Flexible Modeling of Random Effects in Linear Mixed Models using Skew-Normal Distribution

Clécio da Silva Ferreira[a], **Celso R. B. Cabral**[b] **and Victor H. Lachos**[c] [1]

a Departamento de Estatística, Universidade Federal de Juiz de Fora, Minas Gerais, Brazil
b Departamento de Estatística, Universidade Federal de Amazonas, Amazonas, Brazil
c Departamento de Estatística, Universidade Estadual de Campinas, São Paulo, Brazil

Abstract

Flexible modelling of random effects in linear mixed models has attracted some attention recently. Following Verbeke and Lesaffre (1996), we propose the use of finite skew-normal mixtures which includes the finite normal mixtures as a special case and provides flexibility in capturing a broad range of non-normal behavior, controlled by a tuning parameter that controls the skewness of the mixture components. Likelihood based inference is adopted since the marginal likelihood may be expressed in closed form and an EM algorithm for maximum likelihood estimation is developed which results in an analytically tractable maximization step. In addition, we offer a general information-based method for obtaining the asymptotic covariance matrix of maximum likelihood estimates. Numerical results on simulated and real data sets are provided to demonstrate the usefulness of the proposed methodology.

Key Words: *EM algorithm; Linear Mixed Models; Mixing proportion; Skew-normal distributions.*

1 Introduction

Traditionally, the random effects in linear mixed models (LMM) are assumed to follow a multivariate Gaussian distribution for mathematical convenience. Although some studies suggest that inference on fixed effects may be robust to non-normality of random effects (Verbeke and Lesaffre, 1997), there are also findings of inconsistencies in fixed and random effects estimations under misspecification of random effects distributions (Neuhaus et al., 1992). Moreover, modelling the random effects with less restrictive distributional assumptions may provide important insights. A skewed or even multimodal random effect may indicate exclusion of important factors and suggest improvements in model settings. Thus, considerable interest has focused on relaxing the normality assumption and jointly estimating the random effects distribution and model parameters. Examples, in nonparametric or semi-parametric estimations of the random effects include, the discrete nonparametric MLE (Laird, 1978; Lindsay, 1983), the smooth nonparametric MLE (Magder and Zeger, 1996) and predictive recursive estimation (Tao et al., 1999). From a parametric point of view Verbeke and Lesaffre (1996) use a mixture of normals via EM algorithm, Ma et al. (2004) consider a generalized flexible skew–elliptical distribution for the random effects density. Arellano–Valle et al. (2005), Lin and Lee (2008) and Lachos et al. (2007) proposed a skew–normal linear mixed model (SN–LMM) based on multivariate skew-normal (SN) distribution introduced by Azzalini and Dalla–Valle (1996). More recently, Lachos et al. (2010) propose the skew-normal/independent linear mixed model and Ho and Hu (2008) propose the use of finite

[1] *Address for correspondence:* Víctor Hugo Lachos Dávila, Departamento de Estatística, IMECC, Universidade de Campinas, Caixa Postal 6065, CEP 13083-859, Campinas, SP, Brazil.
E-mail: hlachos@ime.unicamp.br.

Gaussian mixtures as in Verbeke and Lesaffre (1996).

In this paper, we provide a flexible hierarchical mixture modeling framework for the random effects and the model errors using the multivariate skew-normal distribution. In particular, we assume a finite mixture of skew-normal (FM-SN) distributions for the random effects and a normal distribution for the within–subject errors. Together, the observed responses follow a FM-SN distribution and define what we call finite mixture of skew–normal linear mixed models (FM–SNLMM). The marginal density of the observed quantities is obtained analytically by integrating out the random effects, leading to an observed (marginal) likelihood function. The hierarchical representation of the proposed model makes possible the implementation of an efficient EM–type algorithm for carrying out maximum likelihood estimation. The EM algorithm for FM–SNLMM presented here has analytically tractable E and M steps.

