2'5/12

Exam Part 2

[redacted]

Monday, June 08, 2015

We have

$$Y_i = \begin{cases} 0 \text{ if } -1 \leq X_i \leq 1 \\ X_i \text{ if } |X_i| > 1 \end{cases}$$

$X_1, ..., X_n$ are the complete case and $Y_1, ..., Y_n$ the observed data.

0

1.

If $y_i = 0$ the contribution of the i-th experiment result is

?

1

2.

We use this notation:

- $\phi(z)$ $\Phi(z)$ the density function and the distribution functions of a *N(0,1)*, respectively.
- $\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$ the density function of a *N(μ,σ^2)*
- $\Phi(\frac{x-\mu}{\sigma})$ the distribution function of a *N(μ,σ^2)*

For the complete case we have the likelihood function such as

$\Pi \frac{1}{\sigma}\phi(\frac{x_i-\mu}{\sigma})$

We can consider the log-likelihood for the complete case:

$l^c(\mu, \sigma) = -n(ln(\sigma) + ln(\sqrt{2\pi})) - \frac{1}{2\sigma^2}\sum(x_i - \mu)^2$

For the observed case

$l^o(\mu, \sigma) = \cancel{-s(ln(\sigma) + ln(\sqrt{2\pi})) - \frac{1}{2\sigma^2}\sum(y_i - \mu)^2}$

1'5

3.

$f_{X,\cancel{Y}}(X, \cancel{Y}, \mu, \sigma)$ is the probability density of the complete data.

$Q(\mu, \sigma|\mu_m, \sigma_m) = E_{\mu_m,\sigma_m}(log f_{X,\cancel{Y}}(X, \cancel{Y}, \mu, \sigma)| \cancel{-1 \leq X \leq 1})$ $Y_1 = y_1, ? \neq 0, Y_n = y_n$

With the indication about the variance of a truncated normal distribution we have that

$E_{\mu,\sigma}((X - \tau)^2| - 1 \leq X \leq 1) = Var_{\mu,\sigma}(X| - 1 \leq X \leq 1) + (E_{\mu,\sigma}(X| - 1 \leq X \leq 1) - \tau)^2$

We also have

$$f_{X|-1\leq X\leq 1} = \begin{cases} \frac{f(x)}{F(b)-F(a)} \text{ if } -1 \leq X \leq 1 \\ 0 \text{ otherwise} \end{cases}$$

So now we have to write $E_{\mu_m,\sigma_m}(log f_{X,Y}(X, Y, \mu, \sigma)| -1 \leq X \leq 1)$ in the form of $E_{\mu,\sigma}((X - \tau)^2| -1 \leq X \leq 1)$

We have from **2.** $log f_{X,Y}(X, Y, \mu, \sigma) = -n(ln(\sigma) + ln(\sqrt{2\pi})) - \frac{1}{2\sigma^2}\sum(x_i - \mu)^2$

So

$$E_{\mu,\sigma}((X-\mu)^2|-1 \leq X \leq 1) = Var_{\mu,\sigma}(X|-1 \leq X \leq 1) + (E_{\mu,\sigma}(X|-1 \leq X \leq 1) - \mu)^2$$

With the expression of those terms in the subject,

$$Q(\mu,\sigma|\mu_m,\sigma_m) = n(ln(\sigma) + ln(\sqrt{2\pi})) - \frac{1}{2\sigma^2}\sum \sigma^2\left[1 + \frac{\frac{-1-\mu}{\sigma}\phi(\frac{-1-\mu}{\sigma}) - \frac{1-\mu}{\sigma}\phi(\frac{1-\mu}{\sigma})}{\Phi(\frac{1-\mu}{\sigma}) - \Phi(\frac{-1-\mu}{\sigma})} - \left(\frac{\phi(\frac{-1-\mu}{\sigma}) - \phi(\frac{1-\mu}{\sigma})}{\Phi(\frac{1-\mu}{\sigma}) - \Phi(\frac{-1-\mu}{\sigma})}\right)^2\right] + \left(\frac{\phi(\frac{-1-\mu}{\sigma}) - \phi(\frac{1-\mu}{\sigma})}{\Phi(\frac{1-\mu}{\sigma}) - \Phi(\frac{-1-\mu}{\sigma})}\sigma\right)^2$$

$$\cancel{Q(\mu,\sigma|\mu_m,\sigma_m) = n(ln(\sigma) + ln(\sqrt{2\pi})) - \frac{1}{2\sigma^2}\sum \sigma^2\left[1 + \frac{\frac{-1-\mu}{\sigma}\phi(\frac{-1-\mu}{\sigma}) - \frac{1-\mu}{\sigma}\phi(\frac{1-\mu}{\sigma})}{\Phi(\frac{1-\mu}{\sigma}) - \Phi(\frac{-1-\mu}{\sigma})}\right]}$$

✓ You have to differentied two cases:

$$\begin{cases} y_i = 0 \\ y_i \neq 0 \end{cases}$$

0/

4.

