Comparing the Exponential Distribution with the Central Limit Theorem

Andrew William Judd

March 20, 2016

Overview

This project is used in order to investigate and experiment with the exponential distribution in `R` while comparing it to the Central Limit Theorem.

We are simulating the exponential distribution by using the `R`'s built in `rexp(n, lambda)` function. In this example, `lambda = rate`.

We will be running this test with a total of 1000 simulations.

Simulations

Start by defining some constants that we will be using throughout the analysis.

```
# Set the simulation seed value so that the data is reproducible
set.seed(200)

# Setup the variables that we need
simulation.random_uniform <- 40
simulation.rate <- 0.2
simulation.count <- 1000
```

Now that the environment is set up, let's run the simulation.

```
# Generate the sample data
simulation.data <- matrix(
    rexp(
        simulation.random_uniform * simulation.count,
        simulation.rate
    ),
    simulation.count
)
```

Questions

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

With the simulation data present, we need to compare the actual versus the theoretical.

```r
# Mean = 1 / lambda
theoretical.mean <- 1 / simulation.rate

# Calculate the mean of each row
row.mean <- apply(
    simulation.data,
    1,
    mean
)

# Calculate the actual mean
actual.mean <- mean(row.mean)
```

- Theoretical Mean: `5`
- Actual Mean: `4.9841258`

As you can see above, both the Theoretical and Actual means are very close. This proves that the Central Limit Theorem is accurate when we are trying to calculate means.

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

Now we need to compare the variance between the theoretical and actual.

```r
# Variance = ((1 / lambda) ^ 2) / n
theoretical.variance <- ( (1 / simulation.rate) ^ 2) / simulation.random_uniform

# Calculate the actual variance
actual.variance <- var(row.mean)
```

- Theoretical Variance: `0.625`
- Actual Variance: `0.6288244`

As you can see above, both the Theoretical and Actual variances are very close. This proves that the Central Limit Theorem is accurate when we are trying to calculate variances

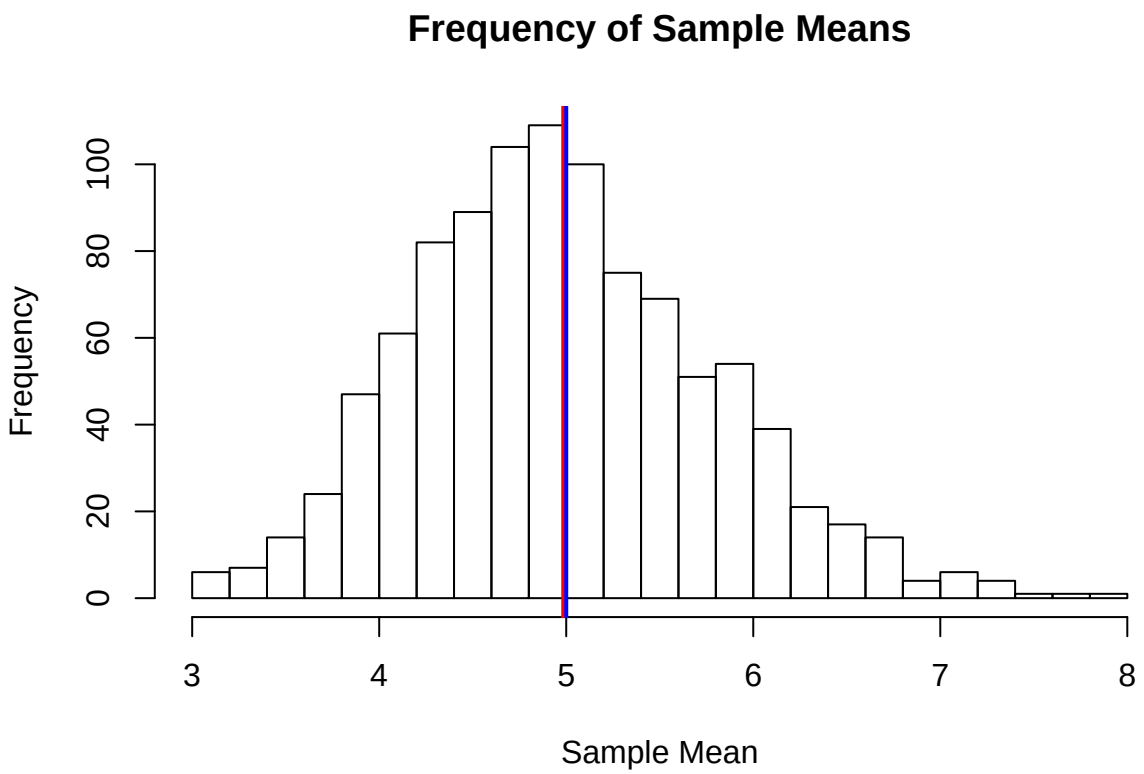
3. Show that the distribution is approximately normal.

Draw a histogram that contains the frequency of the row means withing the derived data.

```r
# Draw a histogram with the data points that we collected
hist(
    row.mean,
    xlab = 'Sample Mean',
    main='Frequency of Sample Means',
    breaks = 20
)

# Add in a vertical line at the actual mean
abline(
    v = actual.mean,
    col = 'red',
    lwd = 2
)
```

```r
# Add in a vertical line at the theoretical mean
abline(
    v = theoretical.mean,
    col = 'blue',
    lwd = 2
)
```



By looking at this graph, we are able to see that the curve is almost normal, and based on our previous two questions, we know that the means and variances are similar. In order to provide a good understanding of the graph, both the theoretical (blue) and actual (red) means have been added to the plot.

Because of that, we are able to conclude that we are able to use the exponential distribution to simulate a normal distribution.