Exponential Distribution Simulation

Matthew Richards

October 2, 2016

Overview

In this brief report, we will investigate the exponential distribution by running 1000 simulations of n = 40, allowing us to construct distributions of sample means and variances. The overall goal is to compare this distribution with the Central Limit Theorem, and we will do so in 3 specific ways:

1. Investigate how well the distribution of 1000 exponential sample means approximates the theoretical mean of the exponential distribution
2. Investigate how well the distribution of 1000 exponential sample standard deviations approximates the theoretical standard deviation of the exponential distribution.
3. Compare the distribution of 1000 exponential sample means to the normal distribution to demonstrate that it is approximately normal.

Create Distributions of 1000 Sample Means and 1000 Sample Standard Deviations

Using the exponential distribution, we can run 1000 simulations of sample size n = 40 using the *rexp()* function. In these simulations, we'll set lambda = 0.2 as directed in the assignment and for each simulation, we'll capture the mean and standard deviation of the sample

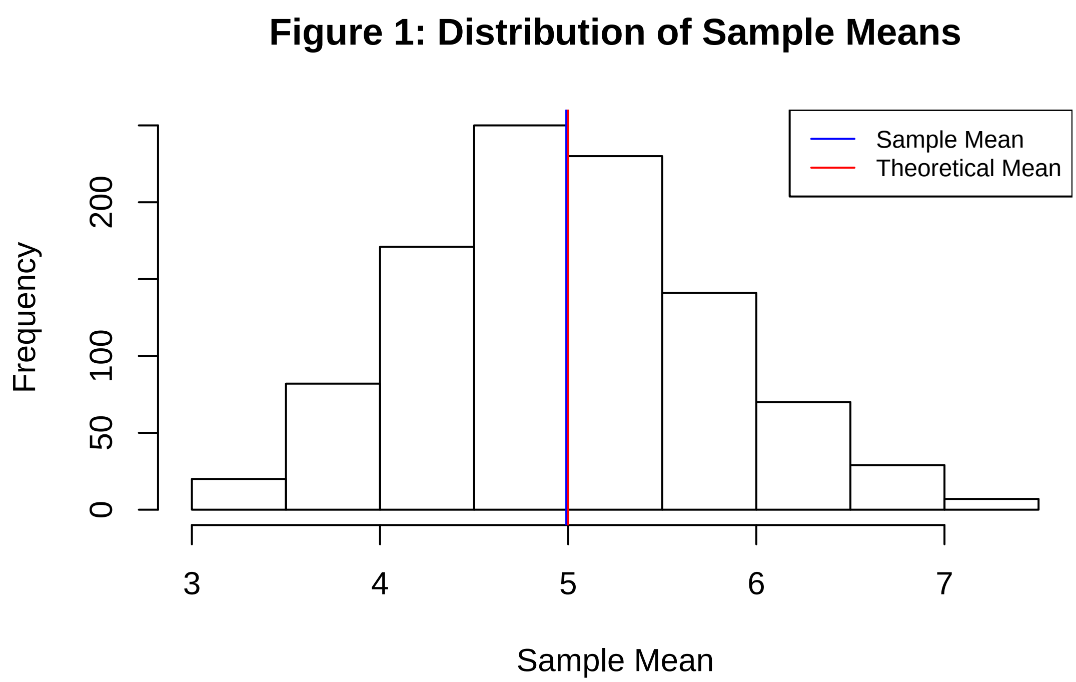
```
set.seed(1)
exp_means <- NULL
exp_sds <- NULL
for (i in 1:1000) {
  exp_sample <- rexp(40,0.2)
  exp_means = c(exp_means, mean(exp_sample))
  exp_sds = c(exp_sds, sd(exp_sample))
}
```

Now we have our two distributions, in *exp_means* and *exp_sds*, respectively. These two numeric vectors will allow us explore the properties of these simulations and how they compare with the theoretical values.

