Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature

Sebastian Ueckert*, France Mentré

INSERM, IAME, UMR 1137, F-75018 Paris, France
Univ Paris Diderot, Sorbonne Paris Cité, F-75018 Paris, France

ARTICLE INFO

Article history:
Received 6 October 2015
Received in revised form 13 October 2016
Accepted 14 October 2016
Available online xxxx

Keywords:
Optimal design
Fisher information matrix
Generalized linear mixed model
Nonlinear mixed effect models
Discrete data

ABSTRACT

The design of experiments for discrete mixed effect models is challenging due to the unavailability of a closed-form expression for the Fisher information matrix (FIM), on which most optimality criteria depend. Existing approaches for the computation of the FIM for those models are all based on approximations of the likelihood. A new method is presented which is based on derivatives of the exact conditional likelihood and which uses Monte Carlo (MC) simulations as well as adaptive Gaussian quadrature (AGQ) to integrate those derivatives over the data and random effects. The method is implemented in R and evaluated with respect to the influence of the tuning parameter, the accuracy of the FIM approximation, and computational complexity. The accuracy evaluation is performed by comparing the expected relative standard errors (RSE) from the MC/AGQ FIM with RSE obtained in a simulation study with four different discrete data models (two binary, one count and one repeated time-to-event model) and three different estimation algorithms. Additionally, the results from the MC/AGQ FIM are compared with expected RSE from a marginal quasi-likelihood (MQL) approximated FIM. The comparison resulted in close agreement between the MC/AGQ-based RSE and empirical RSE for all models investigated, and better performance of MC/AGQ than the MQL approximated FIM for variance parameters. The MC/AGQ method also proved to be well suited to calculate the expected power to detect a group effect for a model with binary outcomes.

1. Introduction

Mixed effect models allow to naturally capture data features arising in inherently longitudinal experiments with heterogeneity between experimental blocks. These design properties are found in many practical contexts (most notably clinical trials), which has led to an increasing utilization of mixed effect models for both continuous (often normally distributed) and discrete responses. The design of experiments for the discrete response mixed effect case is especially challenging due to the lack of a closed-form expression for the Fisher information matrix (FIM), for both discrete response generalized linear mixed effect models (GLMMs) and for discrete response nonlinear mixed effect models (NLMEMs) (this is similar to the normal case, but here effective approximations are available (Nyberg et al., 2015)). In this work, we describe a new method for evaluation of the FIM for this class of models.

There is a relatively large body of work on calculation of the FIM for GLMM. However, most articles describe methods that are based on approximate inferential methods, such as quasi-likelihood and generalized estimating equations, and are

* Corresponding author at: INSERM, IAME, UMR 1137, F-75018 Paris, France.
E-mail address: sebastian.ueckert@inserm.fr (S. Ueckert).

http://dx.doi.org/10.1016/j.csda.2016.10.011
0167-9473/

Please cite this article in press as: Ueckert, S., Mentré, F., A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.10.011

response type-specific. Examples include the work by Tekle et al., which focuses on the case of longitudinal binary data (Tekle et al., 2008), and Niaparast's article, which is limited to Poisson models with random intercepts (Niaparast, 2009). Also based on generalized estimating equations inference, but more general in terms of the distribution of the responses, is the approach presented by Woods and van de Ven (2011). For likelihood estimation of general GLMM with random intercepts, FIM approximations have been presented and evaluated in detail by Waite and Woods using marginal quasi-likelihood (MQL), penalized quasi-likelihood (PQL) or new complete enumeration-based methods, as well as Monte Carlo (MC) approximations thereof (Waite, 2012; Waite and Woods, 2015).

For discrete NLMEM, there is considerably less preexisting work concerning the calculation of the FIM. Ogungbenro and Aarons describe a method based on generalized estimating equations and the MQL approximation, adapted to binary, ordinal and count responses (Ogungbenro and Aarons, 2011). Nyberg et al. presented a method utilizing a second-order approximation of the likelihood and applied it to binary and count responses (Nyberg et al., 2009).

