Statistical Inference Project Part 1

Jamison Stevens

6/10/2020

Overview

This project involves looking at the exponential distribution in R and comparing it with the Central Limit Theorem. One thousand simulations will be done on the averages of forty exponential distributions. The sample mean and variance will be compared to their theoretical values, and the distribution will be shown to be approximately normal.

Simulations

In the simulations code, the constants for the simulations are defined. Next, the program enters a for loop that executes the code it contains one thousand times. The code inside the for loop uses rexp to perform one simulation of forty exponential distributions and the mean of the simulation result is put into a numeric vector. After completing the for loop, the simulation number and result vector are used to create a data frame, named simDF. The data inside of simDF is then summarized to get an idea of what is contained.

```r
# Define Constants
lambda <- 0.2
numExp <- 40
simulations <- 1000

# Simulations
meanExpSims <- c()
for (i in 1:simulations) {
  sam <- rexp(numExp, lambda)
  meanExpSims <- c(meanExpSims, mean(sam))
}
simDF <- data.frame(1:simulations, meanExpSims)
colnames(simDF) <- c("Simulation_Number", "Exponential_Distribution_Mean")

# Summary of Exponential_Distribution_Mean data in simDF
summary(simDF$Exponential_Distribution_Mean)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.513   4.518   5.009   5.042   5.509   8.043
```

Sample Mean versus Theoretical Mean

The first step in this code is calculating the theoretical mean of the exponential distribution, which is the reciprocal of lambda. Next, the sample mean is calculated by taking the mean of the resulting value from each simulation. After the theoretical and sample means are calculated, their values are printed, then the difference between the two values is displayed. The purpose of this comparison is to explore whether the actual and theoretical centers of the data are approximately the same.

```r
# Theoretical Mean
theoreticalMean <- 1 / lambda

# Compare Theoretical and Sample Means
sampleMean <- mean(simDF$Exponential_Distribution_Mean)

# Print Sample Mean
paste("Sample Mean:", sampleMean)
```

```
## [1] "Sample Mean: 5.04191280667716"
```

```r
# Print Theoretical Mean
paste("Theoretical Mean:", theoreticalMean)
```

```
## [1] "Theoretical Mean: 5"
```

```r
# Difference Between Sample and Theoretical Mean
paste("Difference in Means:", sampleMean - theoreticalMean)
```

```
## [1] "Difference in Means: 0.0419128066771579"
```

The difference between the sample and theoretical means is close to zero, so the simulations closely approximate the expected value of forty exponential distributions.

Sample Variance versus Theoretical Variance

The first step in this code is calculating the theoretical sample standard deviation and variance of the exponential distribution, which is the reciprocal of lambda and reciprocal of lambda squared, respectively. Next, the theoretical variance is calculated by dividing the theoretical sample variance by the number of observations in one simulation. The sample variance is calculated by finding the variance of the average values for the simulations. After the theoretical and sample variances are calculated, their values are printed, then the difference between the two values is displayed. The purpose of this comparison is to explore whether the actual and theoretical variabilities of the data are approximately the same.

```r
# Theoretical Sample Standard Deviation and Variance
theoreticalSamStDev <- 1 / lambda
theoreticalSamVariance <- theoreticalSamStDev^2

# Theoretical Population Standard Deviation and Variance
theoreticalVariance <- theoreticalSamVariance / numExp

# Compare Theoretical and Sample Variance
sampleVariance <- var(simDF$Exponential_Distribution_Mean)

# Print Sample Variance
paste("Sample Variance:", sampleVariance)
```

```
## [1] "Sample Variance: 0.607216004361014"
```

```r
# Print Theoretical Variance
paste("Theoretical Variance:", theoreticalVariance)
```

```
## [1] "Theoretical Variance: 0.625"
```

```r
# Difference Between Sample and Theoretical Variance
paste("Difference in Variance:", sampleVariance - theoreticalVariance)
```

