



Bayesian computational strategies for multivariate t linear mixed models with missing outcomes

Tsung-I Lin*

Institute of Statistics, National Chung Hsing University, Taichung, Taiwan – tilin@nchu.edu.tw

Wan-Lun Wang

Department of Statistics, Feng Chia University, Taichung, Taiwan – wlunwang@fcu.edu.tw

Abstract

Multivariate t linear mixed models (MtLMM) have been recently proposed as a robust tool for analyzing multivariate longitudinal data with outlying observations and missing values simultaneously. As a powerful alternative to the traditional EM-type algorithm employing only single imputed values, this paper provides a Bayesian imputation method for the MtLMM to account for the uncertainties of model parameters as well as missing outcomes. Natural conjugate and weakly informative priors are employed to ensure properness and robustness in the sense that the posterior inference would not be influenced too much by the choice of hyperparameters. An inverse Bayes formulas (IBF) sampler coupled with Metropolis-within-Gibbs scheme is used to effectively draw the posterior distributions of latent data and model parameters. The techniques for multiple imputation of missing values, estimation of random effects, and diagnostics of potential outliers are investigated as well. The proposed methodology is illustrated through a simulation study and an application to AIDS/HIV data.

Keywords: conditional conjugate priors; damped exponential correlation; data augmentation; IBF sampler; missing data.

1. Introduction

Multivariate linear mixed models (MLMMs) originally proposed by Shah et al. (1997) have been widely used for analyzing multi-outcome longitudinal data which arise in many areas such as clinical trials, biomedical studies, epidemiology, and sociology. However, the MLMMs, usually suffering from a lack of robustness against departures from normality assumptions, may be too restrictive to provide an accurate representation of the data in the presence of outliers or heavy-tail noises. To remedy such weakness, Wang and Fan (2011) proposed a robust extension of MLMM, called the multivariate t linear mixed model (MtLMM), by imposing a joint multivariate t distribution (2004) for the random effects and within-subject errors.

In longitudinal studies, missing values commonly occur due to a variety of reasons such as missed scheduled visits, withdrawal from a study, loss to follow-up, and death or disabling event. Statistical analysis of longitudinal data with missing data has received increasing attention in recent years. It is well known that ignoring the missing information could hamper the representativeness of the remaining observations and likely leads to biased inferences. To eliminate such biases, imputation of the missing part of the data through an appropriate statistical modeling approach becomes an important task. Most procedures rely on assumptions of missing data mechanisms in the sense of Rubin (1976), including missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).

The aim of this paper is to develop a fully Bayesian approach to the MtLMM for analyzing irregularly and partially observed longitudinal data in the presence of outliers under the MAR condition. To describe the autocorrelation among irregular appointments of each characteristic, a damped exponential correlation (DEC) structure (Muñoz et al. 1992) is used in the scale-covariance structure of the within-subject errors. To make posterior inference of interest more reliable, we develop a hybrid Bayesian sampling procedure which combines two stages: imputation stage and posterior stage. In the imputation stage, we perform a non-iterative inverse Bayes formulae (IBF) sampler (Tan et al. 2003) to impute all latent data by simulating a set of independent and identically distributed (iid) samples approximately from the posterior distributions of the underlying latent variables. As such, multiple imputation in the sense of replacing each missing value with

a set of reasonable samples generated from the corresponding posterior distributions of missing responses is implemented to properly reflect the uncertainty due to missing information. In the posterior stage, we apply an iterative Metropolis-within-Gibbs sampler (Geman and Geman 1984; Hastings 1970) to draw posterior samples of model parameters.

2. Bayesian formulation of MtLMM with DEC errors

2.1. Model and likelihood function

Suppose that there are N subjects in the study, with subject i having s_i observations on each of the r outcome variables. Let $\mathbf{Y}_i = [\mathbf{y}_{i1} : \mathbf{y}_{i2} : \dots : \mathbf{y}_{ir}]$ be a $s_i \times r$ matrix of responses for subject i ($i = 1, \dots, N$), where $\mathbf{y}_{ij} = (y_{ij,1}, \dots, y_{ij,s_i})^\top$ is the response vector for outcome j ($j = 1, \dots, r$). Let $\mathbf{E}_i = [\mathbf{e}_{i1}, \dots, \mathbf{e}_{ir}]$ be a $s_i \times r$ matrix of within-subject errors corresponding to \mathbf{Y}_i . Let $\mathbf{y}_i = \text{vec}(\mathbf{Y}_i)$ and $\mathbf{e}_i = \text{vec}(\mathbf{E}_i)$. Following the notation of Wang (2013), the MtLMM for each vector \mathbf{y}_i is