Here we take a slightly different approach: rather than deriving the FIM for an approximate likelihood, we determine an expression for the FIM based on the exact conditional likelihood and subsequently compute the resulting expressions using MC simulations and adaptive Gaussian quadrature (AGQ). Provided there are a sufficient number of quadrature grid points, our method evaluates the exact likelihood, it is adapted for likelihood estimation of general discrete mixed effect models (GLMM as well as NLMEM) and requires only the specification of the expression for the conditional likelihood. The proposed method is an extension of the work by Nguyen and Mentré which describes the application of an MC/AGQ-based approach to approximate the FIM for NLMEM with normally distributed responses (Nguyen and Mentré, 2014). In their work, Nguyen and Mentré show the superiority of the Monte Carlo/Adaptive Gaussian quadrature (MC/AGQ)-based approach to a linearization-based calculation of the FIM when the model nonlinearity is increasing (Nguyen and Mentré, 2014), as measured, for example, through the absolute value of the second derivative (Smyth, 2006). This finding, together with the performance of AGQ when used for parameter estimation in NLMEM (Plan et al., 2012; Jönsson et al., 2004), was our motivation to extend the approach of Nguyen and Mentré and apply it to discrete response mixed effect models.

We evaluate our method by comparing the expected relative standard errors (RSE) from the MC/AGQ-based FIM with the parameter precisions obtained by repeatedly simulating responses from a model and subsequently re-estimating the model parameters. We will refer to this procedure as clinical trial simulation (CTS) as our primary interests are models originating from the analysis of clinical trials. The use of CTS is a common approach to evaluate FIM approximations for NLMEM, due to the lack of an analytic reference expression (Ogungbenro and Aarons, 2011; Nyberg et al., 2009; Nguyen and Mentré, 2014). However, rather than limiting the comparison to a reference value obtained with one estimation algorithm, as is generally done, we obtain reference values with three different estimation algorithms: the Laplace approximation (Geweke, 1989), importance sampling (Tierney and Kadane, 1986) and stochastic approximation expectation maximization (SAEM) (Kuhn and Lavielle, 2004). These algorithms represent different approaches to handling the analytically intractable NLMEM likelihood as well as its maximization. The Laplace approximation-based estimation algorithm uses a second-order approximation of the likelihood and a gradient-based algorithm for maximization. Importance sampling approximates the likelihood through MC sampling and performs maximization using the expectation maximization algorithm. SAEM relies on Markov-Chain MC sampling in combination with the expectation maximization algorithm. All three algorithms are commonly used in practice and have been proven suitable for a wide range of models (Jönsson et al., 2004; Plan et al., 2009; Karlsson et al., 2011; Savic et al., 2010; Savic and Lavielle, 2009).

We structured our paper in three parts. Part one (Section 2) describes the derivation of an expression of the FIM for discrete mixed effect models as well as the proposed way of approximating this expression using MC simulations and AGQ. In part two (Section 3), the proposed method is evaluated in regard to the influence of tuning parameters, its performance in comparison with CTS or an MQL approximation, and the computational complexity of the method. Part three (Section 4) presents an application example that uses the MC/AGQ method to calculate the expected power to detect a group effect for a clinical trial with binary outcomes under cost constraints.

2. Fisher information matrix for discrete nonlinear mixed effect models

2.1. Notation

2.1.1. Design

Let ξ_i denote the vector of design variables for individual i ($i = 1, \ldots, N$). In most cases, a limited number K of groups of subjects with identical design variables $\xi_{\{k\}}$ exist. Hence, the population design, i.e. the superset of design variables from all subjects, can be efficiently written as the set of pairs $\Xi = \{[\xi_{\{1\}}, N_{\{1\}}], \ldots, [\xi_{\{K\}}, N_{\{K\}}]\}$ where $\xi_{\{k\}}$ is the design variable for group k and $N_{\{k\}}$ is the number of subjects in that group.

2.1.2. Nonlinear mixed effect models

This work considers NLMEM for discrete data where the conditional probability for observation j from subject i can be written as

$$P(y_{ij}|b_i) = h(y_{ij}, \xi_i, g(\mu, b_i, z_i)). \tag{1}$$