Practical Analysis of the CLM

Kate Prendergast

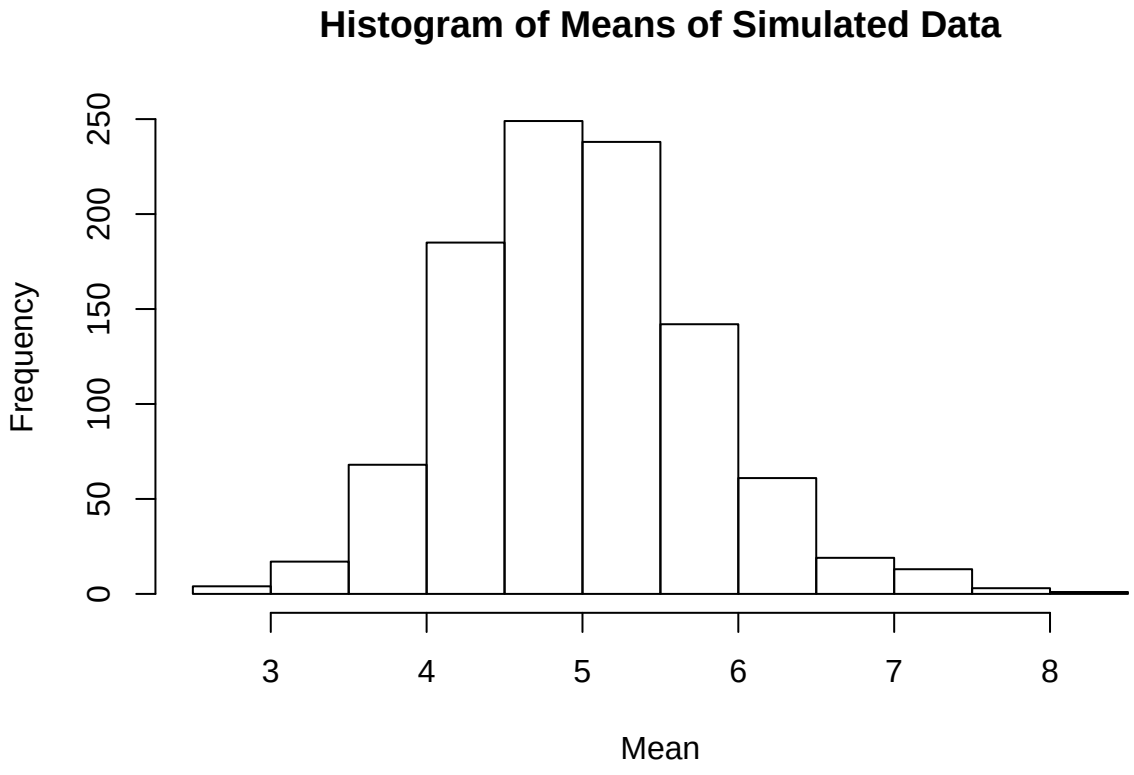
Overview

In this report, we will examine the distribution of means of 40 random exponentials, using the exponential distribution. This will be based on running 1000 simulations, where each one takes the mean of 40 random exponentials. We will analyze the mean and variance of our simulated data and look at the distribution.

Simulations

We first create our 1000 simulations, taking the mean of 40 exponentials in each. For this analysis, we will use a lambda of 0.2.

```
## set the seed for the sake of reproducibility of this report
set.seed(22)
mns = replicate(1000, mean(rexp(40, 0.2)))
hist(mns, xlab = "Mean", main = "Histogram of Means of Simulated Data")
```



We can see that the sample means are centered around 5.

Sample Mean vs. Theoretical Mean

```
simMean <- mean(mns)
```

We now calcualte the exact mean of our data, which gives us **4.989191**. The theoretical mean of the exponential distribution (1/lambda) gives an expected mean of **5**. Hence the mean of our simulated data is very close to the expected value.