2 The Model and Main Notation

2.1 Preliminaries

In this section we present the skew-normal distribution and some interesting properties. We say that a $p \times 1$ random vector $\mathbf{Y}$ follows a SN-distribution with $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive definite dispersion matrix $\boldsymbol{\Sigma}$ and $p \times 1$ skewness parameter vector $\boldsymbol{\lambda}$, and we write $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, if its probability density function (pdf) is given by

$$f(\mathbf{y}) \quad = \quad 2\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \tag{1}$$

where $\phi_p(.; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the pdf of the p–variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$, $N_{\mathrm{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ say, and $\Phi(.)$ represents the cumulative distribution function (cdf) of the standard univariate normal distribution.

2.2 Skew-normal mixture random effects

As in Ho and Hu (2008), we consider a two stages linear mixed model that incorporates a finite mixture model of SN distributions within the family defined in (1). The model can be written as

$$\mathbf{Y}_i|\mathbf{b}_i, \quad \overset{\text{ind}}{\sim} \quad N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \sigma_e^2\mathbf{I}_{n_i}), \tag{2}$$

$$\mathbf{b}_i \quad \overset{\text{iid}}{\sim} \quad \sum_{k=1}^{G} w_k SN_q(\boldsymbol{\nu}_k, \mathbf{D}_k, \boldsymbol{\lambda}_k), \tag{3}$$

where $i = 1, \ldots, n$, $k = 1, \ldots, G$, $\boldsymbol{\nu}_k = k_1\boldsymbol{\Delta}_k + \boldsymbol{\mu}_k$, $\mathbf{Y}_i$ is a $(n_i \times 1)$ vector of observed continuous responses for sample unit i, $\mathbf{X}_i$ of dimension $(n_i \times p)$ is the design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$ of dimension $(p \times 1)$ is a vector of population-averaged regression coefficients called fixed effects, $\mathbf{Z}_i$ of dimension $(n_i \times q)$ is the design matrix corresponding to the $(q \times 1)$ random effects vector $\mathbf{b}_i$. At the second stage given by (3), a finite mixture model with G multivariate skew-normal components is used to describe the population distribution. The weight w_k are nonnegative number summing to one, denoting the relative size of each mixing component (subpopulation), $\boldsymbol{\mu}_k \in \mathbb{R}^q$ is the mean vector, $\mathbf{D}_k \in \mathbb{R}^{q \times q}$ is the positive definite scale matrix that depend on unknown and reduced parameters $\boldsymbol{\alpha}_k$ and $\boldsymbol{\lambda}_k \in \mathbb{R}^q$ is skewness vector . Letting $\boldsymbol{\theta}$ represent the collection of parameters $\{\boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\theta}^k, k = 1, \dots, G\}$, with $\boldsymbol{\theta}^k = (w_k, \boldsymbol{\mu}_k, \boldsymbol{\alpha}_k, \boldsymbol{\lambda}_k)$, the population problem involves estimating the overall data likelihood L with respect to $\boldsymbol{\theta}$. Under the *i.i.d.* assumption of individual parameters $\mathbf{b}_i$ and the result of Proposition 1 given in Lachos et al. (2007), the log-likelihood function for $\boldsymbol{\theta}$ given the observed sample $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$ is given by

$$\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} \log \left\{ \sum_{k=1}^{G} w_k 2\phi_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\nu}_k, \boldsymbol{\Psi}_{ik})\Phi_1\left(\bar{\boldsymbol{\lambda}}_{ik}^\top \boldsymbol{\Psi}_{ik}^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\nu}_k)\right)\right\}, \quad (4)$$

