Homework 6

Ziyang (Bob) Ding

05 March, 2020

Problem 1

In the exponential family, EM computations are somewhat simplified. Show that if the complete data density f is of the form

$$f(y,z|\theta)=h(y,z)\exp\{\eta(\theta)T(y,z)-C(\theta)\}$$

then we can write

$$Q\left(\theta|\theta^{(t)},y\right)=E_{\theta^{(t)}}[\log(h(y,Z))]+\sum\eta_i(\theta)E_{\theta^{(t)}}\left[T_i|y\right]-C(\theta)$$

so that calculating the complete data MLE only involves the simpler expectation $E_{\theta^{(t)}}\left[T_i|y\right]$, the expected sufficient statistics.}

Proof

$$\begin{aligned}
Q(\theta|\theta^t,y) &= \mathbb{E}_{z|y,\theta^t}[\log f(y,z|\theta)] \\
&= \mathbb{E}_{z|y,\theta^t}[\log h(y,z)\exp\{\eta(\theta)T(y,z)-C(\theta)\}] \\
&= \mathbb{E}_{z|y,\theta^t}[\log h(y,z)+\eta(\theta)T(y,z)-C(\theta)] \\
&= \mathbb{E}_{\theta^t}[\log h(y,z)]+\sum\eta_i(\theta)\mathbb{E}_{\theta^t}[T_i|y]-C(\theta)
\end{aligned}$$

Problem 2

For the situation of Example 5.21, **data** $(x_1,x_2,x_3,x_4)=(125,18,20,34)$ **are collected.**

*(a) **Use the EM algorithm to find the MLE of θ.**

*(b) **(b) Use the Monte Carlo EM algorithm to find the MLE of θ. Compare your results to those of part (a).**

Proof

- (a)

$$\theta^{(t+1)}=\frac{\frac{\theta^{(t)}x_1}{2+\theta_0}+x_4}{\frac{\theta^{(t)}x_1}{2+\theta_0}+x_2+x_3+x_4}$$

```r
x1 = 125
x2 = 18
x3 = 20
x4 = 34
EM = function(n){
theta = rep(0, n)
for(i in 2:n){
tmp = theta[i-1]*x1/(2+theta[i-1])
theta[i] = (tmp+x4)/(tmp+x2+x3+x4)
}
return (theta[n])
}
t = EM(100)
print(t)
```

```
## [1] 0.6268215
```

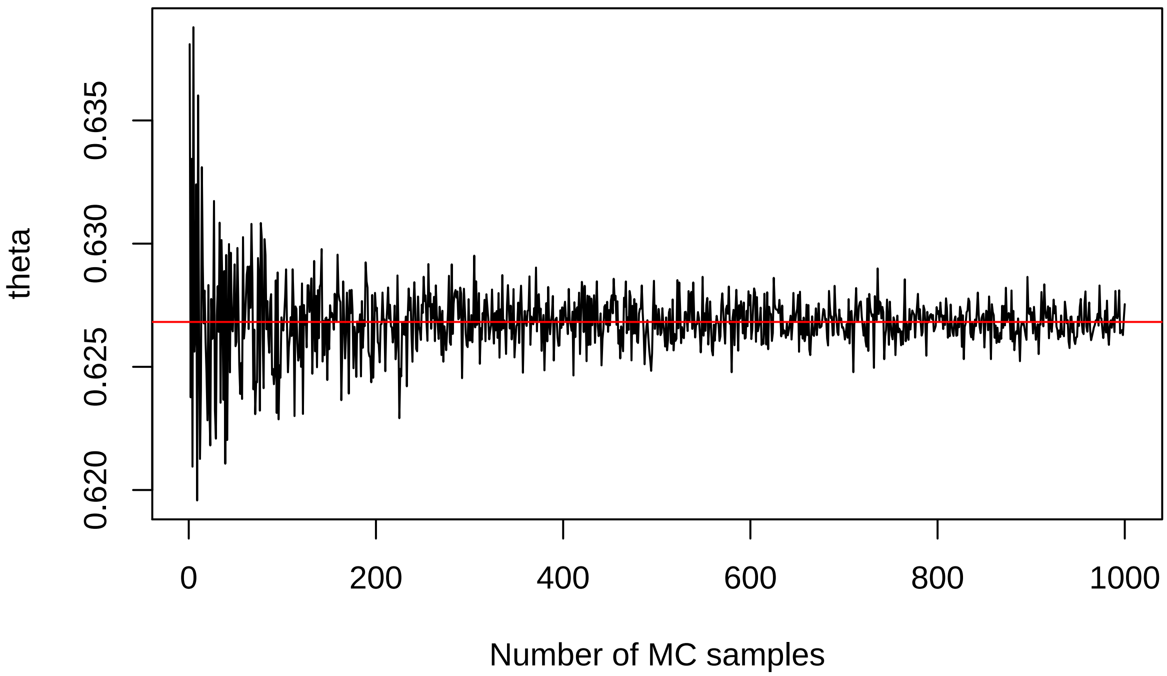
- (b)

$$\theta^{(t+1)} = \frac{\frac{1}{m}\sum_{i=1}^{m} z_i + x_4}{\frac{1}{m}\sum_{i=1}^{m} z_i + x_2 + x_3 + x_4}$$

```r
MCEM = function(n,m){
theta = rep(0, n)
for(i in 2:n){
zm = sum(rbinom(m, x1, theta[i-1]/(2+theta[i-1])))/m
theta[i] = (zm+x4)/(zm+x2+x3+x4)
}
return(theta[n])
}
MCEM(100, 100)
```

```
## [1] 0.6249507
```

```r
th = rep(0,1000)
for(m in 1:1000){
th[m] = MCEM(100, m)
}
plot(1:1000, th, xlab = 'Number of MC samples', ylab = 'theta', type = 'l')
abline(t,0, col = 'red')
```



