Semiparametric Nonlinear Mixed Effects Models

Yuedong Wang and Chunlei Ke
Department of Statistics and Applied Probability
University of California, Santa Barbara, California 93106

Abstract

NonLinear Mixed effects Models (NLMM) and SElf-MOdeling nonlinear Regression models (SEMOR) are often used to fit repeated measurement data. These models are based on the assumption that the mean response functions for all subjects have the same shape. Variation between subjects is modeled using some fixed and/or random parameters. The parametric NLMM may be too restrictive and the semi-parametric SEMOR ignores correlations within each individual. In this paper we propose a class of Semi-parametric Nonlinear Mixed effects Models (SNMM) which extends NLMM, SEMOR and many other existing models in a natural way. A SNMM contains parameters which provide interpretable data summary and some nonparametric functions which provide flexibility to leave some unknown/uncertain functions unspecified. A second stage model with fixed and random effects is used to model the parameters. Smoothing splines are used to model the nonparametric functions. Covariate effects on parameters can be built into the second stage model and covariate effects on nonparametric functions can be modeled using smoothing spline ANOVA decompositions. We extend the two-step procedure in the NLMM to estimate parameters and nonparametric functions using penalized likelihood. We illustrate usefulness of the SNMM with analyses of two real datasets.

Keywords: Curve registration; Functional data analysis; Longitudinal data; Nonlinear mixed effects models; Penalized likelihood; Repeated measurement; Self-modeling nonlinear regression; Smoothing spline; Smoothing spline ANOVA.

1 Introduction

Repeated measurement data arise in many areas of investigation, such as agriculture, pharmacokinetics, epidemiology, medicine and social science. They are generated by observing each of a number of subjects repeatedly under varying conditions where the subjects are assumed to constitute a random sample from a population of interest.

Lindstorm and Bates (1990) proposed the following NLMM:

$$\begin{aligned} y_{ij} &= \eta(\boldsymbol{\phi}_i; t_{ij}) + e_{ij}, \;\; i = 1, \ldots, m, \;\; j = 1, \cdots, n_i, \\ \boldsymbol{\phi}_i &= \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{B}_i \boldsymbol{b}_i, \;\; \boldsymbol{b}_i \overset{iid}{\sim} N(\mathbf{0}, \sigma^2 \mathbf{D}), \end{aligned} \tag{1}$$

where y_{ij} is the jth observation on the ith subject; η is a known function of a covariate t_{ij}; $\boldsymbol{\phi}_i$ is a vector of parameters which may act nonlinearly on η; e_{ij}'s are random errors, $\boldsymbol{e}_i = (e_{i1}, \cdots, e_{in_i})^T \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda}_i)$ and $\boldsymbol{e}_i$ are independent of $\boldsymbol{e}_j$ for $i \neq j$; $\mathbf{A}_i$ and $\mathbf{B}_i$ are design matrices; $\boldsymbol{\beta}$ is the population parameter vector (fixed effects); $\boldsymbol{b}_i$ are random effects for the ith subject, and $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ are assumed to be uncorrelated.

An alternative approach to model longitudinal data is SEMOR (Lawton, Sylvestre, and Maggio 1972). A SEMOR model assumes that a *common* curve f exists for all subjects and that a particular subject's response curve is some parametric transformation of the common curve:

$$y_{ij} = \eta(\boldsymbol{\phi}_i, f; t_{ij}) + e_{ij}, \quad i = 1, \cdots, m, \;\; j = 1, \cdots, n_i, \tag{2}$$

where f is the common curve; η is a known mean function; $\boldsymbol{\phi}_i$ are deterministic parameters which determine the transformation for the ith subject.

The function η in a NLMM are assumed to be known. Usually they are derived from an approximation to a mechanistic model such as a compartmental model under some simplified assumptions. The resulting model may or may not be appropriate for a dataset because the assumptions may be too restrictive and/or the approximations may be too crude. The drawbacks of a SEMOR model are that it ignores correlation among observations and has the number of parameters proportional to the number of subjects due to the deterministic nature of $\boldsymbol{\phi}_i$. Furthermore, it is difficult to investigate covariate effects on parameters and/or the common curve.

Usually the observed subjects are a random sample from a population and inferences are on the population. Therefore, as in a NLMM, it is more natural to treat the parameters $\boldsymbol{\phi}_i$ in (2) as random effects. The model we propose is a combination and extension of models (1) and (2). We include mixed effects in a SEMOR to model the within- and between-subject variation, covariate effects and variance-covariance structure naturally. Often, certain components such as the common shape function are unknown or difficult to specify. We will leave those components unspecified and estimate them nonparametrically.

2 Semi-parametric Nonlinear Mixed Effects Models

We define a class of SNMMs as

$$\begin{aligned} y_{ij} &= \eta(\boldsymbol{\phi}_i, f; t_{ij}) + e_{ij}, \quad j = 1, \cdots, n_i, \quad i = 1, \cdots, m, \\ \boldsymbol{\phi}_i &= \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \boldsymbol{b}_i, \quad \boldsymbol{b}_i \overset{iid}{\sim} N(\mathbf{0}, \sigma^2 \mathbf{D}), \end{aligned} \tag{3}$$

where t_{ij} is a covariate, which could be a scalar or a vector; η is a known function of t_{ij}, which depends on a parameter vector $\boldsymbol{\phi}_i$ of length r and an unknown function f; $\boldsymbol{\beta}$ is a deterministic p-vector of population parameters; $\boldsymbol{b}_i$ is q-vector of random effects associated with individual i; $\mathbf{A}_i$ and $\mathbf{B}_i$ are design matrices of sizes $r \times p$ and $r \times q$ for the fixed and random effects respectively; $\boldsymbol{e}_i = (e_{i1}, \cdots, e_{in_i})^T \sim N(\mathbf{0}, \sigma^2 \mathbf{\Lambda_i})$, $\boldsymbol{e}_i$ are independent of $\boldsymbol{b}_i$, and $\boldsymbol{e}_i$ are also independent of $\boldsymbol{b}_j$ and $\boldsymbol{e}_j$ for $i \neq j$.

We use smoothing spline and smoothing spline ANOVA methods to model f nonparametrically. Specifically, f belongs to an infinite dimensional model space. Which space to use depends on factors such as domain of the function, prior knowledge and purpose of the analysis.

The estimation and inference are difficult because (1) η may depends on both $\boldsymbol{\phi}_i$ and f nonlinearly; (2) The solutions of $\boldsymbol{\phi}_i$ and f depend on each other in a complicated way; (3) SNMM does not reduce to any existing model thus no software can be used directly; (4) We have extended the two-step procedure for NLMMs to fit SNMMs using penalized likelihood; (5) We have implemented our algorithms using the interface of Splus with Fortran and C.

We will demonstrate the application of SNMMs using a real data set to investigate human circadian rhythms.

Acknowledgements

This research was supported by NIH Grants R01 GM58533.

References

[1] Lawton, W. H., Sylvestre, E.A. and Maggio, M.S. (1972). Self modeling nonlinear regression. *Technometrics* **14**, 513-532.

[2] Lindstorm, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**, 673-687.