Computational Statistics and Data Analysis 71 (2014) 211–219

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

# Nonparametric estimation of the mixing distribution in logistic regression mixed models with random intercepts and slopes

Mary Lesperance [a,*], Rabih Saab [a], John Neuhaus [b]

[a] Department of Mathematics and Statistics, University of Victoria, Victoria, BC, V8W 3R4, Canada

[b] Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94143-0560, USA

ARTICLE INFO

*Article history:*
Received 30 June 2012
Received in revised form 15 May 2013
Accepted 15 May 2013
Available online 24 May 2013

*Keywords:*
Generalized linear mixed models with binary outcomes
Random effects
Direct search method
Nonparametric maximum likelihood estimation

ABSTRACT

An algorithm that computes nonparametric maximum likelihood estimates of a mixing distribution for a logistic regression model containing random intercepts and slopes is proposed. The algorithm identifies mixing distribution support points as the maxima of the gradient function using a direct search method. The mixing proportions are then estimated through a quadratically convergent method. Two methods for computing the joint maximum likelihood estimates of the fixed effects parameters and the mixing distribution are compared. A simulation study demonstrates the performance of the algorithms and an example using National Basketball Association data is provided.



## 1. Introduction

Generalized linear mixed models (GLMM's) are an extension of generalized linear models that introduce random effects to the linear predictor. The presence of packages that fit GLMM's in widely available software such as SAS (NLMIXED), Stata and R (lme4) is indicative of the extensive use of these models in modern research and data analyses. GLMM's are useful models for analysing repeated measurements and clustered observations, however, the majority of estimation methods that have been developed are based on the normality assumption of random effects (Breslow and Clayton, 1993). This assumption provides a robust and convenient way to estimate the fixed effects but may compromise estimation efficiency (Tao et al., 1999). In addition, Neuhaus et al. (1992, 2012) demonstrate that misspecification of the random effects distribution can lead to biased estimation of the intercept and covariates associated with the random effects.

A GLMM is typically specified as a conditional distribution of the data vector $\mathbf{y}$ given the random (possibly vector) effects $\boldsymbol{\gamma}$, the random effect covariate vector $\mathbf{x}$ and the fixed effect covariate vector $\mathbf{z}$. In a clustered data setting, the $n_i$ observations on the $i$th cluster, $y_{i1}, \ldots, y_{in_i}$, are conditionally independent and modelled as

$$Y_{i1}, \ldots, Y_{in_i} | \boldsymbol{\gamma}_i, \mathbf{z}_{i1}, \ldots, \mathbf{z}_{in_i}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i} \sim \prod_{j}^{n_i} f(y_{ij} | \boldsymbol{\gamma}_i, \mathbf{z}_{ij}, \mathbf{x}_{ij}), \tag{1}$$

where $\mathbf{z}_{ij}$ and $\mathbf{x}_{ij}$ are the fixed and random covariate vectors respectively characterizing observation $y_{ij}$, $i = 1, \ldots, n$.

* Corresponding author. Tel.: +1 250 721 7461; fax: +1 250 721 8962.
*E-mail address:* mlespera@uvic.ca (M. Lesperance).

0167-9473/$ – see front matter 
http://dx.doi.org/10.1016/j.csda.2013.05.014

GLMM's connect the mean $\mu_{ij} = E[y_{ij}|\boldsymbol{\gamma}_i, \mathbf{z}_{ij}, \mathbf{x}_{ij}]$ to the covariate vectors $\mathbf{x}$ and $\mathbf{z}$ through

$$g(\mu_{ij}) = \mathbf{z}_{ij}^T\boldsymbol{\beta} + \mathbf{x}_{ij}^T\boldsymbol{\gamma}_i, \tag{2}$$

where $g$ is the link function. We refer to the distribution of the random effects $G_{\boldsymbol{\gamma}}$ as the mixing distribution. Parametric mixture models assign $G_{\boldsymbol{\gamma}}$ a parametric distribution, often a normal distribution as in Stiratelli et al. (1984) and McCulloch et al. (2008). Semiparametric mixture models traditionally estimate the mixing distribution $G_{\boldsymbol{\gamma}}$ using nonparametric maximum likelihood estimation methods (NPMLE) (Laird, 1978), but these approaches have typically considered random intercepts only.

Clustered binary data arise frequently in clinical studies where repeated measurements are gathered on experimental units. An example is the study by Bent et al. (2006) in which 225 men with moderate to severe symptoms of benign prostatic hyperplasia were randomized to receive one year of treatment with either saw palmetto extract ($n = 112$) or placebo ($n = 113$). The study measured outcomes and predictors at 8 visits over a 14 month period. The outcome of interest is a binary indicator of a severe symptom, that is, a level of the American Urological Association Symptom Index score $> 20$. The predictors of interest are: the binary indicator of treatment group, month of the visit, and the month by treatment group interaction. Here it is natural to model the outcomes using patient-specific random effects.

