Simulation

Bohdan Horak

Saturday, August 23, 2014

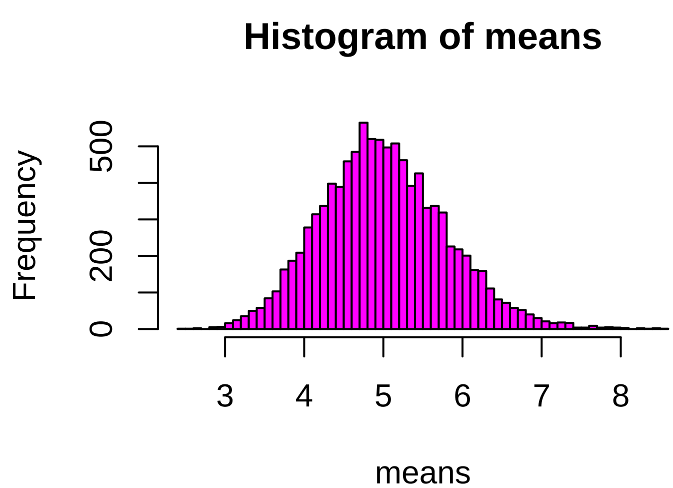
Summary

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should ## Question 1

1.Show where the distribution is centered at and compare it to the theoretical center of the distribution.

```
nosim <- 10000
n <- 40
set.seed(1)
means<-apply(matrix( rexp(nosim * n, rate=0.2), nosim), 1, mean)
mean_val<-mean(means)
sds<-sd(means)
theo_sd<-5/sqrt(40)
hist(means,breaks=50,col="magenta")
```

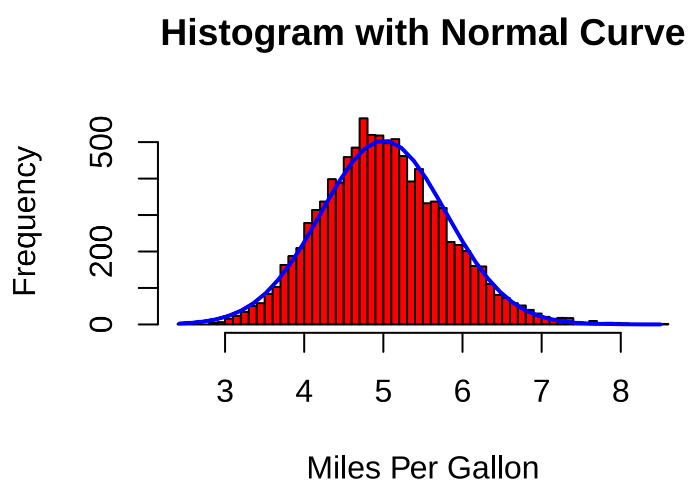


From the resuling simulation the distribution has a mean 5.0029 which is very close to the theoretical mean of 1/lambda=1/0.2=5.

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

The statndard deviation of the simulation is 0.7913 which is also close to the theoretical standard deviation of SD=s/sqrt(n)=5/sqrt(40)=0.7906.

3. Show that the distribution is approximately normal.



This is the same histogram as before but overlayed with a normal curve. Clearly is approximatly normal, even with a small bin size. This is a very good example of the central limit theorem which confirms that the distribution of the sample means is approximetly normal as the sample size gets larger.

4. Evaluate the coverage of the confidence interval for 1/lambda

```
nosim<-10000
count<-0
for (i in 1:nosim) {
    temp<-rexp(n,rate=0.2)
    x_bar<-mean(temp)
    sd_bar<-sd(temp)
    ci_l<-x_bar-1.96*sd_bar/sqrt(40)
    ci_u<-x_bar+1.96*sd_bar/sqrt(40)
    if (ci_l<5 & 5<ci_u) {
      count<-count+1
    }
}
coverage<-100*count/nosim
```

Simulated coverage for a sample size of 40 for the confidence interval is about 92.13 %.Note that the intervals provided in the assignment are biased downwards, that's why the coverage is relatively low.

For the more complete .rmd code please visit my Github repo