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i \quad \text{with} \quad (\mathbf{b}_i^\top, \mathbf{e}_i^\top)^\top \stackrel{\text{ind}}{\sim} t_{q+n_i}(\mathbf{0}, \text{diag}\{\mathbf{D}, \mathbf{R}_i\}, \nu), \quad (1)$$

where $t_a(\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$ denotes the a -variate t -distribution with location $\boldsymbol{\mu}$, scale covariance $\boldsymbol{\Omega}$ and degrees of freedom (df) ν ; $n_i = s_i r$; $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_r^\top)^\top$ is a $p \times 1$ vector of fixed effects, with $p = \sum_{j=1}^r p_j$ and each $p_j \times 1$ sub-vector $\boldsymbol{\beta}_j$ used to describe the mean profile of outcome j ; $\mathbf{b}_i = (\mathbf{b}_{i1}^\top, \dots, \mathbf{b}_{ir}^\top)^\top$ is a $q \times 1$ vector of random effects, with $q = \sum_{j=1}^r q_j$ and each $q_j \times 1$ sub-vector \mathbf{b}_{ij} corresponding to subject differences among \mathbf{y}_{ij} . Moreover, the variance components in (1) involve an unstructured $q \times q$ positive-definite matrix $\mathbf{D} = [D_{jj'}]$ in which D_{jj} is the dispersion matrix of \mathbf{b}_{ij} , and an $n_i \times n_i$ scale covariance matrix \mathbf{R}_i .

To simultaneously account for variance-covariances within and between outcomes at the same occasion and serial correlations of a given outcome over occasions, we adopt a Kronecker product (KP) structure on scale covariance matrix \mathbf{R}_i . Let $\boldsymbol{\Sigma} = [\sigma_{jj'}] \in \mathbb{R}^{r \times r}$ and $\mathbf{C}_i \in [-1, 1]^{s_i \times s_i}$ be the scale-covariance and time-dependence correlation matrices of each row and column of \mathbf{E}_i , respectively. Then we write $\mathbf{R}_i = \boldsymbol{\Sigma} \otimes \mathbf{C}_i$, where \otimes denotes the KP. Here we consider a parsimonious DEC structure defined as

$$\mathbf{C}_i = \mathbf{C}_i(\phi, \gamma; \mathbf{t}_i) = [\phi^{|t_{ik} - t_{ik'}|^\gamma}],$$

where the autoregressive (AR) coefficient $\phi \in [0, 1)$ describes the autocorrelation between observations separated by the absolute length of two occasions, and the damping parameter $\gamma \in [0, \infty)$ permits acceleration of the exponential decay of the autocorrelation function defining a continuous-time AR model.

To handle possibly missing responses, we partition \mathbf{y}_i of each subject into two components for ease of exposition, say the observed response vector \mathbf{y}_i^o and the missing response vector \mathbf{y}_i^m . To facilitate computation, we introduce two permutation matrices \mathbf{O}_i and \mathbf{M}_i with dimensions $n_i^o \times n_i$ and $(n_i - n_i^o) \times n_i$, which are extracted from an identity matrix of order n_i corresponding to the respective row positions of \mathbf{y}_i^o and \mathbf{y}_i^m in \mathbf{y}_i . Accordingly, we have $\mathbf{y}_i^o = \mathbf{O}_i \mathbf{y}_i$, $\mathbf{y}_i^m = \mathbf{M}_i \mathbf{y}_i$, and $\mathbf{y}_i = \mathbf{O}_i^\top \mathbf{y}_i^o + \mathbf{M}_i^\top \mathbf{y}_i^m$. Define $\mathbf{X}_i^o = \mathbf{O}_i \mathbf{X}_i$, $\mathbf{X}_i^m = \mathbf{M}_i \mathbf{X}_i$, $\mathbf{Z}_i^o = \mathbf{O}_i \mathbf{Z}_i$, and $\mathbf{Z}_i^m = \mathbf{M}_i \mathbf{Z}_i$. It follows that marginally $\mathbf{y}_i^o \sim t_{n_i^o}(\mathbf{X}_i^o \boldsymbol{\beta}, \boldsymbol{\Lambda}_i^o, \nu)$, where $\boldsymbol{\Lambda}_i^o = \mathbf{O}_i \boldsymbol{\Lambda}_i \mathbf{O}_i^\top$ with $\boldsymbol{\Lambda}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \boldsymbol{\Sigma} \otimes \mathbf{C}_i$.

