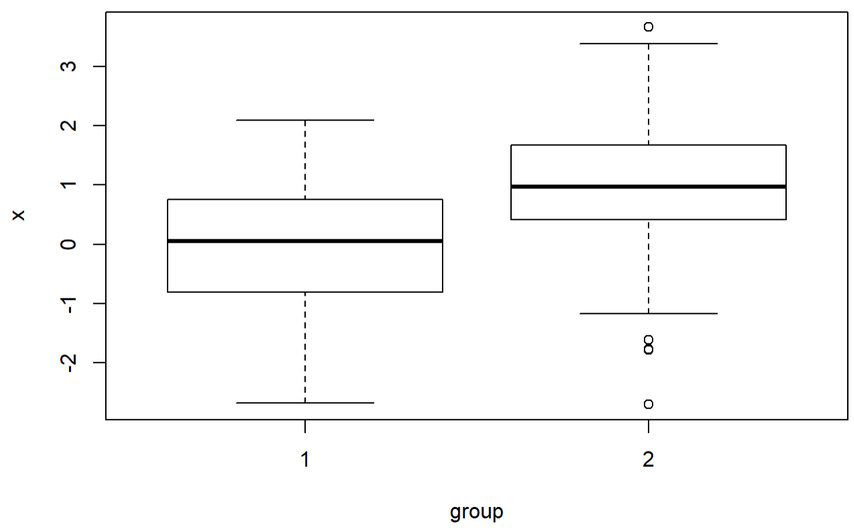
hw12

Enguang Fan

5/10/2020

1

```
set.seed(1288)
group=sample(c(1,2),400,replace=TRUE,prob=c(.25,.75))
rnorm1=rnorm(400,0,1)
rnorm2=rnorm(400,1,1)
x=rnorm1*(group==1)+rnorm2*(group==2)
boxplot(x~group)
```



Given data x and starting values of π = .1 and μ = .5, run 40 iterations of the EM-algorithm.

Now let us implement the EM algorithm

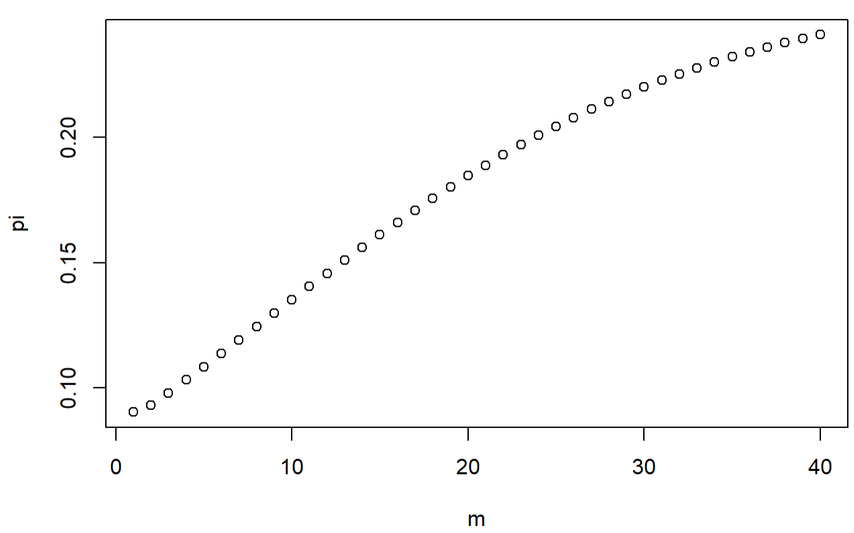
x is the all observation value

```
p1 = 0.1
m1 = 0.5
n=length(x)
resultmtx = matrix(0,40,2)
```

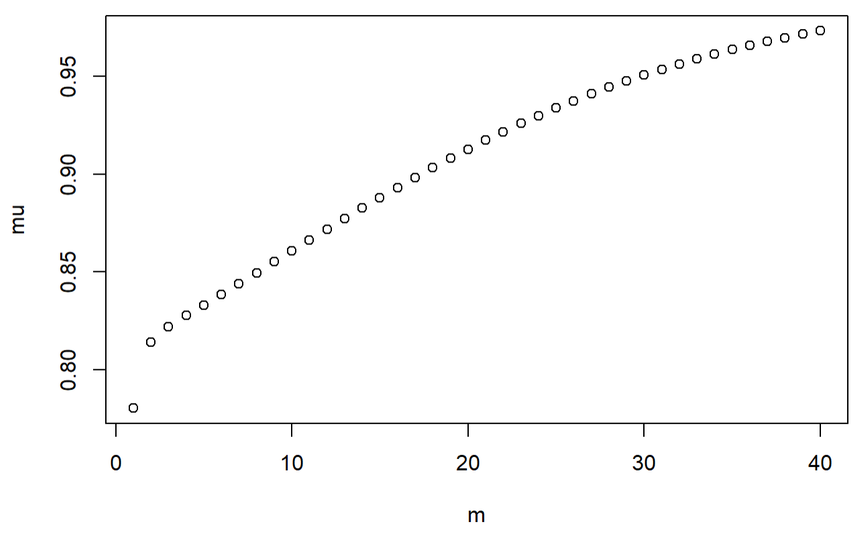
```
#two w vector to store lists of weights for 2 catagory of distribution
w1 = rep(0,400)
w2 = rep(0,400)
for (i in 1:40) {
 for (j in 1:400) {
   w1[j] = p1*dnorm(x[j]) / ( p1*dnorm(x[j]) + (1-p1) * dnorm(x[j],m1) )
   w2[j] = (1-p1) * dnorm(x[j],m1) / ( p1*dnorm(x[j]) + (1-p1) * dnorm(x[j],m1) )
 }
  #n1: how many distribution 1 out of 400 observations
  #n2: how many distribution 2 out of 400 observations
  n1 = sum(w1)
  n2 = sum(w2)
  p1 = n1/(n1+n2)
  m1 = sum(w2*x) / n2
  #first column for p, second for m
  resultmtx[i,1] = p1
  resultmtx[i,2] = m1


}
```

```
colnames(resultmtx) = c("pi","u")
#resultmtx
```

```
m=1:40
plot(m,resultmtx[,1],xlab = "m",ylab = "pi")
```




```
plot(m,resultmtx[,2],xlab = "m",ylab = "mu")
```



2

```
resultmtx[40,]
```

```
##        pi         u
## 0.2410009 0.9731355
```

So π_{40} = 0.2611534 and μ_{40} = 0.9952051 which are close enough to real value