



Published in final edited form as:

*Aust N Z J Stat.* 2010 September ; 52(3): 275–288. doi:10.1111/j.1467-842X.2010.00581.x.

## A Semiparametric Bayesian Approach to Multivariate Longitudinal Data

Pulak Ghosh\* and Timothy Hanson†

Pulak Ghosh: pulakghosh@gmail.com

\*Department of Quantitative Methods & Information Systems, Indian Institute of Management, Bangalore, India

### Summary

We extend the standard multivariate mixed model by incorporating a smooth time effect and relaxing distributional assumptions. We propose a semiparametric Bayesian approach to multivariate longitudinal data using a mixture of Polya trees prior distribution. Usually, the distribution of random effects in a longitudinal data model is assumed to be Gaussian. However, the normality assumption may be suspect, particularly if the estimated longitudinal trajectory parameters exhibit multimodality and skewness. In this paper we propose a mixture of Polya trees prior density to address the limitations of the parametric random effects distribution. We illustrate the methodology by analyzing data from a recent HIV-AIDS study.

### Keywords

Conditional predictive ordinate; Longitudinal data; Mixture of Polya trees; Penalized spline

## 1 Introduction

Longitudinal data analysis, in which repeated measurements are taken on subjects at various time points, plays an important role in applied statistics, especially in biomedical research involving clinical trials. Rather than considering a single outcome, it is common nowadays to measure multiple outcomes on a subject to characterize an effect of interest, thus resulting in multivariate longitudinal data. Although statistical methods and estimation techniques are well developed for univariate longitudinal data (Grizzle & Allen, 1969; Khatri, 1966; Pothoff & Roy, 1964; Laird & Ware, 1982; Kleinman & Ibrahim, 1998; Shah et al., 1997; Zhang et al., 1998; Zeger & Diggle, 1994; Diggle et al., 2002), relatively little attention has been concentrated on longitudinal data of a multivariate response. Reinsel (1984) considered a multivariate random intercept covariance structure for balanced data. However, balanced data are quite unlikely in a longitudinal setting. Shah et al. (1997) developed a multivariate longitudinal model for unbalanced data with inference carried out via the EM algorithm. They assumed multivariate longitudinal data with Gaussian random effects and allowed for missing observations. They considered two responses (log CD4 count, log CD8 count) in their example; each was modelled linearly, and so the subject-specific random effect was 4-dimensional. Roy & Lin (2002) also developed a multivariate model for longitudinal data that allows for missing observations, assuming a latent variable approach. Recently, Fieuws & Verbeke (2006) developed a pairwise approach in which all possible bivariate models are fitted.

†Author to whom correspondence should be addressed. Department of Statistics, University of South Carolina, Columbia, SC 29208 USA. hansonst@stat.sc.edu.

In this paper we take a Bayesian approach to multivariate longitudinal data. Our method generalizes existing methods on several issues. Most of the mixed models above provide a flexible likelihood framework to model continuous correlated outcomes parametrically. However, in practice the parametric assumption on the covariate effects may not always be appropriate; in our example CD4 and CD8 counts vary in a complicated manner over time (see Figure 1). To model longitudinal data of this kind, Zhang et al. (1998) and Zeger & Diggle (1994) considered semiparametric mixed models. In this paper we extend their semiparametric approach to the multivariate setting. We specify fixed covariate effects, but develop a smooth function of time to model the overall mean population effect, and account for the within-subject correlations using random effects. The unknown function is modelled using a linear combination of B-spline basis terms (de Boor, 1978), where the coefficients are modelled using a pairwise shrinkage prior distribution. In multivariate longitudinal data, random effects are usually assumed to have a normal distribution, and it may well happen that this distributions does not correctly fit the data at hand. This is particularly true if the data are multimodal, skewed, have outliers, or consist of diverse populations. Specifically, in our application we consider data on two widely used markers in HIV-AIDS studies: CD4 and CD8 counts. These data are collected from a new AIDS clinical trial (ACTG 398). It is well known that these markers usually have a high amount of between-subject variability (Figure 1). Thus, the normal distribution may not effectively model this extra heterogeneity in the samples. Also, observations from each patient's profile may not have the same distribution. For example, in a clinical study some patients may drop out of the study before the study ends. The longitudinal observations of the patients who left the study may not have the same distribution as those who completed the study (Brown & Ibrahim, 2003). Based on these observations, it is desirable to take the random effects from a sufficiently large class to capture such possibilities. Thus, we consider a model for multivariate random effects which generalizes the parametric normal distributional assumption to include a class of distributions that allows for features such as skewness and multimodality as well as generalizes the parametric models.

One can use a simple extension of the existing model that employs a finite mixture of normal distributions for random effects. However, determining the number of mixands can be problematic; identifiability issues also arise. In contrast, continuous mixtures alleviate these problems, with nonparametric specification of the mixing distribution being most attractive. Thus, in this paper we assume a multivariate mixture of Polya trees (MPT) prior distribution for  $q$ -dimensional random effects. Univariate Polya tree and MPT models have been recently developed (Lavine, 1992, 1994; Walker & Mallick, 1997, 1999; Berger & Guglielmi, 2001; Hanson & Johnson, 2002; Hanson, 2006), but relatively little has been done for the multivariate case. We model the random effects distribution for our multivariate longitudinal data as a mixture of absolutely continuous finite Polya trees in  $\mathbb{R}^q$  centred at the family of mean-zero  $q$ -variate Gaussian distributions. This new framework allows the incorporation of the existing prior information based on the parametric model. This robust generalization has the potential to capture departures from a normal distribution, while having a good performance if the actual distribution is normal. One of the major advantages of using Polya tree prior distributions is that, in contrast to Dirichlet process prior distributions, they can place probability one on the set of absolutely continuous distributions. We perform the analysis within a Bayesian framework, thus providing a coherent means of inference and allowing for easy computation by standard Markov chain Monte Carlo (MCMC) techniques (Gelfand & Smith, 1990).

The paper is organized as follows. In the next section we develop a model for multivariate longitudinal data. In Section 3 we develop the Bayesian analysis of the parametric model. We relax the normality assumption with multivariate MPT and Polya tree prior distributions

in Section 4. In Section 5 we describe and analyze a clinical trial data set, ACTG 398; Section 6 draws conclusions.

## 2 Multivariate mixed model

Here we introduce our notation for the form of the multivariate longitudinal model. Let the sample in a longitudinal study consists of  $K$  characteristics and  $m$  subjects, with the  $i$ th subject having  $n_i$  measurements over time. Let  $y_{ijk}$  denote the measurement of the  $i$ th subject at time  $t_{ij}$  for characteristic  $k$ ,  $i = 1, 2, \dots, m$ ;  $k = 1, 2, \dots, K$ ;  $j = 1, 2, \dots, n_i$ . Then the semiparametric mixed model for outcome  $y_{ijk}$  is given by

$$y_{ijk} = \mathbf{x}_{ijk}^\top \boldsymbol{\beta}_k + f_k(t_{ij}) + \mathbf{z}_{ijk}^\top \boldsymbol{\gamma}_{ik} + e_{ijk}, \quad (1)$$

where  $\boldsymbol{\beta}_k$  is the  $p_k \times 1$  vector of regression coefficients associated with covariates  $\mathbf{x}_{ijk}$  ( $p_k \times 1$ ), and  $f_k(t)$  is a twice-differentiable smooth function of time for characteristic  $k$ . Thus, we allow each characteristic to have a unique nonparametric population mean function in time. The  $\boldsymbol{\gamma}_{ik}$  are independent  $q_k \times 1$  vectors of random effects associated with covariates  $\mathbf{z}_{ijk}$  ( $q_k \times 1$ ), and the  $e_{ijk}$  are independent measurement errors.

