Talk on 2010-06-24: The EM-REML algorithm

Model:

$$Y_i = X_i\beta + Z_i b_i + e_i$$

where $b_i \sim N(0, D)$ and $e_i \sim N(0, \sigma I)$.

Define $U = H_1^T Y_i = H_1^T(X_i\beta + Z_i b_i + e_i) = H_1^T X_i\beta + H_1^T Z_i b_i + H_1^T e_i$, we have,

$$\begin{aligned}
E(U) &= E(H_1^T X_i\beta + H_1^T Z_i b_i + H_1^T e_i) = E(H_1^T X_i\beta) \\
(H_1^T X_i\beta)^T(H_1^T X_i\beta) &= \beta^T X_i^T H_1 H_1^T X_i\beta \\
&= \beta^T X_i^T \underbrace{(I - P_X)}_{0} X_i\beta \\
&= 0 \\
=> H_1^T X_i\beta &= 0 \\
=> E(U) &= 0
\end{aligned}$$

Let $\text{Var}(Y) = V = Z_i D Z_i^T + \sigma I$, so we have $\text{Var}(U) = H_1^V H_1$. Since U and b are normal distributed, the joint distribution of them is also normal distributed, the covariance of U and b is

$$\begin{aligned}
Cov(U, b) &= Cov(H_1^T X_i\beta + H_1^T Z_i b_i + H_1^T e_i, \quad b) \\
&= Cov(H_1^T Z_i b_i, \quad b) \\
&= H_1^T Z_i Var(b) = H_1^T Z_i D
\end{aligned}$$

Therefore, the distribution of $(U, b)^T$ is

$$\begin{pmatrix} U \\ b \end{pmatrix} \sim MN\Big(\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} H_1^T V H_1 & H_1^T Z_i D \\ D Z_i^T H_1 & D \end{pmatrix}\Big)$$

therefore, we have

$$\begin{aligned}
E(b|U) &= 0 + (D Z_i^T H_1)(H_1^T V H_1)^{-1}(H_1 Y_i) \\
&= D Z_i^T \underbrace{H_1 H_1^T V H_1)^{-1} H_1}_{P} Y_i \\
&= D Z_i^T P Y_i
\end{aligned} \tag{1}$$

and

$$\begin{aligned}
Var(b|U) &= D - D Z_i^T \underbrace{H_1 H_1^T V H_1)^{-1} H_1}_{P} Z_i D \\
&= D - D Z_i^T P Z_i D
\end{aligned} \tag{2}$$

In the same way, we have the conditional expectation and covariance for e_i,

$$E(e_i|U) = \sigma P Y_i \tag{3}$$

$$Var(e|U) = \sigma I - \sigma^2 P \tag{4}$$

Question: how to get the REML estimation of $\hat{D}$ in terms of b? We have the REML log-likelihood function:

$$L_{REML} = -\frac{1}{2}(ln|V| + ln|X^T V^{-1} X| + Y^T P Y)$$

To get the estimation, we should set

$$\frac{\partial log L_{REML}}{\partial \hat{D}} = 0$$

but how is the above equation related to b?

According to the paper "Random-Effect Model for Longitudinal Data" by Nan M. Laird(1982), we can use the sufficient estimator $\mathbf{t_1}$ to estimate $\hat{D}$

$$\hat{D} = \frac{\sum_1^n b_i b_i^T}{n} = \frac{\mathbf{t_1}}{n} \tag{5}$$

where

$$\begin{aligned} \hat{\mathbf{t_1}} &= E(\sum_1^n b_i b_i^T | y, D, \sigma) \\ &= \sum_1^n [E(b_i|y, D, \sigma) E(b_i|y, D, \sigma^T] + var(b|y, D, \sigma) \end{aligned} \tag{6}$$

In the same, we can use the sufficient estimator $\mathbf{t_2}$ to estimate σ by

$$\hat{\sigma} = \frac{\sum_1^n e_i^T e_i}{n} = \frac{\mathbf{t_2}}{n} \tag{7}$$

where

$$\begin{aligned} \hat{\mathbf{t_2}} &= E(\sum_1^n e_i e_i^T | y, D, \sigma) \\ &= \sum_1^n [E(e_i|y, D, \sigma) E(e_i|y, D, \sigma^T] + tr(var(e|y, D, \sigma)) \end{aligned} \tag{8}$$

So the EM algorithm can be described as:

1. Set the initial value for D and σ.
2. Compute the conditional expectation and variance of b_i and e_i using (1)-(4).
3. Compute the sufficient estimator $\mathbf{t_1}$ and $\mathbf{t_2}$ using (6) and (8).
4. Compute the estimation of $\hat{D}$ and $\hat{\sigma}$.
5. Repeat step 2-4 until $\hat{D}$ and $\hat{\sigma}$ converge.