Exponential Distribution Simulation

Mark Culp

March 25, 2017

Overview:

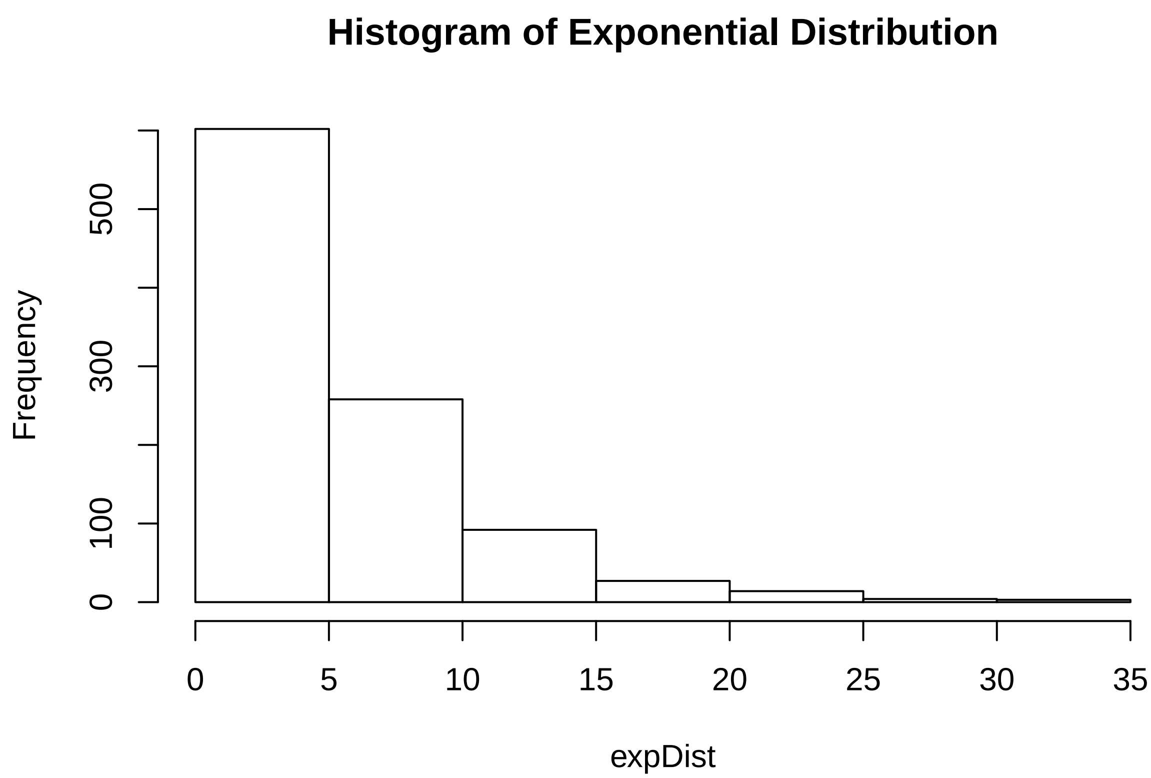
This paper explores simulations on an exponential distribution, a distribution that describes the time between events in a Poisson process. We use R's "rexp(n, lambda)" function to create samples from an exponential distribution, and compare it with the Central Limit Theorem (CLT). The CLT is one of the most important theorems in statistics which states that the distribution or sum of a large number of independent, random variables tends toward a normal distribution regardless of the underlying distribution.