Let  $\mathbf{y}_{ik} = (y_{i1k}, y_{i2k}, \dots, y_{in_i k})^\top$ . Then we define the following notation:

$$\mathbf{y}_{ik} = \mathbf{X}_{ik} \boldsymbol{\beta}_k + \mathbf{N}_{ik} \mathbf{f}_k + \mathbf{Z}_{ik} \boldsymbol{\gamma}_{ik} + \mathbf{e}_{ik}. \quad (2)$$

We write  $\mathbf{t}^0$  for the vector of the  $r$  ordered distinct time points of  $\{t_{ij}\}$ .  $\mathbf{N}_{ik}$  is the incidence matrix for the  $i$ th subject and  $k$ th characteristic connecting  $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})$  to  $\mathbf{t}^0$  such that the  $(j, h)$ th element of  $\mathbf{N}_{ik}$  is 1 if  $t_{ij} = t_h^0$  and 0 otherwise for  $j = 1, 2, \dots, n_i$ ,  $h = 1, 2, \dots, r$ . Finally,  $\mathbf{f}_k = (f_k(t_1^0), \dots, f_k(t_r^0))^\top$ , the subject-specific design matrix  $\mathbf{X}_{ik}$  contains the  $n_i$  rows  $\mathbf{x}_{ijk}^\top$ , and  $\mathbf{Z}_{ik}$  contains the  $n_i$  rows  $\mathbf{z}_{ijk}^\top$ .

We assume that  $\boldsymbol{\gamma}_i = (\boldsymbol{\gamma}_{i1}^\top, \boldsymbol{\gamma}_{i2}^\top, \dots, \boldsymbol{\gamma}_{iK}^\top)^\top$ , where  $\boldsymbol{\gamma}_{ij} \in \mathbb{R}^{q_j}$ ,  $q_j = \sum_{j=1}^K q_j$ . For initial model development,  $\boldsymbol{\gamma}_i$  is assumed to have a  $N_q(\mathbf{0}, \mathbf{D})$  distribution, where  $\mathbf{D}$  is a symmetric positive definite matrix. The normality assumption for  $\boldsymbol{\gamma}_i$  will be relaxed in Section 4, where we will assume a nonparametric distribution for  $\boldsymbol{\gamma}_i$ . The measurement error  $\mathbf{e}_{ik}$  is assumed to be distributed as  $N(0, \mathbf{R}_{ik})$ , where  $\mathbf{R}_{ik} = \sigma_k^2 \mathbf{I}_{n_i}$ . We also assume independence between  $\boldsymbol{\gamma}_i$  and  $\mathbf{e}_{ik}$ .

## 3 Smoothing spline and priors

Let  $\theta = (\boldsymbol{\beta}_{1:K}, \sigma_{1:K}^2, \mathbf{f}_{1:K}, \mathbf{D})$  be the set of parameters for model (2), where we generically write  $\mathbf{v}_{1:n} = \{\mathbf{v}_i\}_{i=1}^n$  for any matrices  $\mathbf{v}_1, \dots, \mathbf{v}_n$ . We assume conditionally conjugate, independent prior distributions that lead to simpler updating schemes in the Markov chain sampling methodology. In particular,

$$\boldsymbol{\beta}_k \sim N_p(\boldsymbol{\mu}_k, \sum_k^{-1}), \mathbf{D}^{-1} \sim \text{Wishart}(\eta_0, \sum_\eta), \sigma_k^{-2} \sim \text{gamma}(v_{0k}/2, \delta_{0k}/2).$$

We model each of the  $K$  flexible nonparametric time effects in (2) as a linear decomposition of basis terms. Suppose that  $\mathbf{f}_k = (f_k(t_1^0), \dots, f_k(t_r^0))^T$ . Let us suppress the notation  $k$  for the moment. Then  $f$  is modelled as

$$f(t) = \sum_{l=1}^L \varphi_l b_l(t).$$

Here  $\{b_1, \dots, b_L\}$  are  $B$ -spline basis functions of degree 2 with equidistant knots, and  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_L)^T$  are the associated coefficients. To ensure that the function is flexible, yet smooth, we follow the methodology of Lang & Brezger (2004) and place a first order random walk prior distribution on the coefficients, so that

$$p(\varphi_1) \propto 1, \quad \varphi_l = \varphi_{l-1} + u_l, \quad l=2, \dots, L, \quad u_2, \dots, u_L | \tau \stackrel{iid}{\sim} N(0, \tau^2).$$

Here the random walk variance  $\tau^2$  can be interpreted as a smoothing parameter. Thus the resulting conditional prior density for  $\boldsymbol{\phi}$  satisfies

$$p(\boldsymbol{\phi} | \tau) \propto \exp\left(-\frac{1}{2\tau^2} \boldsymbol{\phi}^T \boldsymbol{\Omega} \boldsymbol{\phi}\right),$$

with a banded penalty matrix

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}.$$

Now let us generalize the above prior density for  $f_k(t)$ . Write the vector of function evaluations  $\mathbf{f}_k = (f_k(t_1^0), \dots, f_k(t_r^0))^T$  as the product  $\mathbf{f}_k = \mathbf{B}_k \boldsymbol{\phi}_k$ , where  $\mathbf{B}_k$  is an  $L \times r$  design matrix. Then the prior density on the penalized spline coefficients can be cast into a general form as follows:

$$p(\boldsymbol{\phi}_k | \tau_k) \propto \exp\left(-\frac{1}{2\tau_k^2} \boldsymbol{\phi}_k^T \boldsymbol{\Omega} \boldsymbol{\phi}_k\right).$$

For the variance parameter we assume an inverse gamma distribution, i.e.,  $\tau_k^2 \sim \text{gamma}^{-1}(a_k, b_k)$ . This prior structure is a Bayesian interpretation of the work of Eilers & Marx (1996), who constructed a difference penalty from the random walk prior and took a penalized maximum likelihood approach.

Because the marginal and joint posterior distributions are analytically intractable, an MCMC sampling scheme is used to produce iterates from the joint posterior distribution, from which the full spectrum of inference can be constructed. Full details of the evaluation of conditional posterior distributions are provided in the Appendix.

#### 4 Nonparametric random effects distribution

We consider a MPT prior distribution for the distribution  $G$  on the  $\gamma_i$ s in order to relax distributional assumptions. Many authors have investigated nonparametric distributions for random effects. Kleinman & Ibrahim (1998) considered a univariate linear mixed effects model where the random effects  $\gamma_1, \dots, \gamma_m \in \mathbb{R}^q$  arise from a probability measure  $G$  assigned a Dirichlet process prior distribution. Zhang & Davidian (2001) considered modelling the random effects distribution in a standard linear mixed model as the square of a polynomial of order  $K$  multiplied by a bivariate Gaussian density. Song, Davidian & Tsiatis (2002) extended this approach to the joint modelling of survival and longitudinal data, and modelled a bivariate random effects distribution for CD4 counts in this manner. More recently, Ghidry, Lesaffre & Eilers (2004) considered modelling a bivariate random effects distribution using a mixture of bivariate Gaussian densities with fixed locations.

