Simulation of the Exponential Distribution

Statistical Inference Course Project - Part 1

Brian Gulbis

October 03, 2016

Overview

In this report, the exponential distribution will be explored using simulations to determine if it follows the Central Limit Theorem. The sample mean will be compared with the theoretical mean, and the sample variance will be compared with the theoretical variance. Finally, it will be shown that the exponential distribution is, in fact, approximately normally distributed.

Simulation

The exponential distribution will be simulated by taking the mean and variance of 40 random exponentials. The rate parameter, or lambda (λ), will be set to 0.2. The simulation will be repeated 1,000 times. To properly normalize the data, the expected mean will be subtracted from the sample mean and then divided by the standard error, using the following equation from the Central Limit Theorem:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

Sample Mean versus Theoretical Mean

The mean of the exponential distribution is $1/\lambda$, which in this case is expected to be $1/0.2 = 5$.

In the sample of random exponentials, the mean of the mean for the 1,000 samples is 4.98, which is very close to the expected mean, differing by just 0.34%.

A plot of the distribution of sample means shows that they are essentially centered around the expected mean (see figure 1). The sample mean is indicated by the black line, while the expected mean is indicated by the dashed blue line.

Sample Variance versus Theoretical Variance

The variance of the exponential distribution is $1/\lambda^2$, which in this case is expected to be $1/0.22 = 25$. In the sample of random exponentials, the mean of the variance for each of the 1,000 samples is 24.77, which is quite close to the expected variance, differing by only 0.01%. A plot of the distribution of sample variances shows that they are essentially centered around the expected variance (see figure 2). The sample variance is indicated by the black line, while the expected variance is indicated by the dashed blue line.

The variance of the sample mean is σ^2/n. In this case, $\sigma^2 = 1/\lambda^2 = 1/0.2^2 = 25$, and $n = 40$. Therefore, the expected variance of the sample mean is $25/40 = 0.625$. In the sample of random exponentials, the variance of the sample mean is 0.64, which is quite close to the expected variance of the sample mean, differing by only 2.46%.

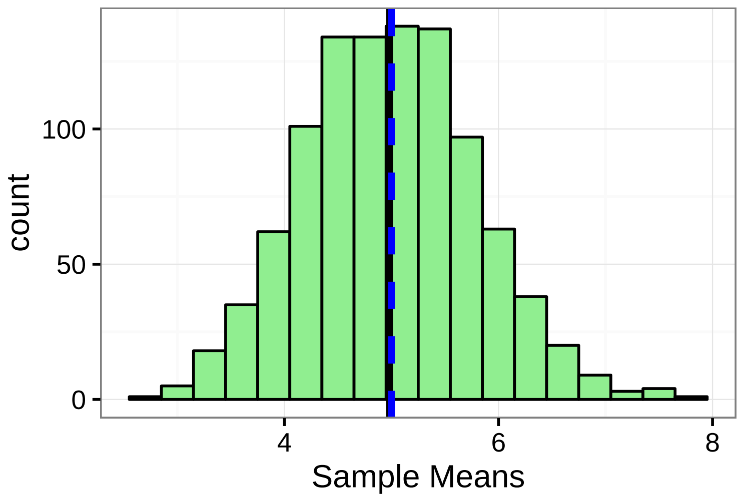

Figure 1: Sample Mean vs. Expected Mean of Random Exponentials

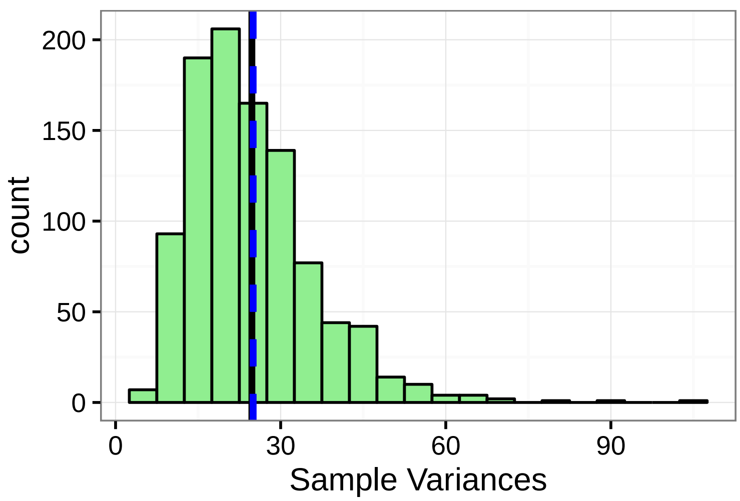

Figure 2: Distribution of Sample Variances Compared with Expected Variance

Distribution

Initially, the exponential distribution may not seem to be normally distributed. If a sample of 1,000 properly normalized random exponentials with $\lambda = 0.2$ is plotted, the distribution appears to be skewed to the left of the mean and is not bell-shaped (see figure 3).