Problem 3

The EM algorithm can also be implemented in a Bayesian hierarchical model to find a posterior mode. Suppose that we have the hierarchical model $X|\theta \sim f(x|\theta)$

$$\theta|\lambda \sim \pi(\theta|\lambda)$$
$$\lambda \sim \gamma(\lambda)$$

where interest would be in estimating quantities from $\pi(\theta|x)$. **since**

$$\pi(\theta|x) = \int \pi(\theta, \lambda|x) d\lambda$$

where $\pi(\theta, \lambda|x) = \pi(\theta|\lambda, x)\pi(\lambda|x)$, **the EM algorithm is a candidate method for finding the mode of** $\pi(\theta|x)$**, where** λ **would be used as the augmented data.**

- (a) **Define** $k(\lambda|\theta, x) = \pi(\theta, \lambda|x)/\pi(\theta|x)$ **and show that**

$$\log \pi(\theta|x) = \int \log \pi(\theta, \lambda|x) k\left(\lambda|\theta^*, x\right) d\lambda - \int \log k(\lambda|\theta, x) k\left(\lambda|\theta^*, x\right) d\lambda$$

- (b) **If the sequence** $\left(\hat{\theta}_{(j)}\right)$ **satisfies**

$$\max_\theta \int \log \pi(\theta, \lambda|x) k\left(\lambda|\theta_{(j)}, x\right) d\lambda = \int \log \pi\left(\theta_{(j+1)}, \lambda|x\right) k\left(\lambda|\theta_{(j)}, x\right) d\lambda$$

show that $\log \pi\left(\theta_{(j+1)}|x\right) \geq \log \pi\left(\theta_{(j)}|x\right)$. **Under what conditions will the sequence** $\left(\hat{\theta}_{(j)}\right)$ **converge to the mode of** $\pi(\theta|x)$?

- (c) **For the hierarchy**

$$\begin{aligned} X|\theta &\sim \mathcal{N}(\theta, 1) \\ \theta|\lambda &\sim \mathcal{N}(\lambda, 1) \end{aligned}$$

with $\pi(\lambda) = 1$, show how to use the EM algorithm to calculate the posterior mode of $\pi(\theta|x)$

Proof

- (a)

$$\begin{aligned} \log \pi(\theta|x) &= \log \pi(\theta, \lambda|x) - \log k(\lambda|\theta, x) \\ \log \pi(\theta|x) k\left(\lambda|\theta^*, x\right) &= \log \pi(\theta, \lambda|x) k\left(\lambda|\theta^*, x\right) - \log k(\lambda|\theta, x) k\left(\lambda|\theta^*, x\right) \\ \int \log \pi(\theta|x) k\left(\lambda|\theta^*, x\right) d\lambda &= \log \pi(\theta, \lambda|x) k\left(\lambda|\theta^*, x\right) d\lambda - \log k(\lambda|\theta, x) k\left(\lambda|\theta^*, x\right) d\lambda \\ \log \pi(\theta, \lambda|x) &= \log \pi(\theta, \lambda|x) k\left(\lambda|\theta^*, x\right) d\lambda - \log k(\lambda|\theta, x) k\left(\lambda|\theta^*, x\right) d\lambda \end{aligned}$$

- (b)

$$\begin{aligned} &\int \log k(\lambda|\theta, x) k\left(\lambda|\theta_{(j)}, x\right) d\lambda - \int \log k\left(\lambda|\theta_{(j)}, x\right) k\left(\lambda|\theta_{(j)}, x\right) d\lambda \\ &= \int \log \tfrac{k(\lambda|\theta, x)}{k\left(\lambda|\theta_{(j)}, x\right)} k\left(\lambda|\theta_{(j)}, x\right) d\lambda \\ &\le \log \int \tfrac{k(\lambda|\theta, x)}{k\left(\lambda|\theta_{(j)}, x\right)} k\left(\lambda|\theta_{(j)}, x\right) d\lambda \\ &= 0 \end{aligned}$$

Therefore, we know that EM always converge to a zero-gradient point given that $\log \pi(\theta|x)$ is concave (local maximum, local mimimum, saddle point).

- (c)

$$\begin{aligned} k(\lambda|\theta, x) &= \pi(\lambda|\theta, x) \\ &= \pi(\lambda|\theta) \\ &\propto \pi(\theta|\lambda) \\ &\propto e^{-\frac{(\lambda-\theta)^2}{2}} \\ &\sim N(\theta, 1) \\ \pi(\theta, \lambda|x) &\propto \pi(\theta, \lambda)\pi(x|\theta, \lambda) \\ &\propto e^{-\frac{(\lambda-\theta)^2}{2}} e^{-\frac{(x-\theta)^2}{2}} \\ Q\left(\theta|\theta_{(j)}\right) &= C + E_{\lambda|\theta_{(j)}}\left(-\frac{(\lambda-\theta)^2}{2} - \frac{(x-\theta)^2}{2}\right) \\ &= C' - \frac{(\theta-x)^2}{2} - \frac{\left(\theta-\theta_j\right)^2}{2} \\ \theta_{j+1} &= \left(x + \theta_j\right)/2 \end{aligned}$$

Problem 4

Let M be a random matrix having standard Wishart distribution with ν d.o.f., and let $\Sigma = LL'$ be symmetric positive definite. Show that $\tilde{M} = LML'$ has distribution $W_p(\Sigma, \nu)$

Proof

$$
\begin{aligned}
f(M) &\sim \mathrm{IW}(I, \nu) \\
f(LML^T) &\sim \mathrm{IW}(LIL^T, \nu) = \mathrm{IW}(\Sigma, \nu)
\end{aligned}
$$