First, let's create an exponential distribution, and examine its properties:

```r
# Set parameters for our exponential distribution
n <- 1000
lambda <- 0.2

set.seed(0)

# Generate 1,000 random deviates for an
# exponential distribution
expDist <- rexp(n, lambda)

# Check the range of numbers in the distribution
range(expDist)
```

```
## [1]  0.008504873 32.922664795
```

```r
# Create a histogram of the distribution of exponentials
hist(expDist, main = "Histogram of Exponential Distribution")
```



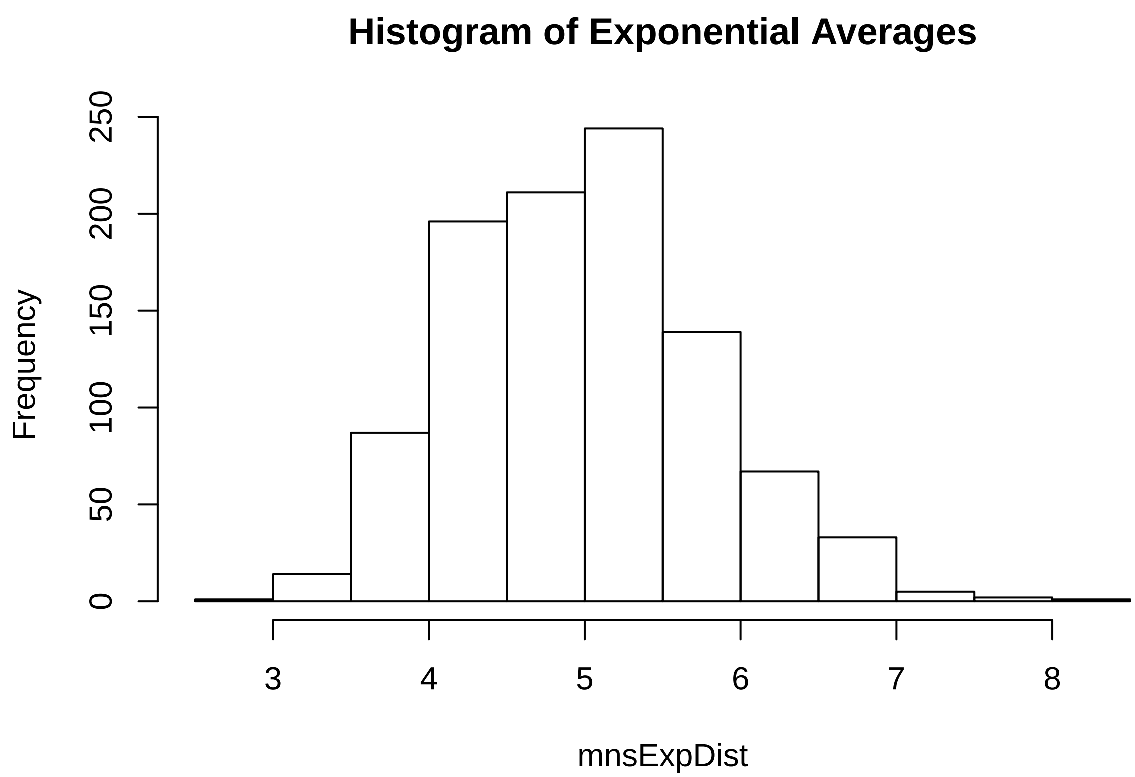
This histogram shows how the density of the exponential distribution changes according to its rate parameter. It is downward sloping, at a decreasing rate. It does not look anything like a normal, bell-shaped distribution.

Simulations:

Now let's run a simulation where we create 1,000 samples of 40 exponential distribution deviates, and calculate the means of the 1,000 samples. These means or averages from the exponential distribution will create a very different distribution.

```
set.seed(0)

# Reset parameters for our exponential distribution
n <- 40
lambda <- 0.2

# Generate a distribution of 1000 averages of 40
# random deviates for an exponential distribution
# and calculate the variances in each sample
expSample = NULL
mnsExpDist = NULL
varExpDist = NULL
for (i in 1 : 1000) {
  expSample = rexp(n, lambda)
  mnsExpDist = c(mnsExpDist, mean(expSample))
  varExpDist = c(varExpDist, var(expSample))
}
```

```r
# Create histogram of the distribution of exponential averages
hist(mnsExpDist, main = "Histogram of Exponential Averages")
```



This distribution appears to be more normal and bell-shaped. This is the Central Limits Theorem at work. Lets now compare the means and variances of the two distributions with the theoretical mean and variance of an exponential distribution.

