Mean and Variance of The Exponential Distribution

Barret Miller

## Overview

In this project we will look at sample means and sample vairances from the Exponential Distribution. We'll compare these values found empirically to what theory tells us about the mean and variance for this distribuion. We know that the Exponential Distribution has mean 1/lambda and also standard deviation 1/lambda.

## Simulations

For this piece, I'll generate a matrix of 1000 rows by 40 columns. Each row will represent 1 set of 40 random samples from the exponential distribution. Each of the sets of 40 will be averaged, taking the sample mean. Then we will have 1000 means of sets of size 40 from the exponential distribution. We will take the mean of these sets, which will represent our empirical sample mean. We will also take the variance and standard error of this set of 1000 means of 40. For our simulations, we will generate the random exponentials with the rexp() function using a rate (lambda) of .2.

expSamples = matrix(rexp(40000, rate=.2), 1000)  
  
#Display Our Matrix  
str(expSamples)

## num [1:1000, 1:40] 6.17 6.89 0.29 11.2 1.14 ...

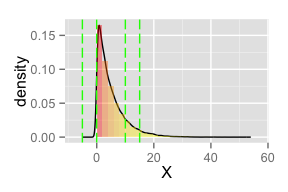
## Sample Mean vs. Theoretical Mean

Let's look at the mean and standard deviation of all of our samples. As you can see below, the mean is about 5, and so is the standard error, so this seems to be in line with what we know about the distribution, with some sample noise causing the values to not be exact. This is expected.

suppressMessages(library(ggplot2))  
m <- mean(as.vector(expSamples))  
v <- var(as.vector(expSamples))  
s <- sd(as.vector(expSamples))  
sprintf("Mean:%f, var:%f, sd:%f", m, v, s)

## [1] "Mean:5.006010, var:25.173823, sd:5.017352"

p0 <- qplot(as.vector(expSamples), geom = "blank", xlab="X") +  
 geom\_line(aes(y = ..density..), stat = 'density') +   
 geom\_histogram(aes(y=..density.., fill = ..density..), alpha = 0.4) +  
 scale\_fill\_gradient("Count", low = "yellow", high = "red", guide = F) +  
 geom\_vline(xintercept = m+c(-1\*s, s, -2\*s, 2\*s),   
 colour="green", linetype = "longdash")  
suppressMessages(print(p0))



Now, we apply the mean function to all rows, returning a single vector of means. This will be our new random variable, X', which is comprised of the means of size n=40 random samples from the underlying exponential distribution.

meansOfFortySamps = apply(expSamples, 1, FUN=mean)  
str(meansOfFortySamps)

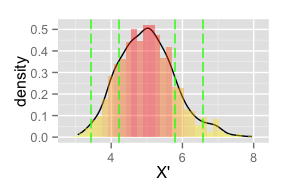
## num [1:1000] 5.03 4.79 3.49 4.12 4.79 ...

Now that we have our new random variable, we'll take the mean and standard error of the sample means. Stats-on-stats. Notice that the mean of the distribution of means of samples of size 40 is still about 5. However, the standard error is now much smaller, not even close to the mean. Is that what you expected? What about 1/lambda and what we know about the SD of the exponential distribution being equal to the mean? At first, this may throw you, but it's critical to remember here that this is the standard deviation of a new distribution, X'. This is the distribution of the sample means taken from samples of size 40.

suppressMessages(library(ggplot2))  
m <- mean(meansOfFortySamps)  
v <- var(meansOfFortySamps)  
s <- sd(meansOfFortySamps)  
sprintf("Mean:%f, var:%f, sd:%f", m, v, s)

## [1] "Mean:5.006010, var:0.617927, sd:0.786083"

p1 <- qplot(meansOfFortySamps, geom = "blank", xlab="X'") +  
 geom\_line(aes(y = ..density..), stat = 'density') +   
 geom\_histogram(aes(y=..density.., fill = ..density..), alpha = 0.4) +  
 scale\_fill\_gradient("Count", low = "yellow", high = "red", guide = F) +   
 geom\_vline(xintercept = m+c(-1\*s, s, -2\*s, 2\*s),   
 colour="green", linetype = "longdash")  
suppressMessages(print(p1))



## Sample Variance vs. Theoretical Variance

## Distribution and Normality