Advanced Data Analysis

Whiteboard Lecture

Longitudinal Data Analysis:

Whiteboard Interlude IV

The random intercept model is

$$y_{ij} = U_i + \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}$$

$$U_i \overset{\text{ind.}}{\sim} N(0, \sigma_U^2), \qquad \varepsilon_{ij} \overset{\text{ind.}}{\sim} N(0, \sigma_\varepsilon^2)$$

and the U_i and ε_{ij} are independent of each other.

It follows that the distribution of y_{ij} is

$$y_{ij} \sim N(\beta_0 + \beta_1 x_{ij}, \sigma_U^2 x_{ij}^2 + \sigma_\varepsilon^2).$$

But since the y_{ij}'s are correlated, we can't just take the product of the individual densities to get the likelihood.

Proper likelihood calculations get very messy if we don't use matrices.

Let $n = n_1 + \ldots + n_m$,

$$\boldsymbol{y} = \begin{bmatrix} y_{11} \\ \cdot \\ \cdot \\ y_{1n_1} \\ \hline y_{21} \\ \cdot \\ \cdot \\ y_{2n_2} \\ \hline \cdot \\ \cdot \\ \hline y_{m1} \\ \cdot \\ \cdot \\ y_{mn_m} \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_{1n_1} \\ \hline 1 & x_{21} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_{2n_2} \\ \hline \cdot & \cdot \\ \cdot & \cdot \\ \hline 1 & x_{m1} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_{mn_m} \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \ \boldsymbol{u} = \begin{bmatrix} U_1 \\ \cdot \\ \cdot \\ U_m \end{bmatrix}$$

and let $\boldsymbol{Z}$ be the $n \times m$ matrix:

$$\boldsymbol{Z} = \left[\begin{array}{cccccc} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \vdots & \vdots & \ddots & & & \vdots \\ 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \hline 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \vdots & \vdots & \ddots & & & \vdots \\ 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline 0 & 0 & \cdot & \cdot & \cdot & 1 \\ \vdots & \vdots & \ddots & & & \vdots \\ 0 & 0 & \cdot & \cdot & \cdot & 1 \end{array}\right].$$

In this $\boldsymbol{Z}$ matrix, the first column of ones is of length n_1, the second of length n_2, and so on.

Then the random intercept model can be written in matrix notation as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$$

This is the general form of the linear mixed model.

Using results for multivariate means and covariance matrices we get

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{V})$$

where

$$\boldsymbol{V} = \sigma_U^2 \boldsymbol{Z}\boldsymbol{Z}^T + \sigma_\varepsilon^2 \boldsymbol{I}.$$

Hence, the likelihood function is:

$$\mathcal{L}(\beta_0, \beta_1, \sigma_U^2, \sigma_\varepsilon^2) = (2\pi)^{-n/2} |\boldsymbol{V}|^{-1/2} \times \exp\{-\tfrac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X\beta})\}.$$

The log-likelihood is then

$$\ell(\beta_0, \beta_1, \sigma_U^2, \sigma_\varepsilon^2) = -\tfrac{1}{2}\log|\boldsymbol{V}| - \tfrac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X\beta}) - \frac{n}{2}\log(2\pi).$$

Differentiation with respect to $\boldsymbol{\beta}$ leads to the maximum likelihood estimators:

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix}_{\text{ML}} = (\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{y}.$$

However,

$$\boldsymbol{V} = \sigma_U^2 \boldsymbol{Z}\boldsymbol{Z}^T + \sigma_\varepsilon^2 \boldsymbol{I}$$

still depends on the unknown
variance components σ_U^2 and σ_ε^2.

Substitution of the above solution for $\boldsymbol{\beta}$ into the log-likelihood expression leads to the profile

likelihood

$$\ell_P(\sigma_U^2, \sigma_\varepsilon^2) = -\frac{n}{2}\log(2\pi) - \tfrac{1}{2}[\log|\boldsymbol{V}| \\ + \boldsymbol{y}^T\boldsymbol{V}^{-1}\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\}\boldsymbol{y}].$$

Maximum likelihood estimation of σ_U^2 and σ_ε^2 involves maximising $\ell_P(\sigma_U^2, \sigma_\varepsilon^2)$ over $(\sigma_U^2, \sigma_\varepsilon^2)$.

There is no closed form solution, so it has to be done numerically.

However, the default method used by mixed model packages for variance component estimation is restricted maximum likelihood – commonly known by its acronym REML.

REML estimation involves maximisation of the restricted log-likelihood:

$$\ell_R(\sigma_U^2, \sigma_\varepsilon^2) = \ell_P(\sigma_U^2, \sigma_\varepsilon^2) - \tfrac{1}{2}\log|\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}|.$$

The justification for the additional term is based on "contrast" arguments that account for estimation of the fixed effects vector $\boldsymbol{\beta}$.

The 2000 Wiley book by McCulloch and Searle has the details.

♦ ♦ ♦