Talk on 2010-07-09: The EM-REML algorithm(3)

The simple model:

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

where

$Y : N \times 1$,

$X : N \times p$,

$\gamma : p \times 1$, γ is the fixed effect,

$H : N \times q$,

$\beta : q \times 1$, $\beta \sim MN(0, \tau R)$,

and $e : N \times 1$, $e \sim MN(0, \sigma I)$.

Let $U = A^T Y$ with the condition that$A^T A = I_{N-p}$ and $AA^T = I - P_X$, we have

$$\begin{aligned} E(U) &= E(A^T Y) \\ &= E(A^T X\gamma + A^T H\beta + A^T e) \\ &= E(A^T X\gamma) \end{aligned}$$

Consider that,

$$\begin{aligned} (A^T X\gamma)^T (A^T X\gamma) &= \gamma^T X^T AA^T X\gamma \\ &= \gamma^T X^T \underbrace{(I - P_X)}_{0} X\gamma \\ &= 0 \end{aligned}$$

Therefore $A^T X\gamma = 0$, so $E(U) = 0$.

Define $\mathrm{Var}(Y) = V = \tau HRH^T + \sigma I$, so we have $\mathrm{Var}(U) = A^T V A$.

Since U and β are normal distributed, the joint distribution of them is also normal distributed, the covariance of U and β is

$$\begin{aligned} Cov(U, \beta) &= Cov(A^T X\gamma + A^T H\beta + A^T e, \quad \beta) \\ &= Cov(A^T H\beta, \quad \beta) \\ &= A^T H Var(\beta) = \tau A^T HR \end{aligned}$$

Therefore, the joint distribution of $(U,\beta)^T$ is

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

According to $E(\mu_1|\mu_2) = \mu_1 + \Sigma_{12}(\Sigma_{22})^{-1}\mu_2$ and $Var(\mu_1|\mu_2) = \Sigma_{11}(\Sigma_{22})^{-1}\Sigma_{21}$.

$$\begin{aligned}
E(U|\beta) &= 0 + \tau A^T HR\tau^{-1}R^{-1}\beta \\
&= A^T H\beta \\
Var(U|\beta) &= A^TVA - \tau A^T HR\tau^{-1}R^{-1}\tau RH^T A \\
&= A^TVA - A^T(\tau HRH^T)A \\
&= A^T(V - \tau HRH^T)A \\
&= A^T \cdot \sigma I \cdot A \\
&= \sigma I
\end{aligned}$$

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

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

For $\beta|U$, we have

$$\begin{aligned}
E(\beta|U) &= 0 + \tau RH^T A \cdot (A^TVA)^{-1} \cdot A^T Y \\
&= \tau RH^T \cdot A(A^TVA)^{-1}A^T \cdot Y \\
Var(\beta|U) &= \tau R - \tau RH^T A \cdot (A^TVA)^{-1} \cdot \tau A^T HR \\
&= \tau R - \tau^2 RH^T \cdot A(A^TVA)^{-1}A^T \cdot \tau HR
\end{aligned}$$

To show that $A(A^TVA)^{-1}A^T = P = V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}$, we constructing the $n \times p$ matrix $\mathbf{G} = \mathbf{V^{-1}X(X^TV^{-1}X)^{-1}}$ and the $n \times n$ matrix $\mathbf{B} = [\mathbf{A}|\mathbf{G}]$. First the product,

$$\mathbf{B}^T\mathbf{B} = \begin{bmatrix} \mathbf{A}^T \\ \mathbf{G}^T \end{bmatrix} [\mathbf{A} \quad \mathbf{G}] = \begin{bmatrix} \mathbf{A}^T\mathbf{A} & \mathbf{A}^T\mathbf{G} \\ \mathbf{G}^T\mathbf{A} & \mathbf{G}^T\mathbf{G} \end{bmatrix}$$