Lavine (1992) introduced MPT models and noted that random partitions smooth out partitioning effects seen in simple Polya tree models. Berger & Guglielmi (2001) considered a univariate MPT model with a fixed partition to test whether data arise from a parametric family versus a nonparametric alternative. The marginal model is fitted using an importance sampling distribution derived from the parametric family. Hanson & Johnson (2002) considered a median regression model with a univariate MPT prior distribution on the error density centred at a scaled family of Gaussian distributions.

Paddock (1999) and Paddock et al. (2003) considered multivariate Polya trees centred at a multivariate uniform distribution. They randomly jittered the partition, smoothing out the partition effects in a manner similar to an MPT. Hanson (2006) proposed a class of multivariate MPT prior distributions that generalized the multivariate Gaussian family. In either approach, the Polya tree is fully specified up to a finite level  $J$  and suffers from an explosion of parameters when the dimension  $q$  climbs past 2 or 3. For a  $q$ -dimensional MPT prior distribution, the number of  $2^q$ -dimensional Dirichlet vectors needed to define the

model is  $\sum_{j=0}^{J-1} 2^{jq} = (2^{qJ} - 1)/(2^q - 1)$ . For example, for a finite 4-dimensional tree with  $J = 5$  levels, 69 905 16-dimensional Dirichlet vectors are required, along with the 10 parameters in  $\mathbf{D}$ . For this reason, we instead develop associated marginal models which have only the parameters in the covariance matrix  $\mathbf{D}$ .

A random probability measure  $G$  given a Polya tree prior distribution is defined by a class of partitions  $\{\Pi^j\}_{j=1}^J$  and a class of Dirichlet distribution parameters  $\mathcal{A}$ . We consider one method to extend the ‘‘canonical’’ partitions of described in Lavine (1992) to multiple dimensions.

The space  $\mathbb{R}^q$  is partitioned at level  $j$  by sets  $B_0(j; k_1, \dots, k_q)$  where  $k_l = 1, \dots, 2^j$ ,  $l = 1, \dots, q$ . Define

$$B_0(j; k_1, \dots, k_q) = (\Phi^{-1}((k_1 - 1)2^{-j}), \Phi^{-1}(k_1 2^{-j})) \times \dots \times (\Phi^{-1}((k_q - 1)2^{-j}), \Phi^{-1}(k_q 2^{-j})). \quad (3)$$

Let  $\Pi_0^j$  denote the class of partitions at level  $j$ ,  $\Pi_0^j = \{B_0(j; k_1, \dots, k_q)\}$ . The transformed partitions  $\Pi_D^j$  are obtained from (3) through

$$B_D(j; k_1, \dots, k_q) = \{\mathbf{D}^{1/2} \mathbf{y} : \mathbf{y} \in B_0(j; k_1, \dots, k_q)\}. \tag{4}$$

Clearly, if  $\mathbf{y} \sim N_q(\mathbf{0}, \mathbf{D})$ , then  $\Pr(\mathbf{y} \in B_D(j; k_1, \dots, k_q)) = 2^{-qj}$ .

Let  $S_j = \{1, \dots, 2^{j-1}\}^q$ . Define vectors of conditional probabilities  $\mathbf{C}_{j;\mathbf{s}} = (C(j; k_1, \dots, k_q) : k_l = 2_{s_l} - 1, 2_{s_l}; l = 1, \dots, q)$ , where  $\mathbf{s} = (s_1, \dots, s_q)^T \in S_j$ , by

$$\mathbf{C}_{j;\mathbf{s}} | \alpha \sim \text{Dirichlet}(\alpha \rho(j) \mathbf{1}_{2^q}). \tag{5}$$

For example,  $C_{2;(1,2)} = (C(2; 1, 3), C(2; 1, 4), C(2; 2, 3), C(2; 2, 4))^T$ . These are the conditional probabilities of, respectively, sets  $B_D(2; 1, 3)$ ,  $B_D(2; 1, 4)$ ,  $B_D(2; 2, 3)$ ,  $B_D(2; 2, 4)$  given  $B_D(1; 1, 2)$ . The  $G$ -measure of a set in  $\Pi_D^j$  given  $\mathcal{C} = \{C_{j;\mathbf{s}} : j = 1, \dots, J; \mathbf{s} \in S_j\}$  is defined to be

$$G\{B_D(j; k_1, \dots, k_q)\} = \prod_{l=1}^j C(l; \lceil k_1 2^{l-j} \rceil, \dots, \lceil k_q 2^{l-j} \rceil), \tag{6}$$

where  $\lceil \cdot \rceil$  is the ceiling function. On sets at level  $J$ , we assume that  $G$  follows a  $N_q(\mathbf{0}, \mathbf{D})$  distribution given  $\mathbf{D}$ . That is, for  $A \subset B_D(J, k_1, \dots, k_q)$ ,  $G(A) | G(B_D(J, k_1, \dots, k_q)) = 2^{qJ} \Phi_D(A) G\{B_D(J, k_1, \dots, k_q)\}$ .

Denote the prior Polya tree (PT) model on the measure  $G$ , given by (4)-(6), as  $G \sim \text{PT}_q(\alpha, \rho, \Phi_D)$  and further assume the hyperprior  $\mathbf{D}^{-1} \sim \text{Wishart}(\eta_0, \Sigma_\eta)$ . The random effects model is then

$$\gamma_1, \dots, \gamma_m | G \stackrel{iid}{\sim} G, \quad G | \mathbf{D} \sim \text{PT}_q(\alpha, \rho, \Phi_D), \quad \mathbf{D}^{-1} \sim \text{Wishart}(\eta_0, \Sigma_\eta). \tag{7}$$

It is easy to verify, given  $\mathbf{D}$ , for  $\mathbf{y} | G \sim G$ ,  $G \sim \text{PT}_q(c, \rho, \Phi_D)$ , that  $E(\mathbf{y}) = \mathbf{0}$ ,  $\text{cov}(\mathbf{y}) = \mathbf{D}$ , and  $\Pr(\mathbf{y} \in A) = \int_A \phi_D(\mathbf{x}) d\mathbf{x}$ , where  $\phi_D(\cdot)$  is the probability density function (pdf) of a  $N_q(\mathbf{0}, \mathbf{D})$  vector. The precision  $\alpha$  determines how ‘‘close’’ the random probability measure  $G$  is to the centering  $\Phi_D$ , with larger values giving  $G \approx \Phi_D$  (Hanson, 2006).

Given a vector  $\mathbf{y} \in \mathbb{R}^q$ , we need to know which set  $\mathbf{k} = (k_1, \dots, k_q)^T$  in  $\Pi_D^j$  at level  $j$ ,  $B_D(j; k_1, \dots, k_q)$ , contains it. This is given by

$$\mathbf{k}_D(j, \mathbf{y}) = \begin{bmatrix} \text{Int}\{2^j \Phi(x_1) + 1\} \\ \vdots \\ \text{Int}\{2^j \Phi(x_q) + 1\} \end{bmatrix}, \quad \text{where } \mathbf{x} = \mathbf{D}^{-1/2} \mathbf{y}.$$

The  $k$ th index of  $k_{\mathbf{D}}(j-1, w)$  is recursively given by the ceiling of one half of the  $k$ th index of  $k_{\mathbf{D}}(j, w)$ .

