RT-MAE 2004-04

ASYMMETRIC MIXED LINEAR MODELS

by

R.B., Arellano-Valle,
H., Bolfarine
and
V.H., Lachos

Palavras-Chave: Maximum likelihood; EM algorithm; marginal likehood; multivariate skew-normal distribution.
Classificação AMS: 62F05, 62J12.
(AMS Classification)

- Março de 2004 -

ASYMMETRIC MIXED LINEAR MODELS

R. B. Arellano-Valle*, H. Bolfarine, V. H. Lachos****

(*) Pontificia Universidad Católica de Chile
Casilla 306, Correo 22
Santiago-Chile

(**) Universidade de São Paulo
Caixa Postal 66281 - CEP 05315 970
São Paulo - S.P. - Brasil

Abstract

In this paper, we consider linear mixed models with random effects and model errors following asymmetric normal distributions. The marginal distribution for the observed quantity is derived which is shown to follow an asymmetric distribution belonging to a general class of asymmetric normal distributions. The likelihood function that follows from the marginal distribution can be maximized directly by using existing statistical software. We also implement an EM type algorithm which seems to provide some advantages over direct maximization. Results of simulation studies and applications to real data sets are reported.

Key Words: *Maximum likelihood; EM algorithm; marginal likelihood; multivariate skew-normal distribution.*

1 Introduction

The *linear mixed-effects* model (LMM, Laird and Ware, 1982) is typically used in analyzing repeated measurement (or individuals sharing a same trait) and longitudinal data, which is defined by

$$\mathbf{Y}_j = \mathbf{X}_j\beta + \mathbf{Z}_j\mathbf{b}_j + \epsilon_j, \quad j = 1, \ldots, m, \tag{1}$$

where $\mathbf{Y}_j$ is a $n_j \times 1$ vector of responses from m subjects or individuals, $\mathbf{X}_j$ of dimension $(n_j \times p)$ is the design matrix corresponding to the fixed effects, β of dimension $p \times 1$ is a vector of population-averaged regression coefficients called fixed effects, $\mathbf{Z}_j$ of dimension $(n_j \times q)$ is the design matrix

corresponding to the $(q \times 1)$ random effects vector $\mathbf{b}_j$, and $\boldsymbol{\epsilon}_j$ of dimension $(n_j \times 1)$ is the vector of random errors. Typically, it is assumed that the random effects $\mathbf{b}_j$ and the residual components $\boldsymbol{\epsilon}_j$ are independent with

$$\mathbf{b}_j \overset{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{D}), \quad \boldsymbol{\epsilon}_j \overset{\text{ind.}}{\sim} N_{n_j}(\mathbf{0}, \boldsymbol{\psi}_j), \quad j = 1, \ldots, m, \tag{2}$$

where $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ and $\boldsymbol{\psi}_j = \boldsymbol{\psi}_j(\boldsymbol{\gamma})$ are dispersion matrices, usually associated with the variability between and within units, which are depending on unknown parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, respectively. Note that under the above assumptions, the model can be written hierarchically as

$$\mathbf{Y}_j|\mathbf{b}_j \overset{\text{ind.}}{\sim} N_{n_j}(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j, \boldsymbol{\psi}_j), \tag{3}$$

$$\mathbf{b}_j \overset{\text{ind.}}{\sim} N_q(\mathbf{0}, \mathbf{D}), \quad j = 1, \ldots, m. \tag{4}$$

The main interest in this model is to make inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T$, which gives information on the relationship between $\mathbf{Y}$ and $\mathbf{X}$ and also on the degree of dependence within individuals in the same group (or measurements from the same individual). The normal linear models (LMM) is typically simple to analyze, either trough direct expressions (such as ANOVA) in the balanced case or through equation solving algorithms (Henderson, 1956) in unbalanced situations. Typically, the fundamental assumption in the standard version of the model are that within-unit error and random effects are normally distributed, even though this model offers great flexibility for modeling these effects, it suffers from the same lack of robustness against departures from distributional assumptions as other statistical models based on the Gaussian distribution and may be too restrictive to provide an accurate representation of the structure that is present in repeated measures and clusters data. From a practical point of view, the most commonly adopted approach to achieve multivariate normality involves variables transformation. Although such methods may give reasonable empirical results, it should be avoided if a more suitable theoretical model can be found. By introducing a more flexible parametric family capable to accommodating such departures, in this paper we extend the above normal mixed model by considering that $\boldsymbol{\epsilon}_j$ and $\mathbf{b}_j$, $j = 1, \ldots, m$, follow a multivariate skew-normal distributions, which contains the normal distribution as special case.

As considered in Azzalini (1985), a random variable Z follows a univariate skew normal distribution with location parameter μ, scale parameter σ^2 and

skewness parameter λ if the probability density function (pdf) of Y is given by

$$f_Y(y) = \frac{2}{\sigma}\phi_1\left(\frac{y-\mu}{\sigma}\right)\Phi_1\left(\lambda\frac{y-\mu}{\sigma}\right), \tag{5}$$

where $\phi_1(.)$ and $\Phi_1(.)$, denote the probability density function (pdf) and cumulative distribution function (cdf), respectively, of the standard univariate normal distribution. Note that if $\lambda = 0$ then the density of Y in (5) reduces to the density of the normal distribution. We use the notation $Y \sim SN_1(\mu, \sigma^2, \lambda)$ to denote this distribution, which will be reduced to $Y \sim SN_1(\lambda)$ when is assumed that $\mu = 0$ and $\sigma^2 = 1$. Some properties of this distribution includes:

$$E[Y] = \mu + \sqrt{\frac{2}{\pi}}\delta\sigma \quad \text{and} \quad Var[Y] = (1 - \frac{2\delta^2}{\pi})\sigma^2, \quad \text{where} \quad \delta = \frac{\lambda}{\sqrt{1+\lambda^2}},$$

with asymmetry (γ) and kurtosis (κ) indexes such that:

$$-0.9953 < \gamma < 0,9953 \quad \text{and} \quad 3.0000 < \kappa < 3.8692.$$

All these properties may be obtained easily using that (Henze, 1986; Azzalini, 1986) if $Y \sim SN_1(\lambda)$ then

$$Y \stackrel{d}{=} \delta|X_0| + (1-\delta^2)^{1/2}X_1, \tag{6}$$

where $X_i \stackrel{iid}{\sim} N(0,1)$, $i = 0, 1$, and "$\stackrel{d}{=}$" meaning "distributed as".

Multivariate skew-normal distributions are considered in Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), Branco and Dey (2001), among others. Genton et al. (2001) derive the moments of a random vector with multivariate skew-normal distribution and their quadratic forms. Arellano-Valle, del Pino and San Martin (2002) show that many of the properties of the multivariate skew-normal distribution hold for a general class of skewed distributions obtained from a symmetric class, defined in terms of independence conditions on signs and absolute values and give general formulae to obtain skewed pdf's. From these results, Arellano-Valle and Genton (2003) introduce the class of fundamental skewed distributions, giving an unified approach to obtain multivariate skew distributions starting

from symmetric ones. See also Arellano-Valle and del Pino (2003).

In this paper, we consider a multivariate extension of the univariate skew-normal distribution defined by (5) and using this formulation a mixed skew-normal model is defined extending the usual normal mixed model. The marginal distribution of the observed data (observed likelihood) is obtained by integrating out the random effects. Although maximum likelihood estimation can be obtained by directly maximizing the likelihood function by using statistical software such as Ox or Matlab, a stochastic representation is proposed which allows implementation of an EM type algorithm. The EM algorithm seems to be more effective in terms of starting values and, moreover, the complete likelihood does not involve complex expressions as is the case with the observed likelihood function.

