GEOG606 - HW1

Anika Cartas

Saturday, February 14, 2015

#### 1. Solve the following linear equations:

-x + 2y = 1

3x + y = 3

left = matrix(c(-1,3,2,1,1,3),2,2); left

## [,1] [,2]  
## [1,] -1 2  
## [2,] 3 1

right = c(1,3); right

## [1] 1 3

solve(left,right)

## [1] 0.7142857 0.8571429

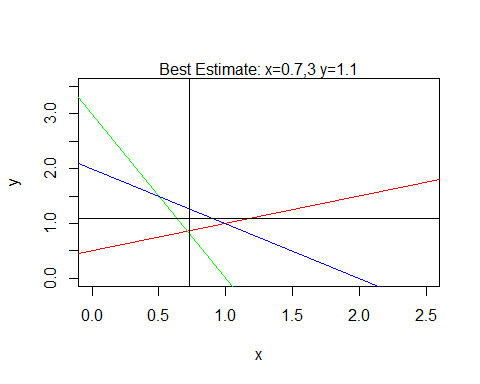
If we add another equation, x + y = 2, is there a solution for the set of 3 equations?

* No. We can only solve for a square matrix.

If not, could you give a best estimate of x and y that satisfy the equations?

* To find a best estimate, plot all three lines and find the centerpoint of the triangle their intersections create.

#Plot all three lines  
plot(NULL,xlim=c(0,2.5),ylim=c(0,3.5), xlab="x",ylab="y")  
abline(1/2,1/2, col="red") #first first equation  
abline(3,-3,col="green") #second equation  
abline(2,-1,col="blue") #third equation  
abline(h=1.1)  
abline(v=0.73)  
mtext("Best Estimate: x=0.7,3 y=1.1")



#### 2. Calculate the covariance matrix of the following 3\*2 data matrix:

First, calculate by hand:

$\frac{1}{n} `sum\_{i=i}^{n} x\_{i}$

x = matrix(c(1,2,3,3,2,1),3,2); x

## [,1] [,2]  
## [1,] 1 3  
## [2,] 2 2  
## [3,] 3 1

cov(x)

## [,1] [,2]  
## [1,] 1 -1  
## [2,] -1 1

What does the covariance tell you about the data?

* The covariance matrix shows the variances of the two variables (x,y) along the diagonal, and the covariances between them in the other positions.
* In this case, the variance of both x and y is 1, while their covariance is -1, meaning that as x grows y shrinks, and vice versa.

#### 3.Use the following dataset to demonstrate the Central Tendency Theorem:

data = c(3,2,3,4,3,2,5,4,5,2,3,3,10,4,5,4,12,15,10,17,15,18,16,14,9,10,11,18,17,18,16,17,18)

3.1. Randomly select 5 values and calculate their mean.

s1 = sample(data,5); s1

## [1] 15 2 3 10 14

mean(s1)

## [1] 8.8

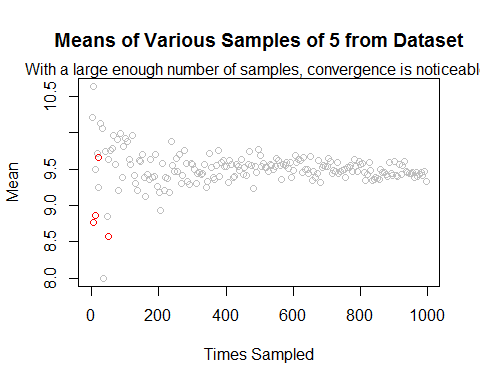
Repeat 3 times, and calculate the mean of the sampling mean. Then, repeat 5, 10, 20, and 50 times.