Define  $\mathbf{n}_{\mathbf{D}}(j, \mathbf{s}; \boldsymbol{\gamma}_{1:j})$  to be the  $2^q$ -dimensional vector  $\{\#\{\gamma_i\}_{i=1}^l \text{ in } B_{\mathbf{D}}(j; k_1, \dots, k_q)\}$  where each  $k_l, l = 1, \dots, q$ , ranges over the two values  $2_{s_l} - 1$  and  $2_{s_l}$ . Then the full conditional distribution for  $\mathbf{C}_{j,\mathbf{s}}$  is given by

$$\mathbf{C}_{j,\mathbf{s}}|\alpha, \boldsymbol{\gamma}_{1:m}, \mathbf{D} \sim \text{Dirichlet}(\alpha\rho(j)\mathbf{1}_{2^q} + \mathbf{n}_{\mathbf{D}}(j, \mathbf{s}; \boldsymbol{\gamma}_{1:m})).$$

Let  $\mathbf{y} \in \mathbb{R}^q$ . Given  $\mathcal{C}$  and  $\mathbf{D}$ , the pdf of  $G$  is given by

$$g(\mathbf{y}|\mathcal{C}, \mathbf{D}) = 2^{Jq} p_{\mathcal{C}}\{\mathbf{k}_{\mathbf{D}}(J, \mathbf{y})\} \varphi_{\mathbf{D}}(\mathbf{y}),$$

where  $p_{\mathcal{C}}(k_1, \dots, k_q) = G\{B_{\mathbf{D}}(J; k_1, \dots, k_q)\}$ , which is obtained from (6).

Given  $\boldsymbol{\gamma}_{1:j}$  and  $\mathbf{D}$ , the expected value of the pdf of  $G$  is given by

$$g(\mathbf{y}|\boldsymbol{\gamma}_{1:j}, \mathbf{D}) = \varphi_{\mathbf{D}}(\mathbf{y}) \prod_{j=1}^J \frac{2^q \alpha \rho(j) + 2^q \mathbf{n}_{\mathbf{D}}(j, \mathbf{k}_{\mathbf{D}}(j, \mathbf{y}); \boldsymbol{\gamma}_{1:j})}{2^q \alpha \rho(j) + \mathbf{n}_{\mathbf{D}}(j-1, \mathbf{k}_{\mathbf{D}}(j-1, \mathbf{y}); \boldsymbol{\gamma}_{1:j})}. \tag{8}$$

Through (8), the marginal pdf of  $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_m$  is thus given by

$$g(\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_m|\mathbf{D}) = \left[ \prod_{i=1}^m \varphi_{\mathbf{D}}(\boldsymbol{\gamma}_i) \right] \left[ \prod_{i=2}^m \prod_{j=1}^J \frac{2^q \alpha \rho(j) + 2^q \mathbf{n}_{\mathbf{D}}(j, \mathbf{k}_{\mathbf{D}}(j, \boldsymbol{\gamma}_i); \boldsymbol{\gamma}_{1:i-1})}{2^q \alpha \rho(j) + \mathbf{n}_{\mathbf{D}}(j-1, \mathbf{k}_{\mathbf{D}}(j-1, \boldsymbol{\gamma}_i); \boldsymbol{\gamma}_{1:i-1})} \right] \\ = \left[ \prod_{i=1}^m \varphi_{\mathbf{D}}(\boldsymbol{\gamma}_i) \right] q_{\mathbf{D}}(\boldsymbol{\gamma}_{1:m}; \alpha), \text{ say.}$$

When the vectors  $\{\boldsymbol{\gamma}_i\}_{i=1}^m$  are evenly distributed among sets  $B_{\mathbf{D}}(j|k_1, \dots, k_q)$  on each level  $j$ ,  $q_{\mathbf{D}}(\boldsymbol{\gamma}_{1:m}; \alpha) = 1$  and the parametric density is obtained. We exploit this fact in the construction of Metropolis-Hastings proposal distributions for sampling  $\mathbf{D}$  given the remaining parameters as well as sampling the  $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_m$ .

Assuming the model (7), the full conditional distribution for  $\mathbf{D}$  given  $\mathcal{C}$  and  $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_m$  is

$$p(\mathbf{D}^{-1}|\mathcal{C}, \boldsymbol{\gamma}_{1:m}) \propto |\mathbf{D}^{-1}|^{-(m+\eta_0-q-1)/2} \exp \left\{ -0.5 \text{tr} \mathbf{D}^{-1} \left( \sum_{i=1}^m \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T + \sum_{\eta}^{-1} \right) \right\} q_{\mathbf{D}}(\boldsymbol{\gamma}_{1:m}; \alpha).$$

By experimenting with various types of Metropolis-Hastings proposal distributions, we have found a Metropolis-Hastings step centred at the current value  $\mathbf{D}^{-1}$  updates  $\mathbf{D}^{-1}$  fairly well when the estimated random effects are not highly “non-normal.” Updating  $\mathbf{D}^{-1}$  proceeds by sampling  $\mathbf{D}^{*-1} \sim \text{Wishart}(df, \mathbf{D}^{-1}/df)$  and accepting the proposal  $\mathbf{D}^{*-1}$  with probability

$$\rho = \min \left\{ 1, \frac{q_{\mathbf{D}^*}(\gamma_{1:m}; \alpha) \exp\{-0.5 \operatorname{tr}(\mathbf{D}^{*-1}(\sum_{i=1}^m \gamma_i \gamma_i^\top + \sum_{\eta}^{-1}) + df \mathbf{D}^*)\} |\mathbf{D}^*|^{\frac{n_0+m}{2} - df}}{q_{\mathbf{D}}(\gamma_{1:m}; \alpha) \exp\{-0.5 \operatorname{tr}(\mathbf{D}^{-1}(\sum_{i=1}^m \gamma_i \gamma_i^\top + \sum_{\eta}^{-1}) + df \mathbf{D})\} |\mathbf{D}|^{\frac{n_0+m}{2} - df}} \right\}.$$

The random effects  $\gamma_1, \dots, \gamma_m$  are also updated via Metropolis-Hastings steps by generating  $\gamma_i^* \sim N_q(\mathbf{V}_i^{-1}[\mathbf{Z}_i^\top \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{N}_i \mathbf{f})], \mathbf{V}_i^{-1})$ , where  $\mathbf{V}_i = \mathbf{Z}_i^\top \mathbf{R}_i^{-1} \mathbf{Z}_i + \mathbf{D}^{-1}$ ,  $\mathbf{R}_i = \operatorname{diag}(\mathbf{R}_{i1}, \dots, \mathbf{R}_{iK})$  and  $\mathbf{Z}_i = \operatorname{diag}(\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{iK})$  (see Appendix), and accepting the proposal  $\gamma_i^*$  with probability

$$\rho = \min \left\{ 1, \frac{q_{\mathbf{D}}(\gamma_{1:m}^*; \alpha)}{q_{\mathbf{D}^*}(\gamma_{1:m}; \alpha)} \right\},$$

where  $\gamma_{1:m}^*$  is the set  $\gamma_{1:m}$  with  $\gamma_i^*$  replacing  $\gamma_i$ . When a prior density  $p(\alpha)$  is placed on  $\alpha$ , the full conditional distribution is given by  $p(\alpha; \gamma_{1:m}, \mathbf{D}) \propto p(\alpha) q_{\mathbf{D}}(\gamma_{1:m}; \alpha)$ . Updates proceed via Metropolis-Hastings steps as before.