The paper is organized as follows. In Section 2, for the sake of completeness, we consider a multivariate extension of the skew-normal distribution used for defining the mixed model. Properties like moments and stochastic representation of this multivariate distribution are also discussed. In Section 3 the mixed model is defined and the marginal density of $\mathbf{Y}_j$ is obtained by integrating out the random effects $\mathbf{b}_j$, $j = 1, \ldots, m$, leading to the observed (marginal) likelihood function that can be maximized directly by using existing statistical software. The asymptotic covariance matrix can be estimated by using the observed information matrix (Hessian). Section 4 presents an EM type algorithm which seems to present advantages over the direct maximization approach, specially in terms of robustness with respect to starting values. Section 5 reports results of a simulation study indicating good performance of the maximum likelihood approach and Section 6 reports applications to a real data set indicating the usefulness of the approach.

2 A multivariate skew-normal distribution

As is discussed in Arellano-Valle and Genton (2003) and Arellano-Valle and del Pino (2003), there are many definitions of the multivariate skew-normal distribution. In this work, we consider a definition given next of an n-variate skew-normal distribution (see also Azzalini and Dalla-Valle, 1996 and Azzalini and Capitanio, 1999).

Let $\phi_n(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $\Phi_n(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ be the pdf and the cdf, respectively, of the $N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$ distribution evaluated at $\mathbf{x}$. When $\boldsymbol{\mu}=\mathbf{0}$ and $\boldsymbol{\Sigma}=\mathbf{I}_n$ (the $n\times n$ identity matrix), we denote these functions as $\phi_n(\mathbf{x})$ and $\Phi_n(\mathbf{x})$.

Definition 1. *An n-dimensional random vector $\mathbf{Y}$ follows a skew-normal distribution with location vector $\boldsymbol{\mu}\in\mathbb{R}^n$, dispersion matrix $\boldsymbol{\Sigma}$ (a $n\times n$ positive definite matrix) and skewness vector $\boldsymbol{\lambda}\in\mathbb{R}^n$, if its pdf is given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})), \quad \mathbf{y}\in\mathbb{R}^n. \tag{7}$$

We denote this by $\mathbf{Y}\sim SN_n(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda})$ and by $\mathbf{Y}\sim SN_n(\boldsymbol{\lambda})$ when $\boldsymbol{\mu}=\mathbf{0}$ and $\boldsymbol{\Sigma}=\mathbf{I}_n$, the n-dimensional identity matrix.

Remark 1. *Since the condition that $\Phi_1(-w)=1-\Phi_1(w)$ for all $w\in\mathbb{R}$ is sufficient to guarantee that (7) is a pdf, we can then use different reparameterizations to represent the asymmetric parameter $\boldsymbol{\lambda}$, as for example:*

$$\boldsymbol{\lambda} = \frac{\boldsymbol{\Delta}^{-1/2}\boldsymbol{\delta}}{\sqrt{1-\boldsymbol{\delta}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\delta}}}, \tag{8}$$

for some $\boldsymbol{\delta}\in\mathbb{R}^n$ and positive definite $n\times n$ matrix $\boldsymbol{\Delta}$ such that $\sqrt{\boldsymbol{\delta}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\delta}}<1$. Two special case are $\boldsymbol{\Delta}=\boldsymbol{\Sigma}$, which is just the reparameterization used by Azzalini and Dalla-Valle (1996), and $\boldsymbol{\Delta}=\mathbf{I}_n$, which is used in Arellano-Valle and Genton (2003). In a more general way, we can replace in (7) the asymmetric part (or skewing function; see Genton and Loperfido, 2002) $\Phi_1(\cdot)$ by an arbitrary function $Q(\cdot)$ on $[0,1]$, which depends on $\mathbf{y}$ trough an even real function (or antisymmetric function; see Arellano-Valle and del Pino, 2003), say $w(\mathbf{y})$, and is such that $Q(w(-\mathbf{y}))=Q(-w(\mathbf{y}))=1-Q(w(\mathbf{y}))$. Thus, the skew-normal distribution in (7) can be extended by considering

$$f_{\mathbf{Y}}(\mathbf{y}) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})Q(w(\mathbf{y})), \quad \mathbf{y}\in\mathbb{R}^n.$$

The following lemma will be used in the following. It can also be used to show that (7) is just a pdf on $\mathbb{R}^n$.

Lemma 1. *Let $\mathbf{Y}\sim N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Then for any fixed k-dimensional vector $\mathbf{a}$ and $k\times n$ matrix $\mathbf{B}$,*

$$E[\Phi_k(\mathbf{a}+\mathbf{B}\mathbf{Y}|\boldsymbol{\eta},\boldsymbol{\Omega})] = \Phi_k(\mathbf{a}|\boldsymbol{\eta}-\mathbf{B}\boldsymbol{\mu},\boldsymbol{\Omega}+\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T).$$

Proof. The proof follows by noticing that

$$E[\Phi_k(\mathbf{a}+\mathbf{B}\mathbf{Y}|\boldsymbol{\eta},\boldsymbol{\Omega})] = E[P(\mathbf{U}\leq\mathbf{a}|\mathbf{Y})] = P(\mathbf{U}\leq\mathbf{a}),$$

where $\mathbf{U}|\mathbf{Y}=\mathbf{y}\sim N_k(\boldsymbol{\eta}-\mathbf{B}\mathbf{y},\boldsymbol{\Omega})$, so that $\mathbf{U}\sim N_k(\boldsymbol{\eta}-\mathbf{B}\boldsymbol{\mu},\boldsymbol{\Omega}+\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T)$. □

Many properties of the above skew-normal distribution may be derived from the results developed by Arellano-Valle and Genton (2003) (see also Arellano-Valle et al., 2002 and Arellano-Valle and del Pino, 2003). From there it follows, for example, the stochastic representation given next for an standardized skew-normal random vector. The proof is based in the following lemma

Lemma 2. *Let* $\mathbf{Y}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *and* $\mathbf{X}\sim N_q(\boldsymbol{\eta},\boldsymbol{\Omega})$. *Then,*

$$\begin{aligned}\phi_p(\mathbf{y}|\boldsymbol{\mu}+\mathbf{A}\boldsymbol{\eta},\boldsymbol{\Sigma})\phi_q(\mathbf{x}|\boldsymbol{\eta},\boldsymbol{\Omega}) &= \phi_p(\mathbf{y}|\boldsymbol{\mu}+\mathbf{A}\boldsymbol{\eta},\boldsymbol{\Sigma}+\mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T)\\ &\times\ \phi_q(\mathbf{x}|\boldsymbol{\eta}+\boldsymbol{\Lambda}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}-\mathbf{A}\boldsymbol{\eta}),\boldsymbol{\Lambda}),\end{aligned}$$

where $\boldsymbol{\Lambda}=(\boldsymbol{\Omega}^{-1}+\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}$.