takeSamples <- function(data,sizes, sampleSize){  
 allMeans = matrix(NA,1,2)  
 for (n in sizes){  
 curMeans = vector()  
 for (i in c(1:n)){  
 curMeans = c(curMeans,mean(sample(data,sampleSize)))  
 }  
 allMeans = rbind(allMeans,c(n,mean(curMeans)))  
 }  
 allMeans = allMeans[-1,]  
 colnames(allMeans) = c("Times.Sampled","Mean")   
 allMeans #return value  
}  
  
means.5 = takeSamples(data,c(3,5,10,20,50),sampleSize=5); means.5

## Times.Sampled Mean  
## [1,] 3 10.93333  
## [2,] 5 8.76000  
## [3,] 10 8.86000  
## [4,] 20 9.66000  
## [5,] 50 8.56800

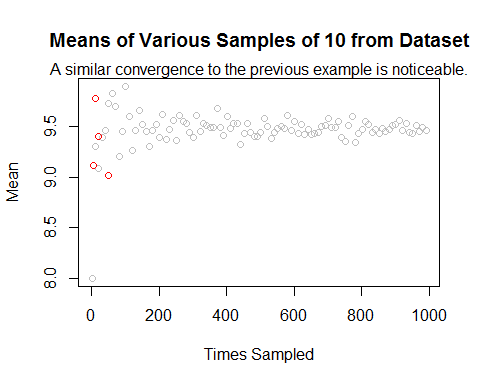
Use a plot to show how the mean of the sampling mean changes as the number of repeats increases.

#Out of curiosity, plotting larger sizes of samples (first, so that axes match up)  
means.5.1000 = takeSamples(data,seq(1,1000,5),sampleSize=5)  
plot(means.5.1000, main="Means of Various Samples of 5 from Dataset", col="grey", xlab="Times Sampled")  
  
#Request from assignment in red  
points(means.5,col="red")  
  
mtext("With a large enough number of samples, convergence is noticeable.")



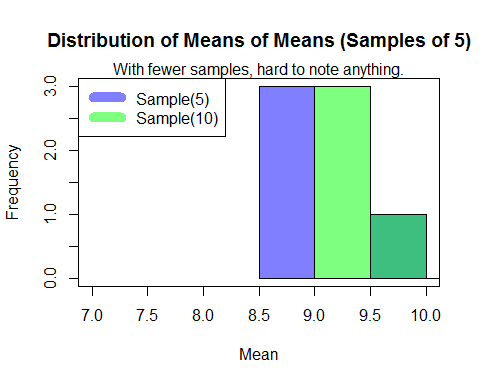
3.2. Repeat 3.1. but select 10 values during each random selection. Explain these plots in relation to the Central Tendency Theorem.

means.10.1000 = takeSamples(data,seq(1,1000,10),sampleSize=10)  
  
plot(means.10.1000, main="Means of Various Samples of 10 from Dataset", col="grey", xlab="Times Sampled")  
  
means.10 = takeSamples(data,c(3,5,10,20,50),sampleSize=10)  
points(means.10, col="red")  
  
mtext("A similar convergence to the previous example is noticeable.")



3.3. How does the sampling mean approach normal distribution in these experiments? Use histograms to show.

hist(means.5[,2], col=rgb(0,0,1,0.5), xlim=c(7,10),main="Distribution of Means of Means (Samples of 5)", xlab="Mean")  
  
hist(means.10[,2], col=rgb(0,1,0,0.5), main="Distribution of Mean of Means (Samples of 10)", xlab="Mean", add=T)  
  
legend("topleft", c("Sample(5)","Sample(10)"),col=c(rgb(0,0,1,0.5),rgb(0,1,0,0.5)), lwd=10)  
box() #decorative box around data  
mtext("With fewer samples, hard to note anything.")



hist(means.5.1000[,2], col=rgb(0,0,1,0.5), main="Distribution of Means of Means (Samples of 5)", xlab="Mean")  
  
hist(means.10.1000[,2], col=rgb(0,1,0,0.5), main="Distribution of Mean of Means (Samples of 10)", xlab="Mean", add=T)  
  
legend("topright", c("Sample(5)","Sample(10)"),col=c(rgb(0,0,1,0.5),rgb(0,1,0,0.5)), lwd=10)  
box() #decorative box around data  
mtext("With more samples, a normal curve with median 9.5 is noticeable.")