To our knowledge, the random effects model we present is the only direct generalization of the Gaussian model considered in the literature. Under Dirichlet process mixing, a Gaussian model is obtained in the limit as the precision goes to zero, but this is at the boundary of the parameter space. A formal Bayes factor can be computed (see, for example, Basu & Chib, 2003), but the models are not formally nested in the sense that one can test  $H_0 : \theta = \theta_0$  where  $\theta_0$  is an interior point of  $\Theta$ . Thus, to compare the different models, we follow Geisser & Eddy (1979), who introduced a predictive criterion termed the *log pseudo marginal likelihood* (LPML). The conditional predictive ordinate (CPO) for the  $(ijk)$ th observation is

$$\text{CPO}_{ijk} = p(y_{ijk} | y_{(-ijk)}) = E\{f(y_{ijk} | \theta, \gamma_{1:m}) | y_{(-ijk)}\},$$

where  $y_{(-ijk)}$  denotes the rest of the data after deleting the  $(ijk)$ th observation. The expectation is taken with respect to the posterior distribution of the model parameter given the cross-validated data  $y_{(-ijk)}$ . Gelfand & Dey (1994) suggested a particularly simple method to obtain estimates of CPO statistics from MCMC output, which we adapt to the situation at hand.

A useful aggregate measure of predictive model performance is the sum of the log CPO statistics:  $\text{LPML} = \log \prod_{i=1}^m \prod_{j=1}^{n_i} \prod_{k=1}^K \text{CPO}_{ijk}$ . The model with the larger LPML value is the better-fitting model from a predictive standpoint. One can compute a pseudo-Bayes factor for comparing model 2 to model 1, defined as  $\text{PBF}_{21} = \exp(\text{LPML}_2 - \text{LPML}_1)$ . Kadane & Lazar (2004) noted that predictive measures, such as the pseudo-Bayes factor, attempt to find “which models give the best predictions of future observations generated from the same process as the original data”.

## 5 Data Analysis

We illustrate our method using data from a large clinical trial on HIV-AIDS, ACTG 398. All regimens included the following ART drugs: abacavir, a nucleoside reverse transcriptase inhibitor (NRTI); adefovir dipivoxil, another NRTI; efavirenz, a non-NRTI (NNRTI); and amprenavir, a protease inhibitor (PI). In addition, three of the four arms included a second PI: saquinavir, indinavir, or nelfinavir, respectively, whereas the fourth arm received a



matched placebo. Thus, three of the treatment arms were similar in that they contained a second PI drug, while the fourth arm had a single PI drug and a matched PI placebo. The results of the primary analysis were reported by Hammer et al. (2002). Based on these results, we combined the three dual PI arms together, versus the single PI arm. Thus, we take two treatment arms in our analysis. The primary objective of the study was to compare the proportion of subjects who had virologic failure after 24 weeks on study between the double-PI arms and the single-PI arm. The subjects in the trial were either NNRTI naive, i.e., they had not previously received drugs from the NNRTI class, or NNRTI experienced, i.e., they had previously received NNRTIs as part of their treatment. NNRTI naive subjects have a better virologic response in the trial. We are interested in analyzing the markers, CD4 and CD8 cell counts of the subjects on the study. These markers were measured at  $r = 47$  distinct time points starting from the 0th week to a maximum 56th week. However, not all subjects have responses at each time point. We discarded patients with two or fewer ( $n_i \leq 2$ ) observations and thus our analysis consisted of  $m = 454$  patients with a total of

$\sum_{i=1}^{454} n_i = 3172$  observations. The covariates we included in this analysis are: treatment (second PI versus placebo), NNRTI experience, and baseline  $\log_{10}$  viral load. We assumed 10 equally-spaced knots between  $t_1^0 = 0$  and  $t_{47}^0 = 56$ , so the number of quadratic B-spline basis functions is 11.

We fitted the following bivariate model:

$$y_{ijk} = \beta_{0k} + \beta_{1k} x_{ijk1} + \beta_{2k} x_{ijk2} + \beta_{3k} x_{ijk3} + f_k(t_{ij}) + \gamma_{i0k} + \gamma_{i1k} t_{ij} + e_{ijk}, \quad (9)$$

for  $i = 1, \dots, 454$ ;  $j = 1, \dots, n_i$  and  $k = 1, 2$ . Again,  $y_{ijk}$  is the response of the  $k$ th marker for the  $i$ th subject at the  $j$ th time point  $t_{ij}$  (in weeks),  $k = 1$  denotes the  $\log(\text{CD4})$  count and  $k = 2$  denotes the  $\log(\text{CD8})$  count. The covariates are fixed for each subject  $i$  across the  $n_i$  time points:  $x_{ij11} = x_{ij21} = \log_{10}$  baseline HIV RNA,  $x_{ij12} = x_{ij22} =$  treatment arm (1=double-PI and 0=single-PI placebo), and  $x_{ij13} = x_{ij23} =$  NNRT experience (1=experienced and 0=naive). The  $\gamma_{i0k}$  and  $\gamma_{i1k}$  are subject  $i$ 's random effects corresponding to each marker. The longitudinal model can be interpreted as follows. The smooth trajectory  $f_k(t)$  is an overall mean response for marker  $k$  that anticipates a time-dependent population response, for example a dip following by a levelling off after a subject was initially enrolled in the trial. The  $i$ th subject's "intercept" term  $\gamma_{i0k}$  allows for a global subject deviation in population mean response and the "slope"  $\gamma_{i1k}$  can be interpreted as the subject-specific difference from the overall population rate of change of the marker at any time  $t$ .

More complex subject-specific trajectories are easily modelled in the existing framework. For example, each trajectory could be modelled with a second nonpenalized spline on top of the population trend  $f_k(t)$ . However given the very limited number of measurements taken on each subject, this would amount to gross overfitting of these data. For example, Brown, Ibrahim & DeGruttola (2005) considered this sort of framework with a normal prior distribution on the subject-specific spline coefficients; their fitted mean trajectories tracked the data quite well, even with measurements totalling as few as  $n_i = 4$ .

Prior values were fixed as follows:  $\alpha = 5, 10, \infty$ ; the  $\mu_k$ s were zero vectors;  $\Sigma_k = 10^{-6} \mathbf{I}_4$

(precision matrices for  $\beta_1$  and  $\beta_2$ );  $\sum_{\eta}^{-1} = \text{diag}(1, 0.00015, 0.5, 0.0001) \times \eta_0$  (precision matrix for Wishart prior);  $\eta_0 = 4$ ;  $v_{01} = v_{02} = \delta_{01} = \delta_{02} = a_1 = b_1 = a_2 = b_2 = 0.001$ . The matrix  $\Sigma_{\eta}$  was chosen to centre  $\mathbf{D}^{-1}$  at precision values obtained by Shah, Laird & Schoenfeld (1997) for a different, but similar, data set on ( $\log(\text{CD4})$ ,  $\log(\text{CD8})$ ) counts using a Gaussian model.

