Comparison of Exponential Distribution & Central Limit Theorem

Chris Keaney

June 18, 2015

Overview

In this project we will be investigating the exponential distribution in R and comparing it with the Central Limit Theorem. The exponential distribution will be simulated in R using ***rexp(n, lambda)***, and we will be setting ***lambda = .2*** for all of our simulations. We will investigate the averages of 40 exponentials over a thousand simulations, and then compare the samples means and variances of the simulation to the theoretical mean and variance of an exponential distribution. We will also show how the simulation distribution is approximately normal.

Simulation

The code for our simulation of 40 exponentials over 1,000 runs is below. In it, we are taking the mean of each simulation through ***mean(rexp(n,lambda)*** and storing the result in the variable ***sample__mean***.

```
n <-  40
lambda <- .2
sim <- 1000

set.seed(8675309)

sample_mean <- NULL

for(i in 1:sim) sample_mean = c(sample_mean,mean(rexp(n,lambda )))

str(sample_mean)
```

```
##  num [1:1000] 4.59 4.71 4.41 4.43 5.58 ...
```

```
summary(sample_mean)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.737   4.459   4.949   5.012   5.528   8.204
```

As you can see from the summary tables above, the variable sample mean has 1,000 values ranging from ***2.737*** to ***8.204***, with a mean of ***5.012***. Now let's compare this mean to the thoeretical mean of an exponential distribution.

Sample Mean vs Theoretical Mean

As stated above, the sample mean from our distribution was:

```
mean(sample_mean)
```

```
## [1] 5.011936
```

According to the CLT we would expect that the theoretical mean of an exponential distribution would be 1/lambda. Our labmda in this case is ***.2***, so our theoretical mean for this exercise is ***1/.2 = 5***. As you can see, the mean from our sample distribution of ***5.012*** very closely approximates the theoretical mean of ***5***.

Sample Variance vs Theoretical Variance

We will now make the same comparison for the variance of the sample distribution vs the theoretical variance. Our sample variance is as follows:

```
sample_var <- var(sample_mean)
sample_var
```

```
## [1] 0.6468625
```

Our theoretical variance is also ***(1/lambda)^2 / n***. So calculating our theoretical variance yields:

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

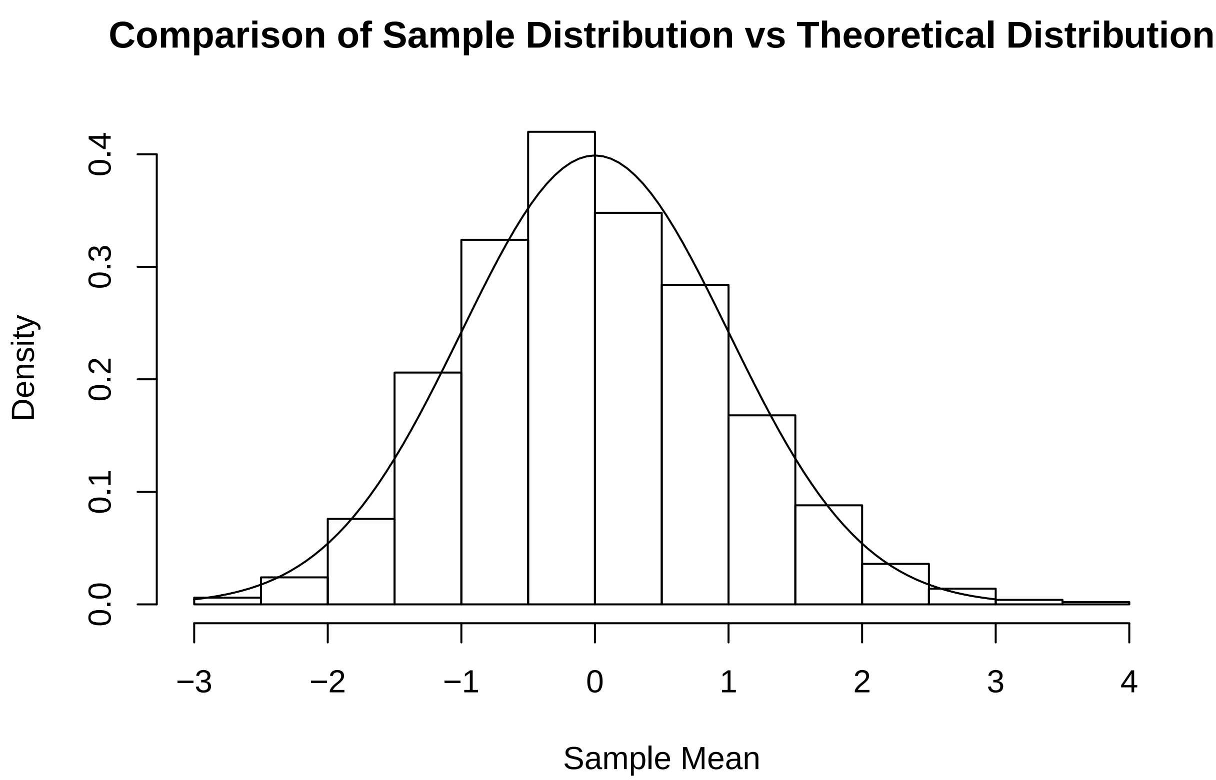
```
## [1] 0.625
```

Again, we have a very close approximation of our sample variance of ***.647*** to our theoretical variance of ***.625***.

Showing That The Sample Distribution is Approximately Normal

Given the above, we will plot out a histogram of our sample distribution and overlay the exponential distribution density fuction to show that we have an approximately normal sample distribution.

```
hist(scale(sample_mean), probability=T, main = "Comparison of Sample Distribution vs Theoretical Distri
curve(dnorm(x,0,1), -3,3, add=T)
```



As you can see, overlaying the normal curve against our data shows a very close fit.

Appendix

Here is all of the code needed to rerun this experiment:

```
# R code for Statistical Inference Project

# Calculate the sample mean of an exponential distribution and compare to theoretical mean.

n <-  40
lambda <- .2
sim <- 1000

set.seed(8675309)

sample_mean <- NULL

for(i in 1:sim) sample_mean = c(sample_mean,mean(rexp(n,lambda )))

str(sample_mean)
summary(sample_mean)

theoretical_mean <- 1/lambda

# Calculate the sample variance of an exponential distrobution and compare to theoretical variance

sample_var <- var(sample_mean)
theoretical_var <- ((1/lambda)^2)/n

# graph comparison of sample distribution to normal distribution.

hist(scale(sample_mean), probability=T)  #add some labels and what not
curve(dnorm(x,0,1), -3,3, add=T)
```