By the essential property of the multivariate t distribution, $\mathbf{y}_i \sim t_{n_i}(\mathbf{X}_i \boldsymbol{\beta}, \boldsymbol{\Lambda}_i, \nu)$ can be hierarchically represented as $\mathbf{y}_i \mid \tau_i \sim \mathcal{N}_{n_i}(\mathbf{X}_i \boldsymbol{\beta}, \tau_i^{-1} \boldsymbol{\Lambda}_i)$ and $\tau_i \sim \text{Gamma}(\nu/2, \nu/2)$. Let $\mathbf{y}^o = (\mathbf{y}_1^o, \dots, \mathbf{y}_N^o)$, $\mathbf{y}^m = (\mathbf{y}_1^m, \dots, \mathbf{y}_N^m)$, $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_N)$, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)$, and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\Sigma}, \phi, \gamma, \nu)$ be the entire model parameters. The likelihood functions of $\boldsymbol{\theta}$ associated with the complete data $(\mathbf{y}^o, \mathbf{y}^m)$, $(\mathbf{y}^o, \mathbf{y}^m, \boldsymbol{\tau})$, and $(\mathbf{y}^o, \mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau})$ can be obtained (Wang 2013).

2.2. Priors and full conditional posteriors

To complete the Bayesian specification, prior distributions have to be assigned to model parameters. Assume that the parameters $\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\Sigma}, \phi, \gamma$ and ν are a priori independent. When the prior information is unavailable, a convenient strategy of avoiding improper posterior distributions is to use diffuse proper priors. The prior distributions adopted are as follows: $\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\beta}_0, \mathbf{F}_0)$, $\mathbf{D} \sim \mathcal{IW}(d_0, \mathbf{G}_0)$, $\boldsymbol{\Sigma} \sim \mathcal{IW}(s_0, \mathbf{H}_0)$, $\phi \sim \mathcal{U}(0, 1)$, $(1 + \gamma)^{-1} \sim \mathcal{U}(0, 1)$, $(1 + \nu)^{-1} \sim \mathcal{U}(0, 1)$, where $\mathcal{U}(0, 1)$ denotes an uniform distribution located between 0 and 1, and $\mathcal{IW}(s, \boldsymbol{\Omega})$ denotes an inverse Wishart distribution with scale matrix $\boldsymbol{\Omega}$ and df s . Multiplying the likelihood functions by the joint prior density, the corresponding joint posterior densities of unknown parameters $\boldsymbol{\theta}$ along with all unobservable quantities (latent data) $(\mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau})$ are obtainable, denoted by $p(\boldsymbol{\theta}, \mathbf{y}^m | \mathbf{y}^o)$, $p(\boldsymbol{\theta}, \mathbf{y}^m, \boldsymbol{\tau} | \mathbf{y}^o)$ and $p(\boldsymbol{\theta}, \mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau} | \mathbf{y}^o)$, respectively.

The full conditional posterior distributions for hidden variables $\{\tau_i, \mathbf{b}_i, \mathbf{y}_i^m\}$ and parameters $\{\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\Sigma}\}$ show explicitly distributional forms. However, none of the three joint posterior densities mentioned previously can lead to the full conditional posteriors of ϕ , γ and ν in standard distributional forms. Fortunately, the M-H algorithm can be implemented to generate posterior samples and thus approximate the target distributions.