where $\boldsymbol{\Psi}_{ik} = \sigma_e^2\mathbf{I}_{n_i} + \mathbf{Z}_i\mathbf{D}_k\mathbf{Z}_i^\top$, $\bar{\boldsymbol{\lambda}}_{ik} = \dfrac{\boldsymbol{\Psi}_{ik}^{-1/2}\mathbf{Z}_i\mathbf{D}_k^{1/2}\boldsymbol{\lambda}_k}{\sqrt{1 + \boldsymbol{\lambda}_k^\top \mathbf{D}_k^{-1/2}\boldsymbol{\Lambda}_{ik}\mathbf{D}_k^{-1/2}\boldsymbol{\lambda}_k}}$ and $\boldsymbol{\Lambda}_{ik} = (\mathbf{D}_k^{-1} + \sigma_e^{-2}\mathbf{Z}_i^\top\mathbf{Z}_i)^{-1}$, i.e., each $\mathbf{y}_i \overset{\text{ind}}{\sim} \sum_{k=1}^{G} w_k SN_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\nu}_k, \boldsymbol{\Psi}_{ik}, \bar{\boldsymbol{\lambda}}_{ik})$, for $i = 1, \dots, n$. **Remarks:**

i) Note that, finite mixture models on the random effects can be used to provide flexible parametric density estimation of the random effects and also to flexible modeling of heterogeneous data with asymmetric behavior. Skew-normal mixture is especially attractive due to the mathematical tractability and its hierarchical Gaussian nature. In a likelihood based context, it means that we can construct a efficient EM algorithm to maximum likelihood estimation as will be seen in the next section.
ii) When $\boldsymbol{\lambda}_k = 0$, for $k = 1, \dots, G$, the FMSN–LMM reduces to the heterogeneous FM-LMM proposed by Verbeke and Lesaffre (1996) and Ho and Hu (2008) since

$$\mathbf{y}_i \overset{\text{ind}}{\sim} \sum_{k=1}^{G} w_k N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\mu}_k, \boldsymbol{\Psi}_{ik}),$$

for $i = 1, \dots, n$.

3 Solution via the EM algorithm

With the inclusion of allocation variables $\mathbf{z}_i = (Z_{i1}, \cdots, Z_{iG})^\top$ and latent variables $t_i's$, $i = 1, \dots, n$, (see Cabral et al., 2009) the hierarchical representation of (2)-(3) is given by

$$\begin{aligned}
\mathbf{Y}_i|\mathbf{b}_i, z_{ik} = 1 &\overset{\text{ind}}{\sim} N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \sigma_e^2\mathbf{I}_{n_i}), && (5)\\
\mathbf{b}_i|T_i = t_i, z_{ik} = 1 &\overset{\text{ind}}{\sim} N_q(\boldsymbol{\mu}_k + \boldsymbol{\Delta}_k t_i, \boldsymbol{\Gamma}_k), && (6)\\
T_i|z_{ik} = 1 &\overset{\text{iid}}{\sim} TN(k_1, 1)I(k_1, \infty), &&\\
\mathbf{z}_i &\sim M(1; w_1, \dots, w_G), && (7)
\end{aligned}$$

$i = 1, \dots, n$, $k = 1, \dots, G$, all independent, where $\boldsymbol{\Delta}_k = \mathbf{D}_k^{1/2}\boldsymbol{\delta}_k$, $\boldsymbol{\Gamma}_k = \mathbf{D}_k - \boldsymbol{\Delta}_k\boldsymbol{\Delta}_k^\top$, $\boldsymbol{\delta}_k = \boldsymbol{\lambda}_k/(1 + \boldsymbol{\lambda}_k^\top\boldsymbol{\lambda}_k)^{1/2}$, $M(1; w_1, \dots, w_G)$ denotes the multinomial distribution and $TN_1(r, s)I(a, b)$ and denotes the truncated univariate normal distribution on (a, b), with parameters values in parenthesis before truncation. The "complete" data are then represented by $\mathbf{Y}_c = \{(\mathbf{Y}_i^\top, \mathbf{b}_i^\top, \mathbf{z}_i^\top, t_i)^\top, i = 1, \dots, n,\}$ with $\{\mathbf{b}_i, \mathbf{z}_i, t_i\}$ representing the "missing" data.

