RT-MAE 2007-12

ROBUST LINEAR MIXED MODEL USING THE SKEW- T DISTRIBUTION

C.L. Bayes
and
M.D. Branco

Palavras-Chave: Skewness,thick-tailed distribution, repeated measures, mixed model, Bayesian inference, Gibbs sampling, model choice and robustness study.

Classificação 62F15, 62G35, 62J05

- Junho de 2007 -

Robust Linear Mixed Model using the Skew-t Distribution

C.L. Bayes and M.D. Branco[1]
University of São Paulo, Brazil

Abstract: *Normal linear mixed models are frequently used in repeated measures data analysis. However, the assumption of normality for the error and the random effects makes inference vulnerable to the presence of outliers, and also depends on the symmetry of the data. Here we propose to consider a skew-t distribution for both error and random effect terms. This flexible model includes as special case severals symmetric and asymmetric models considered in the literature. A Bayesian inference approach is adopted using a MCMC algorithms. A robustness study is carried out showing that the new model is more robust than the normal and skew-normal models to estimate the fixed effects parameters. An example about toxicology study in rats is analyzed and, using different bayesian model choice criteria, we concluded that the best model for this data set considers the skew-t distribution for the residuals.*

Key Words: Skewness, thick-tailed distribution, repeated measures, mixed model, Bayesian inference, Gibbs sampling, model choice and robustness study.

1 Introduction

Linear mixed models have been used to analyze repeated measures data or when it is assumed independence between cluster and dependence between the observations in the same cluster (within-subjects). The popularity of these models can be explained by the flexibility to model the within-subject correlation, by handling of both balanced and unbalanced data. A commonly used linear mixed-effects model for continuous response was proposed by Laird & Ware (1982) and is given by

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i \tag{1}$$

[1]Corresponce Address: M.D. Branco, Departamento de Estatística, IME, Rua do Matão 1010, Cidade Universitária, CEP 05508-090, USP, São Paulo, SP, Brasil. Email: mbranco@ime.usp.br

where, $Y_i(n_i \times 1)$ is a vector of observed responses, $\beta(p \times 1)$ a fixed effects vector, $X_i(n_i \times p)$ is a design matrix for the fixed effects, $Z_i(n_i \times q)$ is a design matrix for the random effects, $b_i(q \times 1)$ is a random effects vector and $\epsilon_i(n_i \times 1)$ a vector of residuals.

In general a normal distribution is assumed for both the random effects and the residuals. The b_i's are assumed to be independent with distribution $N_q(0, \Psi)$ an the ϵ_i's are assumed to be independent with distribution $N_{n_i}(0, \Sigma_i)$. Here, the Σ_i matrices depend on a typically small set of parameters. A special and common case is the homoscedastic specification $\Sigma_i = \sigma_e^2 I_{n_i}$, so we assume independence between the residuals of the same subject. Some authors have suggested the use of thick-tailed distribution to skip the lack of robustness of the normal model. Pinheiro, Liu & Wu (2001) suggest to use a joint multivariate t distribution for the random effects and the residuals. Using the maximum likelihood approach they proposed efficient algorithms to obtain these estimators. Under a Bayesian approach Rosa, Padovani & Gianola (2003) e Rosa, Gianola & Padovani (2004) compared multivariate and univariate specifications for the distribution of the residuals under some thick-tailed distributions. Their results suggest that unless there is a strong reason to believe in the adequacy of normality, it may be safer to use a robust model.

In another direction, the normality assumption is not reasonable when there is evidence of asymmetry in the data set. In this case, a skew distribution can be considered for both random effects and residuals. Under this view, some authors have been developing their works. Arellano-Valle, Bolfarine & Lachos (2005) used the multivariate skew-normal distribution proposed by Azzalini & Dalla Valle (1996) to model the random effects and the residuals. The authors consider the maximum likelihood approach, which they implement via an EM algorithm. Ghosh, Branco & Chakraborty (2007) used the multivariate skew-normal given by Sahu, Dey & Branco (2003) and a Bayesian approach. They applied that model for a HIV-RNA data set, MCMC is used to carry out the posterior analysis and model comparison . Ma, Genton & Davidian (2004) considered a generalized version of the multivariate skew-normal, called flexible skew-normal (Ma & Genton, 2004) for the random effects and normal distribution for the residuals. Since the skew-normal is not a thick-tailed distribution this is not appropriate for robustness interest. An alternative is considered a skew-t distribution, as given by Azzalini & Capitanio (2003), which takes care of both robust and asymmetry problems. Ma et al. (2004) also considered a flexible skew-t distribution for the random effects. However, they have not given much attention for this model, since they concluded that their data set was not well fit for the skew-t and they decided in favor of the skew-normal. Also, they used an improper prior for the degree of the freedom, so it is not clear if they obtained convergence for the MCMC algorithm. In fact, by analogy to the symmetric t-Student model we should avoid to consider improper prior for the degree of freedom (see Fernández & Steel, 1998).

