Analysis of Exponential Distributions

By Zach Christensen

Overview

This report is for the Statistic Inference Course offered as part of the John Hopkins' Data Science specialization on Coursera. This analysis will examine the distribution of random exponentials and their averages compared to the theorectival value predicted by the Central Limit Theorm (CLT). For a full explanantion of the various properties of populations and means used here, please refer to the course's GitHub repository.

Simulations

For this analysis, we will set $\lambda = 0.2$ and run 1,000 simulations, with each simulation containing 40 samples.

```
set.seed(10)
lambda <- 0.2
number_of_simulations <- 1000
number_of_samples <- 40
```

To do this, we will get samples from the `rexp()` function. We will get 40,000 samples (number of simulations * number of samples), and places this data into a 1,000 x 40 matrix. The means of the 40 samples for each simulation is then calculated using the `rowMeans()` function.

```
samples <- matrix(data = rexp((number_of_simulations * number_of_samples), lambda), nrow = number_of_si
row_means <- rowMeans(samples)
```

`samples` now contains 40,000 random samples from the exponential distribution. `row_means` contains 1,000 averages of 40 random samplings of the exponential distribution.

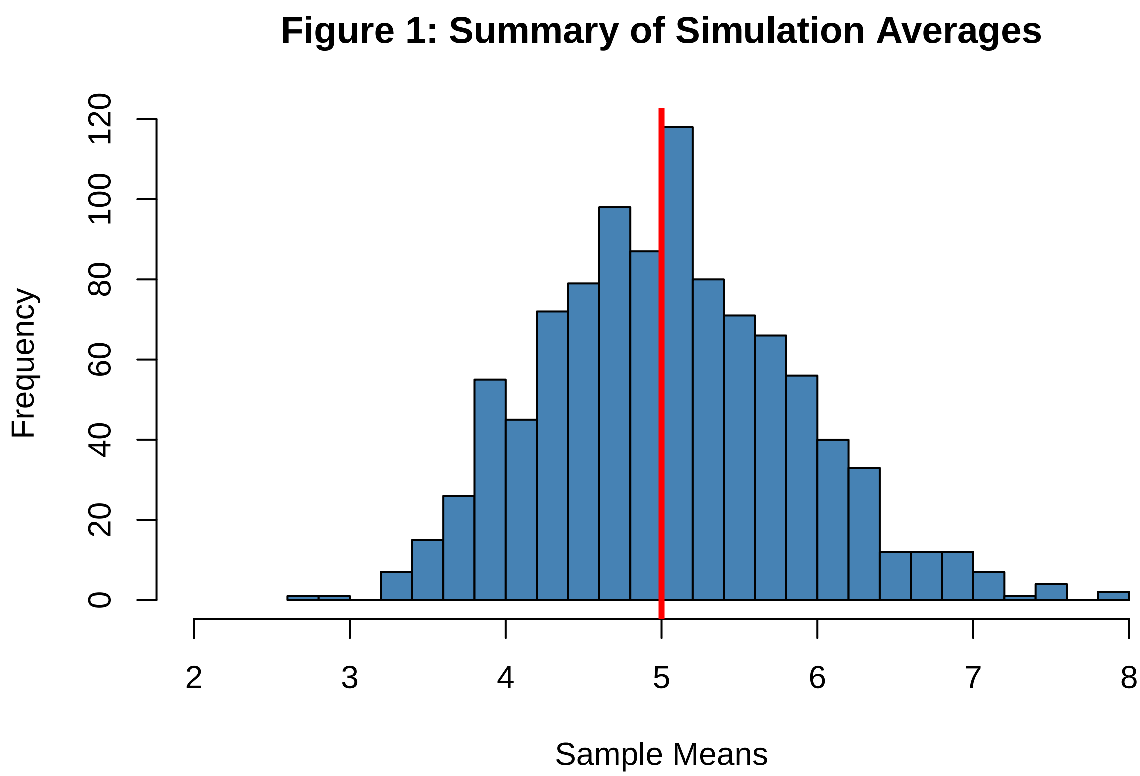
```
summary(row_means)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.772   4.484   5.017   5.045   5.579   7.880
```