The algorithm starts with $\widehat{\boldsymbol{\theta}}^{(0)}$ and moves from $\widehat{\boldsymbol{\theta}}^{(r)}$ to $\widehat{\boldsymbol{\theta}}^{(r+1)}$ at the r-th iteration. At the E-step, define

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)}) = E\{\log L_c(\boldsymbol{\theta})|\mathbf{Y}, \widehat{\boldsymbol{\theta}}^{(r)}\},$$

where $L_c(\boldsymbol{\theta})$ is the complete data likelihood given by

$$\log L_c(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{k=1}^{G} z_{ik}\log w_k f(\mathbf{Y}_i, \mathbf{b}_i, t_i|\boldsymbol{\beta}, \sigma_e^2, \mathbf{D}_k, \boldsymbol{\mu}_k, \boldsymbol{\lambda}_k).$$

Letting $\widehat{z}_{ik}^{(r)} = E\{Z_{ik}|\mathbf{y}_i, \widehat{\boldsymbol{\theta}}^{(r)}\}$, $\widehat{\mathbf{b}}_{ik}^{(r)} = \mathrm{E}\{\mathbf{b}_i|\widehat{\boldsymbol{\theta}}^{(r)}, \mathbf{y}_i, z_{ik}=1\}$, $\widehat{\boldsymbol{\Omega}}_{ik}^{(r)} = \mathrm{Cov}\{\mathbf{b}_i|\widehat{\boldsymbol{\theta}}^{(r)}, \mathbf{y}_i, z_{ik}=1\}$, $\widehat{t}_{ik}^{(r)} = \mathrm{E}\{T_i|\widehat{\boldsymbol{\theta}}^{(r)}, \mathbf{y}_i, z_{ik}=1\}$, $\widehat{t2}_{ik}^{(r)} = \mathrm{E}\{T_i^2|\widehat{\boldsymbol{\theta}}^{(r)}, \mathbf{y}_i, z_{ik}=1\}$ and $\widehat{\mathbf{tb}}_{ik} = \mathrm{E}\{T_i\mathbf{b}_i|\widehat{\boldsymbol{\theta}}^{(r)}, \mathbf{y}_i, z_{ik}=1\}$. Thus, given the current estimate $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}$, the E-step is given by

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)}) = Q_1(\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}^{(r)}) + Q_2(\sigma_e^2, \boldsymbol{\beta}|\widehat{\boldsymbol{\theta}}^{(r)}) + Q_3(\boldsymbol{\theta}^G|\widehat{\boldsymbol{\theta}}^{(r)}),$$

with

$$Q_1(\boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}^{(r)}) = \sum_{i=1}^{n}\sum_{k=1}^{G}\widehat{z}_{ik}^{(r)}\log(w_i),$$

$$\begin{aligned}
Q_2(\sigma_e^2, \boldsymbol{\beta}|\widehat{\boldsymbol{\theta}}^{(r)}) &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{G}\widehat{z}_{ik}^{(r)}\{n_i\log(\sigma_e^2) \\
&\quad + \frac{1}{\sigma_e^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\widehat{\mathbf{b}}_{ik}^{(r)})^\top(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\widehat{\mathbf{b}}_{ik}^{(r)}) + \mathrm{tr}(\sigma_e^{-2}\mathbf{Z}_i\widehat{\boldsymbol{\Omega}}_{ik}^{(r)}\mathbf{Z}_i^\top)\},
\end{aligned}$$

