BIS 557 Final Project Problem 1

```
setwd("/Users/kongwenjing/Desktop")
misReport = read.csv("misreported_counts.csv")

# Initial State
alpha = 1
beta = 1
p = 1/3
q = 1/3
r = 1/3
diff = 1
n = dim(misReport)[1]
a1 = matrix(0,n,10)
a2 = matrix(0,n,10)
b1 = matrix(0,n,10)
c1 = rep(0,n)
c2 = rep(0,n)
c3 = rep(1,n)
iter = 0

while(diff>1e-6){
  iter = iter+1
  lambda = exp(alpha+beta*misReport$z)
  k = seq(0,9,1)
  for(i in 1:n){
    # set the numerators and denominaters for class 1
    numA1 = p*dpois(k+misReport$y[i],lambda[i])
    numB1 = q*dpois(misReport$y[i]-k-1,lambda[i])
    den1 = sum(numA1+numB1)+r*dpois(misReport$y[i],lambda[i])
    # update the value
    a1[i,] = numA1/den1
    b1[i,] = numB1/den1
    c1[i] = r*dpois(misReport$y[i],lambda[i])/den1
    # set the numerators and denominaters for class 2
    numA2 = p*dpois(k,lambda[i])
    den2 = sum(p*dpois(k,lambda[i]))+r*exp(-lambda[i])
    # update value
    a2[i,] = numA2/den2
    c2[i] = r*exp(-lambda[i])/den2
  }
  # indicator functions
  I1 = (misReport$y%%10==0)*(misReport$y>0)
  I2 = (misReport$y==0)*1
  I3 = (misReport$y%%10!=0)*1
  # update class probality
  p_new = t(rowSums(a1))%*%I1 + t(rowSums(a2))%*%I2
  q_new = t(rowSums(b1))%*%I1
  r_new = t(c1)%*%I1+t(c2)%*%I2+t(c3)%*%I3
  s = p_new+q_new+r_new
  p_new = p_new/s
  q_new = q_new/s
```

```
  r_new = r_new/s

  # calculate H1 and G1
  H1 = t(-I1*(c1+rowSums(a1)+rowSums(b1))-I2*(c2+rowSums(a2))-I3*c3)%*%lambda
  # 1st, 4th and 6th term in g1
  g1 = t(I1*c1)%*%(misReport$y-lambda)+t(I2*c2)%*%(-lambda)+t(I3*c3)%*%(misReport$y-lambda)
  # 2nd, 3rd, 5th term in g1
  kmat1 = matrix(rep(k,n),n,10,byrow=T)+misReport$y-lambda
  kmat2 = -matrix(rep(k+1,n),n,10,byrow=T)+misReport$y-lambda
  kmat3 = matrix(rep(k,n),n,10,byrow=T)-lambda
  temp1 = vector()
  temp2 = vector()
  temp3 = vector()
  for(i in 1:n){
    temp1[i] = t(a1[i,])%*%kmat1[i,]
    temp2[i] = t(b1[i,])%*%kmat2[i,]
    temp3[i] = t(a2[i,])%*%kmat3[i,]
  }
  # integrate
  g1 = g1 + t(I1)%*%(temp1+temp2)+t(I2)%*%temp3

  # calculate H2 and G2
  H2 = t(-I1*(c1+rowSums(a1)+rowSums(b1))-I2*(c2+rowSums(a2))-I3*c3)%*%(misReport$z^2*lambda)
  # 1st, 4th and 6th term in g2
  g2 = t(I1*c1)%*%(misReport$y*misReport$z-misReport$z*lambda)
  g2 = g2+ t(I2*c2)%*%(-misReport$z*lambda)+t(I3*c3)%*%(misReport$y*misReport$z-misReport$z*lambda)
  # 2nd, 3rd, 5th term in g2
  kzmat1 = (matrix(rep(k,n),n,10,byrow=T)+misReport$y-lambda)*misReport$z
  kzmat2 = (-matrix(rep(k+1,n),n,10,byrow=T)+misReport$y-lambda)*misReport$z
  kzmat3 = (matrix(rep(k,n),n,10,byrow=T)-lambda)*misReport$z
  tempz1 = vector()
  tempz2 = vector()
  tempz3 = vector()
  for(i in 1:n){
    tempz1[i] = a1[i,]%*%kzmat1[i,]
    tempz2[i] = b1[i,]%*%kzmat2[i,]
    tempz3[i] = a2[i,]%*%kzmat3[i,]
  }
  g2 = g2 + t(I1)%*%(tempz1+tempz2)+t(I2)%*%tempz3

  alpha_new = alpha-1/H1*g1
  beta_new = beta-1/H2*g2

  diff = sum((c(alpha_new,beta_new,p_new,q_new,r_new)-c(alpha,beta,p,q,r))^2)
  alpha = alpha_new
  beta = beta_new
  p = p_new
  q = q_new
  r = r_new
}
c(alpha,beta,p,q,r)
```

```
## [1]  1.6985680 -1.2077996  0.3876160  0.1054451  0.5069389
```

```
###calculate predicted value for x
x_pred1=(max.col(cbind(b1,(cbind(c1,matrix(0,nrow=n,ncol=9))+a1)))+misReport$y-9)*I1
x_pred2=(max.col(cbind(c2,matrix(0,nrow=n,ncol=9))+a2)-1)*I2
x_pred3=misReport$y*I3
x_pred=x_pred1+x_pred2+x_pred3

library(ggplot2)
dat=data.frame(PredictedX=x_pred,
               ObservedY=misReport$y,
               Case=as.character(1*I1+2*I2+3*I3))
ggplot(dat, aes(x=PredictedX, y=ObservedY, color=Case)) +
  geom_point(shape=1)+ ggtitle("Compare Xi and Yi in Scatterplot")
```