3. Computational techniques

3.1. Estimation via the IBF-Gibbs sampler

For posterior inference of the MtLMM with DEC dependence, to alleviate the slow convergence problem of the conventional MCMC methods, we propose a Bayesian sampling procedure, called the IBF-Gibbs sampler, which combines an imputation step (I-step) and a posterior step (P-step). In the I-step, we adopt the IBF sampler to perform the Bayesian data augmentation, which constructs complete data sets $\{\mathbf{y}^{m(h)}, \mathbf{b}^{(h)}, \boldsymbol{\tau}^{(h)}\}_{h=1}^H$. In the P-step, we use Metropolis-within-Gibbs sampler to generate approximately iid samples of a parameter or block of parameters conditioning on the observations and the other parameters from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}^o)$. The P-step is implemented to establish a Markov chain of length H . Stochastically, the simulated Monte Carlo samples $\{\boldsymbol{\theta}^{(h)}\}_{h=h_0+1}^H = \{\boldsymbol{\beta}^{(h)}, \mathbf{D}^{(h)}, \boldsymbol{\Sigma}^{(h)}, \phi^{(h)}, \gamma^{(h)}, \nu^{(h)}\}_{h=h_0+1}^H$ will converge to their associated target distributions after a sufficiently long burn-in period, $h = 1, \dots, h_0$, which is used to remove the effects of the initial sampling values.

The IBF, including the point-wise, function-wise and sampling-wise versions, are used to determine the joint density and thus the marginal densities from the corresponding conditional densities (Tan et al. 2003; Tian and Tan 2003; Tan et al. 2006). Let $\hat{\boldsymbol{\theta}}$ denote the mode or ML estimate of the observed posterior density $p(\boldsymbol{\theta}|\mathbf{y}^o)$, and $\mathcal{S}(\boldsymbol{\theta}, \mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau}|\mathbf{y}^o)$, $\mathcal{S}(\boldsymbol{\theta}|\mathbf{y}^o)$, $\mathcal{S}(\mathbf{y}^m|\mathbf{y}^o)$, $\mathcal{S}(\mathbf{b}|\mathbf{y}^o)$ and $\mathcal{S}(\boldsymbol{\tau}|\mathbf{y}^o)$ denote the joint and conditional supports of $(\boldsymbol{\theta}, \mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau})|\mathbf{y}^o$, $\boldsymbol{\theta}|\mathbf{y}^o$, $\mathbf{y}^m|\mathbf{y}^o$, $\mathbf{b}|\mathbf{y}^o$ and $\boldsymbol{\tau}|\mathbf{y}^o$, respectively. To make posterior inference of the proposed MtLMM, the I-step implements the sampling-wise IBF for simulating samples of $(\mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau})$, which have the fundamental IBF scheme:

$$p(\mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau}|\mathbf{y}^o, \hat{\boldsymbol{\theta}}) = p(\mathbf{y}^m|\mathbf{y}^o, \boldsymbol{\tau}, \hat{\boldsymbol{\theta}})p(\mathbf{b}|\mathbf{y}^o, \boldsymbol{\tau}, \hat{\boldsymbol{\theta}})p(\boldsymbol{\tau}|\mathbf{y}^o, \hat{\boldsymbol{\theta}}), \quad (2)$$

$$p(\mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau}|\mathbf{y}^o) \propto \frac{p(\mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau}|\mathbf{y}^o, \hat{\boldsymbol{\theta}})}{p(\hat{\boldsymbol{\theta}}|\mathbf{y}^o, \mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau})}. \quad (3)$$

Essentially, the implementation of IBF sampler requires two assumptions: (a) the ML estimate $\hat{\boldsymbol{\theta}}$ is already obtained, say, by carrying out the AECM algorithms (Meng and van Dyk 1997); and (b) the joint support is a product space, i.e., $\mathcal{S}(\boldsymbol{\theta}, \mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau}|\mathbf{y}^o) = \mathcal{S}(\boldsymbol{\theta}|\mathbf{y}^o) \times \mathcal{S}(\mathbf{y}^m|\mathbf{y}^o) \times \mathcal{S}(\mathbf{b}|\mathbf{y}^o) \times \mathcal{S}(\boldsymbol{\tau}|\mathbf{y}^o)$.

As pointed out by Tian and Tan (2003), the IBF sampling approach is realized via sampling/importance resampling (SIR) technique. One can refer to Wang and Fan (2012) for details. In summary, we perform

Imputation Step (I-step): Bayesian data augmentation via IBF sampler

I-step 1: Draw K samples such that $\{\mathbf{y}^{m(k)}, \mathbf{b}^{(k)}, \boldsymbol{\tau}^{(k)}\}_{k=1}^K \stackrel{\text{iid}}{\sim} p(\mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau}|\mathbf{y}^o, \hat{\boldsymbol{\theta}})$ via the conditional sampling technique based on (2).

