Statistical Inference Course Project, Part 1: Simulation Exercises

Niraj Patel

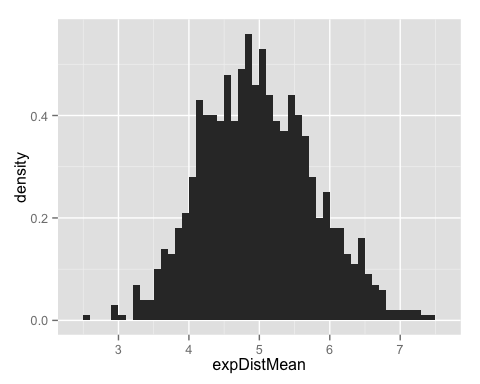
October 24, 2015

# Simulation

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is and the standard deviation is also . For this simulation, we set . In this simulation, we investigate the distribution of averages of 40 numbers sampled from exponential distribution with .

Let's do a thousand simulated averages of 40 exponentials.

# Load necessary libraries  
library(ggplot2)  
  
# Set the simulation constants and seed value   
set.seed(3)  
n=40  
lambda=0.2  
  
# Run the test data   
Mean <- data.frame(sapply(1:1000,function(x) {mean(rexp(n,lambda))}))  
names(Mean) <- c("expDistMean")  
  
# Plot the test results  
  
ggplot(data = Mean, aes(x = expDistMean,y=..density..)) +   
 geom\_histogram(binwidth=0.1) +   
 scale\_x\_continuous(breaks=round(seq(min(Mean$expDistMean), max(Mean$expDistMean), by=1)))



# Sample Mean vs Theoretical Mean

The expected mean of a exponential distribution of rate is

expMean <- 1/lambda  
expMean

## [1] 5

Let be the average sample mean of 1000 simulations of 40 randomly sampled exponential distributions.

sampleMean <- mean(Mean$expDistMean)  
sampleMean

## [1] 4.98662

Due to high number of simulations, it can be observed that the sample mean and expected mean are fairly close to one another.

# Sample Variance vs Theoritical Variance

The expected variance of a exponential distribution of rate is

expsd <- 1/lambda/sqrt(n)  
expsd

## [1] 0.7905694

While the expected variance then is simply the square of

expVar <- (1/lambda/sqrt(n))^2  
expVar

## [1] 0.625

The sample of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution is

samplesd <- sd(Mean$expDistMean)  
samplesd

## [1] 0.7947823

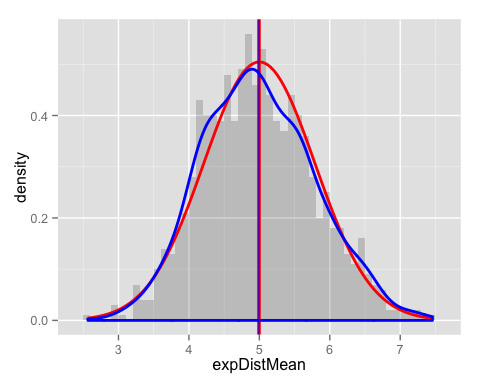
and as before, the sample variance is

sampleVar <- samplesd^2  
sampleVar

## [1] 0.6316789

# Distribution in Graphical Manner

The line in read denotes the values from a normally distributed population while that in blue denotes that of the sampled cohort. It can be seen that due to the fairly large sample size, the sampled cohort closely resembles that of a normal distribution, as seen in the Central Limit Theorem (CLT).



The vertical line in the middle denotes the centre of the distribution and it can be seen that the mean sampled distribution almost directly overlaps that of the normal distribution.