Statistical Inference Project, Part 1: Central Limit Theorem for the Exponential Distribution

Fenton Taylor

August 22, 2016

Overview

This simulation will investigate the Central Limit Theorem as it pertains to the exponential distribution. It will show that the sample mean can estimate the population mean and the sample variance can estimate the population variance. This simulation will also show that the the distribution of sample means is approximately normal.

Simulations

```
# Load libraries
library(ggplot2)
library(knitr)
```

Sample Mean vs. Theoretical Mean

With a sample size size of n, the theoretical mean of an exponential distribution is $1/\lambda$, where λ is the exponential rate. The following simulation generates 1000 trials of 40 random samples from an exponential distribution with $\lambda = 0.2$. Then, the mean of each trial is calculated. Finally, the mean of the trial means is calculated (actual_mean). That value should approximate the theoretical mean (theor_mean).

```
set.seed(17)

nosim = 1000
lambda = 0.2
n = 40

sim_data <- as.data.frame(matrix(rexp(n*nosim, lambda), nosim, n))
sim_data$sampleMean <- rowMeans(sim_data)

actual_mean = mean(sim_data$sampleMean)
theor_mean = 1/lambda
means <- data.frame(MeanType=c("Actual Mean", "Theoretical Mean"),
                    Value=c(actual_mean,theor_mean))
kable(means, digits = 3, align = "c")
```

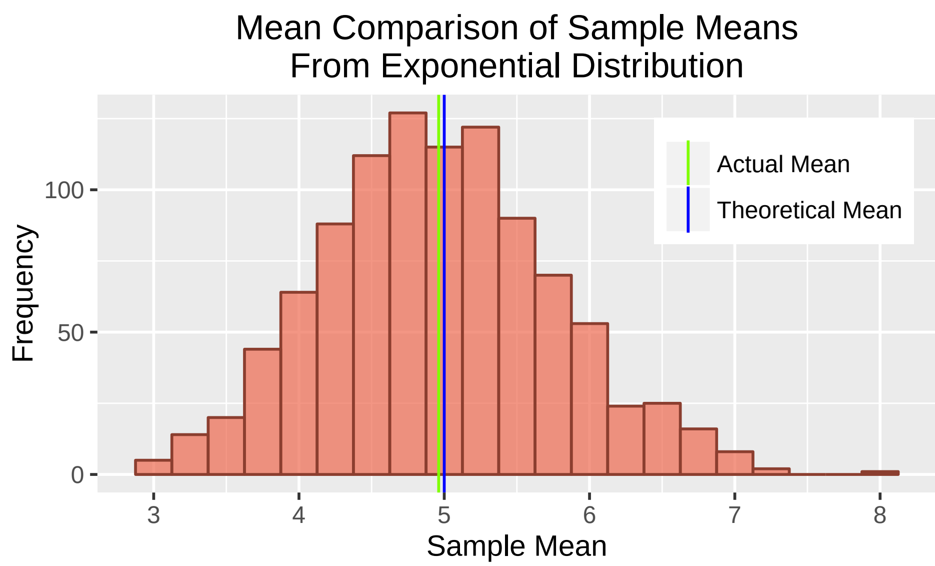
MeanType	Value
Actual Mean	4.962
Theoretical Mean	5.000

Results

The theoretical mean ($1/\lambda$) of this distribution is 5.
The actual mean of the 1000 sample means of 40 samples is 4.9616831.

So, in fact, the sample mean very closely approximates the theoretical, or population, mean. Below is a histogram of the 1000 sample means.

```
ggplot(sim_data, aes(x=sampleMean, fill="Samples")) +
      geom_histogram(binwidth=.25, alpha=0.7, fill="coral2", color="coral4") +
      geom_vline(data = means, aes(xintercept=Value, color=MeanType),
                 show.legend=TRUE) +
      scale_color_manual(values = c("Actual Mean" = "chartreuse",
                         "Theoretical Mean"="blue")) +
      theme(legend.title=element_blank(),legend.justification=c(1,1),
            legend.position=c(1,1)) +
      ggtitle("Mean Comparison of Sample Means\nFrom Exponential Distribution") +
      labs(x="Sample Mean", y = "Frequency")
```



As the figure shows, the actual mean of the samples and the theoretical mean are almost exactly the same and centered at the peak of the distribution.

Sample Variance vs. Theoretical Variance

With a sample size of n, the theoretical variance of an exponential distribution is $(1/\lambda)^2/n$.

```
actual_variance <- var(sim_data$sampleMean)
theor_variance <- (1/lambda)^2/n
vars <- data.frame(Value = c(actual_variance, theor_variance),
                   row.names = c("Actual Variance", "Theoretical Variance"))
kable(vars, digits = 3, align = 'c')
```

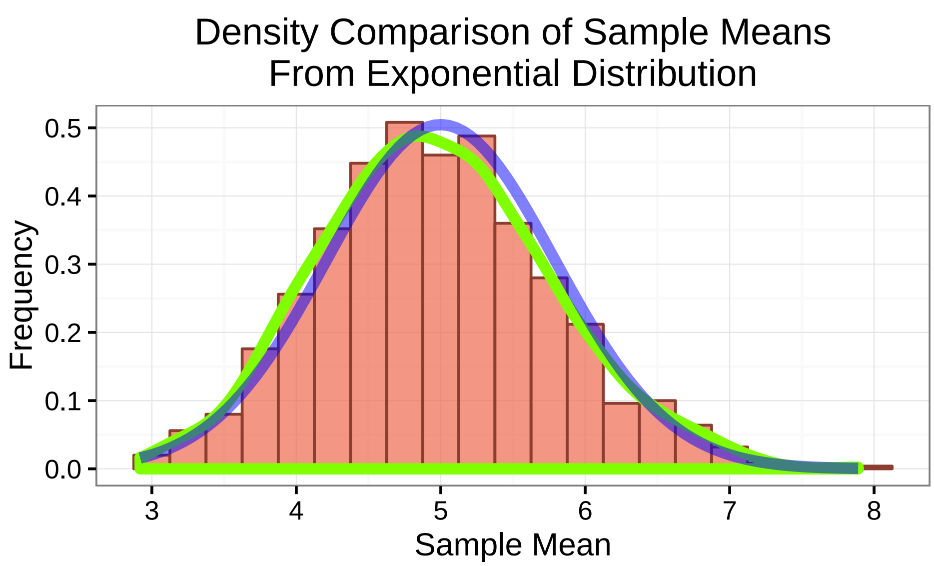
	Value
Actual Variance	0.635
Theoretical Variance	0.625

As the table shows, the actual variance of the samples is nearly identical to the theoretical variance.

Distribution

The Central Limit Theorem posits that the means of a large number of samples should be normally distributed. The following figure shows the same distribution of sample means from above, but with its density curve (green line), and the normal distribution curve (blue line).

```
ggplot(sim_data, aes(x=sampleMean)) +
      theme_bw() +
      geom_histogram(aes(y=..density..), binwidth=.25, alpha=0.7,
                     fill="coral2", color="coral4") +
      geom_density(color="chartreuse", size = 2, alpha = 0.5) +
      stat_function(fun=dnorm, color="blue", size = 2, alpha = 0.5,
                    args = list(mean=theor_mean, sd=sqrt(theor_variance))) +
      ggtitle("Density Comparison of Sample Means\nFrom Exponential Distribution") +
      labs(x="Sample Mean", y = "Frequency")
```



So, the density curve of the sample means (1000 trials of 40 samples) is approximately normal. Larger sample sizes and more trials would shift the curve closer to normal.