Journal of Statistical Planning and Inference 141 (2011) 3345–3355

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Variance least squares estimators for multivariate linear mixed model

Jun Han

Mathematics and Statistics Department, Georgia State University, Atlanta, GA 30303, United States

ARTICLE INFO

Article history:
Received 30 August 2010
Received in revised form
17 April 2011
Accepted 20 April 2011
Available online 28 April 2011

Keywords:
Variance component
Covariance matrix
Variance least squares estimator
Unbiased
Non-iterative
Distribution-free
Multivariate mixed model
Matrix differentiation
Kronecker product

ABSTRACT

Non-iterative, distribution-free, and unbiased estimators of variance components by least squares method are derived for multivariate linear mixed model. A general inter-cluster variance matrix, a same-member only general inter-response variance matrix, and an uncorrelated intra-cluster error structure for each response are assumed. Projection method is suggested when unbiased estimators of variance components are not nonnegative definite matrices. A simulation study is conducted to investigate the properties of the proposed estimators in terms of bias and mean square error with comparison to the Gaussian (restricted) maximum likelihood estimators. The proposed estimators are illustrated by an application of gene expression familial study.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

In biomedical sciences, such as survival analysis and genetic epidemiology, research questions can only be fully answered if multiple repeatedly measured outcomes are jointly modeled simultaneously. In controlled clinical trials, systolic and diastolic blood pressure are often jointly modeled to detect a longitudinal treatment effect on cardiovascular diseases. In familial studies, multiple traits are jointly modeled to investigate both familial association and trait association. In AIDS infection, plasma viral load HIV RNA, marker of the immune system status CD4, and marker of inflammation β_2 microglobuline, can be modeled concurrently to explore relationship between markers, and to improve the likelihood value of joint model compared with those of separate univariate models (Thiébaut et al., 2002). For patients with polycystic kidney disease, the biomarkers for kidney structure and function are simultaneously modeled to study the joint evolution of these response variables over time. In epidemiological studies of hepatitis B virus (HBV) infected liver disease, although many HBV markers such as hepatitis B surface antigen (HBsAg), hepatitis B surface antibody (HBsAb), hepatitis Be antigen (HBeAg), hepatitis Be antibody (HBeAb), hepatitis B core antigen (HBcAg), hepatitis B core antibody (HBcAb), immunoglobulin (Ig) G and M have been identified and used in diagnosing and monitoring the progress of disease, no single marker can unambiguously diagnose the infection (Chan, 2002). Consequently, it becomes imperative to jointly model these markers for diagnostic and therapeutic purposes. For these multivariate clustered data sets, not only the correlation between repeated measurements of one outcome is taken into account, but also the

E-mail address: matjxh@langate.gsu.edu

0378-3758/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2011.04.018

correlation structure between the different outcomes should be allowed for. Therefore, random-effects model (Laird and Ware, 1982) has become a powerful and flexible tool to analyze multiple responses.

We consider the situation that several characteristics are measured on each individual at each occasion under a multivariate random-effects and error structure for repeated measurements or longitudinal observations. Let $l=1,\ldots,k$ index the characteristic or the response, $i=1,\ldots,m$ index the cluster or the subject, and $j=1,\ldots,n_{li}$ index the membership or the occasion within the cluster, for a repeated measurement or a longitudinal observation y_{lij}. In a multivariate linear mixed model, the n_{li} dimensional response vector y_{li}, the $n_{li}\times p_l$ fixed-effects design matrix X_{li} and the $n_{li}\times q_l$ random-effects design matrix Z_{li} satisfy the characteristic-cluster indexed structural relation

$$y_{li}=X_{li}\beta_l+Z_{li}b_{li}+\varepsilon_{li},\quad l=1,\ldots,k,\ i=1,\ldots,m, \tag{1}$$

where β_l is a p_l dimensional fixed-effects coefficient vector, b_{li} is a q_l dimensional random-effects coefficient vector with mean zero and finite covariance matrix D_{ll}, and ε_{li} is an n_{li} dimensional random error vector with mean zero and scalar diagonal covariance matrix Σ_{li}. Here it is assumed that for each fixed $l=1,\ldots,k$, random vectors $\{b_{li},\varepsilon_{li}\}_{i=1}^m$ are mutually uncorrelated. The measurements taken on the same cluster across different characteristics indexed by l are likely to be correlated. The dependence between different types of responses is built through the random effects, where we assume $\mathrm{Cov}(b_{li},b_{ri})=D_{lr}$ for $l\neq r$, and the error terms where $\mathrm{Var}(\varepsilon_{1ij},\ldots,\varepsilon_{kij})^T=\Lambda=(\lambda_{lr})$ for any given i and j. The variance components of random effects and error terms represent the relationships among the response variables and potential correlations among observations from the same cluster or subject. In this article, we assume that there is no serial correlation within the error term of each cluster for each response, i.e., the homogeneous within-group error assumption given by $\Sigma_{li}=\lambda_{ll}I_{n_{li}}$. Additionally, it is assumed that all k characteristics are observed simultaneously for each member (occasion) of a cluster (subject), so that $n_{li}\equiv n_i$ for any $l=1,\ldots,k$.