Proof. By letting $\mathbf{z}=\mathbf{y}-\boldsymbol{\mu}-\mathbf{A}\boldsymbol{\eta}$ and $\mathbf{w}=\mathbf{x}-\boldsymbol{\eta}$, we have after some standard algebraic operations that

$$\begin{aligned}(\mathbf{z}-\mathbf{A}\mathbf{w})^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\mathbf{A}\mathbf{w}) &+ \mathbf{w}^T\boldsymbol{\Omega}^{-1}\mathbf{w} =\\ \mathbf{z}(\boldsymbol{\Sigma}+\mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T)^{-1}\mathbf{z} &+ (\mathbf{w}-\boldsymbol{\Lambda}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{z})^T\boldsymbol{\Lambda}^{-1}(\mathbf{w}-\boldsymbol{\Lambda}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{z}),\end{aligned}$$

and the proof follows by noting also that $|\boldsymbol{\Sigma}+\mathbf{A}\boldsymbol{\Omega}\mathbf{A}^T||\boldsymbol{\Lambda}|=|\boldsymbol{\Sigma}||\boldsymbol{\Omega}|$. □

Proposition 1. *Let* $\mathbf{W}\sim SN_n(\boldsymbol{\lambda})$. *Then*

$$\mathbf{W}\stackrel{d}{=}\boldsymbol{\delta}|X_0|+(\mathbf{I}_n-\boldsymbol{\delta}\boldsymbol{\delta}^T)^{1/2}\mathbf{X}_1,\quad \textit{where}\quad \boldsymbol{\delta}=\frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^T\boldsymbol{\lambda}}},\qquad (9)$$

$X_0\sim N(0,1)$ *and* $\mathbf{X}_1\sim N_n(\mathbf{0},\mathbf{I}_n)$ *and are independent.*

Proof. Let $\mathbf{U}=\boldsymbol{\delta}|X_0|+(\mathbf{I}_n-\boldsymbol{\delta}\boldsymbol{\delta}^T)^{1/2}\mathbf{X}_1$. Since $\mathbf{U}||X_0|=t\sim N_n(\boldsymbol{\delta}t,\mathbf{I}_n-\boldsymbol{\delta}\boldsymbol{\delta}^T)$, where $|X_0|\sim HN(0,1)$ (the standardized half-normal distribution), then by Lemma 2 it follows that

$$\begin{aligned}f_{\mathbf{U}}(\mathbf{w}) &= \int_0^\infty \phi_n(\mathbf{w}|\boldsymbol{\delta}t,\mathbf{I}_n-\boldsymbol{\delta}\boldsymbol{\delta}^T)2\phi(t)dt\\ &= \int_0^\infty \phi_n(\mathbf{w}|\mathbf{0},\mathbf{I}_n)2\phi(t|\boldsymbol{\delta}^T\mathbf{w},1-\boldsymbol{\delta}^T\boldsymbol{\delta})dt\\ &= 2\phi_n(\mathbf{w})\Phi_1\left(\frac{\boldsymbol{\delta}^T\mathbf{w}}{\sqrt{1-\boldsymbol{\delta}^T\boldsymbol{\delta}}}\right),\end{aligned}$$

i.e. $\mathbf{U} \stackrel{d}{=} \mathbf{W} \sim SN_n(\boldsymbol{\lambda})$, with $\boldsymbol{\lambda} = \frac{\boldsymbol{\delta}}{\sqrt{1-\boldsymbol{\delta}^T\boldsymbol{\delta}}}$, which concludes the proof. □

Note that the stochastic representation in (6) is a special case of (9). Two direct consequences of Proposition 1 are given next.

Corollary 1. *Let* $\mathbf{Y} \sim SN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. *Then,* $\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{W}$, *where* $\mathbf{W} \sim SN_n(\boldsymbol{\lambda})$. *Moreover,*

$$E(\mathbf{Y}) = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda} \text{ and } V(\mathbf{Y}) = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^T\boldsymbol{\Sigma}^{1/2}.$$

3 A mixed skew-normal likelihood function

The skew-normal linear mixed model that we consider to extend the normal mixed model in (3)-(4) is defined by the assumptions

$$\mathbf{Y}_j|\mathbf{b}_j \stackrel{\text{ind.}}{\sim} SN_{n_j}(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j, \boldsymbol{\psi}_j, \boldsymbol{\lambda}_{ej}), \tag{10}$$

$$\mathbf{b}_j \stackrel{\text{iid}}{\sim} SN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}_b), \quad j = 1, \ldots, m, \tag{11}$$

which, by Corollary 1, is equivalent to considering (1) with the assumptions that $\mathbf{b}_j \stackrel{\text{iid}}{\sim} SN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}_b)$ and $\boldsymbol{\epsilon}_j \stackrel{\text{ind.}}{\sim} SN_{n_j}(\mathbf{0}, \boldsymbol{\psi}_j, \boldsymbol{\lambda}_{ej})$, $j = 1, \ldots, m$, are all independent. The main interest is to make inference on the parameter vectors $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T$ and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_b^T, \boldsymbol{\lambda}_{e1}^T, \ldots, \boldsymbol{\lambda}_{em}^T)^T$. To obtain the marginal distribution of $\mathbf{Y}_j$, which is required for deriving the likelihood function, we drop the subscript j, corresponding to the j-th group (or individual) to simplify notation. From (10), (11) and the definition of skew-normal multivariate distribution in (7), it follows that the marginal density of $\mathbf{Y}$ is obtained by computing the following integral:

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\lambda}) &= \int_{\mathbb{R}^q} f(\mathbf{y}|\mathbf{b}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}_e) f(\mathbf{b}|\boldsymbol{\alpha}, \boldsymbol{\lambda}_b) d\mathbf{b} \\ &= \int_{\mathbb{R}^q} 2^2 \phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \boldsymbol{\psi}) \Phi_1(\boldsymbol{\lambda}_e^T \boldsymbol{\psi}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})) \\ &\quad \phi_q(\mathbf{b}|\mathbf{0}, \mathbf{D}) \Phi_1(\boldsymbol{\lambda}_b^T \mathbf{D}^{-1/2}\mathbf{b}) d\mathbf{b}. \end{aligned} \tag{12}$$

We make use of the following two propositions that will be used to prove the main result of the paper.

Proposition 2. *Under the notation considered in (12), it follows that*

$$\phi_n(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\boldsymbol{b},\boldsymbol{\psi})\phi_q(\boldsymbol{b}|\boldsymbol{0},\boldsymbol{D}) = \phi_n(\boldsymbol{y}|\boldsymbol{X}\boldsymbol{\beta},\boldsymbol{\Sigma})\phi_q(\boldsymbol{b}|\boldsymbol{\mu}_1,\boldsymbol{\Lambda}) \tag{13}$$

and

$$\Phi_1(\boldsymbol{\lambda}_e^T\boldsymbol{\psi}^{-1/2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{b}))\Phi_1(\boldsymbol{\lambda}_b^T\mathbf{D}^{-1/2}\boldsymbol{b}) = \Phi_2(-\boldsymbol{\Gamma}\boldsymbol{b}|-\boldsymbol{\mu}_2,\mathbf{I}_2), \tag{14}$$

where

$$\boldsymbol{\mu}_1 = \boldsymbol{\Lambda}\mathbf{Z}^T\boldsymbol{\psi}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}), \quad \boldsymbol{\Sigma} = \boldsymbol{\psi}+\mathbf{Z}\mathbf{D}\mathbf{Z}^T, \quad \boldsymbol{\Lambda} = (\mathbf{D}^{-1}+\mathbf{Z}^T\boldsymbol{\psi}^{-1}\mathbf{Z})^{-1}, \tag{15}$$

$$\boldsymbol{\mu}_2 = \begin{pmatrix} \boldsymbol{\lambda}_e^T\boldsymbol{\psi}^{-1/2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\ 0 \end{pmatrix} \quad \textit{and} \quad \boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\lambda}_e^T\boldsymbol{\psi}^{-1/2}\boldsymbol{Z} \\ -\boldsymbol{\lambda}_b^T\mathbf{D}^{-1/2} \end{pmatrix}. \tag{16}$$

Proof. Result (13) follows from Lemma 2. Result (14) is proved by noting that if U and V are i.i.d. $N(0,1)$ random variables, then (14) can be written as

$$P(U \le \boldsymbol{\lambda}_e^T\boldsymbol{\psi}^{-1/2}(\mathbf{y}-\boldsymbol{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b}))P(V \le \boldsymbol{\lambda}_b^T\mathbf{D}^{-1/2}\mathbf{b}) = P(\mathbf{T} \le \boldsymbol{\mu}_2),$$

where $\mathbf{T} = \mathbf{W}+\boldsymbol{\Gamma}\mathbf{b}$, with $\mathbf{W} = (U,V)^T \sim N_2(\mathbf{0},\mathbf{I}_2)$, and $\boldsymbol{\mu}_2$ and $\boldsymbol{\Gamma}$ and defined as in (16). Thus, since $\mathbf{T} \sim N_2(\boldsymbol{\Gamma}\mathbf{b},\mathbf{I}_2)$, we have that

$$P(\mathbf{T} \le \boldsymbol{\mu}_2) = \Phi_2(\boldsymbol{\mu}_2|\boldsymbol{\Gamma}\mathbf{b},\mathbf{I}_2) = \Phi_2(-\boldsymbol{\Gamma}\mathbf{b}|-\boldsymbol{\mu}_2,\mathbf{I}_2),$$

which concludes the proof. □

We prove now the main result of the paper.