Investigation of the Exponential Mean

We know that the theoretical mean of the exponential distribution is 1/lambda, which in this case is 1/0.2 = 5. Using our histogram of the distribution of sample means, let's see how close our distribution's mean comes to the theoretical value.

```
hist(exp_means,main = "Figure 1: Distribution of Sample Means",
      xlab = "Sample Mean")
abline(v=mean(exp_means),col="blue")
abline(v=5,col="red")
legend("topright",legend = c("Sample Mean","Theoretical Mean"),
       col = c("blue","red"),lty = 1,cex=0.75)
```



As shown by this plot, our 1000 simulations of sample means closely approximate the theoretical mean. How close exactly to the theoretical mean of $\mu = 5$?

```r
mean(exp_means)
```

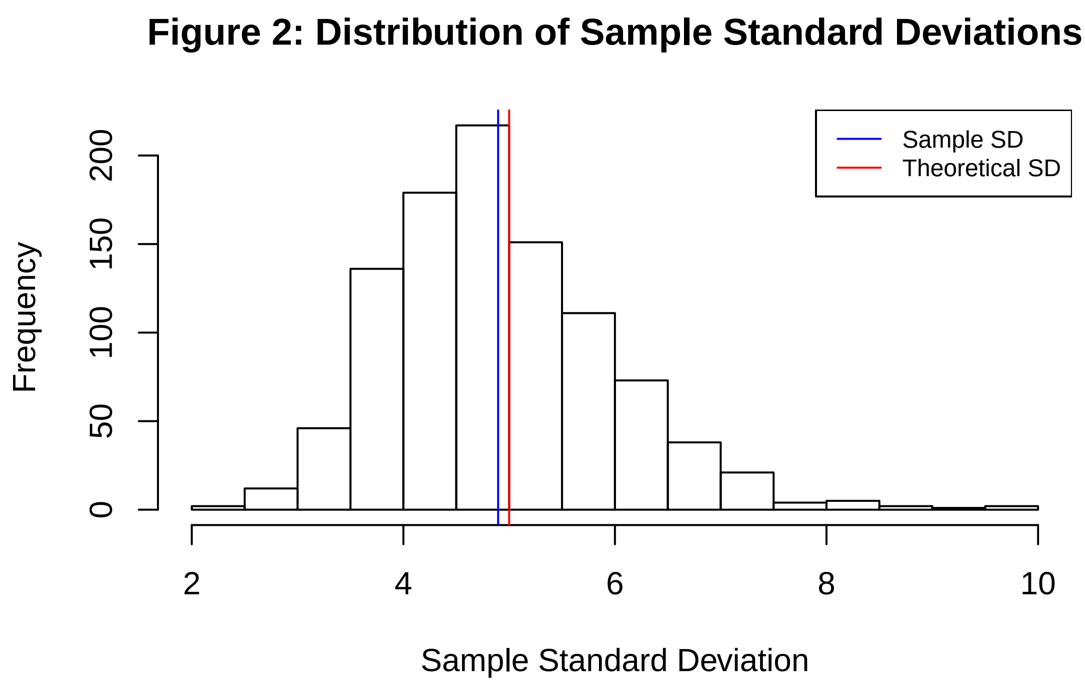
```
## [1] 4.990025
```

It's very close to the theoretical mean and appears to approximate it very well.

Investigation of the Exponential Variance

There are multiple ways to compare the variance of distribution; we could compare the variance of the distribution, s^2, standard deviation of the sample distribution, s, or standard error of the mean, $s/\sqrt{n}$. Given that the theoretical standard deviation of the exponential distribution is 1/lambda, which in this case is also $1/0.2 = 5$, let's work with standard deviation just for numerical simplicity. It's worth noting that we could certainly square our values or divide them by $\sqrt{40}$ if we so chose, but that unnecessarily complicates the math, so let's stick with standard deviation. Just as we did with the means, we can use our histogram of the distribution of sample standard deviations, to see how close our distribution's mean standard deviation comes to the theoretical value.

```r
hist(exp_sds, main = "Figure 2: Distribution of Sample Standard Deviations",
      xlab = "Sample Standard Deviation")
abline(v=mean(exp_sds),col="blue")
abline(v=5,col="red")
legend("topright",legend = c("Sample SD","Theoretical SD"),
       col = c("blue","red"),lty = 1,cex=0.75)
```



Once again, the closeness of the blue and red lines here illustrates that the mean our 1000 simulations of sample standard deviations closely approximate the theoretical standard deviation. For completeness, How close exactly to the theoretical value of $\sigma = 5$?

```
mean(exp_sds)
```

```
## [1] 4.895777
```

As with the mean of our sample means, the mean of our sample standard deviations quite closely approximates the actual value.

