Exponential Distributions and the Central Limit Theorem

Toni Rib

Overview

In this paper, I show that by using a large sample of averages of independent and identically distributed (iid) variables that have been properly normalized, the distribution becomes that of a standard normal. The exponential distribution will be used and 1000 simulations will be run, where in each simulation the average of 40 iid exponentials is taken.

Simulations

To perform the analysis, we start by simulating random exponentials using lambda = 0.2. We calculate 40 random exponentials (n = 40), then take the average of this set. We repeat this 1000 times using the apply function in R to get a reasonably large set of means. To ensure the repeatability of this simulation, we will set the seed prior to performing the calculation.

```
## Set all of the known starting parameters of the problem
n <- 40
lambda <- 0.2
numSim <- 1000
mu <- 1 / lambda
stdDev <- 1 / lambda
var.exp <- stdDev ^ 2
stdErr <- stdDev / sqrt(n)

## Calculate the mean of 40 exponentials, and do this 1000 times
set.seed(8) ## Ensures this is a repeatable calculation
sample <- apply(matrix(rexp(numSim * n, rate = lambda), numSim), 1, mean)
summary(sample)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.762   4.467   4.968   5.006   5.458   7.886
```

We now have a set of 1000 numbers, each number being the average of 40 iid random exponentials using lambda = 0.2.

We can also normalize these samples. Since we have already taken the mean of each set of 40 exponentials, we can use our current data and subtract off the theoretical mean (mu) and then divide by the standard error. This new set of data will reflect a standard normal distribution, centered around 0 (mean = 0) with one standard deviation of 1.

```
sampleCLT <- (sample - mu) / stdErr
summary(sampleCLT)
```

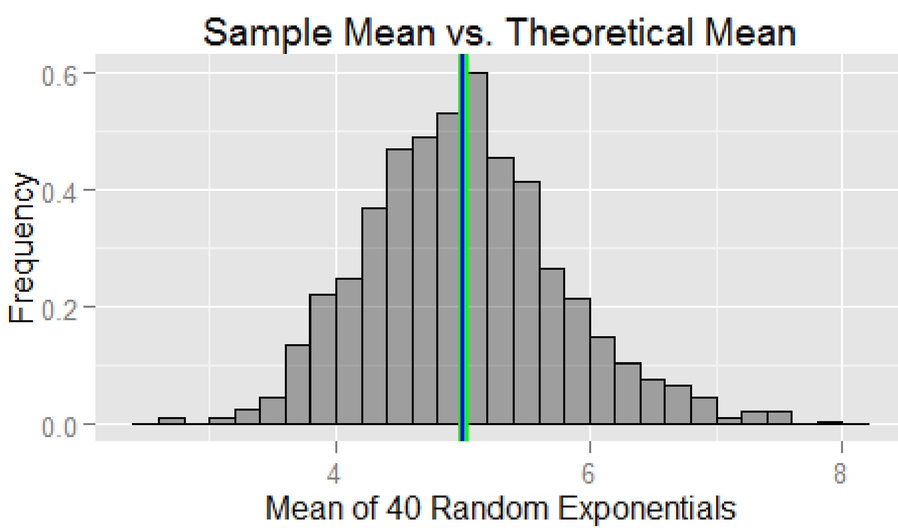
```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2.831000 -0.673600 -0.040830  0.008148  0.579600  3.650000
```

This normalized data set has a mean of 0.008 and a standard deviation of 0.972, very similar to a standard normal.

Sample Mean vs. Theoretical Mean

The theoretical mean of the expoential distribution is known to be 1/lambda, which in the case where lambda = 0.2 is 5. Obvsering our summary output from above, we see that the sample mean from our simulation is 5.006. It is clear that we have a fairly good approximation of the theoretical mean, as our sample mean is only off by 0.006 or 0.13%.

We can also see this from plotting the set of averages from our simulation. In the figure below, the **green** line represents the sample mean while the **blue** line represents the theoretical mean. I have made the green line twice the thickness of the blue line so that it is visible, since it lies so close to the blue line.



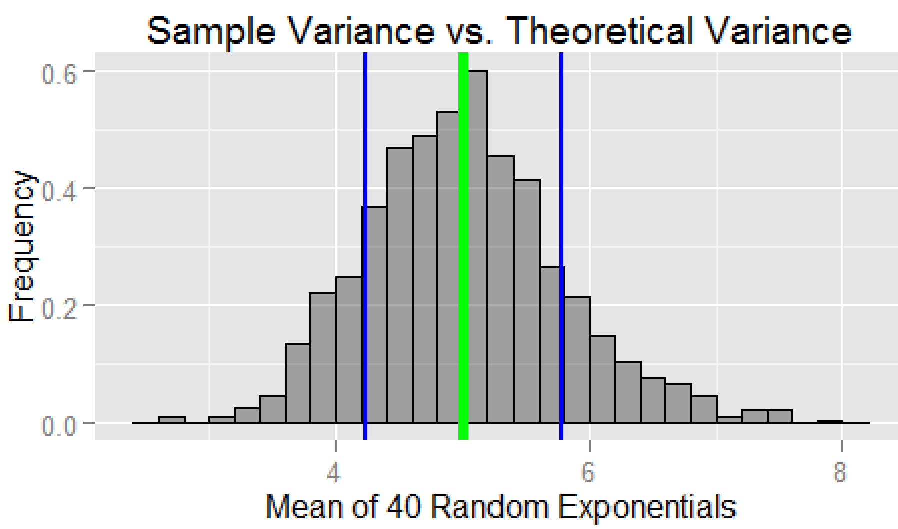
Therefore, our sample mean has converged to the expected (theoretical) mean, which is what we would expect based on the law of large numbers.

