Exploration of Simulated vs. Theoretical Exponential Distribution

Henry Bowers

Sunday, November 23, 2014

This study simulates the exponential distibution and compares it to its theoretical definition.

Simulation of Exp Distribution

For this study we start with the means of 1.000 simulated distributions, each with sample size of 40. The samples are drawn from an exponential distribution with lamba = 0.2.

```
lambda <- 0.2

meanDistExp <- function(reps,size) {
  dist <- NULL
  for (i in 1 : reps) dist = c(dist, mean(rexp(size,lambda)))
  return(dist)
}

mns <- meanDistExp(1000,40)
simMean <- round(mean(mns),2)
simSD <- round(sd(mns),2)
```

Comparison of Simulated and Theoretical Mean

The center of the distribution of simulated means, 5.01, reasonably approximates the theoretical mean (lambda = 0.2) of 5.

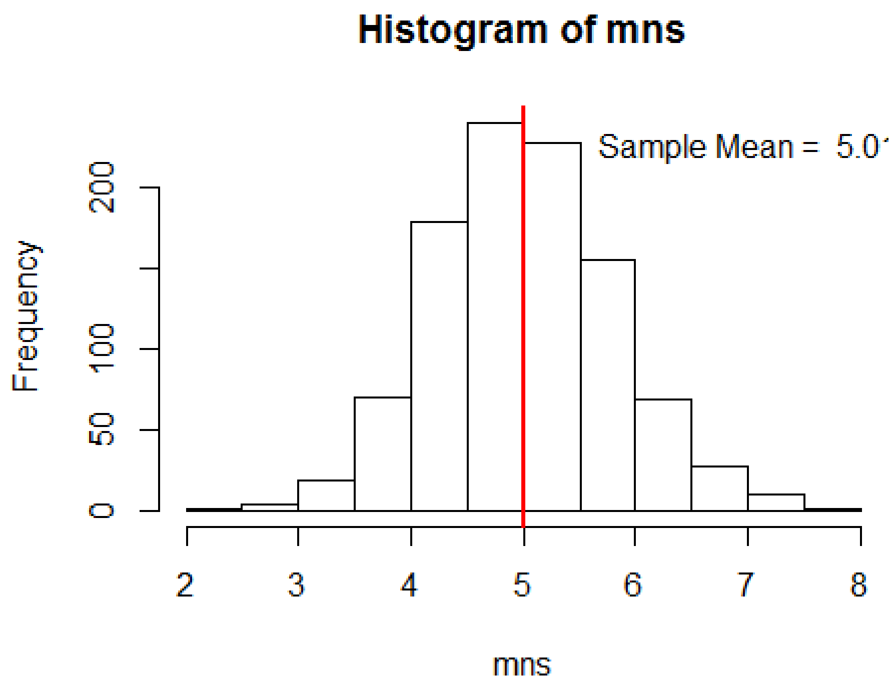
Comparison of Simulated and Theoretical Variance

The standard deviation of the distribution of simulated means, 0.8, indicates the variability of the 1,000 simulated means around the simulated mean. The standard deviation of the simulated mean of means cannot be directly compared to the theoretical standard deviation (lambda = 0.2) of 5.

Approximate Normality of Simulated Distribution of Means

The approximate normality of the simulated distribution of means is evidenced first by observing the shape of the distribution:

```
hist(mns)
abline(v=simMean, col="red", lwd=2)
text(simMean+2,225,paste("Sample Mean = ",round(simMean,2)))
```



We can also look at the density on either side of the mean. If the simulated distribution is normal, thenumber of values to the left and to the right of the mean will be roughly equal.

```
r <- NULL
for(n in c(40,100,1000)) {
  mns <- meanDistExp(1000,n)
  simMean <- mean(mns)
  diffMean <- mns-simMean
  r <- rbind(r, c(n,lessThanZeroPercent <- length(subset(diffMean, diffMean <
0))/length(diffMean), length(subset(diffMean, diffMean >
0))/length(diffMean)))
}
r

##      [,1]  [,2]  [,3]
## [1,]   40 0.523 0.477
## [2,]  100 0.519 0.481
## [3,] 1000 0.503 0.497
```

The table above contains sample size for each calculation of mean, percent of values falling below the mean, and percentage of values falling above the mean, respectively. At the sample size used in the discussion so far, the left side and right side are already approximately equal.

As we increase the sample size, the percentage of values on either side of the mean move closer to 50%.