The degrees of freedom ( $df$ ),  $df = 3$  worked well for  $\alpha = 10$ ,  $df = 6$  worked for  $\alpha = 5$ . However, obtaining reasonable mixing was difficult for  $\alpha = 1$ .

Table 1 gives the posterior means and standard deviations (in parentheses) of model parameters along with the LPML values. Predictively, the MPT model with  $\alpha = 5$  is a better model than either  $\alpha = 10$  or the parametric model ( $\alpha = \infty$ ). The LPML from fitting the model without the splines, but rather an overall population line and with Gaussian random effects is  $-2024$ . The LPML assuming a spline rather than a simple linear function in time was  $-1933$ , indicating vastly increased predictive utility. The baseline HIV-RNA is a significant covariate for both CD4 and CD8, since the 95% credible intervals do not contain zero across models. However, none of the other covariates are significant in any of the models. These results are similar across the various MPT and the normal models; however, there is a general decreasing trend in both estimated effects and standard deviations as  $\alpha$  decreases from  $\infty$  to small values. The estimates of treatment effect are negative for both markers, suggesting a decrease in marker count over the the study period. However, the NNRTI experience was estimated to be negative for CD4 count and positive for CD8 count, although not significant. Hammer et al. (2002) found that, at 48 weeks, there was no significant difference in mean changes in CD4 cell counts relative to baseline in combined double-PI arms versus single-PI placebo, in line with our retaining the hypothesis  $H_0 : \beta_{12} = 0$ .

As  $\alpha$  falls from  $\alpha = \infty$  (parametric model) to  $\alpha = 10$  to  $\alpha = 5$ , the overall predictive performance of the model increases according to the LPML measure. For  $\alpha = 1$ , reasonable MCMC mixing did not occur even after 100 000 iterations. We instead fitted a simple Polya tree model with  $\mathbf{D}^{-1}$  fixed at the posterior mean under the  $\alpha = 10$  model, assigned a prior distribution on  $\alpha$  as  $\alpha \sim \text{gamma}(5, 1)$ . The MCMC mixed very well for this simple PT model. The resulting posterior mean and standard deviation for  $\alpha$  were 0.176 and 0.147, a marked shift towards much smaller values than under the prior, and indicating highly “non-normal” random effects. The LPML for the simple PT model increased by about 22, yielding a pseudo Bayes factor of  $e^{22} \approx 10^9$  in favor of the simple PT model with a prior distribution on  $\alpha$ . The estimated regression coefficients under this simple Polya tree model are also given in Table 1. Under the MPT model, the correlation between the intercepts is estimated to be 0.61, the correlation between the slopes 0.86. But the correlation between intercept and slope is 0.10 for  $\log(\text{CD4})$  and  $-0.13$  for  $\log(\text{CD8})$ . Thus, there is a high degree of similarity between CD4 and CD8 counts within a subject over time.

Figure 2 shows the kernel-smoothed estimated posterior densities for the random effects along with the posterior estimates of the  $m = 454$  values from fitting the PT model with  $\alpha \sim \text{gamma}(5, 1)$ . It is clear from these graphs that the PT ably models the non-normal, skewed characteristics of these densities. Figure 3 shows the estimated mean trajectories for two patients, along with 95% credible intervals; the solid circles represent the observed  $\log$  CD4 and CD8 counts. The solid line is the parametric model with spline mean, and the dashed line is the simple PT model with spline mean and  $\alpha \sim \text{Gamma}(5, 1)$ . The trajectories are not wildly different, but some differences are noticeable, especially for the CD8 counts. Differences in predictive (CD4, CD8) counts for subjects *not* in the study are bound to show more dissimilarities, if interest is rather in predictive inference.

## 6 Conclusion

Normality of random effects is a routine assumption for linear mixed models, but may be unrealistic. In this paper, we have proposed an approach based on a nonparametric Bayesian method for multivariate longitudinal data, where the random effects are assumed to have a flexible MPT prior distribution that generalizes the standard Gaussian assumption. The

model with an MPT prior distribution is shown to outperform the usual Gaussian random effects model. We have found in general that an MPT outperforms a PT with a fixed covariance matrix estimated from either an MPT model or the parametric model, with the same fixed value of  $\alpha$ .

Though our model was built from the features of the data at hand, it has general application to situations where multiple responses are measured over time. As pointed out, our model can have more general subject-specific curves such as splines, but due to data limitations, we instead considered a more parsimonious model.

### Acknowledgments

The authors thank the editor, associate editor, and referees for their valuable suggestions. Dr. Hanson's research was supported in part by USA NIH grant 2-R01-CA95955-05.

### Appendix: Conditional Distributions Under the Gaussian Model

Define matrices  $\mathbf{X}_i = \text{diag}(\mathbf{X}_{i1}, \dots, \mathbf{X}_{iK})$ ,  $\mathbf{Z}_i = \text{diag}(\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{iK})$ ,  $\mathbf{N}_i = \text{diag}(\mathbf{N}_{i1}, \dots, \mathbf{N}_{iK})$ ,  $\mathbf{R}_i = \text{diag}(\mathbf{R}_{i1}, \dots, \mathbf{R}_{iK})$ , and the vectors  $\mathbf{y}_i = (\mathbf{y}_{i1}^\top, \dots, \mathbf{y}_{iK}^\top)^\top$ ,  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_K^\top)^\top$ ,  $\mathbf{f} = (\mathbf{f}_1^\top, \dots, \mathbf{f}_K^\top)^\top$ .

Define  $\mathbf{V}_k = \sum_k + \sum_i \mathbf{X}_{ik}^\top \mathbf{R}_{ik}^{-1} \mathbf{X}_{ik}$ ,  $\mathbf{P}_k^{-1} = \tau_k^{-2} \boldsymbol{\Omega} + \sigma_k^{-2} \sum_i [\mathbf{N}_{ik} \mathbf{B}_k]^\top [\mathbf{N}_{ik} \mathbf{B}_k]$ , and  $\Delta_i = \mathbf{D}^{-1} + \mathbf{Z}_i^\top \mathbf{R}_i^{-1} \mathbf{Z}_i$ .

