# MLE,EM,EM acceleration

*Guanting Wei*

*Nov.3.2018*

Suppose that there are 4 kinds of animals(A,B,C,D) in a farm. And their probability are

$$P(A) = \frac{1}{2} + \frac{\theta}{4}$$

$$P(B) = \frac{1}{4}(1-\theta)$$

$$P(C) = \frac{1}{4}(1-\theta)$$

$$P(D) = \frac{\theta}{4}$$

$$\theta \in (0,1)$$

. Now we randomly choose one animal 197 times and get the result

$$N(A) = 125$$

$$N(B) = 18$$

$$N(C) = 20$$

$$N(D) = 34$$

## MLE

We let $X = (x_1, x_2, x_3, x_4) = (125, 18, 20, 34)$

$$\begin{aligned} L(x_i;\theta) &= (\frac{1}{2} + \frac{\theta}{4})^{x_1}[\frac{1}{4}(1-\theta)]^{x_2}[\frac{1}{4}(1-\theta)]^{x_3}(\frac{\theta}{4})^{x_4} \\ &\propto (2+\theta)^{x_1}(1-\theta)^{x_2+x_3}(\theta)^{x_4} \end{aligned}$$

$$\begin{aligned} l(x_i;\theta) &= \ln[(2+\theta)^{x_1}(1-\theta)^{x_2+x_3}(\theta)^{x_4}] \\ &= x_1 \ln(2+\theta) + (x_2+x_3)\ln(1-\theta) + x_4 \ln\theta \\ &= 125\ln(2+\theta) + 38\ln(1-\theta) + 34\ln\theta \end{aligned}$$

$$l'(x_i;\theta) = \frac{125}{2+\theta} - \frac{38}{1-\theta} + \frac{34}{\theta} = 0$$

We use Newton's Method to find the result

```
f=function(x){
  125/(2+x)-38/(1-x)+34/x
}
f1=function(x){
  -125/(2+x)^2-38/(1-x)^2-34/x^2
}
x=0.1
count=0
  while (abs(f(x))>0.001&&count<1000) {
```

```
    temp=x
    x=temp-f(temp)/f1(temp)
    count=count+1
  }
x
```

```
## [1] 0.6268216
```

We get $\theta = 0.6268216$

# EM

Let's suppose that we have two subspecies of A including $A_1$,$A_2$ and

$$P(A_1) = \frac{1}{2}$$

$$P(A_2) = \frac{\theta}{4}$$

Further more, suppose that

$$N(A_1) = Y$$
$$N(A_2) = x_1 - Y = 125 - Y$$

In this situation, if we want to use MLE to get $\theta$,we will have

$$\begin{aligned} L(X, Y; \theta) &= (\frac{1}{2})^Y (\frac{\theta}{4})^{x_1 - Y} [\frac{1}{4}(1-\theta)]^{x_2} [\frac{1}{4}(1-\theta)]^{x_3} (\frac{\theta}{4})^{x_4} \\ &\propto (1-\theta)^{x_2+x_3} (\theta)^{x_1 - Y + x_4} \end{aligned}$$

We can not calculate $\theta$ through this formula. Then we need EM algorithm

$$\begin{aligned} Q(\theta|\theta^{(i)}) &= E[\ln L(X, Y; \theta|X, \theta^{(i)})] \\ &= E[(x_1 - Y + x_4)\ln\theta + (x_2 + x_3)\ln(1-\theta)|X, \theta^{(i)}] \\ &= (x_1 + x_4)\ln\theta + (x_2 + x_3)\ln(1-\theta) - E(Y|X, \theta^{(i)})\ln\theta \end{aligned}$$

When given $X$ and $\theta^{(i)}$, $(Y|X, \theta^{(i)}) \sim b(x_1, p_1)$,where $p_1 = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{\theta^{(i)}}{4}} = \frac{2}{2+\theta^{(i)}}$

$$E(Y|X, \theta^{(i)}) = np = x_1 p_1 = \frac{2x_1}{2+\theta^{(i)}}$$

So

$$\begin{aligned} Q(\theta|\theta^{(i)}) &= (x_1 + x_4)\ln\theta + (x_2 + x_3)\ln(1-\theta) - E(Y|X, \theta^{(i)})\ln\theta \\ &= (x_1 + x_4 - \frac{2x_1}{2+\theta^{(i)}})\ln\theta + (x_2 + x_3)\ln(1-\theta) \\ &= (159 - \frac{250}{2+\theta^{(i)}})\ln\theta + 38\ln(1-\theta) \end{aligned}$$

Then we calculate the maximum of $Q(\theta|\theta^{(i)})$

$$\begin{aligned} Q'(\theta|\theta^{(i)}) &= (159 - \frac{250}{2+\theta^{(i)}})\frac{1}{\theta} - 38\frac{1}{(1-\theta)} = 0 \\ \Rightarrow \theta^{(i+1)} &= \frac{68 + 159\theta^{(i)}}{144 + 197\theta^{(i)}} \end{aligned}$$

```r
EM=function(theta.init,tol,count.max){
  theta=theta.init
  count=0
  for(i in 1:count.max){
    count=count+1
    theta[i+1]=(68+159*theta[i])/(144+197*theta[i])
  if(abs(theta[i+1]-theta[i])<tol&&count<count.max)break
  }
  theta
}
theta=EM(0.2,1e-5,1000)
count=length(theta)-1
theta
```

```
## [1] 0.2000000 0.5441658 0.6151352 0.6252563 0.6266134 0.6267939 0.6268178
## [8] 0.6268210
```

```r
count
```

```
## [1] 7
```