The plan of the paper is as follows. In section 2 we present a little overview about the multivariate skew-t distribution that will be considered here. Section 3 presents the skew-t linear mixed model, and shows how some other models in the literature can be seen as a special case of the model proposed here. In section 4 the Bayesian inference methodology is presented. In section 5 a robustness study is considered comparing the normal, skew-normal, t-Student and skew-t models. Finally, in section 6 will be presented an application for reproductive toxicology data set and also is developed a model comparison analysis using CPO (Conditional Predictive Ordinate), pseudo-Bayes factor and DIC (Deviance Information Criterion).

2 Skew-t distribution

A random variable Y has a skew-t distribution if its probability density function (pdf) is given by

$$f_Y(y) = 2t_1\left(y;\mu,\sigma^2,\nu\right) T_1\left(\lambda\frac{y-\mu}{\sigma}\left(\frac{\nu+1}{q(y)+\nu}\right)^{1/2};0,1,\nu+1\right), \tag{2}$$

where $t_1(.;\psi,\eta,\tau)$ and $T_1(.;\psi,\eta,\tau)$ are the pdf and cumulated distribution function (cdf) of a t-Student distribution centered in ψ with scale η and τ degree of freedom, respectively, and $q(y) = (y-\mu)/\sigma^2$.

We consider the notation $Y \sim St_1(\mu,\sigma^2,\lambda,\nu)$, where $\mu \in \Re$, $\sigma^2 > 0$ are the location and scale parameters, respectively. Note that, if $\lambda = 0$ then $f_Y(y) = t_1(y;\mu,\sigma^2,\nu)$. λ is a shape parameter, also called skewness parameter, which gives the direction of the asymmetry ($\lambda < 0$ negative skewness and $\lambda > 0$ the skewness is positive). As in the t-Student distribution, the degree of freedom ν is associated with the thickness of the tail of the distribution.

A stochastic representation for the skew-t random variable Y with pdf in (2) is given by

$$Y = \mu + \frac{1}{w^{1/2}}\left(\eta u + \tau z\right), \tag{3}$$

where $\eta = \dfrac{\sigma\lambda}{\sqrt{1+\lambda^2}}$, $\tau = \dfrac{\sigma}{\sqrt{1+\lambda^2}}$, z has a standard normal distribution, u has a positive half normal and w a gamma distribution with parameters $\nu/2$ e $\nu/2$ ($w \sim Gamma(\frac{\nu}{2},\frac{\nu}{2})$) .

If $\nu > 2$, the mean and the variance for Y exist and are given, respectively, by

$$E[Y] = \mu + \sqrt{\frac{\nu}{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \eta \quad \text{and} \quad \text{var}[Y] = \frac{\nu}{\nu-2}\sigma^2 - \frac{\nu}{\pi}\left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)^2 \eta^2.$$

A multivariate version of the skew-t distribution can be constructed using the following stochastic representation

$$Y = \mu + W^{-1/2}\left(diag(\eta)u + \Lambda^{1/2}Z\right). \tag{4}$$

Where u is a vector in $\Re^n$ and each element u_i has a positive half normal distribution; $W = diag((w_1, ..., w_n))$, and each $w_i \sim Gamma(\nu/2, \nu/2)$, $i = 1, ..., n$; Z has a standard n dimensional normal distribution; Λ is a positive definitive matrix; η and μ are vectors in $\Re^n$.

If Λ is a diagonal matrix then the elements of Y are mutually independent and have skew-t univariate distribution. Several asymmetrical and symmetrical multivariate distributions can be represented by (4) depending on the values of u_i e w_i, see table 1, where u has a positive half normal and $w \sim Gamma(\frac{\nu}{2}, \frac{\nu}{2})$.

Table 1: Special cases of stochastic representation (4)

u_i	w_i	Distribution
0	1	Normal
0	w	t-Student
u	1	skew-normal type I
u_i	1	skew-normal type II
u	w	skew-t type I
u_i	w	skew-t type II

The multivariate skew-normal type I was given by Azzalini & Dalla Valle (1996). The multivariate skew-t type I was given by Branco & Dey (2001) and Azzalini & Capitanio (2003). The skew-normal and skew-t type II were given by Sahu et al. (2003).