Teorema 1. *Let* $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b}+\boldsymbol{\epsilon}$, *where* $\mathbf{b} \sim SN_q(\mathbf{0},\mathbf{D},\boldsymbol{\lambda}_b)$ *and* $\boldsymbol{\epsilon} \sim SN_n(\mathbf{0},\boldsymbol{\psi},\boldsymbol{\lambda}_e)$ *are independent. Then, the marginal distribution of* $\mathbf{Y}$ *is given by*

$$f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\lambda}) = 2^2\phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta},\boldsymbol{\Sigma})\Phi_2(\boldsymbol{\mu}_2-\boldsymbol{\Gamma}\boldsymbol{\mu}_1|\mathbf{0},\mathbf{I}_2+\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T), \tag{17}$$

where $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Gamma}$ *and* $\boldsymbol{\Lambda}$ *are given in (15) and (16).*

Proof. From(12), (13) and (14), it follows that

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\lambda}) &= \int_{\mathbb{R}^q} 2^2\phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta},\boldsymbol{\Sigma})\phi_q(\mathbf{b}|\boldsymbol{\mu}_1,\boldsymbol{\Lambda})\Phi_2(-\boldsymbol{\Gamma}\mathbf{b}|-\boldsymbol{\mu}_2,\mathbf{I}_2)d\mathbf{b} \\ &= 2^2\phi_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta},\boldsymbol{\Sigma})E[\Phi_2(-\boldsymbol{\Gamma}\mathbf{W}|-\boldsymbol{\mu}_2,\mathbf{I}_2)], \end{aligned}$$

where $\mathbf{W} \sim N_q(\boldsymbol{\mu}_1,\boldsymbol{\Lambda})$. The proof is concluded by using Lemma 1. □

Note that the likelihood (17) is not in the class of skew multivariate distributions defined in Azzalini and Dalla-Valle (1996) since the skewing function in that expression is of dimension 2. The likelihood function for $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ given the observed sample $\mathbf{y}_1, \ldots, \mathbf{y}_m$ can be written as

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathbf{y}_1, \ldots, \mathbf{y}_m) = \prod_{j=1}^{m} f_{\mathbf{Y}}(\mathbf{y}_j | \boldsymbol{\theta}, \boldsymbol{\lambda}),$$

where $f_{\mathbf{Y}}(\mathbf{y}_j | \boldsymbol{\theta}, \boldsymbol{\lambda})$ is the marginal density that follows from (17) by incorporating again the index j. Thus, Denoting the log-likelihood function by $\ell(\boldsymbol{\theta}, \boldsymbol{\lambda})$, it can be written as

$$\begin{aligned} \ell(\boldsymbol{\theta}, \boldsymbol{\lambda}) \;\propto\; & -\frac{1}{2}\sum_{j=1}^{m} \log|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{j=1}^{m}\{(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})\} \\ & + \sum_{j=1}^{m} \log \Phi_2(\boldsymbol{\mu}_{2j} - \boldsymbol{\Gamma}_j\boldsymbol{\mu}_{1j} | \mathbf{0}, \mathbf{I}_2 + \boldsymbol{\Gamma}_j\boldsymbol{\Lambda}_j\boldsymbol{\Gamma}_j^T), \end{aligned} \tag{18}$$

where $\boldsymbol{\mu}_{1j}$, $\boldsymbol{\mu}_{2j}$, $\boldsymbol{\Sigma}_j$, $\boldsymbol{\Gamma}_j$ and $\boldsymbol{\Lambda}_j$ as defined in (15) and (16), but incorporating the index j on $\mathbf{y}$, $\mathbf{X}$, $\mathbf{Z}$, $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}_e$.

We call attention to the fact that no explicit solution is available for the maximization problem so that the likelihood function has to be maximized numerically. Some special cases may be of interest. For instance, the situation where $\boldsymbol{\lambda}_{e1} = \ldots = \boldsymbol{\lambda}_{em} = \mathbf{0}$ or $\boldsymbol{\lambda}_b = \mathbf{0}$, which are special cases of the above general situation and it clearly implies that the joint distribution of $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ is asymmetric also. These situations are treated next.

Corollary 2. *Under the conditions of Theorem 1, it follows that:*

(i) if $\boldsymbol{\lambda}_e = \mathbf{0}$, *then*

$$f_{\mathbf{Y}}(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\lambda}_b) = 2\phi_n(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})\Phi_1\left(\bar{\boldsymbol{\lambda}}_b^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right), \tag{19}$$

i.e.,

$$\mathbf{Y} \sim SN_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\lambda}}_b), \quad \textit{with} \quad \bar{\boldsymbol{\lambda}}_b = \frac{\boldsymbol{\Sigma}^{-1/2}\mathbf{Z}\mathbf{D}^{1/2}\boldsymbol{\lambda}_b}{\sqrt{1 + \boldsymbol{\lambda}_b^T \mathbf{D}^{-1/2}\boldsymbol{\Lambda}\mathbf{D}^{-1/2}\boldsymbol{\lambda}_b}},$$

(ii) if $\lambda_b = 0$, *then*

$$f_Y(\mathbf{y}|\theta, \lambda_e) = 2\phi_n(\mathbf{y}|\mathbf{X}\beta, \Sigma)\Phi_1\left(\bar{\lambda}_e^T\Sigma^{-1/2}(\mathbf{y} - \mathbf{X}\beta)\right), \qquad (20)$$

i.e.,

$$\mathbf{Y} \sim SN_n(\mathbf{X}\beta, \Sigma, \bar{\lambda}_e), \quad with \quad \bar{\lambda}_e = \frac{\Sigma^{-1/2}\psi^{1/2}\lambda_e}{\sqrt{1 + \lambda_e^T\psi^{-1/2}\mathbf{Z}\Lambda\mathbf{Z}^T\psi^{-1/2}\lambda_e}}.$$