then the determinant can be rewritten as

$$\begin{aligned}
|\mathbf{B}|^2 = |\mathbf{B}'\mathbf{B}| &= |\mathbf{A}'\mathbf{A}| \times |\mathbf{G}'\mathbf{G} - \mathbf{G}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{G}| \\
&= |\mathbf{I}| \times |\mathbf{G}'\mathbf{G} - \mathbf{G}'(\mathbf{I} - \mathbf{P_X})\mathbf{G}| = |\mathbf{G}'P_\mathbf{X}\mathbf{G}| \\
&= |(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}| = |(\mathbf{X}'\mathbf{X})^{-1}|.
\end{aligned}$$

$$
\begin{aligned}
(A^TVG)^T(A^TVG) &= G^TVAA^TVG \\
&= G^TV\underbrace{(I-P_X)VV^{-1}X}_{0}(X^TV^{-1}X)^{-1} \\
&= 0
\end{aligned}
$$

which also says that **B** is nonsingular.

$$
\begin{aligned}
V^{-1} &= B(B^TVB)^{-1}B^T \\
&= [A \quad G]\begin{bmatrix} A^TVA & 0 \\ 0 & G^TVG \end{bmatrix}\begin{bmatrix} A^T \\ G^T \end{bmatrix} \\
&= A(A^TVA)^{-1}A^T + G(G^TVG)^{-1}G^T \\
&= A(A^TVA)-1A^T + G\underbrace{((X^TV^{-1}X)^{-1}X^TV^{-1}VV^{-1}X(X^TV^{-1}X)^{-1})^{-1}}_{X^TV^{-1}X}G^T \\
&= A(A^TVA)-1A^T + V^{-1}X(X^TV^{-1}X)^{-1}\cdot X^TV^{-1}X\cdot(X^TV^{-1}X)^{-1}X^TV^{-1} \\
&= A(A^TVA)-1A^T + V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}
\end{aligned}
$$

so we have $A(A^TVA)^{-1}A^T = P = V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}$. Therefore, the expectation and variance of $\beta|U$ can be simplified as

$$
\begin{aligned}
E(\beta|U) &= \tau RH^TP_XY \\
Var(\beta|U) &= \tau R - \tau^2RH^TP_XHR
\end{aligned}
$$

In another words, we have,

$$
\beta|U \sim MN(\tau RH^TP_XY, \tau R - \tau^2RH^TP_XHR))
$$

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 $\theta^{(t-1)}$,

$$
\begin{aligned}
\log L(\theta^{(t)}|U,\beta) &= \log f(U,\beta|\theta^{(t)}) \\
&= \log\{f(U|\beta,\theta^{(t)})f(\beta|\theta^{(t)})\} \\
&= \log f(U|\beta,\theta^{(t)}) + \log f(\beta|\theta^{(t)})
\end{aligned}
$$

1. Estimate τ Take the expectation of $\log L(\theta^{(t)}|U,\beta)$ under $\beta|U,\theta^{(t-1)}$, we have

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

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

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

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

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

2. Estimate σ We can also get the estimation of σ by Define

$$
\begin{aligned}
Q &= E_{\beta|U,\theta^{(t-1)}}(\log L(\theta^{(t)}|U,\beta)) \\
&= -\frac{1}{2}\Big[(N-p)\log\sigma + 1 + (U - A^T H\beta)^T(U - A^T H\beta)\Big] + \log f(\beta|\theta^{(t)})
\end{aligned}
$$

If we take the expectation based on β, then it is applied to the quadratic form $(U - A^T H\beta)^T(U - A^T H\beta)$, but if we take the expectation based on e, then we have,

$$
\begin{aligned}
E_{\beta|U,\theta^{(t-1)}}((U - A^T H\beta)^T(U - A^T H\beta)) &= E_{\beta|U,\theta^{(t-1)}}[(A^T e)^T(A^T e)] \\
&= E_{\beta|U,\theta^{(t-1)}}[e^T e]
\end{aligned}
$$