We consider here the skew-t type I distribution, for both random effect and residuals. However, it is not difficult to adapt our results for the skew-t type II distribution. The skew-t type I pdf is given by

$$f_Y(y) = 2t_n\left(y;\mu,\Sigma,\nu\right)T_1\left(\lambda\Sigma^{-1/2}(y-\mu)\left(\frac{\nu+n}{q(y)+\nu}\right)^{1/2};\nu+1\right),\quad y\in\Re^n, \tag{5}$$

where $\Sigma = \Lambda + \eta\eta^T$, $\lambda = \dfrac{\Sigma^{1/2}\Lambda^{-1}\eta}{\sqrt{1+\eta^T\Lambda^{-1}\eta}}$ and $q(y) = (y-\mu)^T\Sigma^{-1}(y-\mu)$.

We consider the notation $Y \sim St_n(\mu,\Sigma,\lambda,\nu)$. In this case, the mean and the covariance matrix for Y exist if $\nu > 2$ and are given by

$$E(Y) = \mu + \sqrt{\frac{\nu}{\pi}}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\eta \quad , \quad \nu > 1$$

and

$$Var(Y) = \frac{\nu}{\nu-2}\Sigma - \frac{\nu}{\pi}\left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)^2\eta\eta^T \quad , \quad \nu > 2.$$

An important property of this distribution is to be closed by linear transformations. Then, any linear combination of a skew-t random vector is also a skew-t random vector. This property help us to adopt the skew-t distribution to model an additive error in regression models, as we will do in the next section.

3 Skew t Linear Mixed Model

Let's remember model (1)

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i$$

with all terms as before, but now considering

$$\begin{aligned} b_i &\sim St_q\left(0,\Psi,\lambda_b,\nu_b\right), \\ \epsilon_{ij} &\sim St_1\left(0,\sigma_e^2,\lambda_e,\nu_e\right), \end{aligned} \tag{6}$$

$i = 1,\ldots,m$ and $j = 1,\ldots n_i$; with b_i independent of ϵ_{ij} and the ϵ_{ij}'s are mutually independent. Using the stochastic representation given in (4), this model can be alternatively written in a hierarchical way :

$$
\begin{aligned}
y_{ij} \mid b_i, u_{eij}, w_{eij} &\sim N_1\left(x_{ij}^T\beta + z_{ij}^T b_i + \eta_e \frac{u_{eij}}{w_{eij}^{1/2}}, \frac{\tau_e^2}{w_{eij}}\right) \\
b_i \mid u_{bi}, w_{bi} &\sim N_q\left(\eta_b \frac{u_{bi}}{w_{bi}^{1/2}}, \frac{\Lambda}{w_{bi}}\right) \\
u_{eij}, u_{bi} &\sim HN(0,1) \\
w_{eij} &\sim Gama\left(\frac{\nu_e}{2}, \frac{\nu_e}{2}\right) \\
w_{bi} &\sim Gama\left(\frac{\nu_b}{2}, \frac{\nu_b}{2}\right)
\end{aligned} \tag{7}
$$

where x_{ij} and z_{ij} are the j-row of X_i and Z_i respectively; $\sigma_e^2 = \tau_e^2 + \eta_e^2$; $\lambda_e = \frac{\eta_e}{\tau_e}$; $\Psi = \Lambda + \eta_b \eta_b^T$ and $\lambda_b = \frac{\Psi^{1/2}\Lambda^{-1}\eta_b}{\sqrt{1+\eta_b^T \Lambda_b^{-1}\eta_b}}$.

Some models in the literature can be seen as a special case of the Skew-t Linear Mixed Model (StLMM), depending on the values of w_{eij}, w_{bi}, u_{eij} and u_{bi}. For example, when $u_{eij} = u_{bi} = 0$ and $w_{eij} = w_{bi} = w_i$, where w_i have independent gamma distributions with parameters $\nu/2$ e $\nu/2$, we obtain a joint multivariate t distribution for the random effects and the residuals,

$$
\begin{pmatrix} \epsilon_i \\ b_i \end{pmatrix} \sim t_{n_i+q}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 I_{ni} & 0 \\ 0 & \Psi \end{pmatrix}, \nu\right)
$$

which is the model proposed by Pinheiro et al. (2001). Other examples can be seen in table 2.

Table 2: Special cases of StLMM

u_{eij}	u_{bjh}	w_{eij}	w_{bjh}	Model
0	0	1	1	Normal
0	0	w_i	w_i	Joint multivariate t [a]
0	0	w_{ei}	w_{bj}	Non-correlated errors t [b]
0	0	w_{eij}	w_{bj}	Independent errors t [b]
u_{ei}	u_{bj}	1	1	Skew-normal [c]

[a]Pinheiro et al. (2001)
[b]Rosa et al. (2004)
[c]Arellano-Valle et al. (2005)