Proof. Note first that

$$\mu_2 - \Gamma\mu_1 = \begin{pmatrix} \lambda_e^T\psi^{-1/2}(\psi - \mathbf{Z}\Lambda\mathbf{Z}^T)\psi^{-1}(\mathbf{y} - \mathbf{X}\beta) \\ \lambda_b^T\mathbf{D}^{-1/2}\Lambda\mathbf{Z}^T\psi^{-1}(\mathbf{y} - \mathbf{X}\beta) \end{pmatrix}$$

and

$$\mathbf{I}_2 + \Gamma\Lambda\Gamma^T = \begin{pmatrix} 1 + \lambda_e^T\psi^{-1/2}\mathbf{Z}\Lambda\mathbf{Z}^T\psi^{-1/2}\lambda_e & -\lambda_e^T\psi^{-1/2}\mathbf{Z}\Lambda\mathbf{D}^{-1/2}\lambda_b \\ -\lambda_b^T\mathbf{D}^{-1/2}\Lambda\mathbf{Z}^T\psi^{-1/2}\lambda_e & 1 + \lambda_b^T\mathbf{D}^{-1/2}\Lambda\mathbf{D}^{-1/2}\lambda_b \end{pmatrix},$$

which is a diagonal matrix when $\lambda_e = 0$ or $\lambda_b = 0$. Hence, for $\lambda_e = 0$, the asymmetric part of (17) can be computed as

$$\begin{aligned} \Phi_2(\mu_2 - \Gamma\mu_1|0, \mathbf{I}_2 + \Gamma\Lambda\Gamma^T) &= \Phi_2((\mathbf{I}_2 + \Gamma\Lambda\Gamma^T)^{-1/2}(\mu_2 - \Gamma\mu_1)) \\ &= \frac{1}{2}\Phi_1\left(\frac{\lambda_b^T\mathbf{D}^{-1/2}\Lambda\mathbf{Z}^T\psi^{-1}(\mathbf{y} - \mathbf{X}\beta)}{\sqrt{1 + \lambda_b^T\mathbf{D}^{-1/2}\Lambda\mathbf{D}^{-1/2}\lambda_b}}\right), \end{aligned}$$

where some algebraic manipulations yield $\Lambda\mathbf{Z}^T = \mathbf{D}\mathbf{Z}^T\Sigma^{-1}\psi$. Similarly, for $\lambda_e = 0$, we have that

$$\Phi_2(\mu_2 - \Gamma\mu_1|0, \mathbf{I}_2 + \Gamma\Lambda\Gamma^T) = \frac{1}{2}\Phi_1\left(\frac{\lambda_e^T\Psi^{-1/2}(\psi - \mathbf{Z}\Lambda\mathbf{Z}^T)\psi^{-1}(\mathbf{y} - \mathbf{X}\beta)}{\sqrt{1 + \lambda_e^T\psi^{-1/2}\mathbf{Z}\Lambda\mathbf{Z}^T\psi^{-1/2}\lambda_e}}\right),$$

and the proof concludes by noting that $\psi - \mathbf{Z}\Lambda\mathbf{Z}^T = \psi\Sigma^{-1}\psi$. □

Although simpler the log-likelihood functions that follow by replacing (19) and (20) in (18) must also be maximized numerically. The asymptotic covariance matrix of the maximum likelihood estimators can be estimated by using the Hessian matrix, which can also be computed numerically by using the program Matlab, for example. In the next section we present an EM-type algorithm for computing the maximum likelihood estimator (MLE) of densities obtained in Corollary 2.

3.1 An EM-type algorithm

A direct maximization of the likelihood (19) and (20) may sometimes poses problems since it involves terms like $\log(\Phi_1(w))$, which causes computational problems for w negative ($w < -3$, for example). Further, the approach seems not too robust with respect to starting values, that is, unless good starting values are used, the direct maximization approach will typically not converge. Simulation studies conducted indicate the EM to be more robust in the sense that it may converge more often than the direct maximization approach.

In order to implement the two steeps of the EM-algorithm for maximizing the likelihood from Corollary 2, we need first some additional results.

Proposition 3. *Suppose that* $\mathbf{Y}|T = t \sim N_n(\boldsymbol{\mu} + \mathbf{d}t, \boldsymbol{\Psi})$ *and* $T \sim HN_1(0, 1)$. *Let* $\boldsymbol{\Sigma} = \boldsymbol{\Psi} + \mathbf{d}\mathbf{d}^T$. *Then the joint distribution of* $(\mathbf{Y}^T, T)^T$ *can be written as*

$$f_{\mathbf{Y},T}(\mathbf{y}, t|\boldsymbol{\theta}, \boldsymbol{\lambda}) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\phi_1(t|\eta, \tau^2)\mathbb{I}\{t > 0\}, \tag{21}$$

where

$$\eta = \frac{\mathbf{d}^T\boldsymbol{\Psi}^{-1}(\mathbf{y} - \boldsymbol{\mu})}{1 + \mathbf{d}^T\boldsymbol{\Psi}^{-1}\mathbf{d}} \text{ and } \tau^2 = \frac{1}{1 + \mathbf{d}^T\boldsymbol{\Psi}^{-1}\mathbf{d}}. \tag{22}$$

Proof. In fact, the joint density of $\mathbf{Y}$ and T is

$$f_{\mathbf{Y},T}(\mathbf{y}, t|\boldsymbol{\theta}, \boldsymbol{\lambda}) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu} + \mathbf{d}t, \boldsymbol{\Psi})\phi_1(t)\mathbf{I}\{t > 0\}.$$

After some simple algebraic manipulations we have that

$$\phi_n(\mathbf{y}|\boldsymbol{\mu} + \mathbf{d}t, \boldsymbol{\Psi})\phi_1(t) = \phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\phi_1(t|\eta, \tau^2), \tag{23}$$

concluding the proof. □

Notice that the marginal distribution of $\mathbf{Y}$ follows from (21) after integrating out t and is given by

$$f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\lambda}) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1\left(\frac{\eta}{\tau}\right). \tag{24}$$

The next result is related to properties of the truncated normal distribution.

Lemma 3. *Let $X \sim N(\eta, \tau^2)$. Then, for any real constant a it follows that*

$$E(X|X > a) = \eta + \frac{\phi_1(\frac{a-\eta}{\tau})}{1 - \Phi_1(\frac{a-\eta}{\tau})}\tau,$$

$$E(X^2|X > a) = \eta^2 + \tau^2 + \frac{\phi_1(\frac{a-\eta}{\tau})}{1 - \Phi_1(\frac{a-\eta}{\tau})}(\eta + a)\tau.$$

Proof. See Johnson et al. (1994), Section 10.1. □

Proposition 4. *Under the conditions in Proposition 3,*

$$E(T^k|\mathbf{y}) = E(X^k|X > 0),$$

where $X \sim N_1(\eta, \tau^2)$, with η and τ^2 given in (22). Particularly,

$$E(T|\mathbf{y}) = \eta + \frac{\phi_1(\frac{\eta}{\tau})}{\Phi_1(\frac{\eta}{\tau})}\tau, \tag{25}$$

and

$$E(T^2|\mathbf{y}) = \eta^2 + \tau^2 + \frac{\phi_1(\frac{\eta}{\tau})}{\Phi_1(\frac{\eta}{\tau})}\tau\eta. \tag{26}$$

Proof. Note that we can write

$$E(T^k|\mathbf{y}) = \int_{-\infty}^{\infty} t^k f(t|\mathbf{y})dt = \frac{1}{f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\lambda})} \int_{-\infty}^{\infty} t^k f_{\mathbf{Y},T}(\mathbf{y}, t|\boldsymbol{\theta}, \boldsymbol{\lambda})dt.$$

From (21) and (24), it then follows that

$$E(T^k|\mathbf{y}) = \frac{1}{\Phi_1(\frac{\eta}{\tau})} \int_0^{\infty} t^k \phi_1(t|\eta, \tau^2)dt = E(X^k|X > 0),$$

where $X \sim N_1(\eta, \tau^2)$ and $\Phi_1(\frac{\eta}{\tau}) = P(X > 0)$. Thus, (25) and (26) follow from Lemma 3 with a=0, which concludes the proof. □

To implement the EM-algorithm, we first consider the linear mixed model defined by equations (10)-(11) with $\boldsymbol{\lambda}_{\epsilon 1} = \ldots = \boldsymbol{\lambda}_{\epsilon m} = \mathbf{0}$, that is,

