Central Limit Theorem and the Exponential Distribution

Connor Gooding

8/21/2017

Overview

This report analyzes the mean of exponential distributions and how the statistic adheres to the Central Limit Theorem for large sample sizes. The code runs a simulation of 1000 exponential distributions ($\lambda = .2, n = 40$), takes the mean of each exponential distribution to form a distribution of the sample means, and explores differences between the theoretical mean and standard error and the simulated mean and standard error of this mean distribution.

Base Simulation

```
mns = NULL
for (i in 1 : 1000) mns <- c(mns, mean(rexp(40, rate = .2)))
```

Figure 1 in the Appendix illustrates the distribution of the means of the simulated exponentials.

Comparison of Means - Theoretical vs. Simulated

The theoretical mean of an exponential distribution $= 1/\lambda$. Our λ is .2, so the theoretical mean is calculated as follows:

```
m_theor <- 1/.2
paste("Theoretical Mean:", m_theor)
```

```
## [1] "Theoretical Mean: 5"
```

The Law of Large Numbers states that as the number of trials gets larger, the mean of the sample mean distribution will approach the population mean.

```
m_sim <- mean(mns)
paste("Mean of Sample Distribution:", round(m_sim, 3))
```

```
## [1] "Mean of Sample Distribution: 5.027"
```

Since this distribution consists of 1000 sample means, the distribution's mean is fairly close to the theoretical mean. Figure 2 in the Appendix illustrates the proximity of the simulation's mean to the theoretical mean.

Comparison of Variances - Theoretical vs. Simulated

The theoretical variance of the mean can be calculated as:

$$SE^2 = \frac{\sigma^2}{n}$$

where σ is the standard deviation of the population, n is the sample size, and SE is the standard error of the mean.

For an exponential distribution with $\lambda = 0.2$, the theoretical variance should be:

```
var_theor <- (1/.2)^2 / 40
paste("Theoretical Variance of the Mean:", round(var_theor, 3))
```

```
## [1] "Theoretical Variance of the Mean: 0.625"
```

The Central Limit Theorem tells us that as more and more samples are taken of the exponential distribution, the distribution of the means will resemble a normal distribution with a mean $\bar{X}$ approaching the theoretical mean μ and a variance s^2 approaching the theoretical standard error of the mean, SE^2.

```
var_sim <- sd(mns)^2
paste("Simulated Variance of the Mean:", round(var_sim, 3))
```

```
## [1] "Simulated Variance of the Mean: 0.633"
```

Once again, the large sample size of this distribution of the exponential means adheres to the CLT and gives us a variance close to the theoretical variance. Figure 3 in the Appendix illustrates the accuracy of the simulated standard deviation in predicting the theoretical standard error of the mean.

Distinction - Normality of the Mean Distribution

It must be emphasized that the Central Limit Theorem does not apply to a single exponential distribution with a large sample size but rather the distribution of means of many exponential distributions. Figure 4 in the Appendix is a side-by-side comparison of an exponential distribution with 1000 points and the distribution of means for 1000 exponential distributions of 40 points each.

The exponential plot on the left is not Gaussian, as it is severely skewed. However, the distribution of sample means from 1000 exponential distributions on the right is demonstrably more Gaussian, as it is symmetrical and forming a bell curve.

Appendix

Figure 1

```
g1 <- ggplot(data = data.frame(mns), aes(x = mns)) +
    geom_histogram(binwidth = .1) +
    ggtitle("Distribution of Means, 1000 Exponentials with n = 40, lambda = .2")
g1
```