Using properties of the skew-t distribution, we can see that

$$E[Y_i] = X_i\beta + \sqrt{\frac{\nu_b}{\pi}}\frac{\Gamma\left(\frac{\nu_b-1}{2}\right)}{\Gamma\left(\frac{\nu_b}{2}\right)}Z_i\eta_b + \sqrt{\frac{\nu_e}{\pi}}\frac{\Gamma\left(\frac{\nu_e-1}{2}\right)}{\Gamma\left(\frac{\nu_e}{2}\right)}\eta_e \tag{8}$$

and

$$Var[Y_i] = Z_i Var[b_i] Z_i + Var[\epsilon_i], \tag{9}$$

where

$$Var[b_i] = \frac{\nu_b}{\nu_b - 2}\Psi - \frac{\nu_b}{\pi}\left(\frac{\Gamma\left(\frac{\nu_b-1}{2}\right)}{\Gamma\left(\frac{\nu_b}{2}\right)}\right)^2 \eta_b\eta_b^T$$

and

$$Var[\epsilon_i] = \left[\frac{\nu_e}{\nu_e - 2}\sigma_e^2 - \frac{\nu_e}{\pi}\left(\frac{\Gamma\left(\frac{\nu_e-1}{2}\right)}{\Gamma\left(\frac{\nu_e}{2}\right)}\right)^2 \eta_e^2\right] I_{n_i}.$$

4 Bayesian Inference

To complete the Bayesian specification of the model, it is assumed a prior distribution for all the unknown quantities.

For the univariate skew-normal distribution, Bayes & Branco (2007) in a simulation study demonstrated that assumed a prior that $\lambda \sim t_1(0, 1/2, 2)$ and $p(\sigma^2) \propto 1/\sigma^2$ leads to better point and interval estimates. Considering this prior specifications and using usual variable transformation methods, it induces the following prior distribution for the parametrization used in the stochastic representation (3), $\eta \sim t_1(0, \tau^2/2, 2)$ and $p(\tau^2) \propto 1/\tau^2$. These prior distributions are adopted for the skewness and scale parameters of the error distribution, with a slightly modification, we consider a proper distribution for the scale,

$$\begin{aligned} \eta_e &\sim t_1(0, \tau_e^2/2, 2) \\ \frac{1}{\tau_e^2} &\sim Gamma\left(a_0, b_0\right). \end{aligned} \tag{10}$$

For the random effects skewness and scale parameters we assumed a multivariate version of prior distribution (10) considering the multivariate t and a Inverse Wishart distributions,

$$\begin{aligned} \eta_b &\sim t_q\left(0, \frac{1}{2}\Psi, 2\right) \\ \Psi &\sim Inv - Wishart\left(a_1, B_1\right). \end{aligned} \tag{11}$$

The prior distributions for the degrees of freedom parameters ν_e and ν_b are assumed to be

$$
\begin{aligned}
p(\nu_e) &\propto \nu_e^{-2} I(\nu_e > 3), \\
p(\nu_b) &\propto \nu_b^{-2} I(\nu_b > 3),
\end{aligned} \tag{12}
$$

as in Liu (1995) and Rosa et al. (2003). The truncation point was chosen to assure that the mean and variance of the skew-t distribution are finite.

We assume a multivariate normal distribution for the fixed effects β,

$$
\beta \sim N_p(\mu_\beta, \Omega_\beta). \tag{13}
$$

The augmented joint posterior density is the product of the augmented likelihood function that can be found from (7) and the prior distribution given by (10)-(13). We can apply Markov Chain Monte Carlo (MCMC) algorithms to draw samples from the posterior density. It is not difficult to see that the full conditional distributions have standard form for the prior specification given by (11), (12) and (13) with exception for the degrees of freedom parameters. The Gibbs sampler algorithm also can be easily implemented via BUGS software (Spiegelhalter, Thomas, Best & Lunn, 2004). For obtaining the results in the next section we used this software, and the programs can be obtained by the authors upon request.

5 Robustness study

We considered the data set in Potthoff & Roy (1964) about an orthodontic study of 16 boys and 11 girls between 8 and 14 years. It has been analyzed by Pinheiro et al. (2001) using a joint t-distribution for the random effects and the residuals under a classical approach, and, by Rosa et al. (2004) under a Bayesian framework and they assumed independent t distributions for the random effects and residuals. The response variable is the distance, in milimeters, from the center of the pituitary to the pteryomaxillary fissure, taken at 8, 10, 12 and 14 years. As did by Pinheiro et al. (2001), to our study of the influence of a single outlier, we consider only the girls from the Potthoff & Roy (1964) data set. The model is given by

$$
y_{ij} = \beta_0 + \beta_1 t_j + b_{0i} + b_{1i} t_j + \epsilon_{ij}, \tag{14}
$$

where y_{ij} is the distance in milimeters from the center of the pituitary to the pteryomaxillary fissure at time j of the subject i; t_j is the age measured in years; β_0 e β_1 are the intercept and the slope fixed effects; $b_i = (b_{0i}, b_{1i})^T$ is the random effects vector for the i–th subject; and e_{ij} is the within-subject error, $i = 1, ..., 11$ and $j = 1, ..., 4$.