0/

5.

```r
set.seed(080615)


#data


EM<-function(data,max.iter){

eps.mu <- 1e-10
eps.sigma <- 1e-10
eps.p <- 1e-10
eps.ll <- 1e-10
x<-data
n<- length(x)
p.0 <- c(1/2,1/2)
mu.0 <- sample(x,2)
sigma.0 <- rep(sd(x),2)
k<-2
iter <- 0

ll.EY.0 <- sapply(1:k,function(j,x,mu,sigma,p){
  p[j]*dnorm(x,mean=mu[j],sd=sigma[j])
},x,mu.0,sigma.0,p.0)
ll.0 <- log(prod( apply(ll.EY.0,1,sum)))
stop.criteria <- FALSE
print(" Iter lambda")
## [1] " Iter lambda"
print("  Iter   log.lik    p[1]     p[2]     mu[1]    mu[2]    sigma[1] sigma[2]")

while (!stop.criteria){
  print( c( iter, ll.0, p.0, mu.0, sigma.0), digits=3)
  iter <- iter+1
  # E step: Estimation
  k <- length(p.0)
  E.Y <- ll.EY.0 / apply(ll.EY.0,1,sum)
  # M step: Maximization
```

This code corresponds to a different

problem !

```
  p.1 <- apply(E.Y,2,mean)
  mu.1 <- sapply(1:k,function(j,x,E.Y){
    sum(x*E.Y[,j])/sum(E.Y[,j])
  },x,E.Y)
  sigma.1 <- sapply(1:k,function(j,x,E.Y,mu){
    n <- length(x)
    x.c <- x-mu[j]
    sqrt( sum(E.Y[,j] * x.c^2)/sum(E.Y[,j]))
  },x,E.Y,mu.1)
  # Computing the log-likelihood value
  ll.EY.1 <- sapply(1:k,function(j,x,mu,sigma,p){
    p[j]*dnorm(x,mean=mu[j],sd=sigma[j])
  },x,mu.1,sigma.1,p.1)
  ll.1 <- log(prod( apply(ll.EY.1,1,sum)))
  # checking stopping criteria
  stop.criteria <- (
    (iter==max.iter) |
      (sum((mu.1-mu.0)^2)/length(mu.1) <=  eps.mu) |
      (sum((sigma.1-sigma.0)^2)/length(sigma.1) <=  eps.sigma) |
      (sum((p.1-p.0)^2)/length(p.1) <=  eps.p) |
      ((ll.1-ll.0)^2 <= eps.ll)
  )
  if (!stop.criteria){
    mu.0 <- mu.1
    sigma.0 <- sigma.1
    p.0 <- p.1
    ll.EY.0 <- ll.EY.1
    ll.0 <- ll.1
  }
}

print( c( iter, ll.1, p.1, mu.1, sigma.1), digits=3 )

}
```

