Coursera Statistical Inference Course Project: Part 1

Brad Tweardy

Thursday, March 19, 2015

Introduction

In this project we will analyze the exponential distribution in R and compare it with the Central Limit Theorem. We will investigate the distribution of averages of 40 exponentials.

Simulations

Here we will run the similations and show the code for which we will use to conduct our analysis. For this project we have set lamda = 0.2 for all simulations. Keep in mind that the mean of exponential distribution and the standard deviation is calculated using the formula 1 / lambda. In addition, the exponential distribution is simulated in R using rexp(n, lambda) where lambda is the rate parameter. Our sample for this analysis will come from 1,000 simulations.

The first task that we want to accomplish is to set our variables and calculate the Theoretical Distribution Mean and the mean of 40 exponentials (Exponential Distribution Mean). The simulation of the sample is done with the following code:

```r
lambda = .2
sd_exp_dist <- 1/lambda
n = 40
exp_dist <- rep(NA, 1000)
for (i in 1:1000) {
        exp_dist [i] <- mean(rexp(n, lambda))
}
## Theoretical Distribution Mean
theo_mean <- 1 / lambda
## Exponential Distribution Mean
exp_dist_mean <- mean(exp_dist)
## Theoretical Distribution Variance
theo_var <- ((sd_exp_dist) ^ 2) / n
## Exponential Distribution Variance
exp_dist_var <- var(exp_dist)
```

Results

Now that we have simulated the data we, we can review the results.

Mean Averages of the Distribtutions In our analysis of the first set of data, we will look at the output of the calculations for the Theoretical Distribution Mean and the Exponential Distribution Mean. We will then plot a histogram of the exponential distribtution of the sample of 1,000 simulations and then plot the mean values, represented by a single vertical line, one for each distribution. We will then comment on out observations as to how the results compare.

Theoretical Distribution Mean

```
theo_mean
```

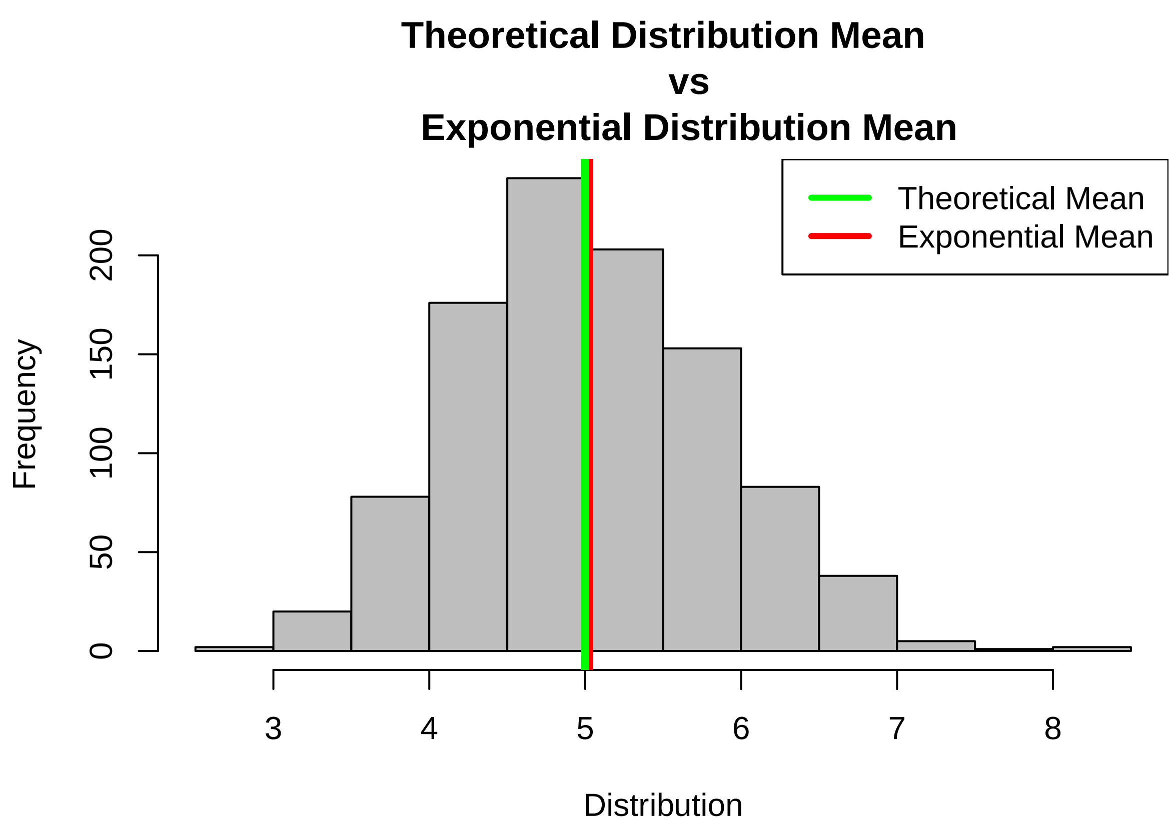
```
## [1] 5
```

Exponential Distribution Mean

```
exp_dist_mean
```

```
## [1] 5.025943
```

```
hist(exp_dist, main = "Theoretical Distribution Mean
     vs
     Exponential Distribution Mean",
     xlab = "Distribution", ylab = "Frequency", "col" = "grey")
abline(v = exp_dist_mean, col = "red", lwd = 4)
abline(v = theo_mean, col = "green", lwd = 4)
legend("topright", c("Theoretical Mean", "Exponential Mean"),
       col = c("green", "red"), lwd = 3)
```



As we can see, based on the calculations of the means, the variance is rather insignificant. This is reflected in the plotted histogram of the exponential values in which the two means are represented by individually plotted vertical lines against the histogram plot. The lines almost directly overlap each other, representing a rather insignificant variance in the two mean averages.

