Statistical Inference Assignment: Part 1

Michael Crocco

Overview

This report will compare the exponential distribution to the Central Limit Theorem. Using a randomly generated exponential distribution, sample mean and standard deviation will be contrasted with those expected from the CLS.

Generate Distribution

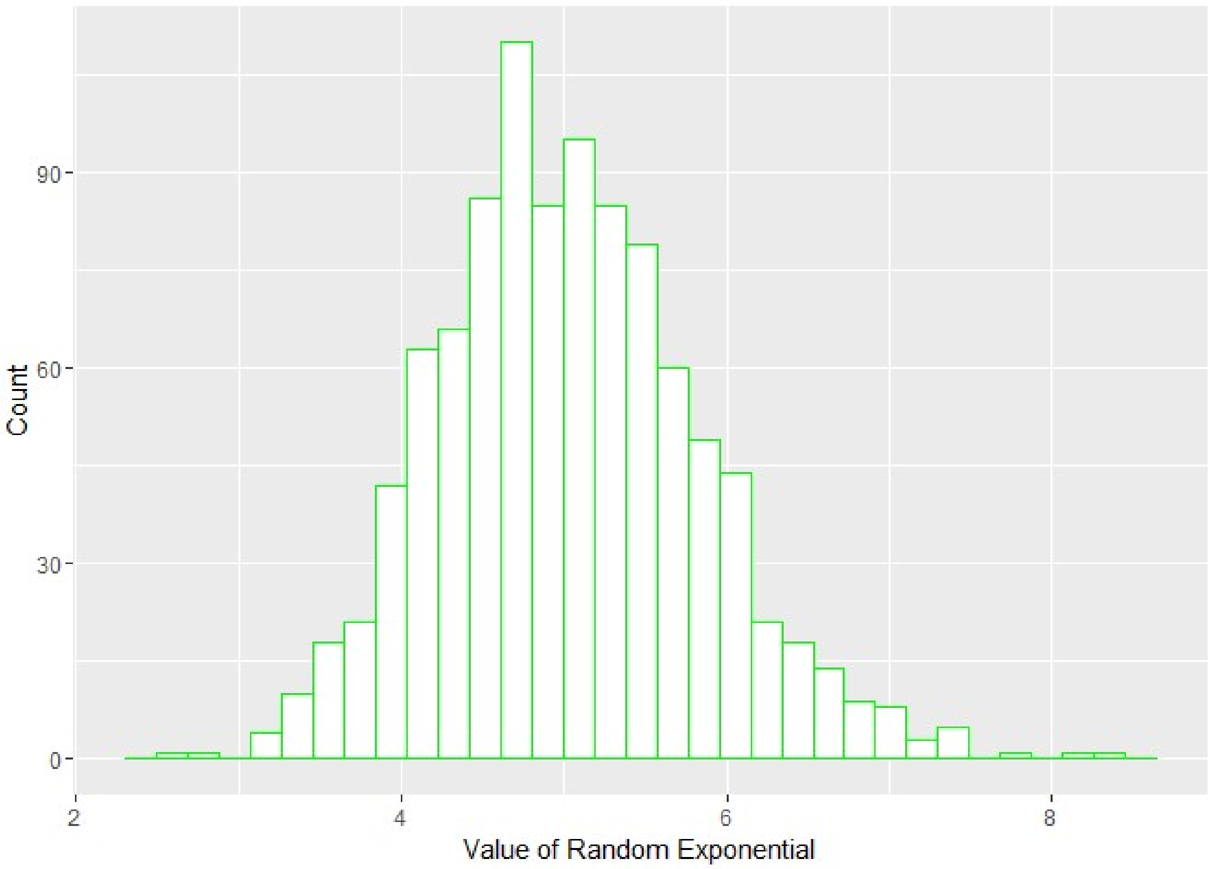
First, we need to set variables and generate some numbers.

```
lambda <- 0.2 # rate parameter in exponential distribution
n = 40 #number of exponentials
means = NULL
for (i in 1 : 1000) means = c(means, mean(rexp(n,lambda))) # take the mean o
f 1000 simulations
```

