Investigation of Exponential Distribution

Matt Dancho

June 8, 2016

Overview

This report investigates sampling of the exponential distribution and compares the samples to the Central Limit Theorem. The investigation uses the distribution of averages of 40 exponentials simulated 1000 times. Specifically, the report investigates the following:

1. The sample mean of the distribution as it compares to the theoretical mean of the exponential distribution.
2. The sample variance of the distribution as it compares to the theoretical variance for the exponential distribution.
3. Whether or not the distribution is approximately normal as per the Central Limit Theorem.

Simulations

First, the necessary libraries are loaded.

```r
# Load libraries
library(ggplot2)
library(scales)
```

Next, a simulation of 1000 sets of 40 random exponentials is computed. Averages of each set of 40 exponentials are taken, and plotted to show the distribution of the means when n=40.

```r
# Set initial values
lambda <- 0.2   # Rate for exponential distribution
n <- 40         # Number of exponentials averaged per simulation
B <- 1000       # Number of simulations, should be large

# Set seed for reproducibility
set.seed(5050)

# Simulate Exponential Distribution Means
expSamples <- rexp(n * B, lambda)
expMatrix <- matrix(expSamples, nrow = B, ncol = n)
expMeans <- rowMeans(expMatrix)
expMeans_df <- data.frame(means=expMeans)
```

Figure 1 shows a histogram of the simulated averages of the exponential distribution.

```
# Plot the sampled distribution
ggplot(data=expMeans_df, aes(x=means)) +
        geom_histogram(bins = 25, color = "black") +
        geom_rug(alpha=.2) +
        scale_x_continuous(limits=c(1,10), breaks = 1:10) +
        ggtitle("Exponential Distribution Simulation: 1000 Averages of 40 Samples") +
        ylab("Frequency") +
        xlab("Mean")
```

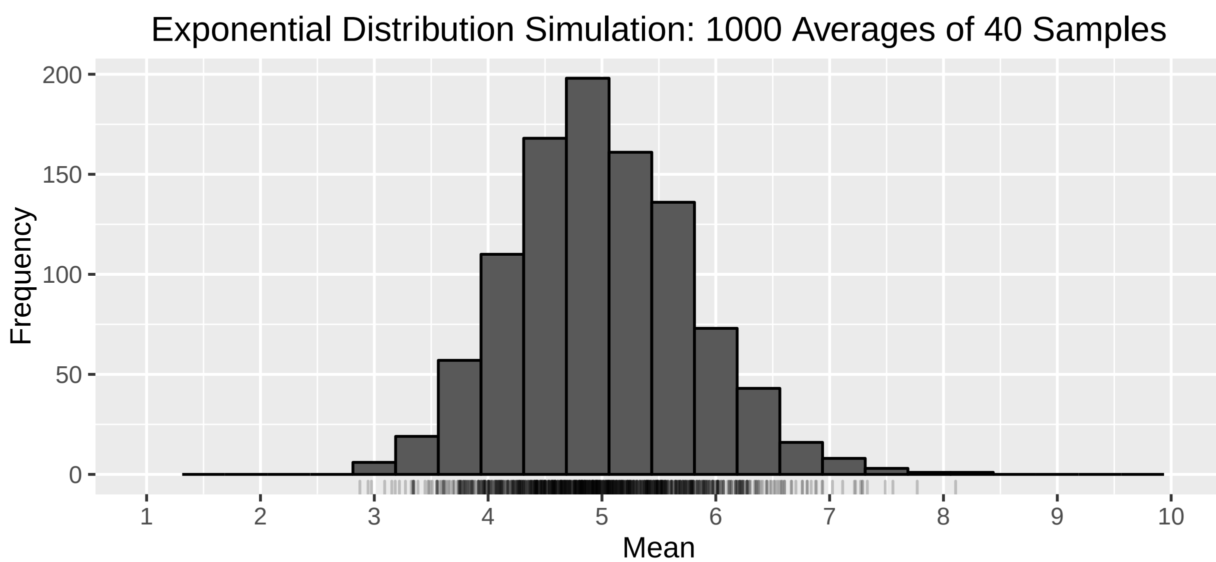

Figure 1: Exponential Distribution Simulation: 1000 Averages of 40 Samples

Sample Mean versus Theoretical Mean

The first part of the investigation compares the sample (experimental) mean to the theoretical mean for the exponential distribution. The expected mean, μ, is given by **equation (1)**:

$$\mu = \frac{1}{\lambda} \tag{1}$$

The following code solves for the theoretical mean in R.

```
theoreticalMean <- 1/lambda
theoreticalMean
```

```
## [1] 5
```

Next, the median of the 1000 samples is computed to yield the experimental mean of the distribution of `n = 40` random exponentials.

```
experimentalMean <- median(expMeans_df$means)
experimentalMean
```

```
## [1] 4.952814
```

As shown, the two means are very close with the absolute difference being 0.047 or 0.944% difference.