I-step 2: Draw $H < K$ samples such that $\{\mathbf{y}^{m(k_h)}, \mathbf{b}^{(k_h)}, \boldsymbol{\tau}^{(k_h)}\}_{h=1}^H \stackrel{\text{iid}}{\sim} p(\mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau}|\mathbf{y}^o)$ via IBF sampler (3).

Posterior Step (P-step): Posterior samples generation via Metropolis-within-Gibbs technique

P-step 1: Generate $\boldsymbol{\beta}^{(h+1)}$, $\mathbf{D}^{(h+1)}$ and $\boldsymbol{\Sigma}^{(h+1)}$ from their full conditional posteriors via Gibbs sampler, in which $\{\mathbf{y}^m, \mathbf{b}, \boldsymbol{\tau}\}$ are replaced by $\{\mathbf{y}^{m(k_h)}, \mathbf{b}^{(k_h)}, \boldsymbol{\tau}^{(k_h)}\}$.

P-step 2: Generate $(\phi^{(h+1)}, \gamma^{(h+1)})$ and $\nu^{(h+1)}$ via the M-H algorithm.

After implementing **I-step** to simulate a set of H samples of missing (latent) data, we subsequently repeat **P-step** H times to generate a required number of Monte Carlo samples of model parameters. By discarding the first h_0 burn-in iterations to eliminate effects from the initial values and then choosing one sample point per (a certain number of) iteration to reduce the autocorrelation within the iterated samples, the remaining samples can be used to draw posterior inference of interest.

“How to estimate the unobservable random effects in the model?” is also a worthwhile issue. Using the post-convergence samples of $\{\mathbf{b}^{(l)}\}_{l=1}^L$, the Bayesian estimate of random effects is $\hat{\mathbf{b}}_i = L^{-1} \sum_{l=1}^L \mathbf{b}_i^{(l)}$, and thus the fitted vector of responses is $\hat{\mathbf{y}}_i = L^{-1} \sum_{l=1}^L (\mathbf{X}_i \boldsymbol{\beta}^{(l)} + \mathbf{Z}_i \mathbf{b}_i^{(l)})$.

3.2. Multiple imputation for missing values

In multiple imputation, we need to generate H independent draws $\mathbf{y}_i^{\text{m}(1)}, \dots, \mathbf{y}_i^{\text{m}(H)}$ from

$$p(\mathbf{y}_i^{\text{m}} | \mathbf{y}_i^{\text{o}}) = \int p(\mathbf{y}_i^{\text{m}} | \mathbf{y}_i^{\text{o}}, \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}_i^{\text{o}}) d\boldsymbol{\theta}, \quad (4)$$

where the posterior density of all model parameters $p(\boldsymbol{\theta} | \mathbf{y}^{\text{o}})$ is proportional to the product of the observed likelihood function $L(\boldsymbol{\theta} | \mathbf{y}^{\text{o}}) = \prod_{i=1}^n f(\mathbf{y}_i^{\text{o}} | \boldsymbol{\theta})$ and the joint prior density $\pi(\boldsymbol{\theta})$. In practical situations, the posterior predictive distribution in (4) cannot be simulated directly. It is possible to create random draws of \mathbf{y}_i^{m} from $p(\mathbf{y}_i^{\text{m}} | \mathbf{y}_i^{\text{o}})$ by subsequently stochastic generation. In the I-step, we generate a sequence of random variates for missing responses $\{\mathbf{y}^{\text{m}(k_h)}\}_{h=h_0}^H$ whose distribution converges to the desired target. From a prediction viewpoint, the imputed values for \mathbf{y}^{m} can be estimated as $\hat{\mathbf{y}}^{\text{m}} = L^{-1} \sum_{l=1}^L \mathbf{y}^{\text{m}(l)}$, where $\{\mathbf{y}^{\text{m}(l)}\}_{l=1}^L$ is a subset of Monte Carlo samples $\{\mathbf{y}^{\text{m}(k_h)}\}_{h=h_0}^H$.

