Available online at www.sciencedirect.com

ScienceDirect

Computational Statistics & Data Analysis 51 (2007) 3244–3256

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

EM algorithms for nonlinear mixed effects models

Jing Wang*

Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803, USA

Received 14 April 2006; received in revised form 15 November 2006; accepted 16 November 2006
Available online 13 December 2006

Abstract

Implementing the Monte Carlo EM algorithm (MCEM) algorithm for finding maximum likelihood estimates (MLEs) in the nonlinear mixed effects model (NLMM) has encountered a great deal of difficulty in obtaining samples used for estimating the E step due to the intractability of the target distribution. Sampling methods such as Markov chain techniques and importance sampling have been used to alleviate such difficulty. The advantage of Markov chains is that they are applicable to a wider range of distributions than the approaches based on independent samples. However, in many cases the computational cost of Markov chains is significantly greater than that of independent samplers. The MCEM algorithms based on independent samples allow for straightforward assessment of Monte Carlo error and can be considerably more efficient than those based on Markov chains when an efficient candidate distribution is chosen, which forms the motivation of this paper. The proposed MCEM algorithm in this paper uses samples obtained from an easy-to-simulate and efficient importance distribution so that the computational intensity and complexity is much reduced. Moreover, the proposed MCEM algorithm preserves the flexibility introduced by independent samples in gauging Monte Carlo error and thus allows the Monte Carlo sample size to increase with the number of EM iterations. We also introduce an EM algorithm using Gaussian quadrature approximations (GQEM) for the E step. In low-dimensional cases, the GQEM algorithm is more efficient than the proposed MCEM algorithm and thus can be used as an alternative. The performances of the proposed EM methods are compared to the existing ML estimators using real data examples and simulations.

Keywords: Nonlinear mixed effects model; Laplace; Linearization; EM; Monte Carlo; Markov chain; Importance sampling; Gauss–Hermite quadrature

1. Introduction

Nonlinear mixed effects models (NLMMs) have become more frequently used for analyzing the data consisting of repeated measurements in pharmacokinetics, growth, and other studies. Repeated measurement data consist of measurements taken on each of a number of individuals repeatedly over time or under different experimental conditions. The NLMM model describes this type of data via a nonlinear function characterizing the systematic dependence of the observations on the covariates and parameters.

Maximum likelihood estimation in the nonlinear mixed effects model brings up a substantial challenge because the likelihood of observations cannot typically be expressed in closed form. Several different approximations to the log-likelihood have been proposed, and they include the linearization approximation (Sheiner and Beal, 1980), the LME approximation (Lindstrom and Bates, 1990), and the Laplace's approximation (Wolfinger, 1993; Vonesh, 1996).

---

* Tel.: +1 225 578 7181; fax: +1 225 578 8344.
E-mail address: jwang@lsu.edu.

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.11.022

Available software packages for these methods include the NONMEM (Beal et al., 1992), the S-PLUS function NLME (Pinheiro and Bates, 1995), and the SAS procedure PROC NLMIXED.

These likelihood approximations often perform well if the number of the intraindividual measurements is not small and the variability of random effects is not large, but when some of the individuals have sparse data or the variability of the random effects is large there are considerable errors in approximating the likelihood function via these approximations (Davidian and Giltinan, 1995; Pinheiro and Bates, 1995; Lindstrom and Bates, 1990). This has motivated the use of exact methods such as Monte Carlo methods. In particular, the Monte Carlo EM algorithm (Wei and Tanner, 1990), in which the E step is approximated using simulated samples from the exact conditional distribution of the random effects given the observed data, has been used for estimation in mixed models. Walker (1996) proposes an MCEM algorithm using approximations based on samples from the distribution of the random effects for ML exact estimation in a specific class of NLMMs. McCulloch (1997) and Booth and Hobert (1999) propose different versions of the MCEM algorithm for estimation in generalized linear mixed effects models, where sampling methods including Markov chain techniques, rejection sampling, and importance sampling are used to simulate samples. A stochastic version of the EM algorithm (SAEM) using stochastic approximations involving samples obtained via Markov chains for fitting NLMMs is proposed by Kuhn and Lavielle (2005). The advantage of Markov chains is that they are applicable to a wider variety of distributions than independent samplers. However, in the NLMM setting the computational cost of Markov chains can be substantially higher than that of independent samplers due to the intractability of the target distribution. Moreover, the Monte Carlo error assessment, which is straightforward for the MCEM algorithm based on independent samples, may be an intensive and complex undertaking for the MCEM algorithm based on Markov chains. An independent sampler such as importance sampling can be considerably more efficient than Markov chains when a proper candidate distribution is chosen. These advantages of independent samplers over Markov chains form the motivation of his paper.

In this paper, we implement the MCEM algorithm using samples obtained via importance sampling from a mixture distribution chosen to be simple in the form, easy to sample from, and efficient (Lai and Shih, 2006). Moreover, because the proposed MCEM algorithm preserves the flexibility introduced by independent samples in assessing Monte Carlo error, the Monte Carlo sample size is increased with the number of EM iterations. Thus, the proposed MCEM algorithm is computationally inexpensive and efficient relative to the EM algorithms based on Markov chains.

Because Monte Carlo method is generally less efficient than quadrature in low dimensions for its inability to take account of the smoothness information built into quadrature, we introduce a Gaussian quadrature based EM method (GQEM) as an alternative to the MCEM algorithm in low-dimensional cases.

The paper is organized as follows. The model and likelihood of the NLMM are presented in Section 2. The MCEM algorithm and the GQEM algorithm are introduced in Section 3. The estimates of the log-likelihood and standard errors for the proposed EM algorithms are computed in Section 4. Section 5 compares the proposed EM algorithms with likelihood approximations and the SAEM algorithm using real data examples and simulations. Conclusions from numerical studies are given in Section 6. The paper concludes with a discussion.

2. Nonlinear mixed effects models

We consider a general class of nonlinear mixed effects models:

$$\boldsymbol{y}_{ij} = f(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{x}_{ij}) + g(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{x}_{ij}, \boldsymbol{\phi})\varepsilon_{ij},$$

where $\boldsymbol{y}_{ij}$ is the jth $(j = 1, \ldots, n_i)$ response for the ith individual $(i = 1, \ldots, n)$, $\boldsymbol{x}_{ij}$ is a vector of covariates, $\boldsymbol{\beta}(p \times 1)$ are fixed effects, $\boldsymbol{b}_i(k \times 1)$ are random effects and assumed i.i.d. $N(\mathbf{0}, \boldsymbol{D})$, and f is a nonlinear function of $(\boldsymbol{\beta}, \boldsymbol{b}_i)$. The intraindividual errors ε_{ij} are assumed i.i.d. $N(\mathbf{0}, \sigma^2)$ and independent of the $\boldsymbol{b}_i$. The variance function g is dependent on f and a parameter vector $\boldsymbol{\phi}$; a common choice of g is the power function, f^{ϕ}, reflecting the possible character of intraindividual variability.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{D}, \boldsymbol{\phi})$ denote the complete vector of unknown parameters. The marginal density of $\boldsymbol{y}_i = (y_{ij})^{\mathrm{T}}$, from which maximum likelihood estimation of $\boldsymbol{\theta}$ is obtained, can be computed by integrating over the distribution of $\boldsymbol{b}_i$:

$$p(\boldsymbol{y}_i) = \int p(\boldsymbol{y}_i, \boldsymbol{b}_i|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{b}_i = \int p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi})\,p(\boldsymbol{b}_i|\boldsymbol{D})\,\mathrm{d}\boldsymbol{b}_i.$$