Four models were considered: skew-t (St), skew-normal (SN), t and normal(N). In all of them the same distribution was considered for both random effects and residuals. Since the parameters β_0 e β_1 do not have the same interpretation for all models, to make the models comparable we consider

$$E[y_{ij}] = \beta_0^* + \beta_1^* t_j$$

where the expression for β_0^* and β_1^* are in table 3.

Table 3: Corrected fixed effects parameters

Model	β_0^*	β_1^*
St	$\beta_0 + \sqrt{\frac{\nu_b}{\pi}} \frac{\Gamma\left(\frac{\nu_b-1}{2}\right)}{\Gamma\left(\frac{\nu_b}{2}\right)} \eta_{b0} + \sqrt{\frac{\nu_e}{\pi}} \frac{\Gamma\left(\frac{\nu_e-1}{2}\right)}{\Gamma\left(\frac{\nu_e}{2}\right)} \eta_e$	$\beta_1 + \sqrt{\frac{\nu_b}{\pi}} \frac{\Gamma\left(\frac{\nu_b-1}{2}\right)}{\Gamma\left(\frac{\nu_b}{2}\right)} \eta_{b1}$
SN	$\beta_0 + \sqrt{\frac{2}{\pi}} \eta_{b0} + \sqrt{\frac{2}{\pi}} \eta_e$	$\beta_1 + \sqrt{\frac{2}{\pi}} \eta_{b1}$
t	β_0	β_1
N	β_0	β_1

We consider the influence of a change of Δ units in a single measurement on the posterior distribution of the corrected fixed effects parameters. That is, we replace a single data point y_{ij} by the contaminated value $y_{ij}(\Delta) = y_{ij} + \Delta$, re-estimate the model and record the posterior mean and HPD interval with probability of 95%. In this example, we contaminated the fourth observation (age of 14 years) of the first girl, and varied Δ between -20mm and 20mm by increments of 2mm.

Figures 1 and 2 present the posterior mean and the 95%-HPD interval for the four models. As expected, the skew-t and t models are less sensitive to variations in Δ than the normal and skew-normal models. The variations in Δ have considerable impact, for the normal models, in both the posterior mean and HPD. Note that, the size of the interval increases as $|\Delta|$ increases.

Table 4 shows the DIC (Spiegelhalter, Best, Carlin & Van der Linde, 2002) and CPO product (Gelfand, 1996) for the four models for the non-contaminated data set; both criteria choose the skew-normal model (bigger CPO's product and smaller DIC). Therefore, a skew model should be considered for this data set.

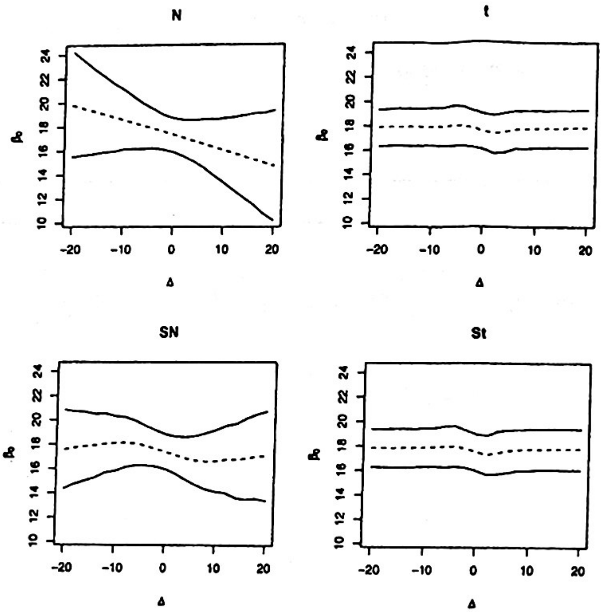

Figure 1: Posterior mean (dashed line) and 95%-HPD interval (solid line) for β_0^* under Skew-t, Skew-normal, t and normal distibution for different contaminations of Δ of a single observation

Table 5 presents the posterior mean and a 95%-HPD interval for β_0^* and β_1^*, we note that the estimates are very close. This suggests that there is not much difference between the models to estimate the fixed effects, however for the skew-normal and the normal models a single outlier observation could have a considerable impact on these estimates. Since this behavior was not observed under the skew-t models, it could be safer to use a robust model instead the skew-normal model.

