

Linear Mixed Model

Geert Molenberghs

I-BioStat, Universiteit Hasselt & Katholieke Universiteit Leuven,
Belgium.

In many studies, data are collected hierarchically. Not always do such data follow balanced, multivariate designs. For example, in repeated measurements may be taken at almost arbitrary time points, resulting in an extremely large number of time points at which only one or only a few measurements have been taken. Many of the parametric covariance models described so far may then contain too many parameters to make them useful in practice, while other, more parsimonious, models may be based on assumptions which are too simplistic to be realistic. A general, and very flexible, class of parametric models for continuous longitudinal data is formulated as follows:

$$\mathbf{y}_i | \mathbf{b}_i \sim N(X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i, \Sigma_i), \quad (1)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, D), \quad (2)$$

where X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ dimensional matrices of known covariates, $\boldsymbol{\beta}$ is a p -dimensional vector of regression parameters, called the fixed effects, D is a general $(q \times q)$ covariance matrix, and Σ_i is a $(n_i \times n_i)$ covariance matrix which depends on i only through its dimension n_i , i.e., the set of unknown parameters in Σ_i will not depend upon i . Finally, \mathbf{b}_i is a vector of subject-specific or random effects.

The above model can be interpreted as a linear regression model for the vector \mathbf{y}_i of repeated measurements for each unit separately, where some of the regression parameters are specific (random effects, \mathbf{b}_i), while others are not (fixed effects, $\boldsymbol{\beta}$). The distributional assumptions in (2) with respect to the random effects can be motivated as follows. First, $E(\mathbf{b}_i) = \mathbf{0}$ implies that the mean of \mathbf{y}_i still equals $X_i \boldsymbol{\beta}$, such that the fixed effects in the random-effects model (1) can also be interpreted marginally. Not only do they reflect the effect of changing covariates within specific units, they also measure the marginal effect in the population of changing the same covariates. Second, the normality assumption immediately implies that, marginally, \mathbf{y}_i also follows a normal distribution with mean vector $X_i \boldsymbol{\beta}$ and with covariance matrix $V_i = Z_i D Z_i' + \Sigma_i$.

Note that the random effects in (1) implicitly imply the marginal covariance matrix V_i of \mathbf{y}_i to be of the very specific form $V_i = Z_i D Z_i' + \Sigma_i$. Let us consider two examples under the assumption of conditional independence, i.e., assuming $\Sigma_i = \sigma^2 I_{n_i}$. First, consider the case where the random effects are univariate and represent unit-specific intercepts. This corresponds to covariates Z_i which are n_i -dimensional vectors containing only ones.

The marginal model implied by expressions (1) and (2) is

$$\mathbf{y}_i \sim N(X_i \boldsymbol{\beta}, V_i), \quad V_i = Z_i D Z_i' + \Sigma_i$$

which can be viewed as a multivariate linear regression model, with a very particular parameterization of the covariance matrix V_i .

With respect to the estimation of unit-specific parameters \mathbf{b}_i , the posterior distribution of \mathbf{b}_i given the observed data \mathbf{y}_i can be shown to be (multivariate) normal with mean vector equal to $DZ_i'V_i^{-1}(\boldsymbol{\alpha})(\mathbf{y}_i - X_i\boldsymbol{\beta})$. Replacing $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by their maximum likelihood estimates, we obtain the so-called empirical Bayes estimates $\widehat{\mathbf{b}}_i$ for the \mathbf{b}_i . A key property of these EB estimates is shrinkage, which is best illustrated by considering the prediction $\widehat{\mathbf{y}}_i \equiv X_i\widehat{\boldsymbol{\beta}} + Z_i\widehat{\mathbf{b}}_i$ of the i th profile. It can easily be shown that

$$\widehat{\mathbf{y}}_i = \Sigma_i V_i^{-1} X_i \widehat{\boldsymbol{\beta}} + (I_{n_i} - \Sigma_i V_i^{-1}) \mathbf{y}_i,$$

which can be interpreted as a weighted average of the population-averaged profile $X_i\widehat{\boldsymbol{\beta}}$ and the observed data \mathbf{y}_i , with weights $\Sigma_i V_i^{-1}$ and $I_{n_i} - \Sigma_i V_i^{-1}$, respectively. Note that the “numerator” of $\Sigma_i V_i^{-1}$ represents within-unit variability and the “denominator” is the overall covariance matrix V_i . Hence, much weight will be given to the overall average profile if the within-unit variability is large in comparison to the between-unit variability (modeled by the random effects), whereas much weight will be given to the observed data if the opposite is true. This phenomenon is referred to as shrinkage toward the average profile $X_i\widehat{\boldsymbol{\beta}}$. An immediate consequence of shrinkage is that the EB estimates show less variability than actually present in the random-effects distribution, i.e., for any linear combination $\boldsymbol{\lambda}$ of the random effects,

$$\text{var}(\boldsymbol{\lambda}'\widehat{\mathbf{b}}_i) \leq \text{var}(\boldsymbol{\lambda}'\mathbf{b}_i) = \boldsymbol{\lambda}'D\boldsymbol{\lambda}.$$

Reference

- Fitzmaurice, G.M., Davidian, M., Verbeke, G., and Molenberghs, G. (2009). *Longitudinal Data Analysis. Handbook*. Hoboken, NJ: John Wiley & Sons.
- Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis*. New York: John Wiley & Sons.
- Henderson, C.R. (1984) *Applications of Linear Models in Animal Breeding*. Guelph, Canada: University of Guelph Press.
- Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Acknowledgment

Based on an article from Lovric, Miodrag (2011), International Encyclopedia of Statistical Science. Heidelberg: Springer Science +Business Media, LLC