Sample Variance vs. Theoretical Variance

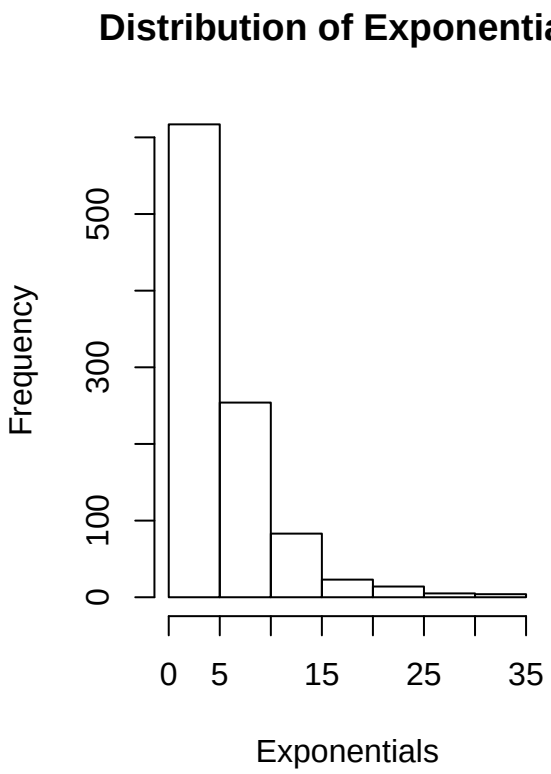
```
simVar <- var(mns)
```

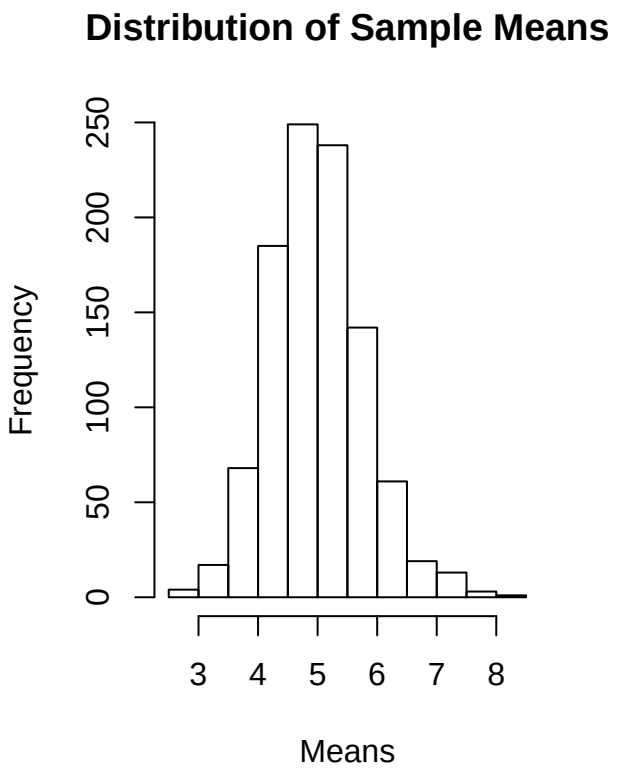
Next, we calculate the variance of our simulated data, which is **0.6337023**. The theoretical variance of the exponential distribution (with n = 40 draws from the population) is (1/n) * (1/lambda^2), which gives us **0.625**. Once again, the variance of our simulated data is very close to the expected value.

Distribution

Finally, we see if the distribution of our sample means is approximately normal. We show a comparison of a plot of 1000 random exponentials (using our same lambda of 0.2) and the same histogram of our means of our simulations of 40 samples.

```
par(mfrow=c(1,2))
hist(rexp(1000, 0.2), xlab = "Exponentials", main = "Distribution of Exponentials")
hist(mns, xlab = "Means", main = "Distribution of Sample Means")
```





As expected, the first plot shows the frequency of values decreasing exponentially, with the majority of values near 5. Our second plot also shows the means centered around 5, however in this plot we see that our distribution of means takes on a Gaussian shape and is normally distributed.

Summary

In summary, our analysis shows that the simulated data does behave as predicted by the CLT. Our mean and variance approach the theoretical values, and the distribution of our sample means is approximately normal.