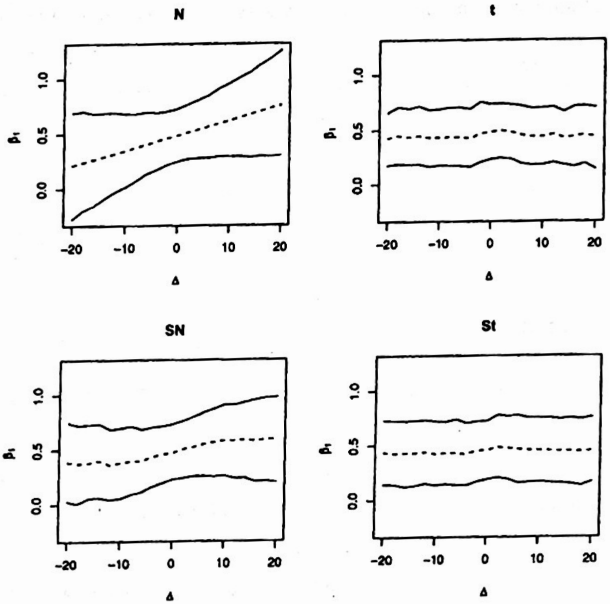

Figure 2: Posterior mean (dashed line) and 95%-HPD interval (solid line) for β_1^* under Skew-t, Skew-normal, t and normal distibution for different contaminations of Δ of a single observation

Table 4: Model comparison measures for the orthodontic data set

Distribution	prod CPO	DIC
St	2.39×10^{-26}	111.45
SN	4.00×10^{-26}	106.17
t	2.06×10^{-26}	115.18
N	2.60×10^{-26}	112.51

Table 5: Posterior mean and 95%-HPD interval for the corrected fixed effects.

	St	*SN*	*t*	N
β_0^*	17.53	17.39	17.54	17.37
	(16.01 , 18.98)	(15.98 , 18.84)	(16.07 , 18.97)	(15.93 , 18.74)
β_1^*	0.45	0.47	0.48	0.48
	(0.19 , 0.72)	(0.22 , 0.73)	(0.23 , 0.73)	(0.24 ,0.73)

6 Application

The data set considered in this section was analyzed by Dempster, Patel, Selwyn & Roth (1984) and Rosa et al. (2003). The data are from a reproductive toxicology study in rats, the outcome variable is birth weights (in grams) of rat pups. In this experiment, 30 dams were randomly allocated into three equal size treatment groups: control, a low dose, and a high dose of the test substance. Data from only 7 litters were available for the high dose group, and the number of pups per litters range from 2 to 18. These data were analyzed according to a mixed model with two sources of random variation: between and within litters. The model also considered the following fixed effects: litter size, sex and treatment. We adopted the same linear mixed model used by Dempster et al. (1984) and Rosa et al. (2003), given by

$$y_{ij} = \beta_0 + \beta_1 LD_i + \beta_2 HD_i + \gamma L_i + \delta S_{ij} + b_i + e_{ij}, \tag{15}$$

where y_{ij} is birth weights (in grams) of rat pup j of litter i; LD_i is a dummy variable for the low dose of the test substance for the litter i; HD_i is a dummy variable for the high dose of the test substance for the litter i; L_i is the size of litter i; S_{ij} is a variable indicator for male rat pups; β_0, β_1 and β_2 are the fixed effects for treatment; γ is the slope fixed effect for the size of the litter; δ is the fixed effect for the sex of the rat pup; b_i is the random effect for the litter i; and e_{ij} is the within-litter error, $i = 1, ..., 47$, $j = 1, ..., L_i$.

Figure 3 presents the histogram for birth weights (in grams) of rat pups, which indicates asymmetry of the data and that could be reasonable to fit a model that takes in account this asymmetry. We assumed $e_{ij} \sim St_1(0, \sigma_e^2, \lambda_e, \nu_e)$ and $b_i \sim St_1(0, \sigma_b^2, \lambda_b, \nu_b)$.

We fitted several models that differ in the distribution for random effects and the errors. Four distributions are considered: skew-t (St), skew-normal (SN), t and normal(N). To carry

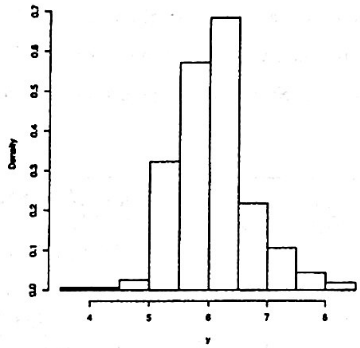

Figure 3: Histogram for birth weights (in grams) of rat pups

out a model comparison, we computed the pseudo Bayes factor (Gelfand, 1996) for each model relative to the normal model. The highest pseudo Bayes factor was obtained when we assume *St* distribution for the residuals, for these models we also calculated the DIC (Spiegelhalter, Best, Carlin & Van der Linde, 2002).