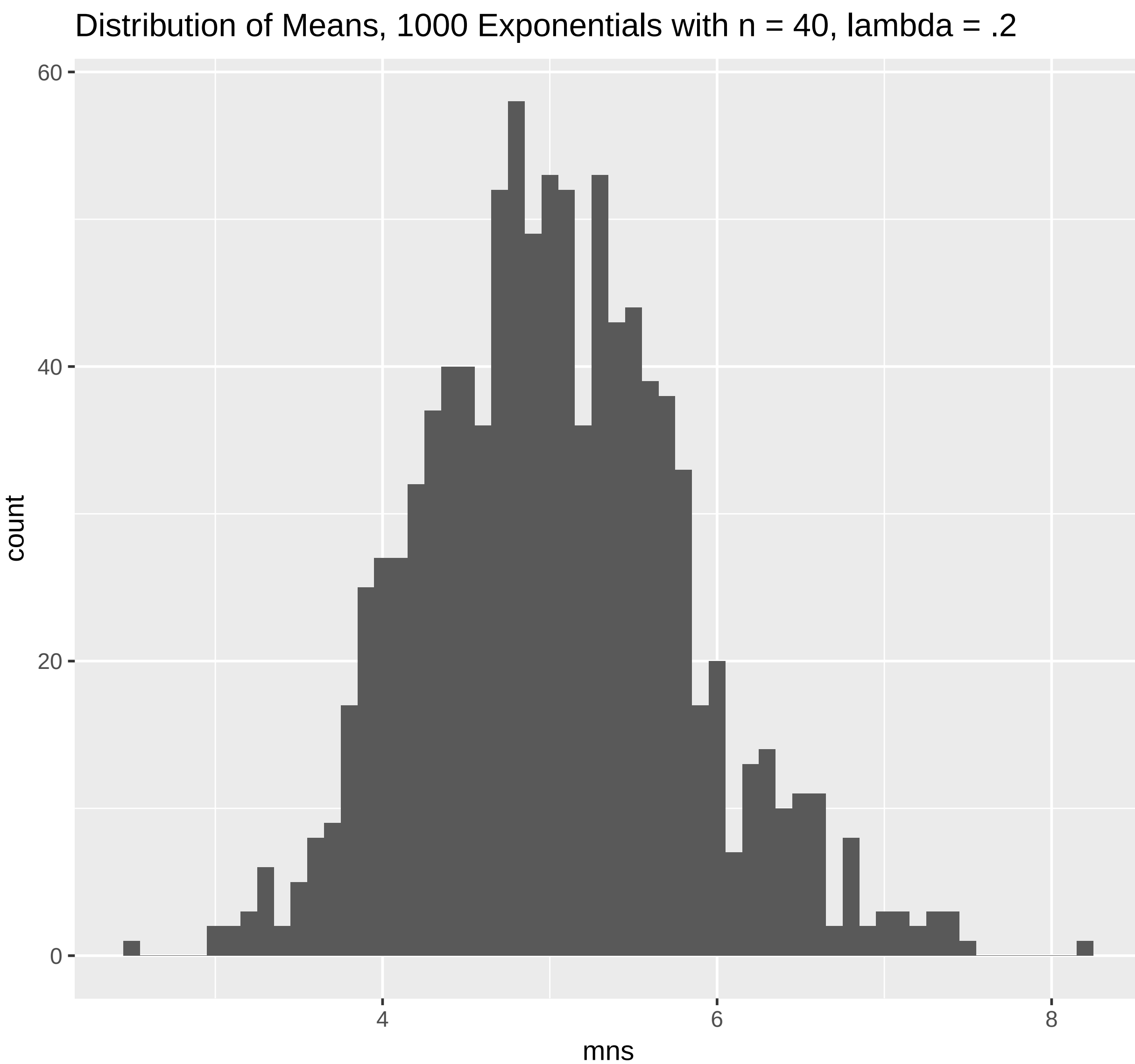

Figure 2

```
g2 <- ggplot(data = data.frame(mns), aes(x = mns)) +
          geom_histogram(binwidth = .1) +
          geom_vline(aes(xintercept=m_sim, color="simulation"), size=1) +
          geom_vline(aes(xintercept=m_theor, color="theoretical"), size = 1) +
          scale_color_manual(name = "Means", values = c( simulation = "blue",
                                                         theoretical = "red")) +
          ggtitle("Exponential Distribution - Sample vs. Theoretical Mean")
g2
```