We consider the binomial model for $f$ in (1) above and use the canonical logistic link, $g$. The random intercept and random slope vary from cluster to cluster so that the $j$th observed response of cluster $i$, $y_{ij}$ has a Bernoulli($p_{ij}$) distribution. The linear model in (2) becomes

$$\text{logit}\left(p_{ij}\right) = a_i + b_i x_{ij} + \sum_l z_{ijl}\beta_l, \tag{3}$$

where $a_i$ and $b_i$ denote the random intercept and slope for cluster $i$, respectively, and $l$ indexes the number of fixed covariates in the model. Here we assume that the random effects vector $\boldsymbol{\gamma}_i = (a_i, b_i)$ has joint distribution $G_{\boldsymbol{\gamma}}$.

Our goal is to compute the NPMLE for $G_{\boldsymbol{\gamma}}$ and the MLE for $\boldsymbol{\beta}$. Earlier methods available in the literature fit univariate random effects and include the expectation–maximization (EM) algorithm (Laird, 1978), the vertex direction method (VDM) (Fedorov, 1972; Wu, 1978a,b), the vertex exchange method (VEM) (Böhning, 1985) and the intra-simplex direction method (ISDM) (Lesperance and Kalbfleisch, 1992). Wang (2007) proposed an algorithm which is a modification of Atwood's (1976) quadratic method. He used a linear regression formulation to solve the quadratic programming sub-problem for estimating the mixing weights for which Atwood did not offer a detailed solution. Like ISDM, Wang added multiple support points at each iteration rather than one as in Atwood's algorithm, and he discarded unwanted support points that have zero mass at each iteration whereas Atwood combined nearby support points. Wang (2007) showed that the algorithm converges at a faster rate than previously published algorithms.

This paper presents a new algorithm, the Direct Search Directional Derivative (DSDD) method, that computes nonparametric maximum likelihood estimates of a mixing distribution for a logistic regression model containing multiple random effects. The algorithm uses a direct search method (Torczon, 1991) to identify maxima of the gradient function to include as mixing distribution support points. Then the algorithm incorporates the quadratically convergent method of Wang (2007) to estimate the mixing proportions.

The structure of the paper is as follows. Section 2 introduces and reviews some of the literature on NPML estimation of mixture models. In Section 3, we describe the DSDD algorithm for computing the NPMLE of $G_{\boldsymbol{\gamma}}$ with $\boldsymbol{\beta}$ fixed and compare it with an alternative. Section 4 compares two methods for computing the joint MLE's of $G_{\boldsymbol{\gamma}}$ and $\boldsymbol{\beta}$. A dataset from the National Basketball Association (NBA) is analysed using a mixed model in Section 5 and Section 6 concludes with a discussion.

## 2. Nonparametric maximum likelihood estimation

Ignoring the dependence on $\boldsymbol{\beta}$, the general mixture model has the form

$$f(y, G_{\boldsymbol{\gamma}}) = \int_{\Omega} f(y|\boldsymbol{\gamma})dG_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}), \tag{4}$$

where $f(y|\boldsymbol{\gamma})$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p) \in \Omega \subset \mathbb{R}^p$, is a density (Lindsay, 1995). Given a random sample $y_1 \cdots y_n$, the log-likelihood (4) takes the form

$$l(G_{\boldsymbol{\gamma}}) = \sum_{i=1}^{n} \log\left\{\int_{\Omega} f(y_i|\boldsymbol{\gamma})dG_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})\right\}. \tag{5}$$

Several suggestions have been provided in the literature to compute the NPMLE of the mixing distribution $G_{\boldsymbol{\gamma}}$ over the set of all possible distributions, $\mathbb{M}$. The geometry of mixture likelihoods in Lindsay (1983) provides the framework for such estimation and we provide a brief review of his work.

Let $\mathbf{L}_{\boldsymbol{\gamma}} = (L_1(\boldsymbol{\gamma}), \ldots, L_n(\boldsymbol{\gamma}))^T$ be the $n$ likelihoods corresponding to $y_1 \cdots y_n$ where $L_i(\boldsymbol{\gamma}) \propto f(y_i|\boldsymbol{\gamma})$. The log-likelihood for a given mixing distribution $G_{\boldsymbol{\gamma}}$ in (5) is

$$l(G_{\boldsymbol{\gamma}}) = \sum_{i=1}^{n} \log \int L_i(\boldsymbol{\gamma})dG_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}). \tag{6}$$