The pseudo Bayes factor (pBF) and DIC for the models with *St* error are presented in table 6. These criteria indicate that the model which assumed a normal distribution for the random effects presents the best fit.

Table 6: Model comparison measures for the rats data set

Random Effects	Residuals	pFB	DIC
N	N	1.00	356.59
St	*St*	1.36×10^{21}	262.74
SN	*St*	2.24×10^{21}	262.14
t	*St*	1.82×10^{21}	262.47
N	*St*	2.32×10^{21}	262.12

We have also considered the values of CPO for each observation over the models (Gelfand, 1996). In figure 4 is presented the CPO for the following four models: (1) normal for both random effects and errors; (2) normal for random effects and t for the errors; (3) normal for random effects and SN for the errors; and (4) normal for random effects and St for the errors. Models 2 and 4 are clearly better than the other models, the pBF for model 4 versus model 2 is 57.14 which leads us to choose the St as the best model for the errors.

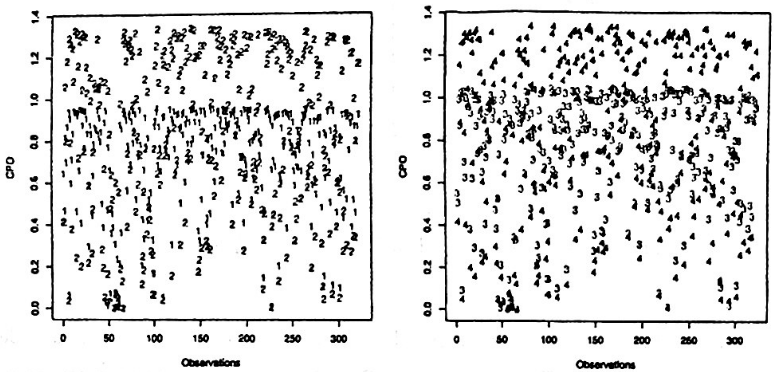

Figure 4: CPO for the models: (1) N–N, (2) N–t, (3) N–SN and (4) N–St

In table 7 we present the posterior mean and a 95%–HPD interval for the models (1) and (4). The HPD for skewness parameter λ_e does not include zero which confirms the negative asymmetry of the data. The narrow interval for the degrees of freedom parameter for the residuals (ν_e) shows us the inadequacy of the normal or skew-normal models in this case. Notice that the effect of sex δ seems to be overestimated under the normal model and also have a wider HPD interval, 23.4% wider than the model N–St.

Table 7: Posterior mean and 95%−HPD interval for models (1) and (4)

Parameter	N $-St$	N − N
β_0^*	8.19	7.95
	(7.56 , 8.79)	(7.41 , 8.51)
β_1	−0.52	−0.43
	(−0.85 , −0.17)	(−0.74 , −0.12)
β_2	−0.94	−0.86
	(−1.34 , −0.53)	(−1.25 , −0.50)
δ	0.30	0.36
	(0.23 , 0.38)	(0.26 , 0.45)
γ	−0.14	−0.13
	(−0.18 , −0.10)	(−0.17 , −0.09)
$Var[b_i]$	0.14	0.11
	(0.06 , 0.23)	(0.04 , 0.19)
$Var[e_{ij}]$	0.16	0.16
	(0.11 , 0.21)	(0.14 , 0.19)
λ_e	−0.77	
	(−1.58 , −0.003)	-
ν_e	4.13	
	(3.00 , 5.81)	-

7 Conclusion

The skew-t linear mixed model shows to be a new robust alternative for modelling repeated measures data. The advantage of this model over the t-Student is the flexibility to work with asymmetrical data, this avoid variables transformation that many times are considered when the condition of symmetry is not observed. Also, many other linear mixed model proposed in the literature can be seen as a special case of the StLMM. Although the model is very flexible to data fitting, there is not much additional effort to do Bayesian inference. Using the hierarchical structure, we showed it can be easily implemented, for example, using the free software BUGS (http://www.mrc-bsu.cam.ac.uk/bugs). However, some inferential points need more attention when the interest is to estimate the parameters associated with the kurtosis (ν) and skewness (λ) of the distribution. For example, we observed that the estimates of the degree of freedom parameter are very sensitive to prior choice, suggesting the importance of the prior specification for this parameter. This sensibility is not observed in the estimates of the fixed effects, which were the main focus of analysis in this paper.

The hierarchical structure of the StLMM given in section 3 is fundamental to make easier the inferential implementation. Similar structure can be easily built if the skew-t type II were considered in the place of the skew-t type I distribution.

8 Acknowledgments

The authors acknowledge the *Fundação de Amparo a Pesquisa do Estado de São Paulo* (FAPESP grants 04/14721 − 2 and 04/15304 − 6) and *Conselho Nacional de Pesquisa* (CNPq grant 310403/2006 − 5) for financial support.