Sample Mean versus Theoretical Mean

```r
# Load libraries
library(knitr)

# Calculate the mean of the exponential distribution
expDistMean <- round(mean(expDist),3)

# Calculate mean of the distribution of exponential averages
mnsDistMean <- round(mean(mnsExpDist),3)

# Calculate hypothetical mean for the exponential distribution
hypoDistMean <- 1/lambda

# Combine means in a row
rowMeans <- rbind(c(expDistMean, mnsDistMean, hypoDistMean))
```

```r
# Display table of calculated means
kable(rowMeans, caption = "Comparison of Sample and Hypothetical Means", col.names = c("ExponDist","Ave
```

Table 1: Comparison of Sample and Hypothetical Means

ExponDist	Averages	Hypothetical
5.148	4.99	5

This distribution of averages is centered at 4.99 while the theoretical center, or mean, is 5. This is consistent with the Law of Large Numbers (LLN) which indicates that a distribution will converge on its theoretical or expected outcomes as the sample size increases.

The distribution of 1,000 means is drawn from a sampling of 40,000 exponential distribution deviates. It approximates the mean better than the sample of 1,000 random deviates drawn directly from the exponential distribution.

Sample Variance versus Theoretical Variance

```r
# Load libraries
library(knitr)

# Calculate variance of the exponential distribution
expDistVariance <- round(var(expDist),3)

# Calculate variance of the distribution of exponential averages
mnsExpDistVariance <- round(mean(varExpDist),3)

# Calculate hypothetical mean for the exponential distribution
hypoDistVariance <- (1/lambda)^2

# Combine variances in a row
rowVar <- rbind(c(expDistVariance, mnsExpDistVariance, hypoDistVariance))

# Display table of calculated variances
kable(rowVar, caption = "Comparison of Sample and Hypothetical Variances", col.names = c("ExponDist","A
```

Table 2: Comparison of Sample and Hypothetical Variances

ExponDist	Averages	Hypothetical
24.418	25.285	25

This variance for the distribution of average's is approximately 25.285 while the theoretical variance of the distribution is 25. Once again, the distribution of averages more closely approximates the theoretical variance than does the sample of 1,000 random deviates.

Distribution

We assumed we were dealing with a properly normalized distribution of independent and identically distributed (IID) random variables. A collection of random variables is independent if they are statistically unrelated from one and another. They are identically distributed if they were drawn from the same population distribution. Use of the rexp() function ensures these properties.

Our initial histogram of the distribution of averages appeared normal. We now conduct some additional tests to assess this fact.

```
# Standard deviation of distribution
# of exponential averages
sdMns <- sd(mnsExpDist)

# 90 quantile for distribution of
# exponential averages
mns90quan <- qnorm(0.90, mean = mnsDistMean, sd = sdMns)

# Actual number of random deviates
# below the 90th quantile
sum(mnsExpDist < mns90quan)
```

```
## [1] 892
```

```
# 75 quantile for distribution of
# exponential averages
mns75quan <- qnorm(0.75, mean = mnsDistMean, sd = sdMns)

# Calculate actual number of random deviates
# below the 75th quantile
sum(mnsExpDist < mns75quan)
```

```
## [1] 761
```

So we've calculated the 75th and 90th quantile for a normal distribution with the same mean and standard deviation as our distribution of averages. Then we find the number of sample means contained in our distribution below these two numbers. Since our distribution has exactly 1,000 means, we would expect the numbers to come out at about 750 and 900. The actual numbers are 761 and 892. Not far from our expectation. Now lets examine a 95% confidence interval.

```
# 95% confidence interval
mnsDistMean + c(-1,1) * qnorm(0.975) * sdMns
```

```
## [1] 3.449017 6.530983
```

