Evaluating the log-likelihood in nonlinear mixed models

Douglas Bates

University of Wisconsin - Madison

<Bates@Wisc.edu>

Slides for this presentation are available at
lme4.R-forge.R-project.org/slides/

Joint Statistical Meetings
Washington, DC
August 6, 2009

Outline

Introduction

- Nonlinear mixed-effects models (NLMMs) are widely used in the analysis of pharmacokinetic/pharmacodynamic data.
- Over the years many algorithms for determining the parameter estimates in these models have been proposed and several implementations of algorithms have been developed.
- Many of these algorithms concentrate on the details of optimization of something like the log-likelihood. It is not always clear that such approaches should be expected to yield estimates close to the maximum likelihood estimates.
- Recent developments in computational methods for linear mixed-effects models (LMMs) provide effective ways of evaluating good approximations to the log-likelihood of an NLMM.

Outline

Evaluating the deviance function

- The *profiled deviance* function for a LMM can be expressed as a function the variance-component parameters only. We can profile out the residual variance and the fixed-effects parameters.
- We describe the probability model in terms of the n-dimensional response random variation, $\boldsymbol{\mathcal{Y}}$, whose value, $\boldsymbol{y}$, is observed, and the q-dimensional, unobserved random effects variable, $\boldsymbol{\mathcal{B}}$, with distributions

$$(\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{B}}=\boldsymbol{b}) \sim \mathcal{N}\left(\boldsymbol{Z}\boldsymbol{b}+\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{I}_n\right), \quad \boldsymbol{\mathcal{B}} \sim \mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Sigma}_\theta\right),$$

- We never really form $\boldsymbol{\Sigma}_\theta$; we always work with the *relative covariance factor*, $\boldsymbol{\Lambda}_\theta$, defined so that

$$\boldsymbol{\Sigma}_\theta = \sigma^2\boldsymbol{\Lambda}_\theta\boldsymbol{\Lambda}_\theta^\intercal.$$

  Note that we must allow for $\boldsymbol{\Lambda}_\theta$ to be less that full rank.

Orthogonal or "unit" random effects

- We will define a q-dimensional "spherical" or "unit" random-effects vector, $\mathcal{U}$, such that

$$\mathcal{U} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}_q\right), \; \mathcal{B} = \boldsymbol{\Lambda}_\theta \mathcal{U} \Rightarrow \mathsf{Var}(\mathcal{B}) = \sigma^2 \boldsymbol{\Lambda}_\theta \boldsymbol{\Lambda}_\theta^\intercal = \boldsymbol{\Sigma}_\theta.$$

- The linear predictor expression becomes

$$\boldsymbol{Z}\boldsymbol{b} + \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{Z}\boldsymbol{\Lambda}_\theta \boldsymbol{u} + \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{U}_\theta \boldsymbol{u} + \boldsymbol{X}\boldsymbol{\beta}$$

  where $\boldsymbol{U}_\theta = \boldsymbol{Z}\boldsymbol{\Lambda}_\theta$.

- The key to evaluating the log-likelihood is the Cholesky factorization

$$\boldsymbol{L}_\theta \boldsymbol{L}_\theta^\intercal = \boldsymbol{P}\left(\boldsymbol{U}_\theta^\intercal \boldsymbol{U}_\theta + \boldsymbol{I}_q\right)\boldsymbol{P}^\intercal$$

  ($\boldsymbol{P}$ is a fixed permutation that has practical importance but can be ignored in theoretical derivations). The sparse, lower-triangular $\boldsymbol{L}_\theta$ can be evaluated and can be updated when $\boldsymbol{\theta}$ is changed, even when q is in the millions and the model involves random effects for several factors.

The profiled deviance

- The Cholesky factor, $\boldsymbol{L}_\theta$, allows evaluation of the conditional mode $\tilde{\boldsymbol{u}}_{\theta,\beta}$ (also the conditional mean for linear mixed models) from

$$\left(\boldsymbol{U}_\theta^\intercal \boldsymbol{U}_\theta + \boldsymbol{I}_q\right) \tilde{\boldsymbol{u}}_{\theta,\beta} = \boldsymbol{P}^\intercal \boldsymbol{L}_\theta \boldsymbol{L}_\theta^\intercal \boldsymbol{P} \tilde{\boldsymbol{u}}_{\theta,\beta} = \boldsymbol{U}_\theta^\intercal (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

  Let $r^2(\boldsymbol{\theta},\boldsymbol{\beta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{U}_\theta\, \tilde{\boldsymbol{u}}_{\theta,\beta}\|^2 + \|\tilde{\boldsymbol{u}}_{\theta,\beta}\|^2$.
- $\ell(\boldsymbol{\theta},\boldsymbol{\beta},\sigma|\boldsymbol{y}) = \log L(\boldsymbol{\theta},\boldsymbol{\beta},\sigma|\boldsymbol{y})$ can be written

$$-2\ell(\boldsymbol{\theta},\boldsymbol{\beta},\sigma|\boldsymbol{y}) = n\log(2\pi\sigma^2) + \frac{r^2(\boldsymbol{\theta},\boldsymbol{\beta})}{\sigma^2} + \log(|\boldsymbol{L}_\theta|^2)$$

- The conditional estimate of σ^2 is

$$\widehat{\sigma^2}(\boldsymbol{\theta},\boldsymbol{\beta}) = \frac{r^2(\boldsymbol{\theta},\boldsymbol{\beta})}{n}$$

  producing the *profiled deviance*

$$-2\tilde{\ell}(\boldsymbol{\theta},\boldsymbol{\beta}|\boldsymbol{y}) = \log(|\boldsymbol{L}_\theta|^2) + n\left[1 + \log\left(\frac{2\pi r^2(\boldsymbol{\theta},\boldsymbol{\beta})}{n}\right)\right]$$

Profiling the deviance with respect to $\boldsymbol{\beta}$

- Because the deviance depends on $\boldsymbol{\beta}$ only through $r^2(\boldsymbol{\theta},\boldsymbol{\beta})$ we can obtain the conditional estimate, $\widehat{\boldsymbol{\beta}}_\theta$, by extending the PLS problem to

$$r^2(\boldsymbol{\theta}) = \min_{\boldsymbol{u},\boldsymbol{\beta}} \left[\left\| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{U}_\theta\, \boldsymbol{u} \right\|^2 + \left\| \boldsymbol{u} \right\|^2 \right]$$

with the solution satisfying the equations

$$\begin{bmatrix} \boldsymbol{U}_\theta^{\intercal}\boldsymbol{U}_\theta + \boldsymbol{I}_q & \boldsymbol{U}_\theta^{\intercal}\boldsymbol{X} \\ \boldsymbol{X}^{\intercal}\boldsymbol{U}_\theta & \boldsymbol{X}^{\intercal}\boldsymbol{X} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{u}}_\theta \\ \widehat{\boldsymbol{\beta}}_\theta \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_\theta^{\intercal}\boldsymbol{y} \\ \boldsymbol{X}^{\intercal}\boldsymbol{y}. \end{bmatrix}$$

- The profiled deviance, which is a function of $\boldsymbol{\theta}$ only, is

$$-2\tilde{\ell}(\boldsymbol{\theta}) = \log(|\boldsymbol{L}_\theta|^2) + n\left[1 + \log\left(\frac{2\pi r^2(\boldsymbol{\theta})}{n}\right)\right]$$