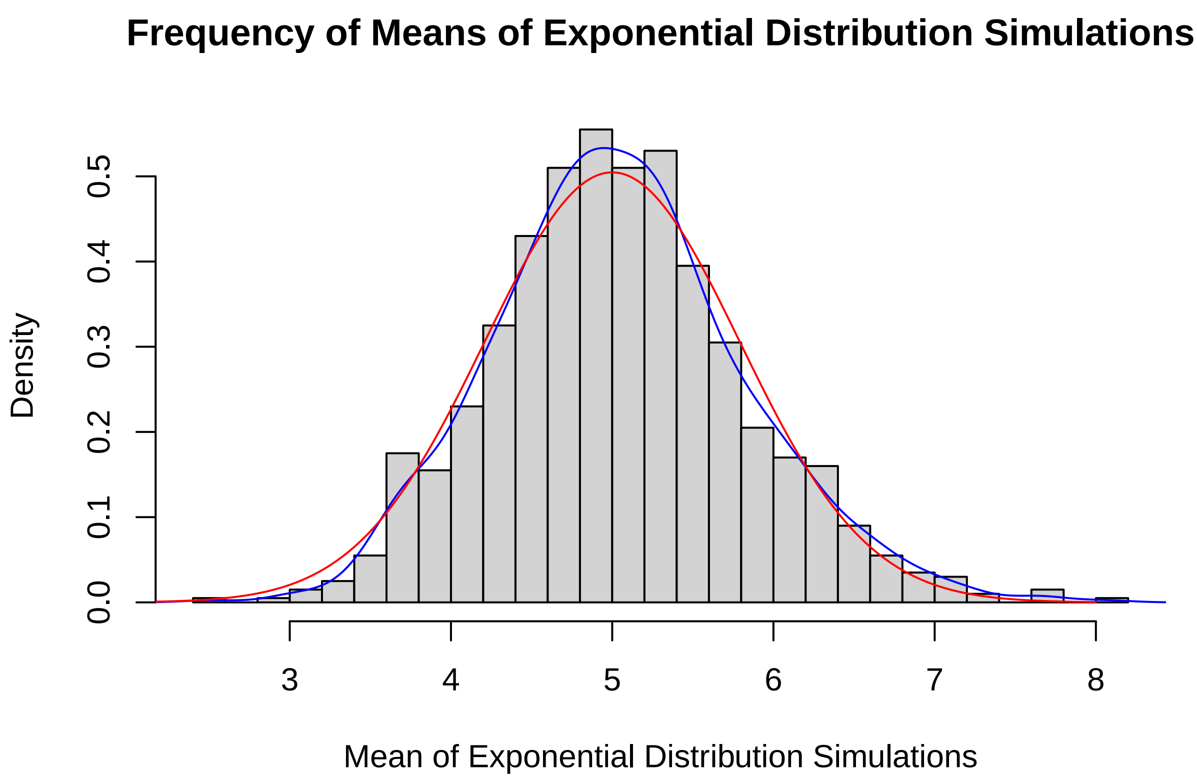
```
## [1] "Difference in Variance: -0.0177839956389857"
```

The difference between the sample variance and theoretical variance is close to zero, so the simulations closely approximate the variance of the expected value of forty exponential distributions.

Distribution

This code creates a histogram of average value for each simulation that shows the frequency in terms of density. Using lines, two curves were added to the graph: a blue density curve for the simulations and a red density curve for the expected distribution.

```r
# Show Distribution is Approximately Normal
plotDensity <- hist(simDF$Exponential_Distribution_Mean, freq = FALSE, breaks = 30,
    main = "Frequency of Means of Exponential Distribution Simulations",
    xlab = "Mean of Exponential Distribution Simulations")
lines(density(simDF$Exponential_Distribution_Mean), col = "blue")
x <- seq(2, 8, length.out= 1000)
y <- with(simDF, dnorm(x, theoreticalMean, sqrt(theoreticalVariance)))
lines(x, y, col = "red")
```



The blue and red curves are nearly identical, meaning the sample simulation approximates a normal distribution.