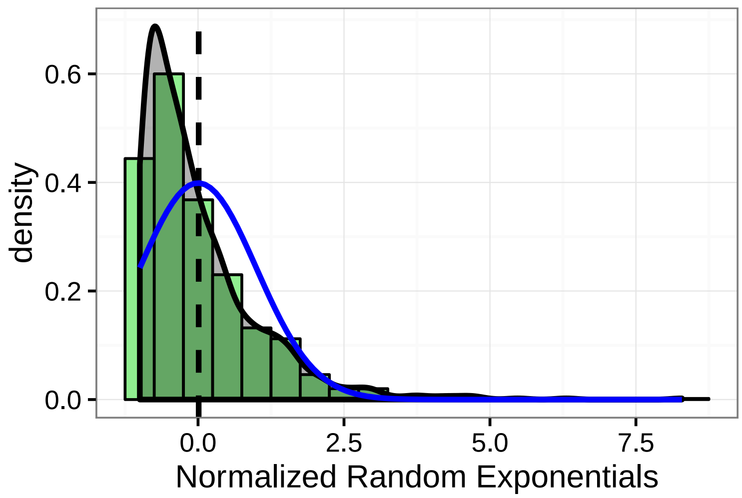

Figure 3: Distribution of 1,000 Normalized Random Exponentials

However, when the exponential distribution is simulated by taking the mean of 40 random exponentials, repeated 1,000 times, and then properly normalized, the distribution does in fact approximate the normal distribution and is centered around the mean (see figure 4). Furthermore, the mean of the normalized distribution is -0.02 and variance is 1.02, which are very close to the expected mean of 0 and variance of 1 for a standard normal distribution.

In both figures 3 and 4, the sample distribution is indicated by the black shaded area, while a standard normal distribution is indicated by the blue line. The sample mean is indicated by the dashed black line.

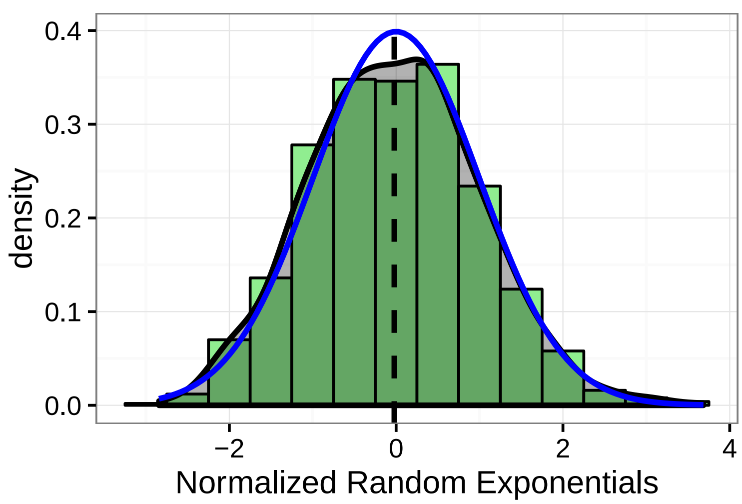

Figure 4: Distribution of 1,000 Means of 40 Random Exponentials

Appendix

The following is the R code used in the simulations and to produce the figures in this report.

Simulation of the exponential distribution

The simulation was done by taking the mean of 40 random exponentials, subtracting off the expected mean, and dividing by the standard error. This was repeated 1,000 times.

```
nosim <- 1000
sim.func <- function(x, n) sqrt(n) * (mean(x) - 5) / 5

m <- matrix(rexp(40 * nosim, 0.2), nosim)

sim.df <- data.frame(clt = apply(m, 1, sim.func, 40), sm = apply(m, 1, mean), sv = apply(m, 1, var))
```

Simulation of a large number of random exponentials

The simulation took 1,000 random exponentials and normalized them by subtracting off the expected mean, then dividing by the standard error.

```
exp.dist <- rexp(1000, 0.2)

exp.df <- data.frame(x = exp.dist, clt = (exp.dist - 5) / 5)
```

Figure 1

```
ggplot(sim.df, aes(x=sm)) +
    geom_histogram(binwidth=0.3, color="black", fill="light green") +
    geom_vline(xintercept=mean(sim.df$sm), size=1.2, color="black") +
    geom_vline(xintercept=5, linetype="dashed", size=1.2, color="blue") +
    xlab("Sample Means") +
    theme_bw()
```

Figure 2

```
ggplot(sim.df, aes(x=sv)) +
    geom_histogram(binwidth=5, colour = "black", fill="light green") +
    geom_vline(xintercept=mean(sim.df$sv), color="black", size=1.2) +
    geom_vline(xintercept=25, color="blue", linetype="dashed", size=1.2) +
    xlab("Sample Variances") +
    theme_bw()
```

Figure 3

```
ggplot(exp.df, aes(x = clt)) +
    geom_histogram(binwidth=0.5, color="black", fill="light green", aes(y = ..density..)) +
    geom_density(size=1, alpha=0.3, fill="black") +
    geom_vline(xintercept=mean(exp.df$clt), size=1, color="black", linetype="dashed") +
    stat_function(fun = dnorm, size = 1, color="blue") +
    xlab("Normalized Random Exponentials") +
    theme_bw()
```

Figure 4

```
ggplot(sim.df, aes(x = clt)) +
    geom_histogram(binwidth=0.5, color="black", fill="light green", aes(y = ..density..)) +
    geom_density(size=1, alpha=0.3, fill="black") +
    geom_vline(xintercept=mean(sim.df$clt), size=1, color="black", linetype="dashed") +
    stat_function(fun=dnorm, size=1, color="blue") +
    xlab("Normalized Random Exponentials") +
    theme_bw()
```