Journal of the Korean Statistical Society 42 (2013) 149–160

Contents lists available at SciVerse ScienceDirect

Journal of the Korean Statistical Society

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

A linear mixed model for analyzing longitudinal skew-normal responses with random dropout

M. Ganjali *, T. Baghfalaki, M. Khazaei

Department of Statistics, Shahid Beheshti University, G. C., Tehran, Iran

ARTICLE INFO

Article history:
Received 19 June 2011
Accepted 25 June 2012
Available online 15 July 2012

AMS 2000 subject classifications:
62P10

Keywords:
Bootstrap
Dropout
ECM algorithm
Empirical Bayes
Linear mixed model
Longitudinal data
Skew-normal distribution

ABSTRACT

In this paper, a linear mixed effects model is used to fit skewed longitudinal data in the presence of dropout. Two distributional assumptions are considered to produce background for heavy tailed models. One is the linear mixed model with skew-normal random effects and normal errors and the other one is the linear mixed model with skew-normal errors and normal random effects. An ECM algorithm is developed to obtain the parameter estimates. Also an empirical Bayes approach is used for estimating random effects. A simulation study is implemented to investigate the performance of the presented algorithm. Results of an application are also reported where standard errors of estimates are calculated using the Bootstrap approach.

1. Introduction

Longitudinal studies represent one of the principal research strategies employed in medical and social research. The defining feature of such studies is that subjects are measured repeatedly through time. The key point in longitudinal data analysis is the important fact that correlations between responses of the same individual should be taken into account. A pervasive problem that arises in the context of analysis of longitudinal data is the presence of missing data. In some cases, a subject may be missing one of several measurement occasions; however, it is more likely that there are missing data due to drop-out, which refers to a subject removing from the study, prior to the end of the study. Consequently, the data record for this subject prematurely terminates.

Rubin (1976) provided a framework for the incomplete data by introducing the important taxonomy of missing data mechanisms, consisting of *missing completely at random* (MCAR), *missing at random* (MAR) and *missing not at random* (MNAR). A mechanism is said MCAR, if missing values are independent of both unobserved and observed data, MAR if, conditional on the observed data, the missing values are independent of the missing measurements and otherwise the missing process is termed MNAR. In addition of the above definitions, Diggle and Kenward (1994) defined a dropout process to be completely random dropout (CRD) if dropout is not dependent on observed and unobserved responses; random dropout (RD) if dropout, given the observed responses, is independent of the missing responses; and otherwise as nonrandom dropout (NRD).

Linear mixed-effects models provide a class of models for the analysis of longitudinal data. The model takes the form

$$Y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i + \boldsymbol{\varepsilon}_{ij}, \quad j : 1, 2, \ldots, T_i, \ i : 1, 2, \ldots, m$$

* Corresponding author. Tel.: +98 21 29902915.
E-mail addresses: m-ganjali@sbu.ac.ir, m_ganjali43@yahoo.com (M. Ganjali), t.baghfalaki@yahoo.com (T. Baghfalaki), m-khazaei@sbu.ac.ir (M. Khazaei).

1226-3192/$ – see front matter
doi:10.1016/j.jkss.2012.06.004

where Y_{ij} is the jth response of the ith individual, b_i is a $q \times 1$ vector of random effects, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects parameter, ε_{ij} are *i.i.d.* $N(0, \sigma^2)$, $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector of covariates and $\boldsymbol{z}_{ij}$ is a $q \times 1$ vector of covariates. The model can also be expressed in matrix notation as follows:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iT_i})'$, $\boldsymbol{\beta}$ $(p \times 1)$ is a vector of fixed effects associated with covariate matrix $\boldsymbol{X}_i$, $\boldsymbol{b}_i$ are q-dimensional, mutually independent subject-specific random effects associated with covariate matrix $\boldsymbol{Z}_i$ and $\boldsymbol{\varepsilon}_i \sim N_{T_i}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ are errors independent of the $\boldsymbol{b}_i$.

In the current literature, the distribution of random effects is routinely assumed to be normal, however, in recent years violation of such normality has been reported in many data analysis. For example, Pinheiro, Liu, and Wu (2001) pointed out that the distribution of random effects appeared to have heavier tails than the normal in the orthodontic data analysis, Zhang and Davidian (2001) found that the random intercept followed a positively skewed distribution in their model for Framingham cholesterol data, and Ishwaran and Takahara (2002) indicated that the distribution of random effects deviated from normality due to negative skewness and positive kurtosis in their analysis of chronic renal disease data.