4. Simulation study

We conduct a simulation study to investigate the speed of convergence of the proposed IBF-Gibbs sampler in comparison with the conventional MCMC procedure under different missing mechanisms. The data of sizes $N = 100$ and 200 are generated from model (1) with bivariate outcomes, where the design sub-matrix \mathbf{X}_{ij} includes an intercept and scheduled visits of time (1 to 7) such that $\mathbf{X}_i = \mathbf{I}_2 \otimes [\mathbf{1}_7 : \mathbf{k}]$, where $\mathbf{1}_7$ is a 7×1 unitary vector and $\mathbf{k} = (1, 2, 3, 4, 5, 6, 7)^{\text{T}}$. For the specification of random effects, we consider (a) the random intercepts (RI) model in which $\mathbf{Z}_i = \mathbf{I}_2 \otimes \mathbf{1}_7$ and (b) the random intercepts plus slopes (RIS) model in which $\mathbf{Z}_i = \mathbf{X}_i$. In this setting, $r = 2$, $p = 4$, $q = 2$ and 4 . The presumed parameters are $\boldsymbol{\beta} = (1, 2, -2, 4)^{\text{T}}$, $\mathbf{D} = [d_{ls}]$ with $d_{ls} = 1$ if $l = s$ and $d_{ls} = 0.25$ otherwise for $l, s = 1, \dots, q$, $\boldsymbol{\Sigma} = \mathbf{I}_2 + \rho(\mathbf{1}_2 \mathbf{1}_2^{\text{T}} - \mathbf{I}_2)$, where $\rho = 0, 0.5$ and 0.9 , and $\nu = 7$. In this experimental study, artificial missing values are generated according to both MCAR and MAR mechanisms. In the MCAR case, missing items are created by deleting at random 5% and 20% of the experimental data, where each timepoint retains at least one observed response. In the MAR case, we consider four nonresponse probabilities based on the observed responses. Let Q_{ji} be the i th quartile of the empirical distribution of the j th outcome \mathbf{y}_j and y_{jk} be the observed value of \mathbf{y}_j at time k , for $j = 1, 2$ and $k = 1, \dots, 7$. The nonresponse probabilities for \mathbf{y}_j are 0.05 if $y_{jk} < Q_{j1}$, 0.10 if $Q_{j1} \leq y_{jk} < Q_{j2}$, 0.15 if $Q_{j2} \leq y_{jk} < Q_{j3}$, and 0.20 if $y_{jk} \geq Q_{j3}$. Each simulated data set with artificial missing responses is fitted through the IBF-Gibbs sampler and the MCMC procedure. A total of 100 replications are implemented for each combination of two random-effect specifications (RI and RIS), two sample sizes, three between-outcome correlations, and three missing settings, namely MCAR 5%, MCAR 20% and MAR.

Table 1: Average iterations of convergence along with the number of non-convergence cases (in parentheses) by implementing IBF-Gibbs and MCMC methods.

Mechanism (Missing rate)	Method	RI		RIS	
		$N = 100$	$N = 200$	$N = 100$	$N = 200$
MCAR (5%)	IBF-Gibbs	1358(0)	1183(0)	1576(0)	1467(0)
	MCMC	1969(5)	1791(0)	4629(79)	4244(84)
MCAR (20%)	IBF-Gibbs	1279(0)	1215(0)	1523(0)	1475(0)
	MCMC	4300(76)	4163(81)	4627(89)	4760(95)
MAR	IBF-Gibbs	1171(0)	1164(0)	1605(0)	1394(0)
	MCMC	2998(35)	2408(20)	4420(95)	4550(84)

For each scenario, the MCMC procedure is run with 5,000 iterations. The IBF-Gibbs sampler is run with $K = 10,000$ to get $H = 5,000$ iid samples of missing (latent) data $\{\mathbf{y}^{\text{m}}, \mathbf{b}, \boldsymbol{\tau}\}$ and then run 5,000 iterations of Metropolis-within-Gibbs sampler. For each method, we run two parallel chains and then evaluate the multivariate potential scale reduction factor (MPSRF; Brooks and Gelman 1998) to monitor the convergence

of Markov chains. Table 1 lists the average number of iterations required to achieve convergence and the frequency of non-convergence cases under the case of $\rho = 0.9$ for saving the space. Obviously, the required iterations for the IBF-Gibbs sampler are much fewer than those for the MCMC method. The IBF-Gibbs sampler achieves the stationary state very quickly, while the MCMC method usually suffers from non-convergence under each considered missingness mechanism. Moreover, the number of converged iterations along with non-convergence frequencies increase with the between-outcome correlation, missing rate, dimension of random effects and sample size. This finding verifies that the IBF-Gibbs sampler can be implemented for the data with missingness and indeed provides a computational efficiency in the sense of a necessity of a small number of “burn-in” and assurance of convergence.

