Nonlinear Mixed-Effects Models

Sung Won Kim
Seminar on Mixed-Effects Models

3rd May 2010

1 Why Nonlinear Mixed-Effects?

So far we learned linear mixed-effects models and nonlinear regression. Now we consider the situation combining these two concepts at the same time: Grouped data with nonlinear expectation function. It means that the random effects are incorporated in the coefficients, and the expectaion function is allowed to be nonlinear in random effects.

2 Model for Nonlinear Mixed-Effects

$$y_{ij} = f(\phi_{ij}, \nu_{ij}) + \varepsilon_{ij} \quad i = 1, \dots, M, \;\; j = 1, \dots, n_i$$

where $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, f is nonlinear differentiable function of group specific parameter vector $\phi_{ij} = A_{ij}\beta + B_{ij}b_i$ and covariate vector ν_{ij}.

Example:

Indomethicin Kinetics is known that it is expressed by a linear combination of two exponentials. This model is nonlinear, and its coefficients vary with subject i.

$$y_{ij} = (\beta_1 + b_{1i}) \exp\left[-\exp\left(\beta_2 + b_{2i}\right) t_j\right] + (\beta_3 + b_{3i}) \exp\left[-\exp\left(\beta_4 + b_{4i}\right) t_j\right] + \varepsilon_{ij}$$

β's are fixed effects representing the mean value at the parameter and b_i's stand for random effects that are representing individual deviations. Using group specific parameter vector ϕ_{ij}

$$y_{ij} = (\phi_{1i}) \exp\left[-\exp\left(\phi_{2i}\right) t_j\right] + (\phi_{3i}) \exp\left[-\exp\left(\phi_{4i}\right) t_j\right] + \varepsilon_{ij}$$

$$\text{where} \quad b_i \sim \mathcal{N}(0, \Psi) \quad \text{and} \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2).$$

In the case $b_{4i} = 0$ (no random effect on β_4),

$$\underbrace{\begin{pmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \\ \phi_{4i} \end{pmatrix}}_{\phi_i} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}}_{B} \underbrace{\begin{pmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \end{pmatrix}}_{b_i}$$

3 Estimation of the Parameter

We want to estimate parameters β, σ^2 and Ψ incorporated in ϕ_{ij}. For computational purpose Ψ is expressed in terms of the *relative prediction factor* Δ satisfying

$$\Delta^T \Delta = \frac{\Psi^{-1}}{\sigma^2}$$

such that $\Psi = \sigma^2(\Delta^T\Delta)^{-1}$. We use likelihood functions.

$$\mathcal{L}(\beta, \sigma^2, \Delta | y) = \underbrace{p(y|\beta, \sigma^2, \Delta)}_{\text{marginal density of} y} = \int \underbrace{p(y|b, \beta, \sigma^2)}_{\text{conditional density of } y \text{given} b} p(b|\Delta, \sigma^2) db$$

Notice that $y_i|b_i \sim \mathcal{N}(f_i(\beta, b_i), \sigma^2 I)$ and $b_i \sim \mathcal{N}(0, \sigma^2(\Delta^T\Delta)^{-1})$ so that

$$= \frac{|\Delta|^M}{(2\pi\sigma^2)^{(N+Mq)/2}} \prod_{i=1}^{M} \int \exp\left(\frac{\|y_i - f_i(\beta, b_i)\|^2 + \|\Delta b_i\|^2}{-2\sigma^2} \right) db_i \quad i = 1, \ldots, M, \; j = 1, \ldots, n_i \tag{1}$$

where $f_i(\beta, b_i) = f_i\left[\phi_i(\beta, b_i), \nu_i\right]$, b_i is q-dim vector. Since f can be nonlinear in random effects the integral in (1) does not have closed form. To make the numerical optimization of this likelihood function tractable three different approximation methods are introduced.