$$\begin{aligned}
Q_3(\boldsymbol{\theta}^k|\widehat{\boldsymbol{\theta}}^{(r)}) &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{G}\widehat{z}_{ik}^{(r)}\{\log|\boldsymbol{\Gamma}_k| + \mathrm{tr}\Big(\boldsymbol{\Gamma}_k^{-1}(\widehat{\boldsymbol{\Omega}}_{ik}^{(r)} + \widehat{\mathbf{b}}_{ik}^{(r)}\widehat{\mathbf{b}}_{ik}^{(r)\top} \\
&\quad -2\widehat{\mathbf{b}}_{ik}^{(r)}\boldsymbol{\mu}_k^\top + 2\widehat{t}_{ik}^{(r)}\boldsymbol{\Delta}_k\boldsymbol{\mu}_k^\top - 2\widehat{\mathbf{tb}}_{ik}^{(r)}\boldsymbol{\Delta}_k^\top + \widehat{t2}_{ik}^{(r)}\boldsymbol{\Delta}_k\boldsymbol{\Delta}_k^\top + \boldsymbol{\mu}_k\boldsymbol{\mu}_k^\top)\Big)\},
\end{aligned}$$

where by Bayes' Theorem,

$$\widehat{z}_{ik}^{(r)} = \frac{\widehat{w}_k^{(r)}2\phi_{n_i}(\mathbf{y}_i|\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(r)} + \mathbf{Z}_i\widehat{\boldsymbol{\nu}}_k^{(r)}, \widehat{\boldsymbol{\Psi}}_{ik}^{(r)})\Phi\left(\widehat{\bar{\boldsymbol{\lambda}}}_{ik}^{(r)}\widehat{\boldsymbol{\Psi}}_{ik}^{(r)-1/2}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(r)} - \mathbf{Z}_i\widehat{\boldsymbol{\nu}}_k^{(r)})\right)}{\sum_{k=1}^{G}\widehat{w}_k^{(r)}2\phi_{n_i}(\mathbf{y}_i|\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(r)} + \mathbf{Z}_i\widehat{\boldsymbol{\nu}}_k^{(r)}, \widehat{\boldsymbol{\Psi}}_{ik}^{(r)})\Phi\left(\bar{\boldsymbol{\lambda}}_{ik}^{(r)}\widehat{\boldsymbol{\Psi}}_{ik}^{(k)-1/2}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(r)} - \mathbf{Z}_i\widehat{\boldsymbol{\nu}}_k^{(r)})\right)},$$

and from Lachos et al. (2007)

$$\begin{aligned}
\widehat{t}_{ik}^{(r)} &= (\widehat{\mu}_{Tik}^{(r)} + k_1) + W_\Phi\Big(\frac{\widehat{\mu}_{T_{ik}}^{(r)}}{\widehat{M}_{T_{ik}}^{(r)}}\Big)\widehat{M}_{T_{ik}}^{(r)}, \\
\widehat{t2}_{ik}^{(r)} &= (\widehat{\mu}_{T_{ik}}^{(r)} + k_1)^2 + \widehat{M}_{T_{ik}}^{2(r)} + W_\Phi\Big(\frac{\widehat{\mu}_{T_{ik}}^{(r)}}{\widehat{M}_{T_{ik}}^{(r)}}\Big)\widehat{M}_{T_{ik}}^{(r)}(\widehat{\mu}_{T_{ik}}^{(r)} + 2k_1), \\
\widehat{\mathbf{b}}_{ik}^{(r)} &= \widehat{\mathbf{r}}_{ik}^{(r)} + \widehat{\mathbf{s}}_{ik}^{(r)}\,\widehat{t}_{ik}^{(r)}, \quad \widehat{\boldsymbol{\Omega}}_{ik}^{(r)} = \widehat{\mathbf{T}}_{b_{ik}}^{2(r)} + \widehat{\mathbf{s}}_{ik}^{(r)}\widehat{\mathbf{s}}_{ik}^{\top(r)}(\widehat{t2}_{ik}^{(r)} - (\widehat{t}_{ik}^{(r)})^2), \\
\widehat{\mathbf{tb}}_{ik}^{(r)} &= \widehat{\mathbf{r}}_{ik}^{(r)}\,\widehat{t}_{ik}^{(r)} + \widehat{\mathbf{s}}_{ik}^{(r)}\,\widehat{t2}_{ik}^{(r)},
\end{aligned} \tag{8}$$