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j + \boldsymbol{\epsilon}_j, \tag{27}$$

with

$$\epsilon_j \overset{\text{ind.}}{\sim} N_{n_j}(\mathbf{0}, \boldsymbol{\psi}_j), \quad \mathbf{b}_j \overset{\text{ind.}}{\sim} SN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}_b), \ j = 1, \ldots, m. \tag{28}$$

Hence, (28) jointly with Proposition 1 imply that

$$\mathbf{b}_j \overset{d}{=} \mathbf{D}^{1/2}\boldsymbol{\delta}|X_{0j}| + \mathbf{D}^{1/2}(\mathbf{I}_q - \boldsymbol{\delta}_b\boldsymbol{\delta}_b^T)^{1/2}\mathbf{X}_{1j}, \ j = 1, \ldots, m, \tag{29}$$

where $X_{0j} \overset{\text{iid}}{\sim} N(0,1)$, $\mathbf{X}_{1j} \overset{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{I}_q)$, with X_{0j} and $\mathbf{X}_{1j}$ independent $j = 1, \ldots, m$, and $\boldsymbol{\delta}_b = \dfrac{\boldsymbol{\lambda}_b}{\sqrt{1 + \boldsymbol{\lambda}_b^T\boldsymbol{\lambda}_b}}$. Moreover, independence between $\mathbf{b}_j$ and ϵ_j, $j = 1, \ldots, m$, imply that $\mathbf{V}_j = (X_{0j}, \mathbf{X}_{1j}^T)^T$ and ϵ_j, are independent, $j = 1, \ldots, m$. Hence, replacing (29) in (27) we have that

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\bar{\boldsymbol{\delta}}_b t_j + \mathbf{r}_j, \tag{30}$$

where

$$\bar{\boldsymbol{\delta}}_b = \mathbf{D}^{1/2}\boldsymbol{\delta}_b, \quad t_j = |X_{0j}| \quad \text{and} \quad \mathbf{r}_j = \epsilon_j + \mathbf{D}^{1/2}(\mathbf{I}_q - \boldsymbol{\delta}_b\boldsymbol{\delta}_b^T)^{1/2}\mathbf{X}_{0j},$$

which are such that

$$\mathbf{r}_j \overset{\text{ind.}}{\sim} N_{n_j}(0, \boldsymbol{\psi}_j + \mathbf{Z}_j(\mathbf{D} - \bar{\boldsymbol{\delta}}_b\bar{\boldsymbol{\delta}}_b^T)\mathbf{Z}_j^T), \qquad t_j \overset{\text{iid}}{\sim} HN(0,1), \tag{31}$$

and are independent, $j = 1, \ldots, m$. Hence, (30) and (31) imply that the model defined by (27)-(28) can be written as

$$\mathbf{Y}_j|t_j \overset{\text{ind.}}{\sim} N_{n_j}(\boldsymbol{\mu}_j + \mathbf{d}_j t_j, \boldsymbol{\Psi}_j) \quad \text{and} \quad t_j \overset{\text{iid}}{\sim} HN_1(0,1), \ j = 1, \ldots, m, \tag{32}$$

where

$$\boldsymbol{\mu}_j = \mathbf{X}_j\boldsymbol{\beta}, \ \mathbf{d}_j = \mathbf{Z}_j\bar{\boldsymbol{\delta}}_b, \ \boldsymbol{\Psi}_j = \boldsymbol{\Sigma}_j - \mathbf{d}_j\mathbf{d}_j^T \text{ and } \boldsymbol{\Sigma}_j = \boldsymbol{\psi}_j + \mathbf{Z}_j\mathbf{D}\mathbf{Z}_j^T. \tag{33}$$

Note that in (33) $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ are the marginal mean vector and covariance matrix, respectively, under the usual linear mixed model. Hence, as a direct consequence of Proposition 3 we have the next result.

Proposition 5. *Under (32) it follows that the complete log-likelihood function associated with* $(\mathbf{y}_j, t_j)$, $j = 1, \ldots, m$, *can be written as*

$$\ell_c(\boldsymbol{\theta}, \boldsymbol{\lambda}_b) \propto -\frac{1}{2}\sum_{j=1}^{m}\log|\boldsymbol{\Psi}_j| - \frac{1}{2}\sum_{j=1}^{m}(\mathbf{y}_j - \boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\mathbf{y}_j - \boldsymbol{\mu}_j) - \frac{1}{2\tau_j^2}\sum_{j=1}^{m}(t_j - \eta_j)^2, \tag{34}$$

where by (22)

$$\eta_j = \frac{\mathbf{d}_j^T \Psi_j^{-1}(\mathbf{y}_j - \mu_j)}{1 + \mathbf{d}_j^T \Psi_j^{-1} \mathbf{d}_j} \text{ and } \tau_j^2 = \frac{1}{1 + \mathbf{d}_j^T \Psi_j^{-1} \mathbf{d}_j}, \tag{35}$$

with μ_j, $\mathbf{d}_j$, Σ_j and Ψ_j as defined in (33).

Likewise, considering the case where $\lambda_b = \mathbf{0}$, that is, the linear mixed model in (27), with the assumption that

$$\epsilon_j \overset{\text{ind.}}{\sim} SN_{n_j}(\mathbf{0}, \psi_j, \lambda_{ej}) \quad \text{and} \quad \mathbf{b}_j \overset{\text{iid}}{\sim} N_q(\mathbf{0}, \mathbf{D}), \quad j = 1, \ldots, m, \tag{36}$$

all independent, we can write

$$\mathbf{Y}_j = \mathbf{X}_j \beta + \psi_j^{1/2} \delta_{ej} t_j + \mathbf{r}_j, \tag{37}$$

where

$$t_j = |X_{1j}|, \ \delta_{ej} = \frac{\lambda_{ej}}{(1 + \lambda_{ej}^T \lambda_{ej}^T)^{1/2}} \text{ and } \mathbf{r}_j = \mathbf{Z}_j \mathbf{b}_j + \psi_j^{1/2} (\mathbf{I}_{n_j} - \delta_{ej} \delta_{ej}^T)^{1/2} \mathbf{X}_{0j},$$

this is

$$\mathbf{Y}_j | t_j \overset{\text{ind.}}{\sim} N_{n_j}(\mu_j + \mathbf{d}_j t_j, \Psi_j) \text{ and } t_j \sim HN(0,1), \ j = 1, \ldots, m, \tag{38}$$

where

$$\mu_j = \mathbf{X}_j \beta, \ \mathbf{d}_j = \psi_j^{1/2} \delta_j, \ \Psi_j = \Sigma_j - \mathbf{d}_j \mathbf{d}_j^T \text{ and } \Sigma_j = \psi_j + \mathbf{Z}_j \mathbf{D} \mathbf{Z}_j^T. \tag{39}$$

As a consequence of the above results, it follows by Proposition 3 that:

Proposition 6. *Under (38) it follows that the complete log-likelihood function associated with $(\mathbf{y}_j, t_j)$, $j = 1, \ldots, m$, can be written as*

$$\ell_c(\theta, \lambda_e) \propto -\frac{1}{2} \sum_{j=1}^{m} \log |\Psi_j| - \frac{1}{2} \sum_{j=1}^{m} (\mathbf{y}_j - \mu_j)^T \Sigma_j^{-1} (\mathbf{y}_j - \mu_j) - \frac{1}{2\tau_j^2} \sum_{j=1}^{m} (t_j - \eta_j)^2, \tag{40}$$

where

$$\eta_j = \frac{\mathbf{d}_j^T \Psi_j^{-1}(\mathbf{y}_j - \mu_j)}{1 + \mathbf{d}_j^T \Psi_j^{-1} \mathbf{d}_j} \text{ and } \tau_j^2 = \frac{1}{1 + \mathbf{d}_j^T \Psi_j^{-1} \mathbf{d}_j}, \tag{41}$$

with μ_j, $\mathbf{d}_j$, Σ_j and Ψ_j defined as in (39).

