Course Project Part1

Casey645

March 6, 2016

# Part 1-Simulation Exercise

## Simulations

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 .

Lets start by doing a thousand simulated averages of 40 exponential numbers.

#Set seed  
set.seed(3)  
lambda <- 0.2  
#We perform 1000 simulations with 40 samples   
sample\_size <- 40  
simulations <- 1000  
  
#Lets do 1000 simulations  
simulated\_exponentials <- matrix(rexp(simulations\*sample\_size, rate=lambda), simulations, sample\_size)  
#Averages of 40 exponentials  
row\_means <- rowMeans(simulated\_exponentials)

## Results

### 1.Show where the discribution is centered, and compare to the theotretical center of distribution.

#mean of distribution of averages of 40 exponentials  
mean(row\_means)

##[1] 4.98662

#mean from analytical expression  
1/lambda

##[1] 5

The distribution of sample means - shown below:



Conclusion - the distribution of averages of 40 exponentials is centered at r mean(row\_means) and the same is close to the theoretical center of the distribution, which is = r 1/lambda.

### 2.Show how variable it is and compare it to the theoretical variance of the distribution.

#standard deviation of distribution of averages of 40 exponentials  
sd(row\_means)

##[1] 0.791

#standard deviation from analytical expression  
(1/lambda)/sqrt(sample\_size)

##[1] 0.7906

#Variance of the sample mean  
var(row\_means)

## [1] 0.6258

#Theoritcal variance of the distribution  
1/((0.2\*0.2) \* 40

##[1] 0.625

Therefore, the variability in distribution of averages of 40 exponentials is close to the theoretical variance of the distribution. The variance of sample means is r var(row\_means) where as the theoretical variance of the distribution is = r 1/(0.04 \* 40).

### 3.Show that the distribution is approximately normal.



Due to the central limit theorem, the averages of samples follow normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. Also, the q-q plot suggests the distribution of averages of 40 exponentials is very close to a normal distribution.

### 4.Evaluate the coverage of the confidence interval for .

#Evaluate coverage of the confidence interval  
lambda\_vals <- seq(4, 6, by=0.01)  
coverage <- sapply(lambda\_vals, function(lamb) {  
 mu\_hats <- rowMeans(matrix(rexp(sample\_size\*simulations, rate=0.2),  
 simulations, sample\_size))  
 ll <- mu\_hats - qnorm(0.975) \* sqrt(1/lambda\*\*2/sample\_size)  
 ul <- mu\_hats + qnorm(0.975) \* sqrt(1/lambda\*\*2/sample\_size)  
 mean(ll < lamb & ul > lamb)  
})



The 95% confidence intervals for the rate parameter () to be estimated () are and . As can be seen from the plot above, for selection of around 5, the average of the sample mean falls within the confidence interval at least 95% of the time. Note that the true rate, is 5.

Since, we consider the distribution of averages of exponentials, the standard deviation of this distribution already incorporates the () term i.e. it is the standard error.

The confidence interval is given by r mean(row\_means) + c(-1, 1) \* 1.96 \* sd(row\_means)