Journal of Multivariate Analysis 115 (2013) 68–83

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Semiparametric Bayesian analysis of nonlinear reproductive dispersion mixed models for longitudinal data[☆]

Nian-Sheng Tang [*], Yuan-Ying Zhao

Department of Statistics, Yunnan University, Kunming 650091, PR China

ARTICLE INFO

Article history:
Received 5 June 2011
Available online 23 September 2012

AMS subject classifications:
62F15
62H25

Keywords:
Bayesian case deletion diagnostic
Dirichlet process prior
Goodness-of-fit statistic
Model comparison
Nonlinear reproductive dispersion mixed models

ABSTRACT

In the development of nonlinear reproductive dispersion mixed models, it is commonly assumed that distribution of random effects is normal. The normality assumption is likely violated in many practical applications. In this paper, we assume that distribution of random effects is specified by a Dirichlet process prior for relaxing this limitation. A semiparametric Bayesian approach combining the stick-breaking prior and the blocked Gibbs sampler as well as the Metropolis–Hastings algorithm is developed for simulating observations from the posterior distributions and producing the joint Bayesian estimates of unknown parameters and random effects. Two goodness-of-fit statistics are presented to assess the plausibility of the posited model, and the procedures for computing the Bayes factor, pseudo-Bayes factor and deviance information criterion for model comparison are given. Also, we propose two Bayesian case deletion influence measures including the ϕ-divergence and Cook's posterior mean distance. Four simulation studies and a real example are presented to illustrate the newly developed Bayesian methodologies.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Longitudinal data [20] frequently arise in economical, biological, agricultural, environmental, medical and clinical studies, as well as many other areas of research. The most commonly used model to analyze longitudinal data is the random-effects model. Various random-effects models have been developed to account for the within-subject correlations for longitudinal data. For example, see linear mixed-effects models [27], nonlinear mixed-effects models [43], generalized linear models with random effects [48], hierarchical generalized linear models [30], semiparametric mixed models [47], nonlinear reproductive dispersion mixed-effects models [41], semiparametric reproductive dispersion mixed-effects models [5]. All the above mentioned mixed-effects models are assumed that the random effects are distributed as a multivariate normal distribution. However, in many practical applications [46,42], restricting the considered mixed-effects model to the normal distributional assumption of random effects may be unreasonable, and may lead to misleading results. To overcome the above mentioned limitations, it is interested in extending the parametric normal distributional assumption for the random effects to a class of distributions that allows for features such as skewness and multi-modality and outliers in a nonlinear reproductive dispersion mixed model (NRDMM), which is a natural extension of exponential family nonlinear mixed models and includes a wider range of random-effects models such as linear mixed-effects models, nonlinear mixed-effects, and generalized linear mixed models as its special cases.

There are considerable methods developed for accommodating normality departures in past years. For example, the normality assumption of the random effects can be relaxed by specifying the random-effects distribution with the more

☆ This work is fully supported by grants from National Natural Science Foundation of China (10961026, 11171017) and the Natural Science Key Project of Yunnan Province (No. 2010CC003).

* Corresponding author.
E-mail address: nstang@ynu.edu.cn (N.-S. Tang).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.09.005

flexible parametric families such as the finite normal mixtures [42] and the multivariate skew-normal/independent distribution [1], or a semi-nonparametric method [49], or some non-parametric methods such as the smoothing method [18] and the Dirichlet process (DP) prior assuming a normal base measure with a zero mean [22]. In particular, semiparametric Bayesian methods have been widely developed for many different models under the DP [12] prior assumption, such as the binary regression model [36], random effects models [25] and structural equation models [29]. However, there is little work done on semiparametric Bayesian inference including goodness-of-fit assessment and case deletion influence analysis for a NRDMM. Hence, the main purpose of this paper is to develop a novel semiparametric Bayesian approach to analyze a NRDMM under the truncated DP prior distribution of the random effects. Also, a hybrid algorithm that combines the stick-breaking prior and the blocked Gibbs sampler [21] as well as the Metropolis–Hastings algorithm [35,19] is developed to sample observations required in making semiparametric Bayesian inference from the posterior distributions in this paper.

Bayesian case deletion methods for identifying influential observations (or sets of observations) have been developed for some specific statistical models such as normal linear models, mixed models and survival models (see e.g., [6,3]) based on the Kullback–Leibler divergence and the conditional predictive ordinate (CPO). But, Bayesian case deletion influence diagnostics for our considered NRDMM have both theoretical and computational challenges (see e.g., Section 3.4) because of the complexity of the considered models and the random effects involved. Therefore, this article proposes Bayesian influence diagnostic measures to assess the influence of a case (or sets of observations) on the posterior distributions and posterior mean of parameter of interest based on the ϕ-divergence and Cook's posterior mean distance.

