Exploration of the Exponential Distribution and the Central Limit Theorem on Mean and Variance

Bill Martersteck

August 09, 2016

Overview

In this project, we will explore how the Exponential Distribution behaves relative to the Central Limit Theorem. We will look at the mean and the variance of large numbers of samples and compare against the theoretical versions of these measures.

Note, the code for all the figures will be in the appendix.

Simulations

```
l = .2; numsims = 1000; samples = 40
RandExp <- matrix(rexp(numsims * samples, rate = l), nrow = numsims)
mns <- apply(RandExp, 1, mean)
```

For this excercise, we will only consider the exponential distribution with $\lambda = 0.2$. We select 40 samples 1000 times and calculate the mean of each group of samples.

Sample Mean versus Theoretical Mean

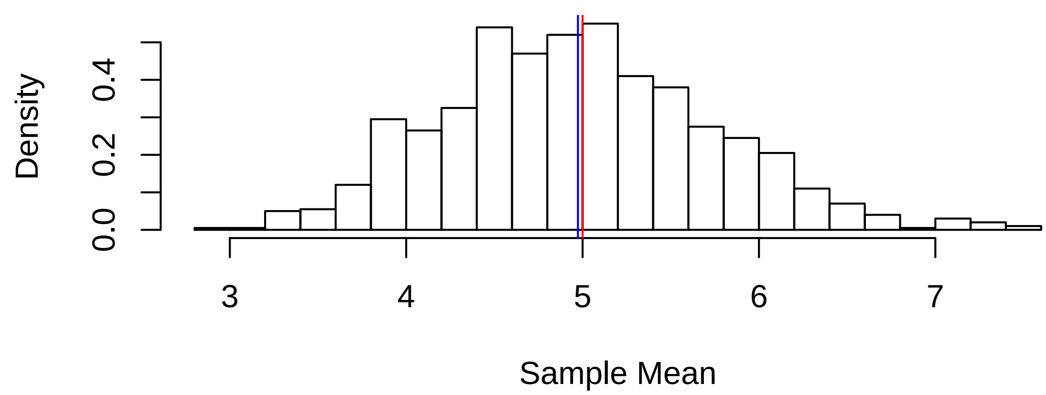

Figure 1: Distribution of Sample Means of Exp Random Variable

The mean of the exponential distribution with $\lambda = 0.2$ is

$$\mu = 1/\lambda = 1/0.2 = 5$$

```
popMean <- 1/l; sampleMean <- mean(mns)
c(popMean, sampleMean)
```

```
## [1] 5.000000 4.973765
```

In figure 1, we show the distribution of the sample means with the sample mean (blue vertical) and population mean (red verticle line) to see the distribution of the sample means. Note, given that the means are so close together, it is difficult to distinguish them on the graph.

The Law of Large Numbers says the mean will limit to what it is approximating. We can see this clearly by plotting the sample means as the number of iterations increases against the population mean:

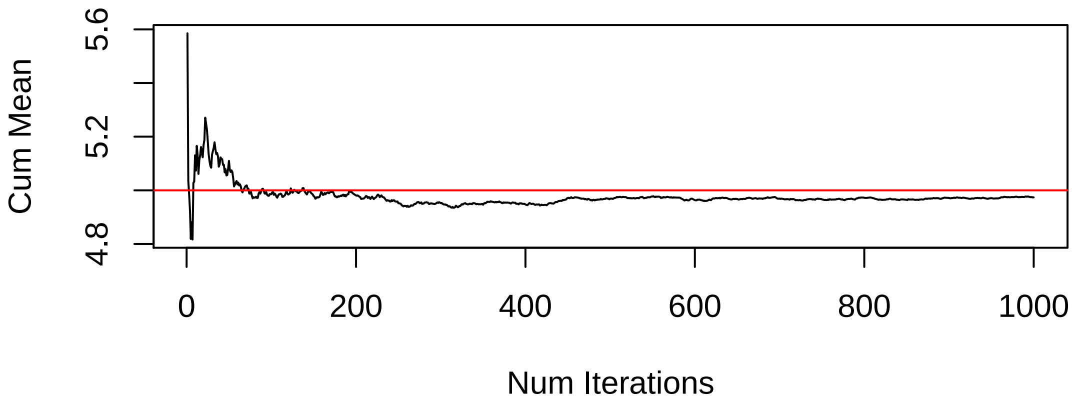

Figure 2: Cumulative Mean as Iterations Increase

The sample mean of 4.9737652 over 1000 iterations and population mean of 5 are very close to each other. We can verify with a t.test that there is no reason that they shouldn't be the same.

$$H_0 : \mu = \bar{x}$$

$$H_a : \mu \neq \bar{x}$$

```
t.test(x = mns, mu = 1/l, alternative = "two.sided")$conf.int
```

```
## [1] 4.926523 5.021007
## attr(,"conf.level")
## [1] 0.95
```

Since the sample mean $\bar{x} = 4.9737652$ is within the 95% confidence interval, we cannot reject H_0.

Sample Variance versus Theoretical Variance

For this project, I assume we are to compare the Variance of the distribution of sample means to what is expected, the sample error.

An exponential random variable has a standard deviation of $\sigma = 1/\lambda$ and a variance of $\sigma^2 = 1/\lambda^2$. The standard error of the variance of means should be $\sigma^2/n = \frac{1/\lambda^2}{n}$.

```
popVar <- (1/(l^2))/samples
sampleVar <- var(mns)
c(popVar, sampleVar)
```

```
## [1] 0.6250000 0.5795727
```

If we plot the the variance as the number of iterations increases, we should see the standard error approximating the Variance of the Sample Means, as it does.