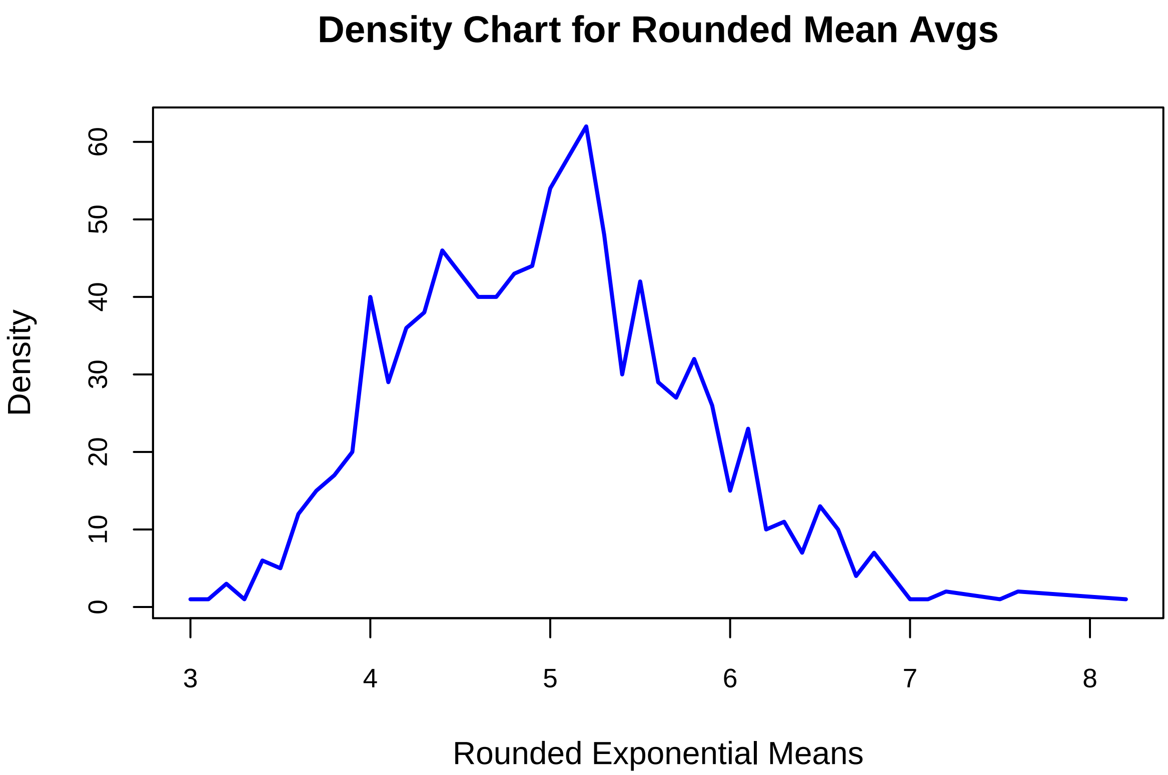
```
# Mean values within the confidence interval
sum(mnsExpDist > 3.449017 & mnsExpDist < 6.530983)
```

```
## [1] 950
```

The 95% confidence interval for the distribution of averages is calculated above, along with the number of means that fall within that interval. This also meets our expectations, 950 values. A normal density chart.

```
# Round means to one decimal place
rndMeans <- round(mnsExpDist,1)

# Sorted unique mean values
x <- sort(unique(rndMeans))

# Create dataframe
df <- as.data.frame(table(rndMeans))

# Plot the densities of the means
```

```
plot(x, df$Freq, col="blue", xlab="Rounded Exponential Means", ylab="Density", type="l",lwd=2, cex=2, ma
```



We know this distribution is normal because it conforms to the characteristics we would expect to see in a normal distribution. We verified this at the 75th and the 90th quantiles, and at the 95% confidence interval. The density chart is not quite normal but it appears to be moving in that direction, as the Central Limits Theorem predicted.