Exponential Distribution in R and the Central Limit Theorem

Elliot Smith

Overview

In this project, I will be investigating the exponential distribution in R and will compare it to the Central Limit Theorem. To do this, I will use the accompanying R functionality to explore this relationship focusing on the distribution of 40 exponentials. I will be running a thousand simulations to clearly enforce my findings.

Simulations

First, I will begin by setting my assigned constant values and loading the necessary packages:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.2
```

```
lambda <- 0.2
n <- 40
numOfSims <- 1000
```

Now I will generate the data we will use to explore the concepts, below you will find the code used to fabricate the data:

```
data <- matrix(rexp(n * numOfSims, lambda), nrow=numOfSims, ncol=n)
data2 <- data.frame(means=apply(data, 1 , mean), sds=apply(data, 1, sd))
```

Sample Mean v. Theoretical Mean

The Theoretical Mean, mu, of an exponential distribution is given by the formula: mu = 1/lambda. Calculating in R will give us the value of mu:

```
mu <- 1/lambda
mu
```

```
## [1] 5
```

We will calculate the Sample Mean, x-bar, by using the data we produced. To arrive at this value, we will take the mean of the 1000 simulations of the means of 40 randomly sampled exponential distributions:

```
samMean <- mean(data2$means)
samMean
```

```
## [1] 5.006051
```

The Sample Mean and Theoretical Mean are quite close! This shows the our sample can be a very good estimator of our population.

Sample Variance v. Theoretical Variance

The Theoretical Standard Deviation, sigma, with rate lambda for a given exponential distribution is given to us by the formula: sigma = (1/lambda)/sqrt (n). Calculating in R will give us the value of sigma:

```
sigma <- (1/lambda)/sqrt(n)
sigma
```

```
## [1] 0.7905694
```

The Theoretical Variance, sigma-squared, is given to us by simply taking the standard deviation (sigma) and squaring it:

```
sigSquared <- sigma^2
sigSquared
```

```
## [1] 0.625
```

Now we will calculate the Sample Variance and the Sample Standard Deviation for the purpose of comparison. First the Sample Variance:

```
samVar <- var(data2$means)
samVar
```

```
## [1] 0.6042312
```

Now the Sample Standard Deviation:

```
samSD <- sd(data2$mean)
samSD
```

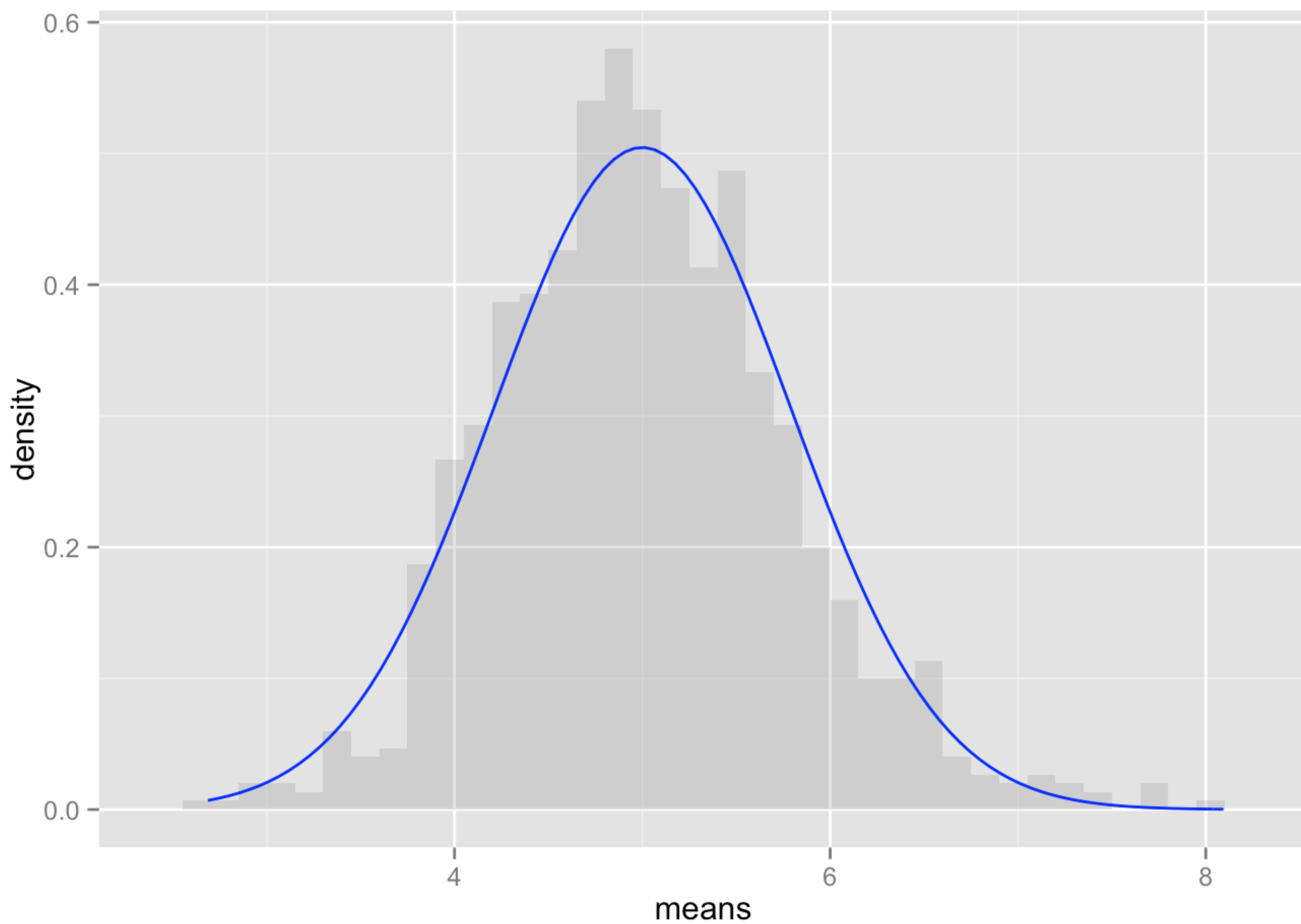
```
## [1] 0.7773231
```

As we can see, both the the Standard Deviations and Variances are quite similar.

Distribution

To prove the normalcy of the distribution, please reference the below graph:

```
ggplot(data=data2, aes(x=means)) +
geom_histogram(binwidth=0.15, alpha=0.15, aes(y=..density..)) +
stat_function(fun=dnorm, arg=list(mean=mu, sd=sigma), colour="blue")
```



As you can visualize (and conclude) from this graph, not only are the Theoretical and Sample Mean and Standard Deviation similar in value, but the sample of means has a very similar distribution to a generated normal distirbution.