Display the Distribution

```
library(ggplot2)
g <- ggplot(data.frame(x = 1:n, y = means), aes(y))
g <- g + geom_histogram(col="green", fill="white")
g <- g + labs(x = "Value of Random Exponential", y = "Count")
g
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

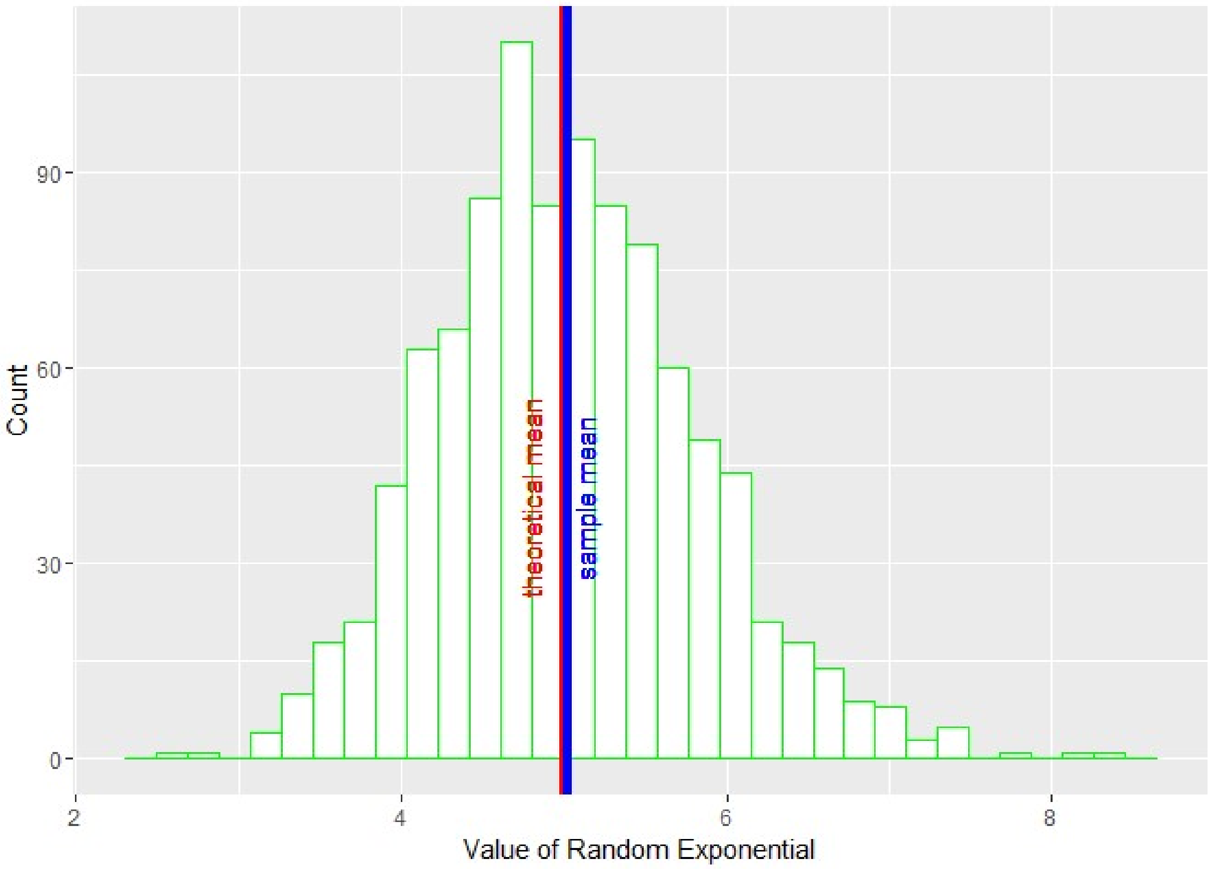


Next, overlay the theoretical and sample means

```
SampleMean <- mean(means)
CLTMean <- 1/lambda

g <- ggplot(data.frame(x = 1:n, y = means), aes(y))
g <- g + geom_histogram(col="green", fill="white")
g <- g + geom_vline(xintercept = CLTMean, col="red", size = 2)
g <- g + geom_vline(xintercept = SampleMean, col="blue", size = 2)
g <- g + geom_text(aes(x=SampleMean, label="sample mean", y=40), colour="blu
e", angle=90, vjust = 1.2)
g <- g +  geom_text(aes(x=(CLTMean), label="theoretical mean", y=40), colour
="red", angle=90, vjust=-1)
g <- g + labs(x = "Value of Random Exponential", y = "Count")
g
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



At 5.0241575, the Sample Mean is very similar to the 5 theoretical mean.

Now, we’ll compare the actual versus the theoretical variance values.

```
SampleVar <- var(means)
CLTVar <- (1/lambda)^2
```

The theoretical variance is:

25

(square of the theoretical standard deviation), but the variance of the distribution of means is:

0.6502169.