The full conditional distributions under the parametric normal model are:

$$\begin{aligned} \beta_k &\sim N_{p_k} \left[ \mathbf{V}_k^{-1} \left( \sum_k \mu_k + \sum_i \mathbf{X}_{ik}^\top \mathbf{R}_{ik}^{-1} (\mathbf{y}_{ik} - \mathbf{N}_{ik} \mathbf{f}_k - \mathbf{Z}_{ik} \gamma_{ik}) \right), \mathbf{V}_k^{-1} \right] \\ \gamma_i &\sim N_q \left[ \Delta_i^{-1} \mathbf{Z}_i^\top \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{N}_i \mathbf{f}), \Delta_i^{-1} \right] \\ \mathbf{D}^{-1} &\sim \text{Wishart} \left( \eta_0 + m, \left( \sum_{\eta}^{-1} + \sum_i \gamma_i \gamma_i^\top \right)^{-1} \right) \\ \sigma_k^{-2} &\sim \text{gamma} \left( \frac{\nu_{0k} + \sum_i n_i}{2}, \frac{\delta_{0k} + \sum_i \|\mathbf{y}_{ik} - \mathbf{X}_{ik} \boldsymbol{\beta}_k - \mathbf{N}_{ik} \mathbf{f}_k - \mathbf{Z}_{ik} \gamma_{ik}\|^2}{2} \right) \\ \boldsymbol{\varphi}_k &\sim N_L \left( \frac{1}{\sigma_k^2} \mathbf{P}_k \mathbf{B}_k^\top \sum_{i=1}^m \mathbf{N}_{ik}^\top (\mathbf{y}_{ik} - \mathbf{X}_{ik} \boldsymbol{\beta}_k - \mathbf{Z}_{ik} \gamma_{ik}), \mathbf{P}_k \right) \\ \tau_k^{-2} &\sim \text{gamma} \left( a_k + \frac{L}{2}, b_k + \frac{1}{2} \boldsymbol{\varphi}_k^\top \boldsymbol{\Omega} \boldsymbol{\varphi}_k \right) \end{aligned}$$

### References

Basu S, Chib S. Marginal likelihood and Bayes factors for Dirichlet process mixture models. *J Amer Statist Assoc.* 2003; 98:224–235.

Berger JO, Guglielmi A. Bayesian testing of a parametric model versus nonparametric alternatives. *J Amer Statist Assoc.* 2001; 96:174–184.

Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics.* 2003; 59:221–228. [PubMed: 12926706]

Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics.* 2005; 61:64–73. [PubMed: 15737079]

de Boor, C. *A Practical Guide to Splines*. New York: Springer-Verlag; 1978.

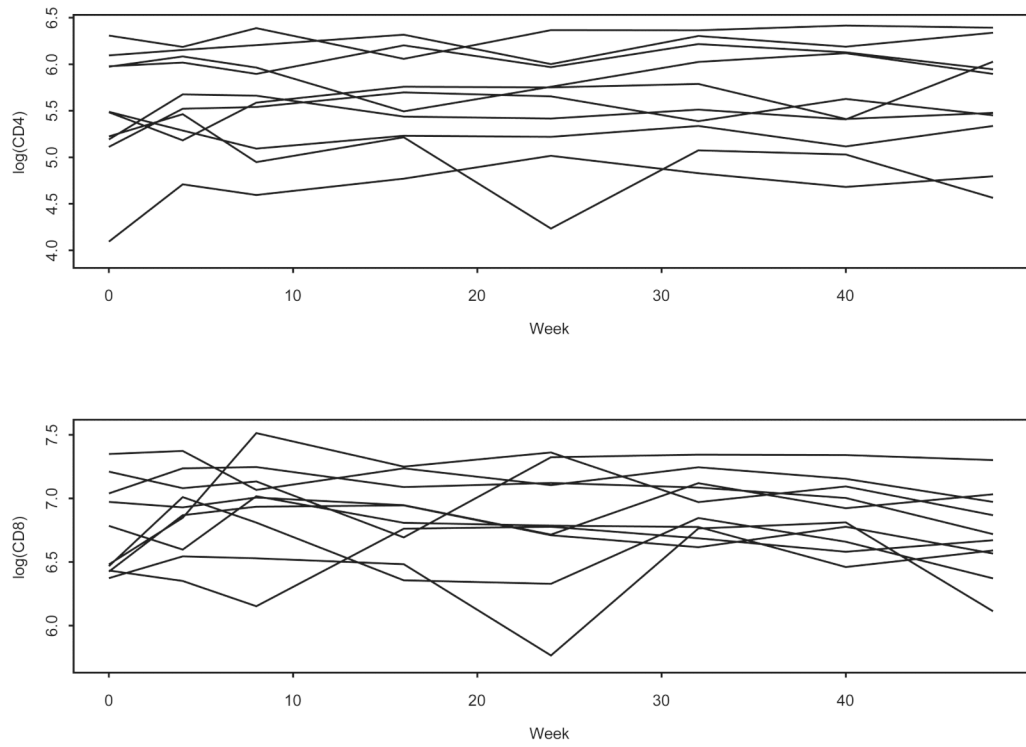
Diggle, P.; Heagerty, P.; Liang, K.; Zeger, S. *Analysis of Longitudinal Data*. Oxford: Oxford University Press; 2002.

Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statist Sci.* 1996; 11:89–121.

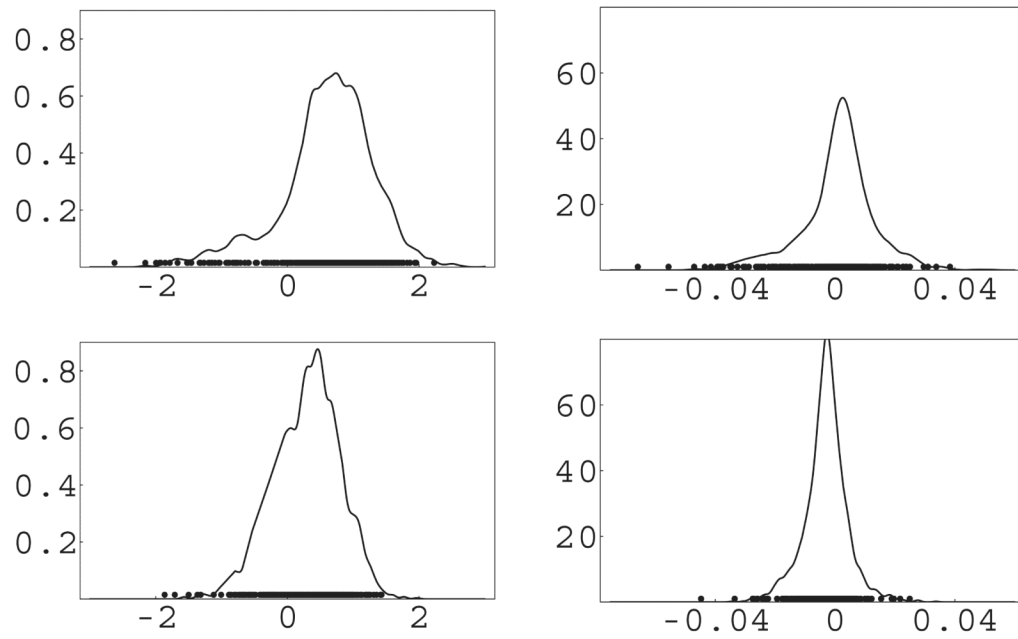
Fieuws S, Verbeke G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics.* 2006; 62:424–431. [PubMed: 16918906]