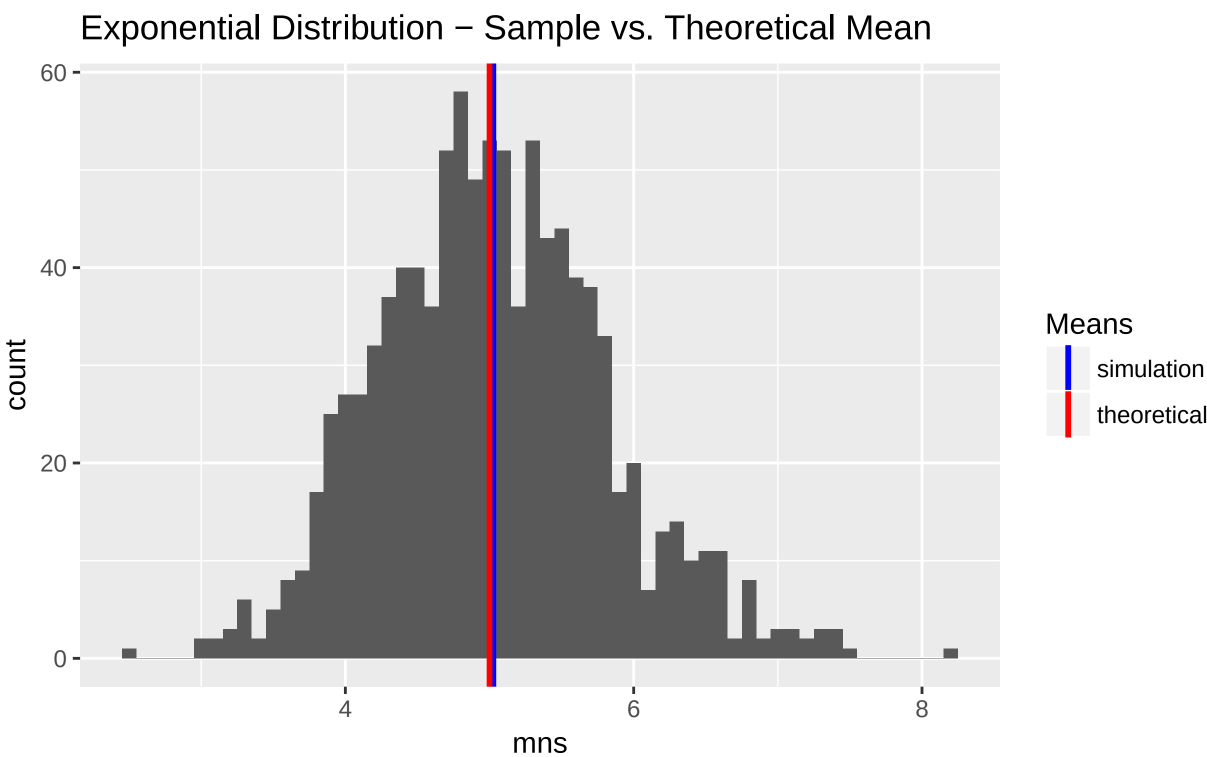

Figure 3

```
g3 <- ggplot(data = data.frame(mns), aes(x = mns)) +
      geom_histogram(binwidth = .1) +
      geom_vline(aes(xintercept=m_sim, color="mean"), size=1) +
      geom_vline(aes(xintercept=m_sim + sqrt(var_theor), color="theoretical"),
                 size = 1) +
      geom_vline(aes(xintercept=m_sim + sqrt(var_sim),
                     color="simulation"), size = 1) +
      scale_color_manual(name = "Means",
            values = c( mean = "black", simulation = "blue", theoretical = "red")) +
      ggtitle("Exponential Distribution - Sample vs. Theoretical Mean")
g3
```

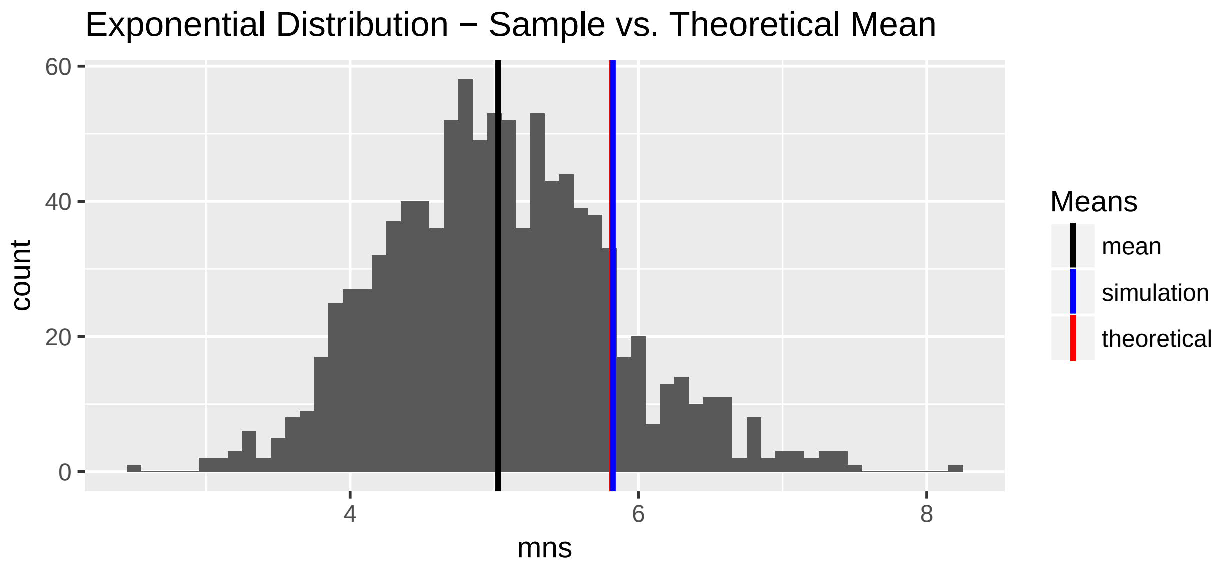

Figure 4

```
gexp <- qplot(rexp(1000), binwidth = .1) + ggtitle("One Exponential Distribution") +
        geom_label(x = "Index")
gnorm <- ggplot(data = data.frame(mns), aes(x = mns)) +
        geom_histogram(binwidth = .1) + ggtitle("Distribution of Means")
grid.arrange(gexp, gnorm, ncol = 2)
```