Statistical Inference Course Project - Part 1

Alex McBride

Friday, February 06, 2015

## Synopsis

Investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should 1. Show the sample mean and compare it to the theoretical mean of the distribution. 2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. 3. Show that the distribution is approximately normal.

## Create the Distribution

# set seed for reproducability  
set.seed(10)  
  
# set lambda to 0.2  
lambda <- 0.2  
  
# 40 samples  
n <- 40  
  
# 1000 simulations  
sim <- 1000  
  
# simulate  
sim\_exp <- replicate(sim, rexp(n, lambda))  
  
# calculate mean of exponentials  
means\_exp <- apply(sim\_exp, 2, mean)

### Show and compare the distribution mean and variance

Theoretical mean is 1/lambda 5

Our analytical mean is mean(means\_exp) 5.0450596

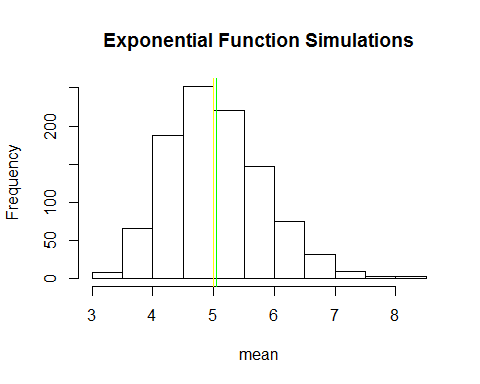
The theoretical standard deviation is 1/lambda/sqrt(n) 0.7905694

Our analytical standard deviation is sd(means\_exp) 0.7982821

The theoretical variance is ((1/lambda)\*(1/sqrt(n)))^2 0.625

Our analytical variance is var(means\_exp) 0.6372544

hist(means\_exp, xlab = "mean", main = "Exponential Function Simulations")  
abline(v = mean(means\_exp), col = "green")  
abline(v = 1/lambda, col = "yellow")



### Show that the distribution is approximately normal

expscale <- scale(means\_exp)  
hist(expscale,probability=T, main="Distribution Density", col = "cyan", ylim=c(0, 0.5))  
lines(density(expscale), col = "purple", pch = 22, lty=5)  
# Compare with the standard normal distribution  
curve(dnorm(x,0,1), -4, 4, col="red", add=T)