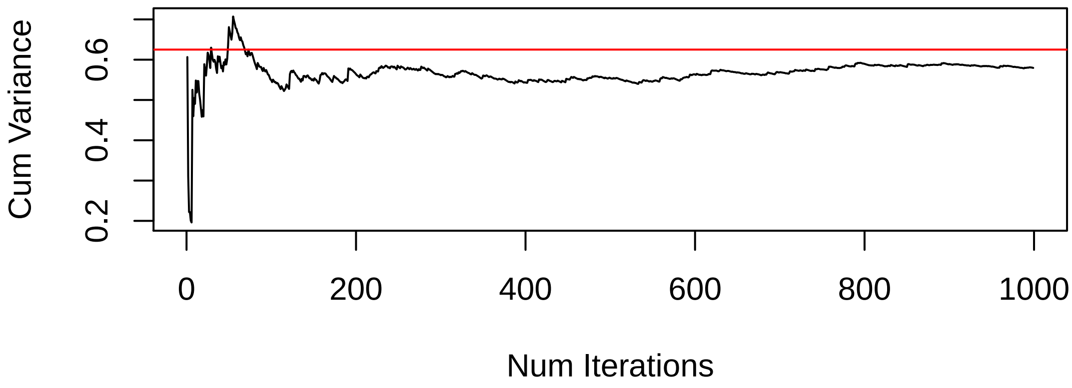

Figure 3: Cumulative Variance as Iterations Increase

Distribution

All our analysis has been done with 1000 iterations of 40 random variables. The figure 2 histogram shows the distribution of sample means looks normal. As we increase the number of samples, the distributions should become more and more normal if the Central Limit Theorem holds with the difference of the means going to zero. Let's see how the distribution density changes with iterations.

Plot the histogram of the density of the means.

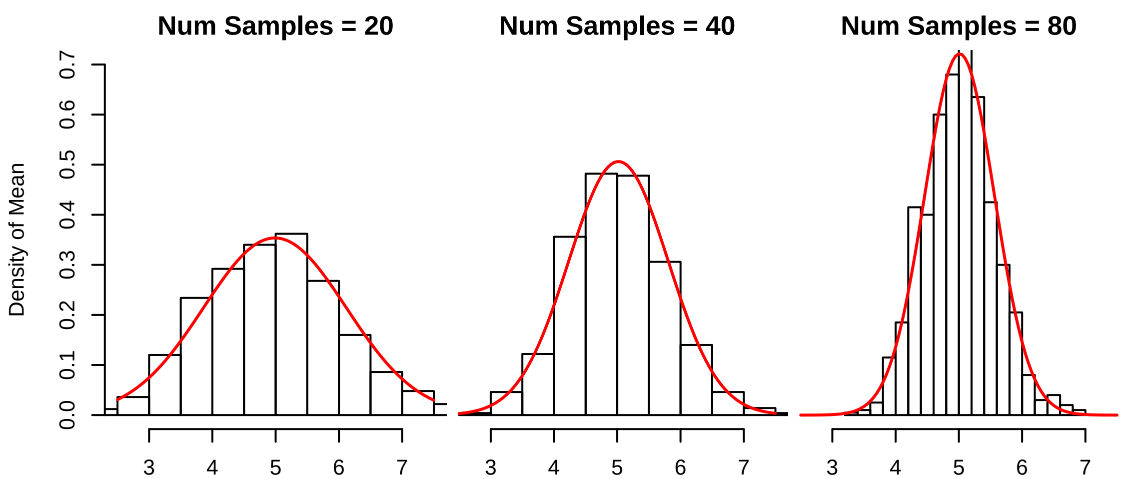

Figure 4: Distribution of Means of N Exp Random Variable

As the number of samples increases, the standard deviation decreases and it appears that the distribution is more normal.

Appendix - Code Producing Plots

Plot Figure 1: Distribution of Sample Means of Exp Random Variable

```r
hist(mns, breaks = 20, freq = FALSE, xlab = "Sample Mean", ylab = "Density", main = "")
abline(v = popMean, col = "red")
abline(v = sampleMean, col = "blue")
```

Plot Figure 2: Cumulative Mean as Iterations Increase

```
x = 1:length(mns)
y = cumsum(mns)/x
plot(x, y, type = "l",xlab = "Num Iterations", ylab = "Cum Mean")
abline(h = popMean, col = "red")
```

Plot Figure 3: Cumulative Variance as Iterations Increase

```
x <- 1:(length(mns)-1)
y <- x
for (i in x) {
        y[i] <- var(mns[1:(i+1)])
}
plot(x, y, type = "l",xlab = "Num Iterations", ylab = "Cum Variance")
abline(h = popVar, col = "red")
```

Plot Figure 4: Distribution of Means of N Exp Random Variable

```
par(mfrow=c(1,length(samples)), mar = c(2,0,2,0), oma = c(0,5,0,0))
for (i in samples) {
        RandExp <- matrix(rexp(numsims * i, rate = 1), nrow = numsims)
        mns <- apply(RandExp, 1, mean)
        x <- seq(2.5, 7.5, length.out = 20)
        hist(mns, breaks = 20, freq = FALSE, xlim = c(2.5, 7.5), ylim = c(0, .7),
             yaxt = 'n', xlab = "Sample Mean", main = paste("Num Samples =", i))
        curve(dnorm(x, mean = mean(mns), sd = sd(mns)), add = TRUE, lwd = 1.5, col = "red")
}
title(ylab = "Density of Mean", outer = TRUE )
axis(2, outer = TRUE )
```