Here are the results of the EM algorithm applied on results.txt

```
y <- read.table("results.txt",col.names=FALSE)[,1]
EM(y,50)
```

```
## [1] " Iter lambda"
## [1] "   Iter   log.lik    p[1]     p[2]     mu[1]    mu[2]    sigma[1] sigma[2]"
## [1]    0.00 -230.19    0.50    0.50    3.24   -1.37    1.98    1.98
## [1]    1.000 -205.889    0.364    0.636    1.932   -0.700    1.780    1.320
## [1]    2.000 -205.701    0.355    0.645    1.952   -0.672    1.855    1.287
## [1]    3.000 -205.667    0.349    0.651    1.961   -0.655    1.884    1.283
## [1]    4.000 -205.653    0.345    0.655    1.968   -0.643    1.898    1.285
## [1]    5.000 -205.644    0.342    0.658    1.974   -0.634    1.905    1.288
## [1]    6.000 -205.637    0.339    0.661    1.982   -0.627    1.908    1.292
## [1]    7.000 -205.632    0.337    0.663    1.991   -0.622    1.910    1.295
## [1]    8.000 -205.627    0.335    0.665    2.000   -0.617    1.910    1.297
## [1]    9.000 -205.623    0.332    0.668    2.011   -0.614    1.908    1.299
## [1]   10.000 -205.619    0.330    0.670    2.022   -0.611    1.905    1.301
```

```
## [1]    11.000 -205.616     0.328     0.672     2.033    -0.608     1.902     1.303
## [1]    12.000 -205.612     0.326     0.674     2.045    -0.606     1.898     1.304
## [1]    13.000 -205.608     0.324     0.676     2.058    -0.604     1.894     1.305
## [1]    14.000 -205.603     0.322     0.678     2.071    -0.602     1.889     1.306
## [1]    15.000 -205.599     0.320     0.680     2.085    -0.601     1.883     1.307
## [1]    16.000 -205.595     0.318     0.682     2.099    -0.599     1.878     1.308
## [1]    17.000 -205.590     0.316     0.684     2.113    -0.598     1.872     1.309
## [1]    18.000 -205.585     0.314     0.686     2.128    -0.596     1.865     1.309
## [1]    19.000 -205.581     0.312     0.688     2.143    -0.595     1.859     1.310
## [1]    20.000 -205.575     0.310     0.690     2.158    -0.593     1.852     1.311
## [1]    21.000 -205.570     0.307     0.693     2.173    -0.592     1.845     1.311
## [1]    22.000 -205.565     0.305     0.695     2.189    -0.590     1.838     1.312
## [1]    23.000 -205.559     0.303     0.697     2.205    -0.588     1.831     1.312
## [1]    24.000 -205.553     0.301     0.699     2.222    -0.587     1.824     1.313
## [1]    25.000 -205.547     0.299     0.701     2.238    -0.585     1.816     1.314
## [1]    26.000 -205.541     0.297     0.703     2.255    -0.583     1.809     1.314
## [1]    27.000 -205.535     0.294     0.706     2.272    -0.581     1.801     1.315
## [1]    28.000 -205.528     0.292     0.708     2.290    -0.579     1.793     1.315
## [1]    29.000 -205.521     0.290     0.710     2.308    -0.577     1.785     1.316
## [1]    30.000 -205.515     0.287     0.713     2.325    -0.575     1.777     1.316
## [1]    31.000 -205.508     0.285     0.715     2.343    -0.573     1.768     1.317
## [1]    32.000 -205.500     0.283     0.717     2.362    -0.571     1.760     1.317
## [1]    33.000 -205.493     0.281     0.719     2.380    -0.569     1.751     1.318
## [1]    34.000 -205.485     0.278     0.722     2.399    -0.566     1.743     1.319
## [1]    35.000 -205.478     0.276     0.724     2.418    -0.564     1.734     1.319
## [1]    36.000 -205.470     0.273     0.727     2.437    -0.561     1.725     1.320
## [1]    37.000 -205.462     0.271     0.729     2.456    -0.559     1.716     1.321
## [1]    38.000 -205.454     0.269     0.731     2.476    -0.556     1.707     1.321
## [1]    39.000 -205.445     0.266     0.734     2.495    -0.553     1.698     1.322
## [1]    40.000 -205.437     0.264     0.736     2.515    -0.550     1.689     1.323
## [1]    41.000 -205.428     0.261     0.739     2.535    -0.547     1.679     1.323
## [1]    42.000 -205.419     0.259     0.741     2.555    -0.544     1.670     1.324
## [1]    43.000 -205.410     0.257     0.743     2.575    -0.541     1.660     1.325
## [1]    44.000 -205.401     0.254     0.746     2.595    -0.538     1.650     1.326
## [1]    45.000 -205.392     0.252     0.748     2.616    -0.535     1.640     1.327
## [1]    46.000 -205.383     0.249     0.751     2.636    -0.532     1.630     1.328
## [1]    47.000 -205.373     0.247     0.753     2.657    -0.529     1.620     1.329
## [1]    48.000 -205.364     0.245     0.755     2.677    -0.525     1.610     1.330
## [1]    49.000 -205.354     0.242     0.758     2.698    -0.522     1.600     1.331
## [1]    50.000 -205.345     0.240     0.760     2.719    -0.518     1.590     1.332
```