Let $y_i=(y_{1i1},\ldots,y_{ki1};\ldots;y_{1in_i},\ldots,y_{kin_i})^T$ be the vector of all the observed response variables for subject i, $\beta=(\beta_1^T,\ldots,\beta_k^T)^T$ be the combined fixed-effects coefficient vector, and $b_i=(b_{1i}^T,\ldots,b_{ki}^T)^T$ be the combined random-effects coefficient vector. Let X_i be the corresponding combined $kn_i\times p$ fixed-effects design matrix with $p=\sum_{l=1}^k p_l$, Z_i be the corresponding combined $kn_i\times q$ random-effects design matrix with $q=\sum_{l=1}^k q_l$, $D=(D_{lr})$ be a $q\times q$ partitioned block matrix, and $\Sigma_i=I_{n_i}\otimes\Lambda$ be a $kn_i\times kn_i$ matrix. By using these definitions, we are able to recast a multivariate linear mixed model as a cluster indexed univariate linear mixed model given by

$$y_i=X_i\beta+Z_ib_i+\varepsilon_i,\quad i=1,\ldots,m, \tag{2}$$

where $E(b_i)=0$, $\mathrm{Var}(b_i)=D$, $E(\varepsilon_i)=0$, and $\mathrm{Var}(\varepsilon_i)=\Sigma_i$. After stacking $\{y_i\}_{i=1}^m$ as y, $\{X_i\}_{i=1}^m$ as X, $\{b_i\}_{i=1}^m$ as b, $\{\varepsilon_i\}_{i=1}^m$ as ε, and diagonalizing $\{Z_i\}_{i=1}^m$ as $Z=\mathrm{diag}(Z_1,\ldots,Z_m)$, we can write the cluster indexed short form model (2) as the index-free long form model given by

$$y=X\beta+Zb+\varepsilon. \tag{3}$$

Multivariate linear mixed model with the Gaussian random effects and error terms has been considered by many authors. For example, Reinsel (1982) derived close-form estimates for multivariate random-effects model with completely observed responses and balanced designs. Shah et al. (1997) studied the multivariate mixed-effects model for longitudinal data via the EM algorithm. Schafer and Yucel (2002) investigated the computational strategies for multivariate linear mixed-effects models with missing values. Fieuws and Verbeke (2006) proposed a pair-wise modeling approach to circumvent the dimensional limitations in multivariate mixed model. The key part of estimation of linear mixed model is that of variance components. However, the normality assumptions imposed on the random effects and error terms may not be appropriate in multivariate setting when the underlying distributions are skewed or multimodal. Calvin and Dykstra (1991) developed an iterated estimation procedure, by least squares criterion and convex analysis, to produce nonnegative definite estimates of the covariance matrices in balanced multivariate variance components models. Han (2011) proposed non-iterative, distribution-free, and unbiased estimators of variance components, including minimum norm quadratic unbiased estimators and method of moments estimators, for multivariate linear mixed model. In this article, we will develop non-iterative, distribution-free, and unbiased estimators of variance matrices via least squares method for multivariate linear mixed model from the perspective of matrix derivatives. A general inter-cluster variance matrix, a same-member only general inter-response variance matrix, and an uncorrelated intra-cluster error structure for each response are assumed.

The outline of this paper is as follows. Section 2 presents the variance least squares estimators and the unbiased variance least squares estimators for variance components. Section 3 performs a simulation study to investigate the properties of the proposed estimators with comparison to the Gaussian (restricted) maximum likelihood estimators. Section 4 illustrates the proposed estimators by a gene expression familial data. We conclude with some remarks in Section 5.

The following matrix identities involving vec operator and Kronecker product (Henderson and Searle, 1979) will prove useful,

$$\mathrm{vec}(ABC)=(C^T\otimes A)\mathrm{vec}(B), \tag{4}$$

$$(A\otimes B)(C\otimes D)=AC\otimes BD, \tag{5}$$

Download English Version:

https://daneshyari.com/en/article/1147632

Download Persian Version:

https://daneshyari.com/article/1147632

Daneshyari.com