Sample Variance vs. Theoretical Variance

The theoretical standard deviation for the exponential distribution is 1/lambda, which is equal to 5 when lambda = 0.2. The theoretical variance is then the standard deviation squared, which is 25 for this example.

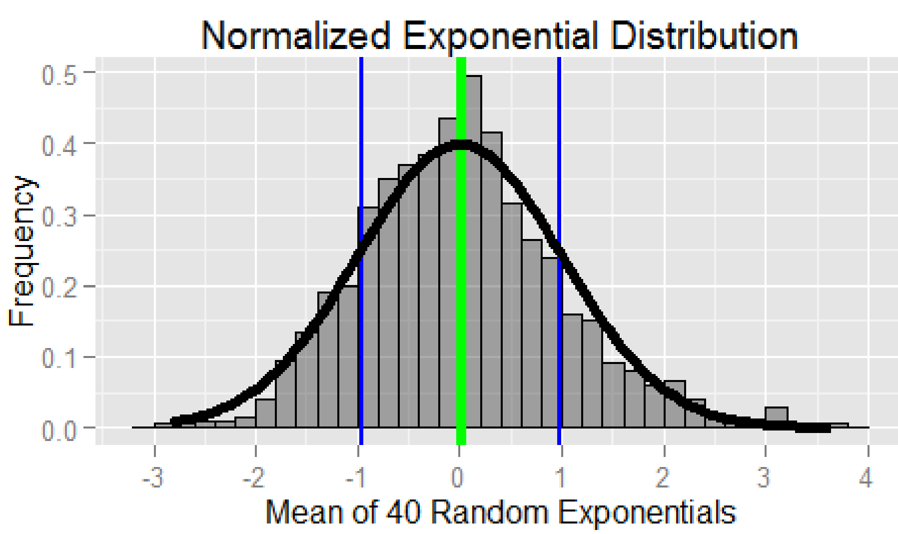
We assume that since we used the average of 40 variables, the sample variance of our new set would be much smaller than the theoretical variance of 25. It turns out this is the case, as the sample variance is 0.591.

We will go back to our graph, keeping the **green** line as the sample mean, but this time using the two **blue** lines to denote plus and minus one standard deviation.



Distribution

For this section, we will use the normalized distribution created in the Simulations section. If we plot our normalized distribution of exponentials and overlay it with the standard normal, we get the following graph. Again, we will use the **green** line as the sample mean and the two **blue** lines to denote plus and minus one standard deviation.



It is easy to see that the standard deviations lie at approximately +1 and -1 and that the mean is approximately 0, which is what we would expect from a standard normal. The histogram looks to give a reasonable approximation of the standard normal, shown by the dark black line.

Appendix

This appendix includes the code for each of the graphs. All of the graphs utilize the ggplot2 package.

Mean Histogram

```
q <- ggplot(data = dat, aes(x = sample)) +
      geom_histogram(alpha = 0.4, binwidth = 0.2, color = "black", aes(y =
..density..)) +
      geom_vline(xintercept = mean(sample), color = "green", size = 2) +
      geom_vline(xintercept = mu, color = "blue", size = 0.8) +
      labs(title = "Sample Mean vs. Theoretical Mean") +
      labs(x = "Mean of 40 Random Exponentials", y = "Frequency")
```

Variance Histogram

```
q <- ggplot(data = dat, aes(x = sample)) +
      geom_histogram(alpha = 0.4, binwidth = 0.2, color = "black", aes(y =
..density..)) +
      geom_vline(xintercept = mean(sample), color = "green", size = 2) +
      geom_vline(xintercept = mu, color = "blue", size = 0.8) +
      labs(title = "Sample Mean vs. Theoretical Mean") +
      labs(x = "Mean of 40 Random Exponentials", y = "Frequency")
```

Normalized Histogram

```
q <- ggplot(data = dat2, aes(x = sampleCLT)) +
      geom_histogram(alpha = 0.4, binwidth = 0.2, color = "black", aes(y =
..density..)) +
      geom_vline(xintercept = mean(sampleCLT), color = "green", size = 2) +
      geom_vline(xintercept = mean(sampleCLT) + sd(sampleCLT), color =
"blue", size = 0.8) +
      geom_vline(xintercept = mean(sampleCLT) - sd(sampleCLT), color =
"blue", size = 0.8) +
      labs(title = "Normalized Exponential Distribution") +
      labs(x = "Mean of 40 Random Exponentials", y = "Frequency") +
      scale_x_continuous(breaks = seq(-6, 6, 1)) +
      stat_function(fun = dnorm, size = 2)
```