It follows from Propositions 5 and 6 (complete likelihood) that to implement the E (or expectation) step is necessary compute the following conditional moments of T_j given $\mathbf{Y}_j = \mathbf{y}_j$:

$$E(T_j^k|\boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{y}_j) = \int_{\infty}^{\infty} t_j^k f(t_j|\boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{y}_j) dt_j \tag{42}$$

$k = 1, 2$, $j = 1, \ldots, m$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}_b$ or $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_{e1}^T, \ldots, \boldsymbol{\lambda}_{em}^T)^T$, which can be obtained easily using (25) and (26) in conjunction with (32) or (38).

The EM algorithm operates as follows. Given starting values $(\widehat{\boldsymbol{\theta}}^{(0)}, \widehat{\boldsymbol{\lambda}}^{(0)})$, compute $\widehat{t}_j^k = E(T_j^k|\boldsymbol{\theta}^{(0)}, \boldsymbol{\lambda}^{(0)}, \mathbf{y}_j)$, $k = 1, 2$, $j = 1, \ldots, m$, by using (42). Replace these values in the complete log-likelihood functions (34) or (40) and maximize it with respect to $(\boldsymbol{\theta}, \boldsymbol{\lambda})$. This maximization step has to proceed numerically, being most easily accomplished by using Matlab, for example, not posing the same difficulties of a direct maximization of the observed likelihood (19) and (20). Further, the approach seems somewhat robust with respect to starting values. It may take more computing time but eventually will lead to the maximum of the observed likelihood.

4 Simulation study

In this section we present results of a small scale simulation study to demonstrate the usefulness of the approach developed in Section 3 in studying linear mixed models defined by (27)-(28), where the random effects $\mathbf{b}_j$, $j = 1, \ldots, m$, are assumed to follow the skew-normal distribution with diagonal dispersion matrix $\mathbf{D}$. Similarly, the dispersion matrices $\boldsymbol{\psi}_j$ are assumed to be equal $\sigma_e^2 \mathbf{I}_{n_j}$. We consider the case with $q = 1$ and $q = 2$ for different values for $\boldsymbol{\lambda}_b$ (see Table 1). The covariate x_{ij} were generated from the N(5,1), the z_{ij} were considered as being 1, and the errors ϵ_{ij} were generated as i.i.d. $N(0, \sigma_e^2)$, with $\sigma_e^2 = 1$. The true location parameters values were $\beta_1 = 1$ and $\beta_2 = 2$. The following combinations of sample sizes were taken for simulation: $(m, n) = (50, 8)$ and $(m, n) = (100, 4)$, which are typical sample sizes in longitudinal studies. For each situation, 1000 simulated samples were considered. The simulation study was conducted using the MATLAB software. The parameter estimates are computed assuming the skew-normal likelihood (skew-normal column) and the incorrect symmetric normal likelihood (normal column). The summary results are presented in Table 1, where the

Table 1: Simulation result for the linear mixed skew-normal model, with $q = 1$ and $q = 2$, based on 1000 samples. The mean and standard error (SE) of each estimator are obtained based on the generated samples. "Skew-normal" ("normal") meaning that estimators are obtained using the corresponding Skew-normal likelihood (the incorrect normal likelihood).

	$m = 50, n = 8$				$m = 100, n = 4$			
	skew-normal		normal		skew-normal		normal	
Parameter	Mean	SE	Mean	SE	Mean	SE	Mean	SE
	$q = 1$							
$\beta_o = 1$	1.0138	0.0387	1.0522	0.0341	1.0138	0.0427	1.0460	0.0400
β_1=2	2.0123	0.0380	2.0431	0.0337	2.0114	0.0421	2.0511	0.0395
$\sigma_e^2 = 1$	1.0058	0.0781	1.0125	0.0812	0.9880	0.0715	1.0029	0.0763
$\sigma_{b1}^2 = 0.5$	0.4139	0.1822	0.2345	0.0779	0.4275	0.1818	0.2326	0.0716
$\lambda_{b1} = 2$	2.4595	3.1800	-	-	1.9600	2.8154	-	-
N.C.	18%				17%			
	$q = 2$ and $\lambda_b = (\lambda_{b1}, \lambda_{b1})^T$							
$\beta_o = 1$	1.0065	0.0444	1.0759	0.0398	1.0055	0.0460	1.0799	0.0410
β_1=2	2.0094	0.0449	2.0809	0.0379	2.0088	0.0451	2.0819	0.0409
$\sigma_e^2 = 1$	0.9897	0.0752	1.0052	0.0760	0.9872	0.0821	1.0019	0.0828
$\sigma_{b1}^2 = 0.5$	0.5745	0.2694	0.2468	0.0748	0.5610	0.2222	0.2486	0.0606
$\sigma_{b1}^2 = 0.8$	0.6772	0.2029	0.3345	0.0869	0.6715	0.1571	0.3287	0.0733
$\lambda_{b1} = 2$	2.9526	3.5258	-	-	2.7291	3.4161	-	-
N.C.	14%				11.6%			
	$q = 2$ and $\lambda_b = (\lambda_{b1}, \lambda_{b2})^T$							
$\beta_o = 1$	1.0069	0.0443	1.0734	0.0381	1.0037	0.0472	1.0753	0.0395
β_1=2	2.0059	0.0447	2.0733	0.0395	2.0078	0.0462	2.0780	0.0402
$\sigma_e^2 = 1$	0.9893	0.0734	1.0009	0.0746	0.9880	0.0807	1.0025	0.0808
$\sigma_{b1}^2 = 0.5$	0.6948	0.2655	0.2638	0.0742	0.5189	0.2109	0.2721	0.0652
$\sigma_{b2}^2 = 0.8$	0.5547	0.1878	0.3644	0.0949	0.7522	0.2103	0.3669	0.0782
$\lambda_{b1} = 1$	2.0702	2.8618	-	-	1.8338	2.0512	-	-
$\lambda_{b2} = 2$	2.3018	2.9892	-	-	2.0759	2.3746	-	-
N.C.	7.4%				5.6%			

entries are the mean values over the 1000 simulated samples and the column SE are the estimated standard errors. For the parameter β_1, for example, $SE = \sqrt{\sum_{i=1}^{1000}(\widehat{\beta}_{1i} - \bar{\widehat{\beta}}_1)^2/1000}$, where $\widehat{\beta}_{1i}$ is the estimator of β_1 computed using the i-th generated sample and $\bar{\widehat{\beta}}_1$ their sample mean. *N.C.* indicates

percentages of samples with any element of $\hat{\lambda}_b = \infty$.

In all cases, the mean values of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are very close to their true values. The approach also performs well for estimating the variance error component σ_e^2. On the other hand, the variances due to the random effect components are severely underestimated if the symmetric normal model is assumed with asymmetric data. Hence, the main conclusion is that if response follows an asymmetric normal distribution and a normal model is fitted, variance of random components will be underestimated, implicating that inter-individual association may not be significant when indeed it is significant.