Comparison to the Normal Distribution

Finally, let's compare to the normal distribution to show that with 1000 simulations, our distribution of sample means is approximately normal. In order to do this, let's convert our sample means into Z-scores by taking each mean, subtracting the theoretical mean $\mu = 5$, and dividing by standard error of the mean, $SE = 5/\sqrt{40}$. Let's plot this against the standard normal distribution to see how they compare (Appendix, Figure 3)

As demonstrated by this plot, the distribution of sample means shown in blue looks awfully close to the normal distribution shown in red. This is much different than comparing the normal distribution to a distribution of random exponentials (Appendix, Figure 4). In this contrasting case, the exponential distribution, shown in green, looks markedly different than the normal distribution shown in red. Thus, this serves to illustrate the Central Limit Theorem, in that even though the exponential distribution itself is not Gaussian, a large collection of sample means from the same distribution does appear Gaussian.

Appendix

Figure 3

```
suppressWarnings(library(ggplot2))
plot_data <- data.frame(exp_means = (exp_means - 5)/(5/sqrt(40)))
plot_data$norm_scores <- rnorm(1000)
g <- ggplot(plot_data,aes(exp_means))
g + stat_density(color="blue",geom = "line",size=2) +
  stat_density(aes(norm_scores),
               color = "red",geom ="line",size=2) +
  labs(title = "1000 Exponential Sample Means (blue) v. Normal Distribution (red)",
       x = "Z-score")
```

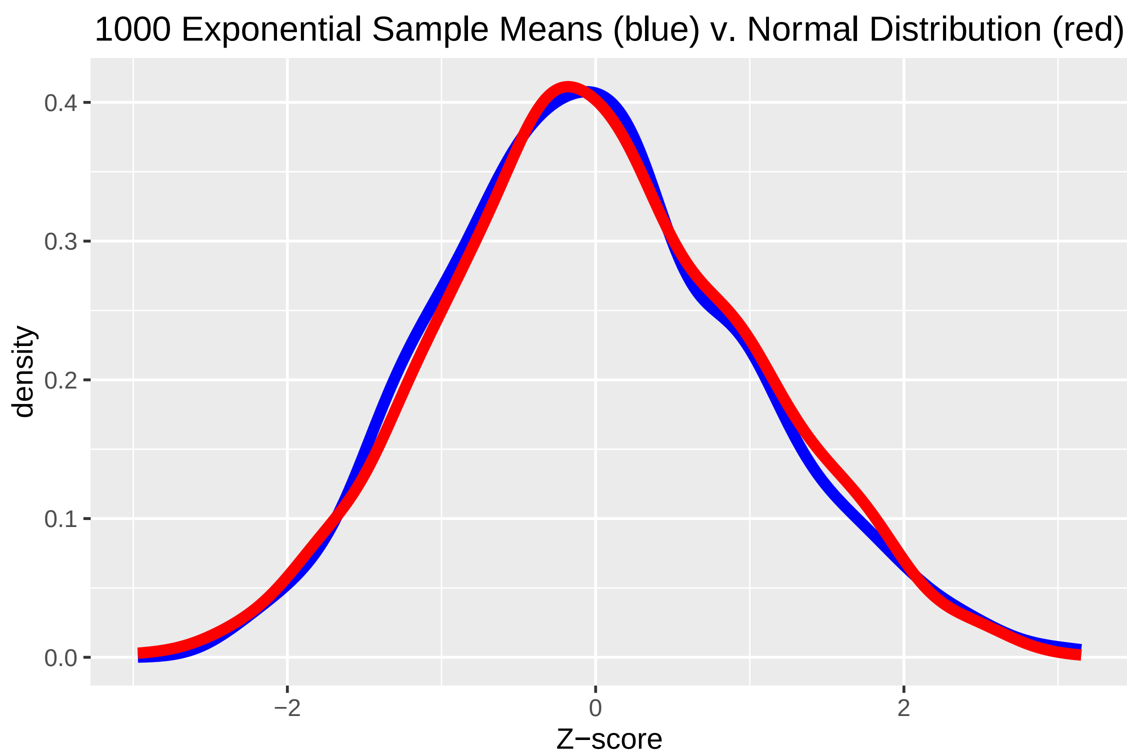


Figure 4

```
plot_data <- data.frame(exp_means = (rexp(1000,0.2)-5)/(5))
plot_data$norm_scores <- rnorm(1000)
g <- ggplot(plot_data,aes(exp_means))
g + stat_density(color="green",geom = "line",size=2) +
  stat_density(aes(norm_scores),geom = "line",
               color = "red",size=2) +
  labs(title = "Exponential Distribution (green) v. Normal Distribution (red)",
       x = "Z-score") +
  scale_fill_discrete()
```