5. Application to a HIV-AIDS study

We apply the proposed technique to the HIV-AIDS data taken by the Seattle research cohort, the Swiss cohort, and University of North Carolina at Chapel Hill (UNCCH) cohort. A bivariate longitudinal data of HIV-1 RNA (count/ml) in seminal and blood plasma of each patient were measured. As considered in Ghosh et al. (2007), we analyze samples collected from baseline to at most 26 week visits with available baseline CD4 values. The data include $N = 149$ patients whose ages varied from 21 to 60 years, where 109 patients were from Seattle cohort, 34 patients were from the Switzerland cohort, and 6 patients were from UNCCH cohort. The total samples were divided into two treatment groups: 106 patients, who were receiving antiviral therapy, belong to group 0; and 43 patients, who were not receiving any antiretroviral therapy or with unknown status of the antiretroviral therapy, belong to group 1. Missingness rates of the responses for group 0 and group 1 are 6.53% and 21.13% for semen RNA and 5.84% and 0.70% for blood RNA, respectively.

Let \mathbf{y}_{i1} and \mathbf{y}_{i2} be the \log_{10} semen RNA and \log_{10} blood RNA repeated measures, respectively, and then write $\mathbf{y}_i = (\mathbf{y}_{i1}^T, \mathbf{y}_{i2}^T)^T$ for patient i . Two design matrices considered for subject-specific random effects include the random intercepts (RI: $\mathbf{Z}_i = \mathbf{I}_2 \otimes \mathbf{1}_{s_i}$) and the random intercepts plus slopes (RIS: $\mathbf{Z}_i = \mathbf{I}_2 \otimes [\mathbf{1}_{s_i} : \mathbf{t}_i]$). To account for the effects of possible autocorrelation, three dependence structures, including uncorrelated (UNC), AR(1), and DEC, are employed for \mathbf{C}_i . To address the effects of different therapies, cohorts, age, and baseline CD4 levels, the design matrix for fixed effects is specified by

$$\mathbf{X}_i = \mathbf{I}_2 \otimes [\mathbf{1}_{s_i} : \mathbf{t}_i : \text{Drug}_i \mathbf{1}_{s_i} : \text{Age}_i \mathbf{1}_{s_i} : \text{CD4}_i : A_{1i} : A_{2i}],$$

where $\mathbf{t}_i = (t_{i1}, \dots, t_{is_i})^T$ with $t_{ik} = \text{week}_{ik}/4$ (months); Drug_i is a treatment indicator (0 = group 0, 1 = group 1); Age_i is the baseline age; CD4_i is the baseline CD4 levels; and (A_{1i}, A_{2i}) are two regional indicators ((1, 0) = the Switzerland cohort, (0, 1) = the Seattle cohort, (0, 0) = the UNCCH cohort).

For Bayesian model choice, we adopt the deviance information criterion (DIC; Spiegelhalter et al. 2002), and a summary statistic of the conditional predictive ordinate (CPO) statistic (Carlin and Louis 2006), i.e., the logarithm of pseudo-marginal likelihood (LPML). Models should be preferred with smaller DIC values. Models with larger LPML are better supported by the data. We find that the performances of MtLMMs are all superior to their normal counterparts in terms of the two criteria. The best fitted MtLMM is ‘RIS-DEC’.

The summary statistics for the posterior inference of the best model, including the posterior means and posterior standard deviations, are summarized in Table 2. For the seminal plasma HIV-RNA, the significant parameters include the intercept β_{10} , baseline CD4 β_{14} and cohort areas (β_{15}, β_{16}) , since their posterior intervals do not contain zero. For the blood plasma HIV-RNA, the significant parameters include the intercept β_{20} , baseline CD4 β_{24} , and cohort area β_{25} , reflecting that the time trend of both responses is inapparent. The posterior median of ν is 4.13 with the 95% posterior interval [2.83, 5.78], indicating a substantial degree of fat tails.