5 An application

We illustrate the usefulness of the proposed algorithms by applying them to the Framingham cholesterol data. The file includes the cholesterol levels over time, age at baseline and gender for $m = 200$ randomly selected individuals. As in Zhang and Davidian (2001), we consider the following mixed linear model

$$y_{ij} = \beta_o + \beta_1 sex_j + \beta_2 age_j + \beta_3 t_{ij} + b_{oj} + b_{1j} t_{ij} + \epsilon_{ij}\,, \qquad (43)$$

where y_{ij} is cholesterol level divided by 100 at the i-th time for subject j and t_{ij} is $(time - 5)/10$, with time measured in years from baseline; age_j is age at baseline; sex_j is the gender indicator ($0 = female$, $1 = male$). Thus, $\mathbf{x}_{ij} = (1,\, age_j,\, sex_j, t_{ij})^T$, $\mathbf{b}_j = (1,\, t_{ij})^T$. Figure 1(a) shows the histogram of cholesterol levels, clearly indicating its asymmetric nature and that it seems adequate fitting a skew-normal model to the data set. Zhang and Davidian (2001) analyzed this data and show that the asymmetric behavior is partially explained by the available covariates and the random effects may not be normally distributed. Based in this conclusion, three statistical models, differing in the error term and random effects distributions, are entertained. These models are:

Model 1: A model with independent multivariate normal distribution for the errors and multivariate skew-normal distribution for random effects with $\boldsymbol{\lambda}_b = (\lambda_{b1},\, \lambda_{b1})^T$;

Model 2: A model with independent multivariate skew -normal distribution for random random errors with common shape parameter between groups and multivariate symmetric normal distribution for the random effects; and

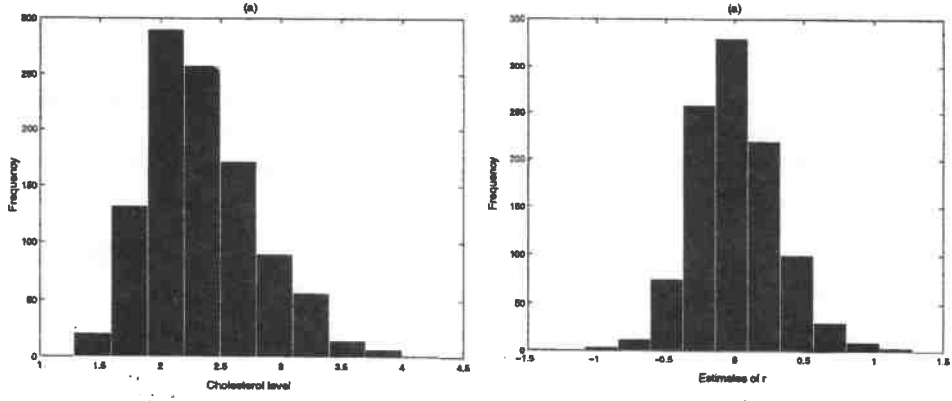
Model 3: A purely Gaussian model.

In all cases we considered $\psi_j = \sigma_e^2 \mathbf{I}_{n_j}, j = 1, \ldots, 200$ (conditional independence). Table 2 presents the results obtained using the EM-type algorithm of the three models described above, SE are the estimated asymptotic standard error based in the Hessian matrix, which was computed numerically by using the program MATLAB. When considering **Model 2** (only random errors are asymmetrically distributed), asymmetry is not detected and parameter estimates are close to the ones obtained under normality (**Model 3**), as expected. The AIC criteria indicate that **Model 1** presents the best fit. Notice that time factor is significant under normality but is not significant under **Model 1**. Figure 1(b) shows the histogram of $\widehat{\tau}$ as (31) for the model 1 indicating that this model fits well the data set.

Table 2: Results of fitting models 1, 2 and 3 to the cholesterol data. d_{11}, d_{12} and d_{22} are the distinct elements of the matrix $\mathbf{D}^{1/2}$

	Model 1		Model 2		Model 3	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
β_o	1.3678	0.1483	1.5725	0.1258	1.5968	0.1543
β_1	-0.0665	0.0549	-0.0632	0.0545	-0.0630	0.0568
β_2	0.0162	0.0036	0.0184	0.0033	0.0184	0.0037
β_3	0.0801	0.0480	0.2817	0.0249	0.2817	0.0242
σ_e	0.2086	0.0058	0.2083	0.0050	0.2084	0.0058
d_{11}	0.4755	0.0385	0.3724	0.0187	0.3715	0.0201
d_{12}	0.1315	0.0306	0.0562	0.0181	0.0563	0.0179
d_{22}	0.2456	0.0419	0.1868	0.0280	0.1868	0.0329
λ_b	2.0273	1.2918	-	-	-	-
λ_e	-	-	0.1563	0.1309	-	-
AIC	-1306.2916		-1301.512		-1303.5104	
iterations	187		144		72	

Figure 1: (a) Histogram of cholesterol levels for 200 subjects of Framingham cholesterol study. (b) Histogram of $\widehat{r}$ as in (31) for **Model 2.**



6 Final Conclusion

In this paper we developed a skew normal mixed model for fitting regression model with dependent data. We believe that this is the first attempt in working in such general distributional structure for mixed models and that the approach used in this paper can be used in treating other multivariate models which will be the subject of incoming papers. An analytical expression (closed form) is obtained for the marginal distribution of the observed response vector which allows carrying out inferences using standard optimization techniques and existing statistical software. For evaluation of the MLE, an EM-type algorithm is developed by exploring statistical properties of the model considered, and as is typical for the EM algorithm, reliability rather than speed is its best feature. An small simulation study is also presented where as observed in other context and approaches (e.g., Zhang and Davidian, 2001), there is potencial to gain efficiency in estimating variances due to the random effect components when the normality assumption does not hold. An additional major advantage of all approaches that relax the assumptions on the random effects density is the insight the estimates provide. We have implemented the approach using MATLAB software, code is available from the authors on request.

7 References

Arellano-Valle, R.B., del Pino, G. and San Martin, E. (2002). Definition and probabilistic properties of skew distributions. *Statistic and Probability Letters* **58**, 111-121.

Arellano-Valle, R.B. and del Pino, G. (2003). From symmetric to asymmetric distributions: A unified approach. *CRC/Chapman-Hall. Ed. Vol. by M.G. Genton on Skew-distributions and their applications: A Journey Beyond Normality.*

Arellano-Valle, R.B. and Genton, M. G. (2003). Fundamental skew distributions. *Institute of Statistics Mimeo Series* #2551, under review.

(http://www.stat.ncsu.edu/∼mggenton/publications.html)

Arellano-Valle, R.B., Ozan S., Bolfarine, H. and Lachos, V.H. (2003). Skew normal measurement error models, RT-MAE 2003-10,IME-USP.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics,* **12**, 171-178.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distributions. *Journal Royal Statistics Society,* **61**, 579-602.

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

Branco, M. and Dey, D. (2001). A general class of multivariate skew-elliptical distribution. *Journal of Multivariate Analysis*, *79*, 93-113.

Genton, M. G. and Loperfido, N. (2001). Generalized skew-elliptical distributions and their quadratic forms. *Institute of Statistics Mimeo Series* #2541, under review.

(http://www.stat.ncsu.edu/∼mggenton/publications.html)

Henze, N., (1986). A probabilistic representation of the "skew normal" distribution. *Scand. J. Statist.* **13**, 271-275.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate distributions,* **Vol. 1.** Wiley, New York.

Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics*, **38**, 963-974.

Zhang, D. and Davidian, M. (2001). Liner mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, **57**, 795-802.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2004-01 – CASTRO, M., CASTILHO, M.V., BOLFARINE, H. Consistent Estimation And Testing in Comparing Analytical Bias Models. 2004.23p. (RT-MAE-2004-01)

2004-02 – TAVARES, H. R., ANDRADE, D. F. Growth Curve Models for item Response Theory. 2004.11p. (RT-MAE-2004-02)

2004-03 – BUENO, V.C., MENEZES, J.E. Pattern's Reliability Importance Under Dependence Condition and Different Information Levels. 2004. 15p. (RT-MAE-2004-03)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05315-970 - São Paulo, Brasil