",color="black") +
        labs(title="Distribution of the mean of 40 Exponentials", x="Mean of 40 Selections",
y="Density") +
        geom_vline(xintercept=actualMean,size=1.0, color="red") # add a line for the actual mean
```




```
      scale_x_continuous(breaks=round(seq(min(exponentialDistributionMeans$means), max(exponentialDistributionMeans$means), by=1)))
```

```
## <ScaleContinuousPosition>
##  Range:
##  Limits:    0 --    1
```

Comparing the theoretical mean to the sample (actual) mean

By definition, theoretical mean is 1/lambda

```
# Calculating the theoretical mean
mu <- 1/lambda
```

Calculating the actual mean

```
actualMean <- mean(exponentialDistributionMeans$means)
```

The theoretical mean is 5 and the sample mean is 4.9896
We can see that these values are very close.

Comparing the theoretical variance to the sample variance

R is equiped with the var function which can give us our sample variance

```
# Compute the sample variance
sampleVariance <- var(as.vector(exponentialDistributionMeans$means))

# Give the theoretical variance
# We get this by dividing the mean by the square root of our number of exponentials (40 in our c
ase)
# to give us the standard deviation, and then squaring that to get the variance.
sd <- 1/lambda/sqrt(nExp)
theoreticalVariance <- sd^2
```

The theoretical variance is 0.625 and the sample variance is 0.6425399
We can see that these values are also very close.

Now lets show that the distribution is approximately normal.

As we can see from the plot below, it does indeed show normality.

```
qqnorm(exponentialDistributionMeans$means)
qqline(exponentialDistributionMeans$means)
```