Talk on 2010-07-01: The EM-REML algorithm(2)

Model:

$$Y = X\gamma + H\beta + e$$

where $\beta \sim MN(0, \tau R)$ and $e \sim MN(0, \sigma I)$.

Let $U = A^T Y$ with the condition that$A^T A = I$ and $AA^T = I - P_X$, according to the "Talk on 2010-06-24", we have

$$\begin{pmatrix} U \\ \beta \end{pmatrix} \sim MN\Big(\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} A^T V A & \tau A^T H R \\ \tau R H^T A & \tau R \end{pmatrix}\Big)$$

Therefore, we have the conditional distribution of $U|\beta$ and $\beta|U$ which is

$$U|\beta \sim MN(A^T H\beta, \sigma(I - P_X))$$

$$\beta|U \sim MN(\tau R H^T P_X Y, \tau R - \tau^2 R H^T P_X H R))$$

Define $\theta = \{\sigma, \tau\}$, according to the EM algorithm, we need to first compute the expectation of $\log \quad L(\theta^{(t)}|U, \beta)$ given U and $\theta^{(t-1)}$,

$$\begin{aligned} \log L(\theta^{(t)}|U, \beta) &= \log f(U, \beta|\theta^{(t)}) = \log\{f(U|\beta, \theta^{(t)}) f(\beta|\theta^{(t)})\} \\ &= \log \underbrace{f(U|\beta, \theta^{(t)})}_{\text{independent from } \tau} - \frac{1}{2}\log(|\tau R|) - \frac{1}{2\tau}\beta^T R^{-1}\beta \end{aligned}$$

so when we take the expectation of $\log L(\theta^{(t)}|U, \beta)$ under $\beta|U, \theta^{(t-1)}$, we have

$$E_{\beta|U,\theta^{(t-1)}}(\log L(\theta^{(t)}|U, \beta)) = \log f(U|\beta, \theta^{(t)}) - \frac{1}{2} q \log \tau - \frac{1}{2}\log|R| - \frac{1}{2\tau} E_{\beta|U,\theta^{(t-1)}}(\beta^T R^{-1}\beta) \quad (1)$$

Since the $E(\varepsilon^T \Lambda \varepsilon) = E(\varepsilon)^T \Lambda E(\varepsilon) + tr(\Lambda Var(\varepsilon))$, (1) can be simplified as

$$\begin{aligned} E_{\beta|U,\theta^{(t-1)}}(\log L(\theta^{(t)}|U, \beta)) =& \log f(U|\beta, \theta^{(t)}) - \frac{1}{2} q \log \tau - \frac{1}{2}\log|R| \\ & - \frac{1}{2\tau}\Big[E(\beta|U, \theta^{(t-1)})^T R^{-1} E(\beta|U, \theta^{(t-1)}) + tr(R^{-1} Var(\beta|U, \theta^{(t-1)}))\Big] \end{aligned} \quad (2)$$

To maximize (2), we let

$$\frac{\partial E_{\beta|U,\theta^{(t-1)}}(\log L(\theta^{(t)}|U, \beta))}{\partial \tau} = 0$$

It is easily to derive that

$$\tau^{(t)} = \frac{1}{q}\Big[E(\beta|U, \theta^{(t-1)})^T R^{-1} E(\beta|U, \theta^{(t-1)}) + tr(R^{-1} Var(\beta|U, \theta^{(t-1)}))\Big] \quad (3)$$

We can also get the estimation of σ by

$$\sigma^{(t)} = \frac{1}{q}\Big[E(\sigma|U,\theta^{(t-1)})^T E(\sigma|U,\theta^{(t-1)}) + tr(Var(\sigma|U,\theta^{(t-1)}))\Big] \tag{4}$$