Distribution Variance In our next analysis, our goal is to compare the Distribution Variance of the Theoretical Distribution and the Exponential Distribution values. We will then confirm a normal distribution of the samples with a plotted histogram of the sample of 1,000 Exponential Distributions. We will then plot two density curves against the histogram to show a normal distribution of the sample set and comment on our findings.

Theoretical Distribution Variance

```
theo_var
```

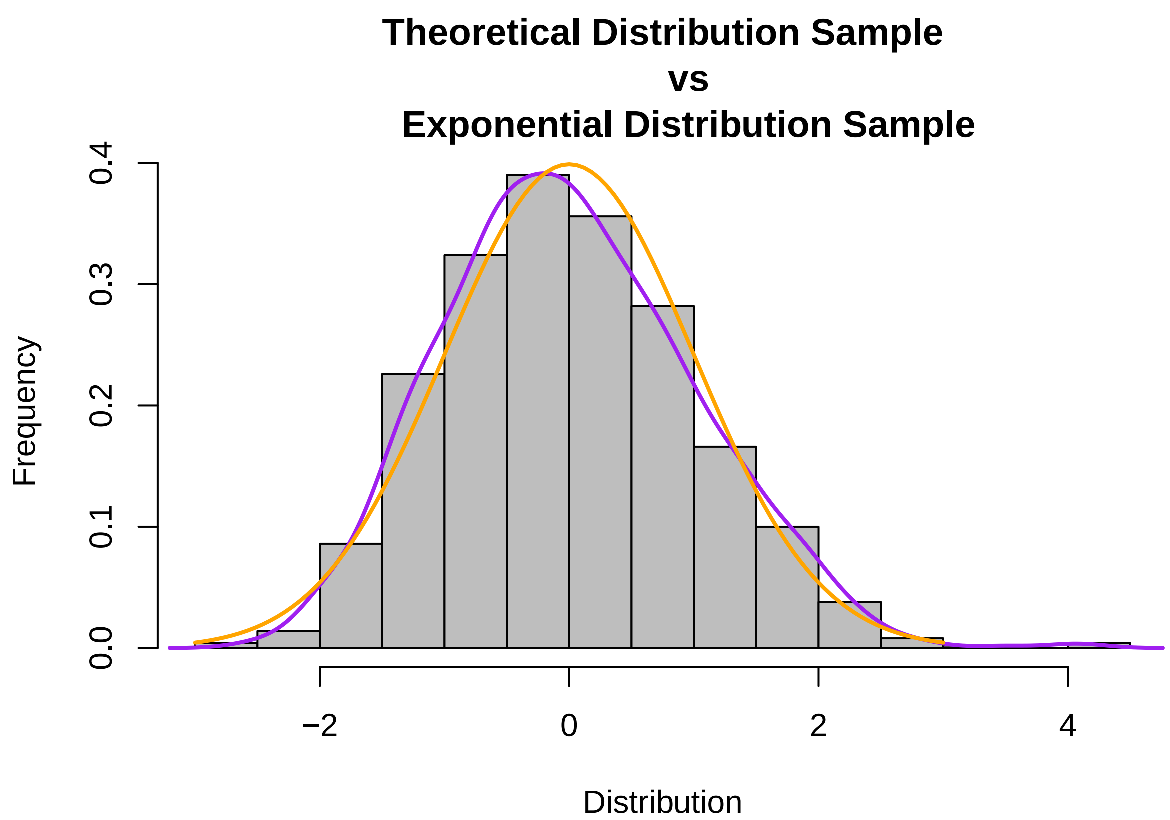
```
## [1] 0.625
```

Exponential Distribution Variance

```
exp_dist_var
```

```
## [1] 0.667329
```

```
exp_dist_scale <- scale(exp_dist)
hist(exp_dist_scale, main = "Theoretical Distribution Sample
     vs
     Exponential Distribution Sample",
     xlab = "Distribution", ylab = "Frequency", "col" = "gray", prob = TRUE)
lines(density(exp_dist_scale), col = "purple", lwd = 2)
curve(dnorm(x, 0, 1), -3, 3, add = TRUE, col = "orange", lwd = 2)
```



As we refer to the Distribution Variances, we can see that the variances are again insignificant. We represent this by creating a histogram plot of the sample distributions. We then added separate density lines to represent the variance in the distributions. The density lines overlap relatively nicely representing an insignificant variance.