This paper is organized as follows. A semiparametric NRDMM is introduced by using the truncated Dirichlet process prior [12,21] to specify the distribution of the random effects in Section 2. In Section 3, we develop a Bayesian algorithm by combining the stick-breaking prior and the blocked Gibbs sampler [21] and the M–H algorithm [35,19] in the context of a NRDMM. Two goodness-of-fit statistics for assessing the plausibility of the posited model, and procedures for calculating the Bayes factor [24], pseudo-Bayes factor [14] and deviance information criterion [40] are also presented in Section 3. In particular, two novel Bayesian case deletion influence diagnostics for detecting influential observations in the considered data set are proposed in Section 3. Four simulation studies and a real example are used to illustrate our proposed methodologies in Section 4. Some concluding remarks are given in Section 5. Technical details are presented in Appendix.

2. Model and notation

Let y_{ij} be the outcome of the jth repeated measurement for the ith individual ($i = 1, \ldots, n, j = 1, \ldots, n_i$), and $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{iq})^T$ be a $q \times 1$ vector of the random effects corresponding to $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$. Given $\boldsymbol{b}_i$, we assume that $y_{i1}, \ldots, y_{in_i}$ are independently distributed and has the following probability density function

$$p(y_{ij}; \mu_{ij}|\boldsymbol{b}_i) = a(y_{ij}; \sigma^2) \exp\left\{-\frac{1}{2\sigma^2} d(y_{ij}; \mu_{ij})\right\}, \tag{1}$$

where μ_{ij} is the location parameter and may represent the mean of y_{ij}, $\sigma^2 \triangleq \gamma^{-1} \in \mathbb{B}(\mathbb{B} \subset \mathbb{R}^+)$ is referred to as the dispersion parameter which is known or can be estimated separately; $a(\cdot; \cdot)$ is a suitable known function; $d(y; \mu)$ is a unit deviance on $\mathbb{C} \times \Omega$ ($\Omega \subseteq \mathbb{C} \subseteq \mathbb{R}$ are open intervals) and satisfies $d(y; y) = 0\ \forall y \in \mathbb{C}$ and $d(y; \mu) > 0\ \forall y \neq \mu$, and is twice continuously differentiable with respect to (y, μ) on $\mathbb{C} \times \Omega$. The model (1) is referred to as a reproductive dispersion distributional family [23] and includes normal distribution, extreme distribution, simplex distribution and exponential family distribution as its special cases. We assume that the systematic part of the model is specified by

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

where $g(\cdot)$ is a monotonic "link" function, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a $p \times 1$ vector of regression coefficients to be estimated, $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are $\kappa \times 1$ and $q \times 1$ vectors of covariates, respectively; and $f(\boldsymbol{x}_{ij}, \boldsymbol{\beta})$ is a twice continuously differentiable function with respect to $\boldsymbol{\beta}$. For simplicity, we only consider the case that $g(\cdot)$ is a canonical link, i.e., $g(\mu_{ij}) = \mu_{ij}$, then (2) reduces to

$$\mu_{ij} = f(\boldsymbol{x}_{ij}, \boldsymbol{\beta}) + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i. \tag{3}$$

In classical random-effects models [27,48,47], it is assumed that the random effects $\boldsymbol{b}_i$ follow a multivariate standard normal distribution. However, in many applications, this assumption may be unreasonable [18,22,36,25]. Given the important role of random effects in longitudinal data analysis, the application of the random effects models with a violation of this basic crucial assumption would lead to biased statistical results [25]. Rather than the traditional normal assumption of the random effects, we specify a DP prior [12] for the distribution of $\boldsymbol{b}_i$. That is, we let $\boldsymbol{b}_i \overset{\text{i.i.d}}{\sim} \mathcal{P}$ and $\mathcal{P} \sim \mathrm{DP}(\alpha F_0)$ in which F_0 is a base distribution that serves as a starting point for constructing the nonparametric distribution, and α is a weight that indicates the researcher's certainty of F_0 as the distribution of $\boldsymbol{b}_i$.

A possible representation for specifying the DP prior $\mathrm{DP}(\alpha F_0)$ is the stick-breaking prior representation [21]. That is, $\mathcal{P} \sim \mathrm{DP}(\alpha F_0)$ can be expressed as

$$\mathcal{P}(\cdot) = \sum_{k=1}^{\infty} \pi_k \delta_{\boldsymbol{H}_k}(\cdot), \quad \boldsymbol{H}_k \overset{\text{i.i.d}}{\sim} F_0, \tag{4}$$

Download English Version:

https://daneshyari.com/en/article/1145850

Download Persian Version:

https://daneshyari.com/article/1145850

Daneshyari.com