3.1 Approximation of Likelihood Function in NLME

3.1.1 LME Approximation (Alternating Algorithm)

It consists of alternating between two steps:

1. Penalized nonlinear Least Squares(PNLS)
2. Linear Mixed-Effects (LME)

Details on 1: For a fixed current estimate of Δ, b_i and β are estimated by minimizing a penalized nonlinear least squares objective function (from (1)):

$$\sum_{i=1}^{M} \left[\|y_i - f_i(\beta, b_i)\|^2 + \|\Delta b_i\|^2 \right] \tag{2}$$

For computation *Gauss-Newton method* is used (see refsec:GN).

Details on 2: We update the estimate of Δ, based on a first order Taylor Expansion of the model function around current estimates $\hat{\beta}^{(w)}$ and $\hat{b}_i^{(w)}$. Let $\hat{\omega}_i^{(w)}(\hat{\beta}, \hat{b}_i) := y_i - f_i(\hat{\beta}^{(w)}, \hat{b}_i^{(w)}) + \hat{X}_i^{(w)}\hat{\beta}^{(w)} + \hat{Z}_i^{(w)}\hat{b}_i^{(w)}$. Then

$$\begin{aligned} \|y_i - f_i(\beta, b_i)\|^2 &\simeq \|\hat{\omega}_i^{(w)}(\hat{\beta}, \hat{b}_i) - \hat{X}_i^{(w)}\beta - \hat{Z}_i^{(w)} b_i\|^2 \\ \text{with} \quad \hat{X}_i^{(w)} &= \left.\frac{\partial f_i}{\partial \beta^T}\right|_{(\hat{\beta}^{(w)}, \hat{b}_i^{(w)})} \quad \text{and} \quad \hat{Z}_i^{(w)} = \left.\frac{\partial f_i}{\partial b_i^T}\right|_{(\hat{\beta}^{(w)}, \hat{b}_i^{(w)})} \end{aligned} \tag{3}$$

Plugging (3) in (1) we obtain

$$\begin{aligned} l_{LME}(\beta, &\sigma^2, \Delta | y) = -\frac{N}{2} \log\left(2\pi\sigma^2\right) \\ &- \frac{1}{2} \sum_{i=1}^{M} \left(\log|\Sigma_i(\Delta)| + \frac{1}{\sigma^2} \left[\hat{\omega}_i^{(w)} - \hat{X}_i^{(w)}\beta\right] \Sigma_i^{-1}(\Delta) \left[\hat{\omega}_i^{(w)} - \hat{X}_i^{(w)}\beta\right] \right) \end{aligned} \tag{4}$$

with $\Sigma_i(\Delta) = I + \hat{Z}_i^{(w)}\Delta^{-1}\Delta^{-T}(\hat{Z}_i^{(w)})^T$.
Now calculate optimal values for $\hat{\beta}(\Delta)$ and $\hat{\sigma}^2(\Delta)$ and then plug them into l_{LME} so as to work with the profiled log-likelihood $l_{LME,p}$ of Δ to estimate Δ.

3.1.2 Laplacian Approximation

Let $g(\beta, \Delta, y_i, b_i) = \|y_i - f(\beta, b_i)\|^2 + \|\Delta b_i\|^2$ from the integral in (1). By second order of Taylor expansion around $\hat{b}_i$ it holds

$$g(\beta, \Delta, y_i, b_i) = g(\beta, \Delta, y_i, \hat{b}_i) + \frac{1}{2}(b_i - \hat{b}_i)^T g''(\beta, \Delta, y_i, \hat{b}_i)(b_i - \hat{b}_i)$$

$$\text{where} \quad \hat{b}_i = \arg\min_{b_i} g(\beta, \Delta, y_i, b_i), \quad g' = \frac{\partial g}{\partial b_i} \quad \text{and} \quad g' = \frac{\partial g}{\partial^2 b_i \partial b_i^T}$$

Then we obtain the *Laplacian Approximation* likelihood function by substitution

$$p(y|\beta, \sigma^2, \Delta) \simeq (2\pi\sigma^2)^{-N/2} |\Delta|^M \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{M} g(\beta, \Delta, y_i, \hat{b}_i) \prod_{i=1}^{M} g''(\beta, \Delta, y_i, \hat{b}_i) \right)^{-1/2}. \tag{5}$$