In most of the clinical studies each individual measurement may be measured repeatedly over time, therefore for each individual, data are collected at multiple time points. In these studies, properties of resources which create random effects may be causes of skew random effects. For example, in mastitis data (which are described in Section 6) animal's genetic characteristics (which can be considered as random effects) may have a skew distribution, also random effects may arise due to some omitted variables such as weight, age, etc. These may make random effects to have skew distribution. Another example in longitudinal data can be found in HIV studies. In these studies, the response variable (in most of them CD4 count measurement or viral load) tends to small quantities over time, that is, each component of it has left skewness.

During the last decade, there has been a growing interest in the construction of flexible parametric classes of distributions exhibiting skewness which is the so-called skew-normal distribution.

The first version of the skew-normal distribution was given by Azzalini (1985, 1986). Multivariate skew-normal distributions are used by Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999) and Branco and Dey (2001). Also some applications of univariate skew-normal distribution are given by Arellano-Valle, Ozan, Bolfarine, and Lachos (2005) where one can find an application of skew-normal in measurement error models. Cancho, Lachos, and Ortega (2008) give applications of this distribution for nonlinear regression models. Besides, one can find some application of multivariate skew-normal distribution in Arellano-Valle, Bolfaine, and Lachos (2005), Lachos, Bolfarine, Arellano-Valle, and Montenegro (2007), Lin and Lee (2008), and Sahu, Dey, and Branco (2003).

In this paper, we would like to use multivariate skew-normal distribution for analysis longitudinal data with random or completely random dropout. The skew-normal distribution that we will use in this work is a slightly modified version of the one proposed by Azzalini and Dalla-Valle (1996), which is a special case of the fundamental skew-normal distribution proposed by Arellano-Valle and Genton (2005). We shall present an Expectation Conditional Maximization (ECM) algorithm to find parameter estimates of the model.

Specifically, we say that a k-dimensional random vector $\boldsymbol{Y}$ has a multivariate skew-normal distribution with skewness vector $\boldsymbol{\lambda}$, location vector $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Psi}$, if its probability density function is given by

$$f(\boldsymbol{y}) = 2\phi_k(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Psi})\Phi_1(\boldsymbol{\lambda}'\boldsymbol{\Psi}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})) \tag{2}$$

where $\phi_k(.|\boldsymbol{\mu}, \boldsymbol{\Psi})$ stands for the pdf of the k-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Psi}$ shown as $N_k(\boldsymbol{\mu}, \boldsymbol{\Psi})$ and $\Phi_1(.)$ is the univariate cumulative standard normal distribution function. The distribution given in (2) will be denoted by $\boldsymbol{Y} \sim SN_k(\boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\lambda})$ which has the following stochastic representation (a form that helps one to easily generate samples of this distribution):

$$\boldsymbol{Y} \overset{d}{=} \boldsymbol{\mu} + \boldsymbol{\Psi}^{1/2}(\delta|X_0| + (\boldsymbol{I}_k - \delta\delta')^{1/2}\boldsymbol{X}_1), \quad \delta = \frac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}'\boldsymbol{\lambda}}} \tag{3}$$

where $X_0 \sim N(0, 1)$ and $\boldsymbol{X}_1 \sim N_k(\boldsymbol{0}, \boldsymbol{I}_k)$ are independent. For more details on this approach see Arellano-Valle and Genton (2005), Arellano-Valle and Bolfaine et al. (2005) and Arellano-Valle and Ozan et al. (2005).

The EM algorithm is a general-purpose iterative algorithm to find maximum likelihood estimates in parametric models for incomplete data, where an algorithm such as the Newton–Raphson method may turn out to be more complicated. Within each iteration of the EM algorithm, there are two steps, called the expectation step, or E-step, and the maximization step, or M-step. The name EM algorithm was given by Dempster, Laird, and Rubin (1977), who provided a general and unified formulation of the EM algorithm, its basic properties, and many examples and applications of it. The books by Little and Rubin (2002), McLachlan and Krishnan (2008), and Schafer (1997) provide detailed descriptions and applications of the EM algorithm.

The ECM algorithm as proposed by Meng and Rubin (1993), is a natural extension of the EM algorithm in situations where the maximization process on the M-step is relatively simple when one conditions on some function of the parameters under estimation. The ECM algorithm therefore replaces the M-step of the EM algorithm by a number of computationally simpler conditional maximization (CM) steps. As a consequence, it typically converges more slowly than the EM algorithm in terms of