- Geisser I, Eddy W. A predictive approach to model selection. *J Amer Statist Assoc.* 1979; 74:153–160.
- Gelfand AE, Dey D. Bayesian model choice: asymptotic and exact calculations. *J Roy Statist Soc Ser B.* 1994; 56:501–514.
- Gelfand AE, Smith AFM. Sampling based approaches to calculating marginal densities. *J Amer Statist Assoc.* 1990; 85:398–409.
- Ghidey W, Lesaffre P, Eilers PHC. P-spline smoothing of the random effects distribution in a linear mixed model. *Biometrics.* 2004; 60:945–953. [PubMed: 15606415]
- Grizzle JE, Allen DM. Analysis of growth and dose-response curves. *Biometrics.* 1969; 25:357–381. [PubMed: 5794105]
- Hanson TE. Inference for mixtures of finite Polya trees models. *J Amer Statist Assoc.* 2006; 101:1548–1565.
- Hanson T, Johnson W. Modeling regression error with a mixture of Polya trees. *J Amer Statist Assoc.* 2002; 97:1020–1033.
- Hammer SM, Vaida F, Bennett KK, et al. Dual vs single protease inhibitor therapy following antiretroviral treatment failure. *J Amer Med Assoc.* 2002; 288:169–180.
- Kadane JB, Lazar NA. Methods and criteria for model selection. *J Amer Statist Assoc.* 2004; 99:279–290.
- Khatri CG. A Note on a MANOVA model applied to problems in growth curve. *Ann Inst Statist Math.* 1966; 8:75–86.
- Kleinman K, Ibrahim JG. A semi-parametric Bayesian approach to the random effects model. *Biometrics.* 1998; 54:921–938. [PubMed: 9750242]
- Laird NM, Ware JH. Random effects model for longitudinal data. *Biometrics.* 1982; 38:963–974. [PubMed: 7168798]
- Lang S, Brezger A. Bayesian P-Splines. *J Comput Graph Statist.* 2004; 13:183–212.
- Lavine M. Some aspects of Polya tree distributions for statistical modelling. *Ann Statist.* 1992; 20:1222–1235.
- Lavine M. More aspects of Polya tree distributions for statistical modelling. *Ann Statist.* 1994; 22:1161–1176.
- O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics.* 1984; 40:1079–1087. [PubMed: 6534410]
- Paddock, SM. Unpublished Doctoral Thesis. Duke University; 1999. Randomized Polya Trees: Bayesian Nonparametrics for Multivariate Data Analysis.
- Paddock S, Ruggeri F, Lavine M, West M. Randomised Polya Tree Models for Nonparametric Bayesian Inference. *Statist Sinica.* 2003; 13:413–460.
- Potthoff R, Roy SN. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika.* 1964; 51:313–326.
- Reinsel G. Estimation and prediction in a multivariate random effects generalized linear model. *J Amer Statist Assoc.* 1984; 79:406–414.
- Roy J, Lin X. Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: changes in methadone treatment practices. *J Amer Statist Assoc.* 2002; 97:40–52.
- Shah A, Laird N, Schoenfeld D. A random-effects model for multiple characteristics with possibly missing data. *J Amer Statist Assoc.* 1997; 92:775–779.
- Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics.* 2002; 58:742–753. [PubMed: 12495128]
- Walker SG, Mallick BK. Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *J Roy Statist Soc Ser B.* 1997; 59:845–860.
- Walker SG, Mallick BK. Semiparametric accelerated life time model. *Biometrics.* 1999; 55:477–483. [PubMed: 11318203]
- Zeger SL, Diggle P. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics.* 1994; 50:689–699. [PubMed: 7981395]
- Zhang D, Lin X, Raz J, Sowers MF. Semiparametric stochastic mixed models for longitudinal data. *J Amer Statist Assoc.* 1998; 93:710–719.

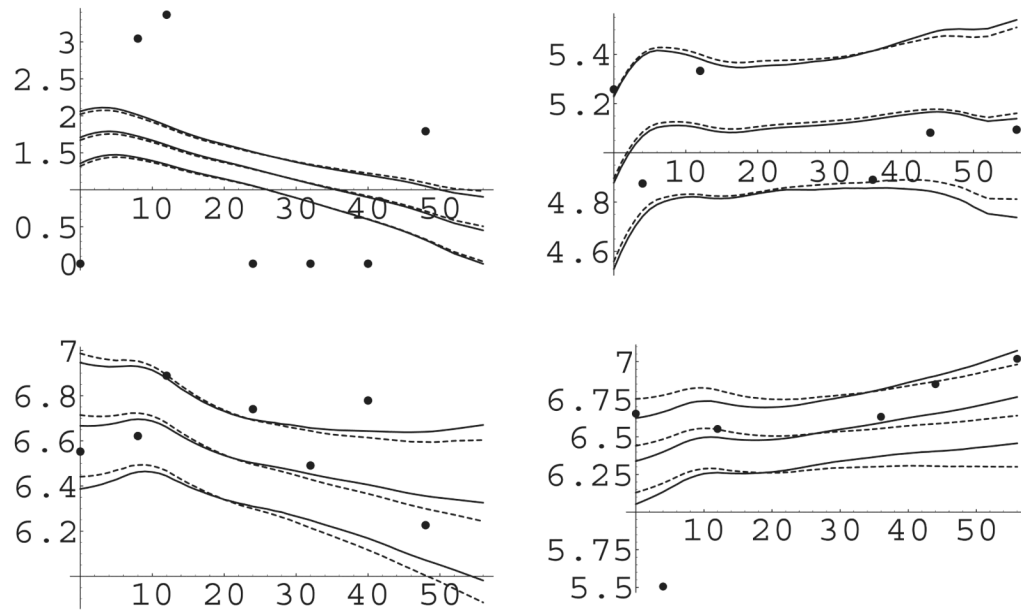
Zhang D, Davidian M. Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*. 2001; 57:795–802. [PubMed: 11550930]



**Figure 1.** Trajectories of log(CD4) and log(CD8) for 10 randomly selected subjects, ACTG 398 data.



**Figure 2.** Estimated marginal densities (lines) and posterior mean random effects (dots) from simple Polya tree model with  $\alpha \sim \text{gamma}(5, 1)$ , ACTG 398 data. Top panels are intercept (left) and slope (right) for  $\log(\text{CD4})$ , bottom panels are for  $\log(\text{CD8})$ .



**Figure 3.** Raw data (dots) and estimated trajectories for two subjects, ACTG 398 data. Solid lines are parametric ( $\alpha \rightarrow \infty$ ) model and dashed lines are simple Polya tree model. Columns are subjects and the top panel is log(CD4), bottom log(CD8).



**Table 1**

Posterior means (standard deviations) of regression effects and LPML values for models (9) and (7), ACTG 398 data.  $\alpha$  is the mixture of Polya trees (MPT) and Polya tree (PT) prior distribution precision; higher  $\alpha$  values indicate a more parametric model.

	$\alpha \sim \Gamma(5, 1)$ PT	$\alpha = 5$ MPT	$\alpha = 10$ MPT	$\alpha = \infty$ Parametric
LPML	-1901	-1924	-1930	-1933
$\beta_{11}$	-0.294 (0.040)	-0.399 (0.0489)	-0.439 (0.0488)	-0.490 (0.0493)
$\beta_{12}$	-0.068 (0.062)	-0.0608 (0.0706)	-0.0246 (0.0744)	-0.00619 (0.0788)
$\beta_{13}$	-0.147 (0.054)	-0.125 (0.0679)	-0.136 (0.0710)	-0.145 (0.0743)
$\beta_{21}$	-0.064 (0.030)	-0.102 (0.0315)	-0.121 (0.0324)	-0.140 (0.0318)
$\beta_{22}$	-0.032 (0.043)	-0.0196 (0.0470)	0.0029 (0.0489)	0.00915 (0.0507)
$\beta_{23}$	0.036 (0.043)	0.0323 (0.0461)	0.0325 (0.0468)	0.0330 (0.0480)