Math 475: Final Exam

Hannah Zmuda

01/08/2021

Problem 1

part a

Show the outputs of the EM algorithm are consistent with the given parameter equations To find the updated parameters (i.e. the maximized value) we fist need to find the Q-function (E step) then maximize the Q function by taking the derivative in regard to each of the parameters (M step). **E step:** Given the likelihood equation, we can work out the log likelihood to be:

$$log[L(\theta|n_{k,i})] = \Sigma_{i=0}^{16} z_0 log(\alpha 1_{i=0}) +$$
$$(t_i)[log(\alpha 1_{i=0}) + log(\beta\mu^i e^{-\mu}) + log((1-\alpha-\beta)\lambda^i e^{-\lambda})] +$$
$$(p_i)[log(\alpha 1_{i=0}) + log(\beta\mu^i e^{-\mu}) + log((1-\alpha-\beta)\lambda^i e^{-\lambda})] - log(i!)$$

where y is the complete data set and z_0, t_i, p_i represent three different groups. These are further broken down into the zero, typical, and promiscuous groups. To find your Q-function, take the expectation of the log likelihood function:

$$Q(\theta|\theta^{(t)}) = \Sigma_{i=0}^{16}[\frac{n_0 z_0 log(\alpha 1_{i=0})}{N} +$$
$$\frac{n_i(t_i)[log(\beta\mu^i e^{-\mu}) + log((1-\alpha-\beta)\lambda^i e^{-\lambda})]}{N} +$$
$$\frac{n_i(p_i)[log(\beta\mu^i e^{-\mu}) + log((1-\alpha-\beta)\lambda^i e^{-\lambda})]}{N} - log(i!)]$$

M step: For the M step of the EM algorithm, we need to maximize the Q function in regard to each parameter then set it equal to zero.

$$\frac{dQ(\theta|\theta^{(t)})}{d\alpha^{(t)}} = \frac{n_0 z_0(\theta^{(t)})}{N}$$

When the derivative is set equal to zero, we find that the updated parameters equal to what we expected:

$$\alpha^{(t+1)} = \frac{n_0 z_0 \theta^t}{N}$$

$$\beta^{(t+1)} = \Sigma_{i=0}^{16} \frac{n_i t_i(\theta^{(t)})}{N}$$

$$\mu^{(t+1)} = \frac{\Sigma_{i=0}^{16} i n_i t_i(\theta^{(t)})}{\Sigma_{i=0}^{16} n_i t_i(\theta^{(t)})}$$

$$\lambda^{(t+1)} = \frac{\Sigma_{i=0}^{16} i n_i p_i(\theta^{(t)})}{\Sigma_{i=0}^{16} n_i p_i(\theta^{(t)})}$$

part b and c

```r
set.seed(475)
#initialize variables
data = data.frame(enc=0:16,
               freq=c(379,299,222,145,109,95,73,59,45,30,24,12,4,2,0,1,1))
N = sum(data$freq)
alpha = 0.5
beta = 0.8
mu = 2
lambda = 15
param = c(alpha,beta,mu,lambda)
tol = 1e-10
tol.cur = 100
time = 0
i = 0:16

#EM Algorithm
while(tol.cur > tol){
  pi = (beta*exp(-mu)*mu^i) + ((1-alpha-beta)*exp(-lambda)*lambda^i)
  pi[1] = pi[1] + alpha
  z.stat = alpha/(pi[1])
  t.stat = (beta*(mu^i)*exp(-mu))/pi
  p.stat = ((1-alpha-beta)*exp(-lambda)*lambda^i)/pi
  alpha = (data$freq[1]*z.stat)/N
  beta = sum(data$freq*t.stat)/N
  mu = sum(i*data$freq*t.stat)/sum(data$freq*t.stat)
  lambda = sum(i*data$freq*p.stat)/sum(data$freq*p.stat)
  new.param = c(alpha,beta,mu,lambda)
  tol.cur = sum(abs(new.param-param))
  param = new.param
  time = time + 1
}

#standard error
#Use log likelihood at theta.hat values (i.e. new parameter values)

#pairwise correlation
cor(x = param, y = param, use = "pairwise.complete.obs")
```

```
## [1] 1
```

Problem 2

part a: Metropolish-Hastings Algorithm

```
set.seed(575)
#use a normal distribution as a starting distribution
```

Problem 3

Problem 4

Problem 5