Linear Mixed Effect Model - Stat 481

A general linear mixed model may be expressed as

$$y = X\beta + Zb + \varepsilon$$

where y is a vector of observations, X is a matrix of known covariates, β is a vector of unknown regression coefficients (fixed effects), Z is a known matrix, b is a vector of random effects, and ε is a vector of errors. Both b and ε are unobservable. Basic assumption is that the random effects and errors have mean zero and finite variances.

Example 1. Height vs weight

```
This is a simple data set to illustrate how the mixed model treats
clustered data. The first column (F) is the family indicator. Each
family is a cluster. For example, the first family consists of 8
adult people. The second and the third columns are height (in
inches) and weight (in pounds). Let H and W be the height and
weight measurement of 75 people from 19 different families.


F  H  W
1 61 120
1 64 155
1 66 165
... ...
19 71 220
19 66.5 134
19 68.5 130
```

Linear regression fit: $W_k = \alpha + \beta H_k + \varepsilon_k$. It is assumed that the errors ε_k are iid, and with constant variances σ^2.

$$\left(\hat{\alpha}, \hat{\beta}\right)_{OLS} = \arg\min \left\{ \sum_{k=1}^{75} \left(W_k - \alpha - \beta H_k\right)^2 \right\}.$$

To consider the familial correlation, we need mixed-effects approach. Define

$$W_i = \left(W_{i1}, ..., W_{in_i}\right)', H_i = \left(H_{i1}, ..., H_{in_i}\right)'.$$

Let within family correlation $\rho = Corr\left(W_{ij}, W_{ik}\right)$ for $j \neq k$.

Each family has its own intercept, $a_i = \alpha + b_i$, with $E\left(b_i\right) = 0, Var\left(b_i\right) = \sigma_d^2$. Linear model with random intercept:

$$W_i = \alpha + b_i + \beta H_i + \varepsilon_i,$$

where $\varepsilon_i \sim (0, \sigma^2 I_{n_i}), i = 1, ...19.$

Linear mixed effects model (LME):

$$W_i = \alpha + \beta H_i + Z_i b_i + \varepsilon_i,$$

where in this example $Z_i = 1_{n_i}$. It is easy to have

$$Var(W_i) = \begin{pmatrix} \sigma^2 + \sigma_d^2 & & \sigma_d^2 \\ & \ddots & \\ \sigma_d^2 & & \sigma^2 + \sigma_d^2 \end{pmatrix}.$$

Note that $\rho = Corr(W_{ij}, W_{ik}) = \frac{\sigma_d^2}{\sigma^2 + \sigma_d^2}$.

Its GLS estimator for α, β is to minimize weighted sum squares

$$\left(\hat{\alpha}, \hat{\beta}\right)_{GLS} = \arg\min \left\{ \sum_i (W_i - \alpha - \beta H_i)' V_i^{-1} (W_i - \alpha - \beta H_i) \right\}.$$

In addition σ_d^2/σ^2 can be estimated via Maximum Likelihood Estimation or Residual MLE if normal distributions are assumed. Variance components σ_d^2 and σ^2 can be estimated by a quadratic function of W_i without normal assumption, MINQUE (Minimum Quardratic Unbiased Estimator).

LME model in matrix format: $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \varepsilon$, or explicitly

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} \beta + \begin{pmatrix} Z_1 & & 0 \\ & \ddots & \\ 0 & & Z_N \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

A. **Log-likelihood Functions (ML method)**

If assume normal distributions, i.e. $\varepsilon_i \sim N(\mathbf{0}, \sigma^2 I_{n_i}), b_i \sim N(\mathbf{0}, \sigma^2 D)$. Then the marginal form of the LME model can be written as

$$y_i \sim N\left(X_i\beta, \sigma^2 (I + Z_i D Z_i')\right).$$

Define scaled covariance $V_i = V_i(D) = I + Z_i D Z_i'$

Log-likelihood function of the LME model (up to a constant):

$$l(\theta) = \frac{-1}{2} \left\{ N_T \log \sigma^2 + \sum_{i=1}^{N} \left[\log |V_i| + \frac{1}{\sigma^2} (y_i - X_i\beta)' V_i^{-1} (y_i - X_i\beta) \right] \right\} \quad (1)$$

where parameter vector $\theta = (\beta', \sigma^2, vech'(D)), \dim(\theta) = m + 1 + k(k+1)/2.$

B. Restricted Maximum Likelihood Method

A general linear model $y \sim N(X\beta, V), V = V(\theta^*)$, then its generalized least square estimator $\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1} X'V^{-1}y$.

Then the RML function is defined as

$$l_R(\beta, \theta^*) = \frac{-1}{2} \left\{ \log \left| X'V^{-1}X \right| + \log |V| + (y - X\beta)' V^{-1} (y - X\beta) \right\}.$$

It differs from the standard log-likelihood function by the term $-\frac{1}{2} \log |X'V^{-1}X|$.