# MCEM

We can also use Monte Carlo Method to do this EM algorithm.

$$
\begin{aligned}
Q^{(i+1)}(\theta|\theta^{(i)}) &= \frac{1}{m}\sum_{j=1}^{m}\ln L(X,Y;\theta) \\
&= \frac{1}{m}\sum_{j=1}^{m}\ln(\theta^{x_1+x_2-Y}(1-\theta)^{x_2+x_3}) \\
&= \frac{1}{m}\sum_{j=1}^{m}(x_1+x_2-Y)\ln\theta+(x_2+x_3)\ln(1-\theta) \\
&= (x_1+x_2-E(Y))\ln\theta+(x_2+x_3)\ln(1-\theta)
\end{aligned}
$$

$$
\begin{aligned}
&Q'^{(i+1)}(\theta|\theta^{(i)}) = \frac{(x_1+x_2-E(Y))}{\theta}-\frac{(x_2+x_3)}{1-\theta}=0 \\
&\Rightarrow \theta^{(i+1)} = \frac{x_1+x_4-E(Y)}{x_1+x_2+x_3+x_4-E(Y)} \\
&Y \sim b(x_1,\frac{2}{2+\theta^{(i)}})
\end{aligned}
$$

```r
x1=125
x2=18
x3=20
x4=34
MCEM=function(theta.init,m,tol,count.max){
  theta=theta.init
  count=0
  for(i in 1:count.max){
    count=count+1
    p1=2/(2+theta.init)
    Y=rbinom(m,x1,p1)
```

```
    theta[i+1]=(x1+x4-mean(Y))/(x1+x2+x3+x4-mean(Y))
    theta.init=theta[i+1]
    if(abs(theta[i+1]-theta[i])<tol&&count<count.max)break
  }
  theta
}
theta=MCEM(0.2,10000,1e-5,1000)
count=length(theta)-1
theta
```

```
##  [1] 0.2000000 0.5442656 0.6150577 0.6254549 0.6266795 0.6266006 0.6266706
##  [8] 0.6268877 0.6271542 0.6268059 0.6268653 0.6268708
```

```
count
```

```
## [1] 11
```

# SQUAREM

```
x1=125
x2=18
x3=20
x4=34
loglik=function(theta){
  y=rbinom(1,x1,2/(2+theta))
  log((1-theta)^(x2+x3)*theta^(x1+x4-y))
  }
em=function(theta){
  theta.new=(68+159*theta)/(144+197*theta)
  theta=theta.new
  return(theta)
}
library(SQUAREM)
theta0=0.2
result=squarem(c(theta=theta0), fixptfn=em, objfn=loglik, control=list(tol=1e-10,mstep=5))
theta=result$par
count=result$iter
theta
```

```
##     theta
## 0.6268215
```

```
count
```

```
## [1] 4
```

# Compare

Let's choose a smaller tol to see the speed of the algorithms we talk above.

```
tol=1e-7
count.max=1000
theta.init=0.2
```

## EM

```
t1=Sys.time()
theta_EM=EM(theta.init,tol,count.max)
t2=Sys.time()
iter_EM=length(theta_EM)-1
theta_EM
```

```
##  [1] 0.2000000 0.5441658 0.6151352 0.6252563 0.6266134 0.6267939 0.6268178
##  [8] 0.6268210 0.6268214 0.6268215
```

```
print(paste("iternation=",iter_EM))
```

```
## [1] "iternation= 9"
```

```
print(paste("time=",t2-t1))
```

```
## [1] "time= 0.00199389457702637"
```

## MCEM

```
t3=Sys.time()
theta_MCEM=MCEM(theta.init,10000,tol,count.max)
t4=Sys.time()
iter_MCEM=length(theta_MCEM)-1
print("MCEM")
```

```
## [1] "MCEM"
```

```
theta_MCEM
```

```
##  [1] 0.2000000 0.5443356 0.6149481 0.6251049 0.6266519 0.6267411 0.6268052
##  [8] 0.6269305 0.6270411 0.6268877 0.6266134 0.6265342 0.6267616 0.6268668
## [15] 0.6271688 0.6266560 0.6268931 0.6268254 0.6269452 0.6268880 0.6267183
## [22] 0.6268495 0.6267044 0.6267044
```

```
print(paste("iternation=",iter_MCEM))
```

```
## [1] "iternation= 23"
```

```
print(paste("time=",t4-t3))
```

```
## [1] "time= 0.0339090824127197"
```

## SQUAREM

```
t5=Sys.time()
result=squarem(c(theta=theta0), fixptfn=em, objfn=loglik, control=list(tol=1e-10,mstep=5))
t6=Sys.time()
theta_SQUAREM=result$par
iter_SQUAREM=result$iter
print("SQUAREM")
```

```
## [1] "SQUAREM"
```

```r
theta_SQUAREM
```

```
##     theta
## 0.6268215
```

```r
print(paste("iternation=",iter_SQUAREM))
```

```
## [1] "iternation= 4"
```

```r
print(paste("time=",t6-t5))
```

```
## [1] "time= 0.0029909610748291"
```