then we can easily get

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

just like τ. If we still take the expectation on β, then

$$
\begin{aligned}
E_\beta[\log(L)] =& E_\beta\Big[-\frac{1}{2}(N-p)\log\sigma + 1 + U^T U - \beta^T H^T A A^T H\beta + \log f(\beta|\theta^{(t)})\Big] \\
=& E_\beta\Big[-\frac{1}{2}(N-p)\log\sigma + 1 + Y^T(I-P_X)Y - \beta^T H^T(I-P_X)^T H\beta\Big] + E_\beta\Big[\log f(\beta|\theta^{(t)})\Big] \\
=& -\frac{1}{2}\Big[(N-p)\log\sigma + Y^T(I-P_X)Y - (H\tilde{\beta})^T(I-P_X)^T(H\tilde{\beta}) + tr(H^T(I-P_X)HVar(\beta|U))\Big] \\
& + E_\beta\Big[\log f(\beta|\theta^{(t)})\Big] + 1
\end{aligned}
$$

Let $\dfrac{\partial Q}{\partial\sigma} = 0$, we have

$$\sigma^{(t)} = \frac{1}{N-p}\Big[(Y - H\tilde{\beta})^T(I-P_X)(Y - H\tilde{\beta}) + tr(H^T(I-P_X)HVar(\beta|U))\Big] \tag{2}$$

If the model is

$$Y = X\gamma + H_A\beta_A + H_B\beta_B + e$$

where

$Y : N \times 1,$

$X : N \times p,$

$\gamma : p \times 1$, γ is the fixed effect,

$H_A : N \times q_A,$

$\beta_A : q_A \times 1$, $\beta \sim MN(0, \tau_A R_A),$

$H_B : N \times q_B,$

$\beta_B : q_B \times 1$, $\beta \sim MN(0, \tau_B R_B),$

construct $H = [H_A H_B]$, $\beta = (\beta_A, \beta_B)^T$ and $R = \begin{bmatrix} \tau_A R_A & 0 \\ 0 & \tau_B R_B \end{bmatrix}$. According to the simple model, we can easily get that

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

and

$$\begin{aligned} E(U|\beta) &= A^T H \beta \\ Var(U|\beta) &= \sigma I_{N-p} \\ E(\beta|U) &= R H^T P_X Y \\ Var(\beta|U) &= R - R H^T P_X H R \end{aligned}$$

The likelihood function is

$$\begin{aligned} \log L(\theta^{(t)}|U, \beta) =& \log f(U, \beta|\theta^{(t)}) \\ =& \log f(U|\beta, \theta^{(t)}) + \log f(\beta|\theta^{(t)}) \\ =& \Big(-\frac{1}{2}\Big)\Big[\log\Big|\begin{bmatrix} \tau_A R_A & 0 \\ 0 & \tau_B R_B \end{bmatrix}\Big| + (\beta_A \quad \beta_B)^T \begin{bmatrix} (\tau_A R_A)^{-1} & 0 \\ 0 & (\tau_B R_B)^{-1} \end{bmatrix} \begin{bmatrix} \beta_A \\ \beta_B \end{bmatrix}\Big] \\ & + \log f(U|\beta, \theta^{(t)}) \\ =& \log f(U|\beta, \theta^{(t)}) + \log f(\beta_A|\theta^{(t)}) + \log f(\beta_B|\theta^{(t)}) \end{aligned}$$

so the estimation of θ is

$$\tau_A^{(t)} = \frac{1}{q_A}\left[E(\beta_A|U,\theta^{(t-1)})^T R_A^{-1} E(\beta_A|U,\theta^{(t-1)}) + tr(R_A^{-1} Var(\beta_A|U,\theta^{(t-1)}))\right] \tag{3}$$

$$\tau_B^{(t)} = \frac{1}{q_B}\left[E(\beta_B|U,\theta^{(t-1)})^T R_B^{-1} E(\beta_B|U,\theta^{(t-1)}) + tr(R_B^{-1} Var(\beta_B|U,\theta^{(t-1)}))\right] \tag{4}$$

$$\sigma^{(t)} = \frac{1}{N-p}\left[(Y-H\tilde{\beta})^T(I-P_X)(Y-H\tilde{\beta}) + tr(H^T(I-P_X)HVar(\beta|U))\right] \tag{5}$$