Sample Mean versus Theoretical Mean

The Theoretical Mean of exponential distributions is equal to 1 / λ. For this analysis, $\lambda = 0.2$, so the Theoretical Mean is 5.00. As shown in the summary above, the averages contained in `row_means` are centered around 5, with a median of 5.0174812 and a mean of `median(row_means)`. However, the ranges of averages is fairly significant, with values falling between 2.7720919 and 7.8802964. A historgram of the averages is presented in Figure 1, with the Theorectical Mean shown by the red line, x = 5.

```
hist(row_means, main = "Figure 1: Summary of Simulation Averages", breaks = 25, xlim = range(2, 8), col
abline(v = c(5), lwd = 3, col = "red")
```



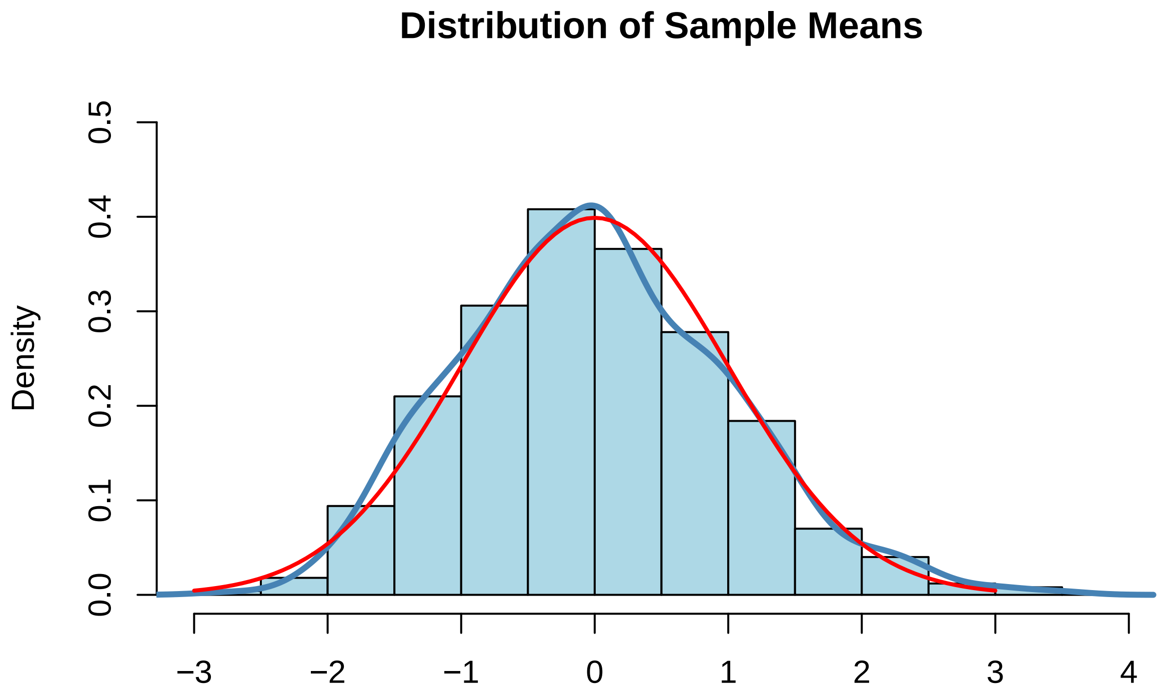
Sample Variance versus Theoretical Variance

The Standard Deviation, σ, of an exponential distribution is equal to 1 / λ. For this analysis, $\lambda = 0.2$, so the Standard Deviation, σ is 5.00 (the same as the mean). The variance of sample means is σ^2 / n, where n is the number of samples. Thus, the Theoretical variance is $(1 / \lambda)^2$ / n, which equals 0.625. The variances of the samples is `var(row_means)` = 0.6541736.

Distribution

In order to investigate the distribution, we will compare the distribution of samples to the normal distribution. We will plot these two alongside each other, allowing easy comparison. To do this, the `scale()` function is used, and the result plotted on top of the histogram of sample means (blue line). Then, the curve of the normal distribution is plotted as well (red line).

```
sample_scale <- scale(row_means)
hist(sample_scale, probability = T, ylim = c(0, 0.5), col = "lightblue", main = "Distribution of Sample
lines(density(sample_scale), col = "steelblue", lwd = 3)
curve(dnorm(x,0,1), -3, 3, col="red", add=T, lwd = 2)
```


Distribution of Sample Means
Density
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
4