Statistical Inference - Simulations and Inferential Data Analysis

Chris Selig

January 5, 2017

# Overview

In the first part of this report, I will perform a simulation exercise that compares a theoretical exponential distribution to a simulated exponential distribution. The analysis will show that using the Central Limit Thereom, the simulated exponential distribution will approach a normal distribution with enough simulations.

In the second part of this report, I will perform hypothesis testing on tooth growth rate in guinea pigs to show that for a dose level of 0.5mg/day or 1.0mg/da, the orange juice delivery method is superior to the ascorbic acid method.

## Part 1

### Simulations

For the simulation exercise, we are going to do 1000 simulations of 40 exponential distributions, using a lambda = 0.2, the rexp function to simulate the exponential distribution. The code used for the simulation is:

set.seed(1264)  
  
###Set paramaters  
n <- 40  
lambda <- 0.2  
numsimulations <-1000  
  
###Simulate 1000 trials of 40 exponential distributions  
simdata <- matrix(rexp(n\*numsimulations,lambda), numsimulations)

### Comparisions of the Theoretical Exponential Distribution to the Simulated Distribution

Next, I calculated the theoretical exponential distribution properties (mean, variance and standard deviation) and compare versus the simulated exponential distribution.

###Calculate Theoretical Properties  
theoreticalmean <- round(1/lambda,2)  
theoreticalsd <- round(1/lambda,2)  
theoreticalvar <- round(1/lambda^2,2)  
  
#Calculate the Simulated Exponential Distribution Properties  
simmeansdata <- round(apply(simdata,1,mean),2)  
simmean <-round(mean(simmeansdata),2)  
simsd <- round(sd(simmeansdata),2)\*sqrt(n)  
simvar <- round(simsd^2,2)

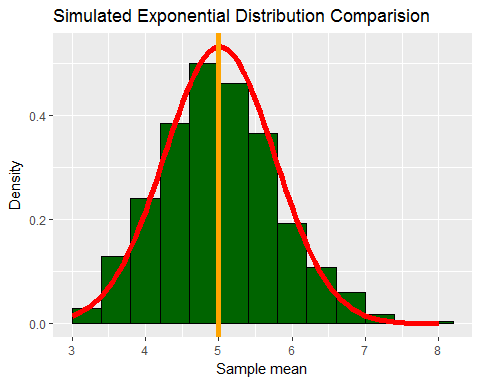
Now, to compare the two:

## Theoretical Simulated  
## Mean 5.000000 5.010000  
## Variance 25.000000 25.600000  
## Standard Deviation 5.000000 5.059644

As you can see, the mean, variance, and standard deviation of both distributions are very close to each other.

## Distribution - Does the Central Limit Thereom Work?

The Central Limit Thereom states that for a large number of independent random variables, their sum will follow a stable (normal distribution). After 1000 simulations, this is what the distribution looks like.



The histogram shows the density of the means, while the red line over lays a normal distribution over top. You can see that the simulated distribution is approximately the same shape as the normal distribution.

Furthermore, you can also see that the theoretical mean of our exponential distribution (mu = 5), is approximately centered on the mean of the simulated distribution.

## Part 1 Appendix

### Entire Code For Part 1

library(ggplot2)  
  
set.seed(1264)  
n <- 40  
lambda <- 0.2  
numsimulations <-1000  
  
theoreticalmean <- round(1/lambda,2)  
theoreticalsd <- round(1/lambda,2)  
theoreticalvar <- round(1/lambda^2,2)  
  
simdata <- matrix(rexp(n\*numsimulations,lambda), numsimulations)   
simmeansdata <- round(apply(simdata,1,mean),2)  
simmean <-round(mean(simmeansdata),2)  
simsd <- round(sd(simmeansdata),2)\*sqrt(n)  
simvar <- round(simsd^2,2)  
  
tablevalues <- matrix(c(theoreticalmean, simmean,theoreticalvar, simvar,   
 theoreticalsd, simsd), ncol = 2, byrow = TRUE )  
colnames(tablevalues) <- c("Theoretical", "Simulated")  
rownames(tablevalues) <- c("Mean", "Variance", "Standard Deviation")  
tablevalues <- as.table(tablevalues)  
tablevalues  
  
simmeansdatadf <- data.frame(simmeansdata)  
  
ggplot(simmeansdatadf, aes(x = simmeansdata)) +   
 geom\_histogram(binwidth = 0.4, color = 'black', fill = 'dark green', aes(y = ..density..)) +  
 stat\_function(fun = dnorm, color = 'red', size = 2, args = list(mean = simmean, sd = .75)) +  
 xlab('Sample mean') + geom\_vline(xintercept = theoreticalmean, color = 'orange', size = 2) +   
 ylab('Density') + ggtitle("Simulated Exponential Distribution Comparision")