By approximating Hessian $g'' \simeq G$ it can be modified as

$$\begin{aligned} l_{LA}(\beta, \sigma^2, \Delta) &\simeq -\frac{N}{2}\log(2\pi\sigma^2) + \\ M \log|\Delta| &- \frac{1}{2}\left(\sum_{i=1}^{M} \log|G(\beta, \Delta, y_i)| + \frac{1}{\sigma^2} \sum_{i=1}^{M} g(\beta, \Delta, y_i, \hat{b}_i) \right)^{-1/2}. \end{aligned} \tag{6}$$

where $g''(\beta, \Delta, y_i, b_i) \simeq G(\beta, \Delta, y_i) = \left.\frac{\partial f_i}{\partial b_i}\right|_{\hat{b}_i} \left.\frac{\partial f_i}{\partial b_i^T}\right|_{\hat{b}_i} + \Delta^T \Delta$. This equation is called *modified Laplacian Approximation to the LL*. Then, for given β and Δ

$$\hat{\sigma}^2_{MLE} = \sum_{i=1}^{M} \frac{1}{N} g(\beta, \Delta, y_i, \hat{b}_i)$$

so that we profile l_{LA} on σ^2.

3.1.3 Adaptive Gaussian Approximation

To improve Laplacian approximation (5) Gaussian quadrature rules are used, which approximates integrals of functions by a weighted average of the integral evaluated predetermined abscisses such that

$$\int_{-1}^{1} f(x)dx = \sum_{i=1}^{n} w_i f(x_i)$$

It is quite technical to derive the likelihood function on this approach.

3.2 Computational Methods for Estimating Parameters

For the moment we concentrate on PNLS step in alternating algorithm. The objective function is given by (1). We add pseudo observations to simplify it. Then

$$\sum_{i=1}^{M} \left[\|y_i - f_i(\beta, b_i)\|^2 + \|\Delta b_i\|^2 \right] = \sum_{i=1}^{M} \left[\|\tilde{y}_i - \tilde{f}_i(\beta, b_i)\|^2 \right]$$

where

$$\tilde{y}_i = \begin{pmatrix} y_i \\ 0 \end{pmatrix} \qquad \text{and} \qquad \tilde{f}_i(\beta, b_i) = \begin{pmatrix} f_i(\beta, b_i) \\ \Delta b_i \end{pmatrix}$$

For this nonlinear square problem, we use *Gaussian-Newton optimization.* Replacing nonlinear $\tilde{f}$ by Taylor approximation around current estimates gives *Least-Squares problem.*

$$\sum_{i=1}^{M} \left[\|Y_i - f_i(\beta, b_i)\|^2 + \|\Delta b_i\|^2 \right] \simeq \sum_{i=1}^{M} \|\tilde{\omega}_i^{(w)} - \tilde{X}_i^{(w)} \beta - \tilde{Z}_i^{(w)} b_i\|^2$$

where

$$\tilde{\omega}_i^{(w)} = \begin{pmatrix} \hat{\omega}_i^{(w)} \\ 0 \end{pmatrix}$$

We can find least squares estimates $\hat{\beta}$ and b_i and obtain then the Gauss Newton increments. The process will be repeated until there is no change according to "step-halving".

4 Extended NLME Model

For general case we extend the basic NLME. We allow the within-group errors ε_i to be heteroscedastic or/and correlated. So our model is given by

$$\begin{gathered} y_i = f_i(\phi_i, \nu_i) + \varepsilon_i \\ \phi_i = A_i\beta + B_i b_i, \quad b_i \sim \mathcal{N}(0, \Psi) \quad \text{and} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2\Lambda_i) \end{gathered} \tag{7}$$

Reference

[1] Pinheiro, J.C. and Bates, D.M. *Mixed-Effects Models in S and S-Plus*, Springer 2000