6. Discussion

This paper studies a fully Bayesian approach to the MtLMM with DEC errors, which offers a great deal of flexibility and robustness to handle incomplete and irregularly observed multiple responses with potential outliers. We provide a hybrid sampling procedure, which combines an IBF-sampler imputation step with a Metropolis-within-Gibbs posterior step. Its computational efficiency relative to the conventional MCMC method has been explored by a simulation study under different missingness mechanisms. One possible extension of current work is to investigate some efficient schemes for Bayesian modeling of asymmetric multivariate longitudinal data in the presence of intermittent missing outcomes.

Table 2: Summary statistics for the MtLMM with RIS and DEC errors.

Fixed effects	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}				
mean	5.5720	-0.0259	-0.1790	-0.0059	-0.0013	-1.5917	-1.6141				
sd	0.5672	0.0270	0.1458	0.0101	0.0004	0.4513	0.4060				
Fixed effects	β_{20}	β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	β_{26}				
mean	5.4394	-0.0054	0.0600	-0.0126	-0.0013	-0.8154	-0.1347				
sd	0.3934	0.0200	0.1066	0.0067	0.0002	0.2892	0.2605				
Random effects	d_{11}	d_{21}	d_{22}	d_{31}	d_{32}	d_{33}	d_{41}	d_{42}	d_{43}	d_{44}	
mean	0.2238	-0.0099	0.0037	0.1175	-0.0094	0.1589	-0.0254	0.0072	-0.0229	0.0190	
sd	0.0440	0.0042	0.0007	0.0297	0.0034	0.0303	0.0093	0.0015	0.0077	0.0036	
Within-subject errors	σ_{11}	σ_{21}	σ_{22}	ϕ	γ	ν					
mean	0.6626	0.0799	0.2215	0.6096	0.2422	4.1336					
sd	0.0778	0.0258	0.0271	0.0470	0.0840	0.7613					

References

- Brooks, S.P., Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Statist.* 7, 434–455.
- Carlin, B.P., Louis, T.A. (2006). *Bayesian Methods for Data Analysis* (Texts in Statistical Science). Chapman & Hall/CRC, New York.
- Dey, D.K., Chen, M.H., Chang, H. (1997). Bayesian approach for the nonlinear random effects models. *Biometrics* 53, 1239–1252.
- Geman, S., Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 721–741.
- Ghosh, P., Branco, M.D., Chakraborty, H. (2007). Bivariate random effect model using skew-normal distribution with application to HIV-RNA. *Statist. Med.* 26, 1255–1267.
- Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Meng, X.L., van Dyk, D. (1997). The EM algorithm – an old folk-song sung to a fast new tune. *J. R. Stat. Soc. Ser. B* 59, 511–567.
- Muñoz, A., Carey, V., Schouten, J.P., Segal, M., Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 48, 733–742.
- Pinheiro, J.C., Liu, C.H., Wu, Y.N. (2001). Efficient algorithms for robust estimation in linear mixed-effects model using the multivariate t distribution. *J. Comput. Graph. Statist.* 10, 249–276.
- Rubin, D.B. (1976). Inference and missing data. *Biometrika* 63, 581–592.
- Shah, A., Laird, N., Schoenfeld, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *J. Amer. Statist. Assoc.* 92, 775–779.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde A.V.D. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* 64, 583–639.
- Tan, M., Tian, G.L., Ng, K.W. (2003). A noniterative sampling method for computing posteriors in the structure of EM-type algorithms. *Statist. Sinica* 13, 625–639.
- Tan, M., Tian, G.L., Ng, K.W. (2006). Hierarchical models for repeated binary data using the IBF sampler. *Comput. Statist. Data Anal.* 50, 1272–1286.
- Tian, G.L., Tan, M. (2003). Exact statistical solutions using the inverse Bayes formulae. *Stat. Probabil. Lett.* 62, 305–315.
- Wang, W.L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biom. J.* 55, 554–571.
- Wang, W. L., Fan, T. H. (2010). ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. *Computational Comput. Statist. Data Anal.* 54, 1328–1341.
- Wang, W. L., Fan, T. H. (2011). Estimation in multivariate t linear mixed models for multiple longitudinal data. *Statistica Sinica*, 21, 1857–1869.
- Wang, W.L., Fan, T.H. (2012). Bayesian analysis of multivariate t linear mixed models using a combination of IBF and Gibbs samplers. *J. Multivar. Anal.* 105, 300–310