Sample Variance versus Theoretical Variance

The second part of the investigation compares the sample (experimental) variance to the theoretical variance for the exponential distribution. The expected variance, σ^2, of `n` measurements is given by **equation (2)**:

$$\sigma^2 = \frac{\frac{1}{\lambda^2}}{n} \tag{2}$$

The following code solves for the theoretical variance in R.

```
theoreticalVar <- (1/lambda^2) / n
theoreticalVar
```

```
## [1] 0.625
```

Next, the variance of the 1000 samples is computed to yield the experimental variance of the distribution of `n = 40` random exponentials.

```
experimentalVar <- var(expMeans_df$means)
experimentalVar
```

```
## [1] 0.6194955
```

As shown, the experimental and theoretical variances are very close with the absolute difference being 0.006 or 0.881% difference.

Is Distribution Normal per Central Limit Theorem?

The third and final part of the investigation evaluates the exponential distribution samples to determine if the distribution is approximately normal as per the Central Limit Theorem.

```
# Create normal dist with same mean and variance as exp dist
x <- seq(1, 10, length.out = 100)
y <- dnorm(x, mean=theoreticalMean, sd=sqrt(theoreticalVar))
norm_df <- data.frame(x, y)

# Plot the sampled exponential distribution and normal distribution
ggplot() +
        geom_histogram(data=expMeans_df, aes(x=means, y=..density..),
                       bins = 25, fill="royalblue", alpha=.3) +
        geom_density(data=expMeans_df, aes(x=means), color="royalblue", size=1) +
        geom_rug(data=expMeans_df, aes(x=means), alpha=.2, color="royalblue") +
        geom_vline(data=expMeans_df, aes(xintercept = median(means),
                                         color="Sampled Exponential"),  size=1) +
        geom_line(data=norm_df, aes(x=x, y=y), color="navy", size = 1) +
        geom_vline(aes(xintercept = theoreticalMean, color="Theoretical Normal"),
                   size = 1) +
        scale_color_manual(name="Distribution",
                           labels=c("Sampled Exponential", "Theoretical Normal"),
                           values=c("royalblue","navy")) +
        scale_x_continuous(limits=c(1,10), breaks = 1:10) +
        ggtitle("Exponential Distribution Simulation Compared to Normal Distribution") +
        theme(plot.title = element_text(size = rel(1))) +
        ylab("Density") +
        xlab("Mean")
```

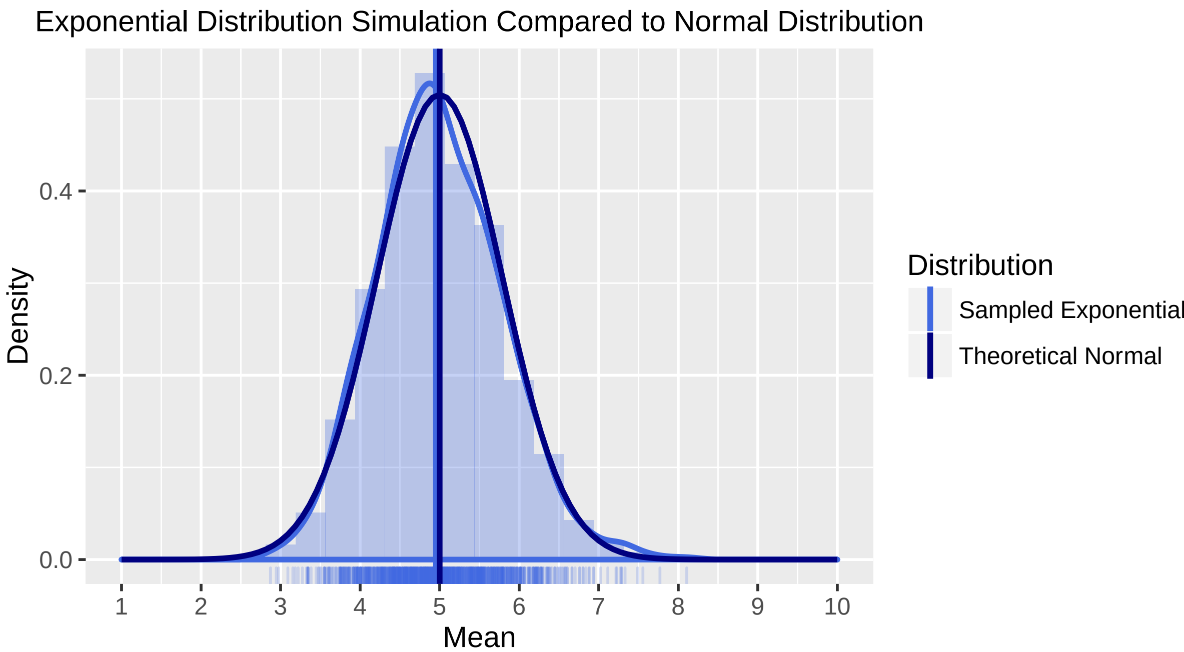

Figure 2: Exponential Distribution Simulation Compared to Normal Distribution

As shown in **Figure 2**, the sampled exponential distribution mimics the normal distribution very closely. With 1000 averages of 40 exponentials, we conclude the distribution is approximately normal as per the Central Limit Theorem.