where, omitting the supraindex (r), $\widehat{M}_{T_{ik}}^2 = [1 + \widehat{\boldsymbol{\Delta}}_k^\top\mathbf{Z}_i^\top(\widehat{\sigma_e^2}\mathbf{I}_{n_i} + \mathbf{Z}_i\widehat{\boldsymbol{\Gamma}}_k\mathbf{Z}_i^\top)^{-1}\mathbf{Z}_i\widehat{\boldsymbol{\Delta}}_k]^{-1}$, $\widehat{\mu}_{T_{ik}} = \widehat{M}_{T_{ik}}^2\widehat{\boldsymbol{\Delta}}_k^\top\mathbf{Z}_i^\top(\widehat{\sigma_e^2}\mathbf{I}_{n_i} + \mathbf{Z}_i\widehat{\boldsymbol{\Gamma}}_k\mathbf{Z}_i^\top)^{-1}(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i\widehat{\boldsymbol{\nu}}_k)$, $\widehat{\mathbf{T}}_{ik}^2 = [\widehat{\boldsymbol{\Gamma}}_k^{-1} + \widehat{\sigma}_e^{-2}\mathbf{Z}_i^\top\mathbf{Z}_i]^{-1}$, $\widehat{\mathbf{r}}_{ik} = \boldsymbol{\mu}_k + \widehat{\sigma}_e^{-2}\widehat{\mathbf{T}}_{ik}^2\mathbf{Z}_i^\top(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \mathbf{Z}_i\widehat{\boldsymbol{\mu}}_k)$, $\widehat{\mathbf{s}}_{ik} = (\mathbf{I}_q - \widehat{\sigma}_e^{-2}\widehat{\mathbf{T}}_{ik}^2\mathbf{Z}_i^\top\mathbf{Z}_i)\widehat{\boldsymbol{\Delta}}_k$.

Then the EM algorithm for ML estimation of $\boldsymbol{\theta}$ is implemented as follows:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}$, compute $\widehat{\mathbf{b}}_{ik}^{(r)}$, $\widehat{\boldsymbol{\Omega}}_{ik}^{(r)}$, $\widehat{t}_{ik}^{(r)}$, $\widehat{t2}_{ik}^{(r)}$, $\widehat{\mathbf{tb}}_{ik}^{(r)}$, for $i = 1, \dots, n$ and $k = 1, \dots, G$.
CM-step: Update $\widehat{\boldsymbol{\theta}}^{(r)}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)})$ over $\boldsymbol{\theta}$, which leads to the closed form expressions (omitted in this summary).

The observed information matrix

We let $\mathbf{J}_o(\boldsymbol{\theta}|\mathbf{y}) = -\partial^2 \ell(\boldsymbol{\theta}|\mathbf{y})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ be the observed information matrix for the FM-SNLMM defined in (2)-(3). Under some regularity conditions, the covariance matrix of ML estimates $\widehat{\boldsymbol{\theta}}$ can be approximated by the inverse of $\mathbf{J}_o(\boldsymbol{\theta}|\mathbf{y})$. We follow Basford et al. (1997) to provide an information-based method to obtain the asymptotic covariance of ML estimates of the FM-SNLMM parameters. Thus we define by

$$\mathbf{J}_o(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^\top \tag{9}$$

the observed information matrix, where $\widehat{\mathbf{s}}_i = \partial \left(\ell(\boldsymbol{\theta}|\mathbf{y})\right) / \partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$, with $\ell(\boldsymbol{\theta}|\mathbf{y})$ as in (4). We can consider now the vector $\widehat{\mathbf{s}}_i$ partitioned into components corresponding with all parameters in $\boldsymbol{\theta}$ as