References

Arellano-Valle, R. B., Bolfarine, H. & Lachos, V. (2005). Skew-normal linear mixed models, *Journal of Data Science* **3**: 415–438.

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistical Society Ser. B* **65**: 367–389.

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

Bayes, C. L. & Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution, *Brazilian Journal of Probability and Statistics . To appear.*

Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions, *Journal of Multivariate Analysis* **79**: 99–114.

Dempster, A. P., Patel, C. M., Selwyn, M. R. & Roth, A. J. (1984). Statistical and computational aspects of mixed model analysis, *Applied Statistics* **33**(2): 203–214.

Fernández, C. & Steel, M. F. J. (1998). On bayesian modelling of fat tails and skewness, *Journal of The American Statistical Association* **93**: 359–371.

Gelfand, A. E. (1996). Model determination using sampling-based methods, *in* W. Gilks, S. Richardson & D. J. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, 1st edn, Chapman Hall, London, chapter 9, pp. 145–161.

Ghosh, P., Branco, M. D. & Chakraborty, H. (2007). Bivariate random effect model using skew-normal distribution with application to HIV-RNA, *Statistics in Medicine* **26**: 1255–1267.

Laird, N. L. & Ware, J. H. (1982). Random-effects model for longitudinal data, *Biometrics* **38**: 963–974.

Liu, C. (1995). Missing data imputation using the multivariate t distribution, *Journal of Multivariate Analysis* **53**: 139–158.

Ma, Y. & Genton, M. G. (2004). A flexible class of skew-symmetric distributions, *Scandinavian Journal of Statistics* **31**: 459–468.

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

Pinheiro, J. C., Liu, C. & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution, *Journal of Computational and Graphical Statistics* **10**: 249–276.

Potthoff, R. F. & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* **51**: 313–326.

Rosa, G. J. M., Gianola, D. & Padovani, C. R. (2004). Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC, *Journal of Applied Statistics* **31**(7): 855–873.

Rosa, G. J. M., Padovani, C. R. & Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and bayesian MCMC implementation, *Biometrical Journal* **45**(5): 573–590.

Sahu, S. K., Dey, D. K. & Branco, M. D. (2003). A new class of multivariate skew distributions with aplications to bayesian regression models, *The Canadian Journal of Statistics* **31**(2): 129–150.

Spiegelhalter, D. J., Best, N., Carlin, B. & Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society, Series B* **64**: 583–640.

Spiegelhalter, D. J., Thomas, A., Best, N. G. & Lunn, D. (2004). *WinBUGS User Manual Version 1.4.1.* http://www.mrc-bsu.cam.ac.uk/bugs.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2007-01 - POLPO, A., PEREIRA, C.A.B. Reliability nonparametric bayesian estimation in parallel systems 2007. 12p. (RT-MAE 2007-01)

2007-02 - GONÇALVES, M., Kolev, N. V., Fabris. A. E. Bounds for quantile-based measures of dependent risks' functions. 2007. 16p. (RT-MAE 2007-02)

2007-03 - MIMBELA, J. A. L, Kolev, N.V. Occupation measure of Markov - modulated risk processes 2007. 12p. (RT-MAE 2007-03)

2007-04 - GONÇALVES, M., Kolev, N. V., Fabris, A. E. Bounds for distorted risk measuress. 2007. 11p. (RT-MAE 2007-04)

2007-05 - BUENO, V.C., A series representation of a parallei systems. 2007. 10p. (RT-MAE 2007-05)

2007-06 - KOLEV, N.V., PAIVA, D. Copula-Based Regression Models. 2007. 27p. (RT-MAE 2007-06)

2007-07 - POLETO, F. Z., SINGER, J.M., PAULINO, C.D., A product-multinomial framework for categorical data analysis with missing responses. 2007. 27p (RT-MAE 2007-07)

2007-08 - KOLEV, N.V., GONCALVES, M. DIMITROV, B. Probabilistic Properties of Sibuya's Dependence Function. 2007. 13p. (RT-MAE-2007-08)

2007-09 - AZEVEDO, C.L.N., ANDRADE, D.F. Latent Trait Estimation in Nominal Response Model: Latent Trait Asymmetry and Hierarchical and Empirical Framework. 2007.20p. (RT-MAE-2007-09)

2007-10 - Figueiredo, C.C., Sandoval, M.C., Bolfarine,H., Lima, C.R.O.P., Skew-Normal Linear Calibration: A Bayesian Perspective. 2007.18p. (RT-MAE-2007-10)

2007-11 - Bayes, C.L., Branco, M.D., Robust Linear Mixed Model using the Skew-t Distribution. 2007.18p (RT-MAE-2007-11.)

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
05311-970 - São Paulo, Brasi