These do not compare well, as the distribution of the exponential means is not exponential, but, rather, normal (we will show this later). Instead, doing a single random exponential simulation, and comparing its variance to the theoretical variance yields a much closer result.

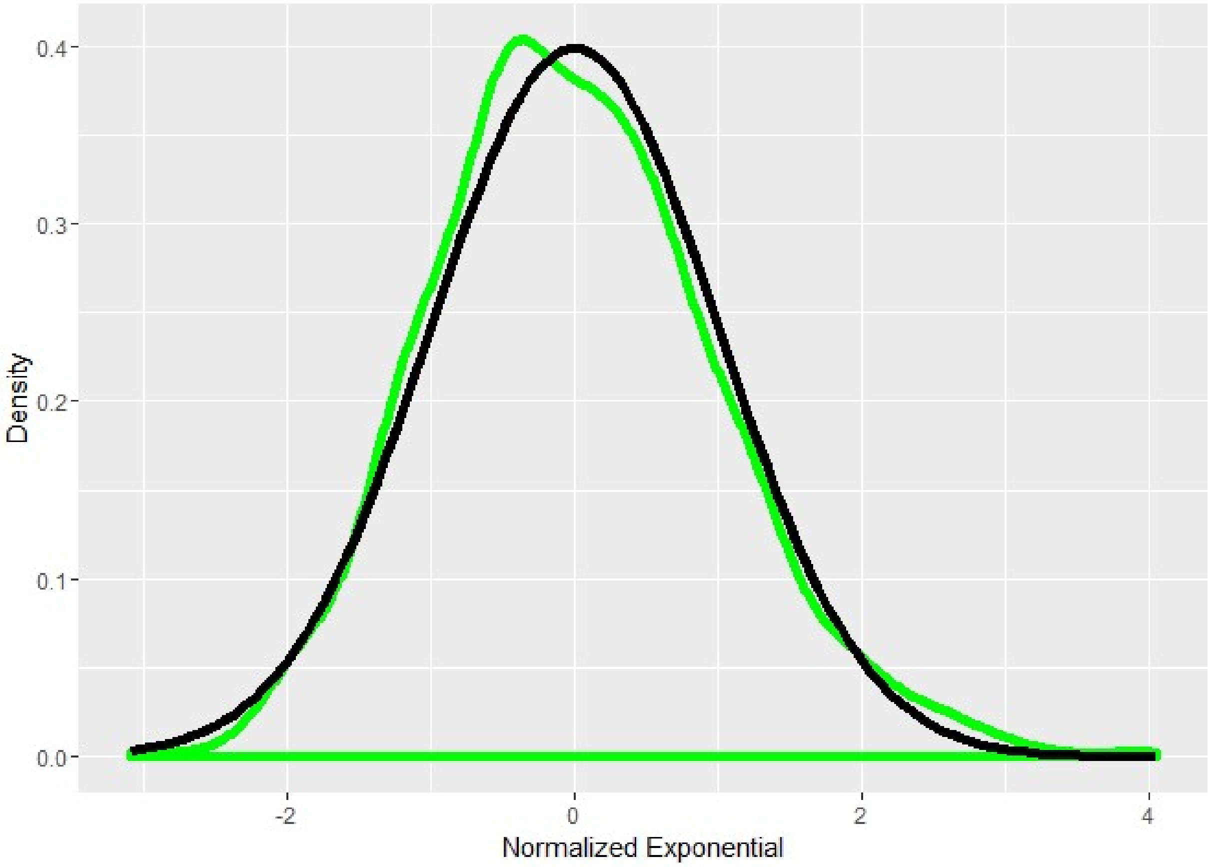
```
SingleSample = NULL
for (i in 1:1000) SingleSample = c(SingleSample, rexp(n, lambda))
SingleSampleVar <- var(SingleSample)
```

The single exponential simulation variance is:

25.1606932.

As suggested, if we overlay a normal distribution curve on our previous histogram, we will see that the distribution of means of 40 simulations of random exponentially distributed values is approximately normal. Let's plot the Normal PMD function on top of our distribution of sample means. First, normalize the means of the exponentials so that it can be compared.

```
normExp <- (means - SampleMean)/sqrt(SampleVar)
g <- ggplot(data.frame(x = 1:n, y = normExp), aes(y))
g <- g + geom_density(size = 2, color = "green")
g <- g + labs(x = "Normalized Exponential", y = "Density")
g <- g + stat_function(fun = dnorm, size = 2)
g
```



As expected, this distribution is very close to the normal distribution.