$$\widehat{\mathbf{s}}_i = (\widehat{s}_{i,\boldsymbol{\beta}}, \widehat{s}_{i,\sigma_e^2}, \widehat{s}_{i,\boldsymbol{\mu}_1}, \dots, \widehat{s}_{i,\boldsymbol{\mu}_G}, \widehat{s}_{i,\boldsymbol{\alpha}_1}, \dots, \widehat{s}_{i,\boldsymbol{\alpha}_Q}, \widehat{s}_{i,\boldsymbol{\lambda}_1}, \dots, \widehat{s}_{i,\boldsymbol{\lambda}_G}, \widehat{s}_{i,\omega_1}, \dots, \widehat{s}_{i,\omega_{G-1}})^\top.$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_Q)^\top$, $Q = \frac{q(q+1)}{2}$, are the elements of non-structured matrix $\mathbf{D}_k^{1/2} \doteq \mathbf{F}_k$ and $\mathbf{s}_{i_j} = \frac{\partial \ell_i(\theta)}{\partial \theta_j}$, $j = 1, ..., K$, $K = p + G + 2qG + QG$.

$f_i(\mathbf{y}_i|\boldsymbol{\theta}_{ik}) = 2\phi_{n_i}(\mathbf{y}_i|\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\nu}_k, \boldsymbol{\Psi}_{ik})\Phi_1\left(\bar{\boldsymbol{\lambda}}_{ik}^\top \boldsymbol{\Psi}_{ik}^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\nu}_k)\right)$, the score functions $\frac{\partial \ell_i(\theta)}{\partial \theta_j}$

Acknowledgments: The authors acknowledges partial financial support from Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

References

[1] Arellano-Valle, R.B.; Bolfarine, H. and Lachos, V.H. Skew-Normal Linear Mixed Models. *Journal of Data Science*, **3**, 415-438, 2005.

[2] Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

[3] Basford, K.E., Greenway, D.R., Mclachlan, G.J. and Peel, D. (1997). Standard errors of fitted component means of normal mixtures. *Computational Statistics*, **12**, 1Ű17.

[4] Cabral, C.R.B Lachos, V.H. and Basso, R. (2009). Robust Mixture Modelling using Scale Mixtures of Skew-Normal Distributions. To appear in *Computational Statistics and Data Analysis*.

[5] Ho, R.K.W. and Hu, I. (2008). Flexible modelling of random effects in linear mixed models. A Bayesian approach. *Computational Statistics & Data Analysis*, **52**, 1347 Ű 1361.

[6] Lachos, V.H.; Bolfarine, H., Arellano-Valle, R.B. and Montenegro, L.C. Likelihood based inference for multivariate skew-normal regression models. *Communications in Statistics: Theory and Methods*, **36**, 1769-1786, 2007.

[7] Lachos, V.H., Dey, D.K. and Cancho, V.G. (2010). Robust linear mixed models with skew-normal independent distributions from a Bayesian perspective. *Journal of Statistical Planning and Inference*, **139**, 4098-4110.

[8] Laird, N. M. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, **73**, 805-811.

[9] Lin, T.I. and Lee, J.C. Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in Medicine*, **27**, p. 1490-1507, 2008.

[10] Lindsay, B.G., 1983. The geometry of mixture likelihoods: a general theory. *Ann. Statist.*, **11**, 86Ũ94.

[11] Ma, Y., Genton, M.G. and Davidian, M. (2004). Linear mixed eőects models with flexible generalized skew-elliptical random effects. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Genton, M.G., ed., Chapman & Hall / CRC, Boca Raton, FL, 339-358.

[12] Magder, L.S. and Zeger, S.L. (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *J. Amer. Statist. Assoc.*, **91**, 1141-1151.

[13] Neuhaus, J.M., Hauck,W.W. and Kalbfleisch, J.D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, **79**, 755Ũ762.

[14] Tao, H., Palta, M., Yandell, B.S., and Newton, M.A. (1999). An estimation method for the semiparametric mixed effects model. *Biometrics*, **55**, 102-110.

[15] Verbeke, G. and Lesaffre, E. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570-582.