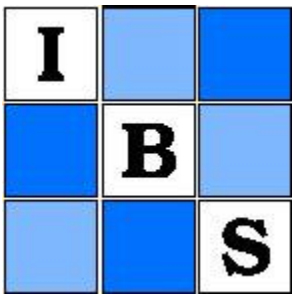


BIOMETRICS 54, 921–938
September 1998

A Semiparametric Bayesian Approach to the Random Effects Model

Ken P. Kleinman and Joseph G. Ibrahim*

Department of Biostatistics, Harvard School of Public Health,
677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

SUMMARY

In longitudinal random effects models, the random effects are typically assumed to have a normal distribution in both Bayesian and classical models. We provide a Bayesian model that allows the random effects to have a nonparametric prior distribution. We propose a Dirichlet process prior for the distribution of the random effects; computation is made possible by the Gibbs sampler. An example using marker data from an AIDS study is given to illustrate the methodology.

1. Introduction

In longitudinal clinical trials, subjects may be evaluated many times over the course of the study. The repeated measurements on each subject are usually correlated within subject, although subjects are usually taken to be independent from each other. One model for this sort of data is known as the random effects model, popularized by Laird and Ware (1982).

For a given individual i with n_i repeated measurements, the random effects model for outcome vector y_i is given by

$$y_i = X_i\beta + Z_i b_i + e_i, \qquad i, \ldots, n,$$

where y_i is $n_i \times 1$; X_i is an $n_i \times p$ matrix of fixed covariates; β is a $p \times 1$ parameter vector of regression coefficients, commonly referred to as fixed effects in these models; Z_i is an $n_i \times v$ matrix of covariates for the $v \times 1$ vector of random effects b_i; and e_i is an $n_i \times 1$ vector of errors. It is standard in implementations of this model to assume e_i and b_i are independent and that both are normally distributed with $e_i \sim \mathrm{N}_{n_i}(0, \sigma^2 I_{n_i})$ and $b_i \sim \mathrm{N}_v(0, D)$, where I_s is the $s \times s$ identity matrix and $\mathrm{N}_s(\mu, \Sigma)$ denotes the s-dimensional multivariate normal distribution with mean μ and variance Σ. Under these assumptions,

$$[y_i \mid \beta, b_i] \sim \mathrm{N}_{n_i}(X_i\beta + Z_i b_i, \sigma^2 I_{n_i}). \tag{1}$$

Throughout, we denote the conditional distribution of A given B by $[A \mid B]$. The model in (1) assumes a distinct set of regression coefficients for each individual once the random effects are known.

For convenience, the distribution of the random effects is usually taken to be normal, as stated above. Large-sample classical (frequentist) inference regarding β should not be much affected asymptotically by changing the distribution of the random effects since the first two moments of the marginal distribution of the outcome variable do not depend on the normality of the distribution of the random effects. To see this, notice that, marginally,

$$[y_i \mid \beta, \sigma^2, D] \sim \mathrm{N}_{n_i}(X_i\beta, Z_i D Z_i^{\mathrm{T}} + \sigma^2 I_{n_i}). \tag{2}$$

For inference regarding β from (2), the individual regression coefficients simply result in a complex

* *Corresponding author's email address:* ibrahim@jimmy.harvard.edu

Key words: Dirichlet process; Gibbs sampler; Longitudinal data; Nonparametric Bayes.

covariance structure. Bayesian inference for β, using the marginal likelihood implied by (2), will depend only on (y, σ^2, D). But the nature of this dependence will be sensitive to the distributional form ascribed to the b_i.

In this article, we present a model for which the random effects have a nonparametric distribution. There has been some work toward this end in the classical setting. In the Bayesian paradigm, it has been accomplished for the repeated measures (West, Müller, and Escobar, 1994) and for the randomized complete block design (Bush and MacEachern, 1996). We provide a general framework for Bayesian analysis of random effects models in which a nonparametric Dirichlet process prior is specified for the random effects. Only recently have tools allowing Bayesian analysis become computationally feasible; here we provide a detailed exploration of a biostatistical application of interest in any situation when the random effects model is appropriate.

From the classical perspective, the distribution of the random effects has an important effect on some quantities of interest. The estimated random effect for each individual will be changed by changing the distribution of the random effects. This point is important because there are many applications in which an estimate of the random effect itself is desired. For example, in Tsiatis, DeGruttola, and Wulfsohn (1995), the estimated random effects are used alternately as covariates themselves in a Cox regression model or to create values for time-varying covariates in such a model. Similarly, Mori, Woodworth, and Woolson (1992), DeGruttola and Tu (1994), and Wu and Carroll (1988) all present complex models in which the random effects are both estimated and used in predicting other pieces of the model. In such applications, unbiased estimation of the random effects is crucial and the assumption of normality may introduce bias.

From the Bayesian perspective, inferential interest focuses on the posterior distribution of the regression coefficients β as well as the random effects b_i. Allowing distributions other than the normal for the random effects may more accurately model our prior beliefs, or it may allow us to better express our uncertainty about the true distribution of the random effects. It is also important to accurately model the distribution of the random effects when prediction for a future observation from a given subject is desired. Another situation in which it would be desirable to relax the assumption of normality is when inference is to be made about the distribution of the random effects itself. For example, in randomized complete block designs, the distribution of the random block effects is usually of inferential interest (see John, 1971).

One major question arising in Bayesian analysis concerns the sensitivity of the results to the chosen prior. In Section 4, we give an example demonstrating how the posterior distribution of the regression coefficients can be affected by the prior distribution of the random effects. For a further example of the impact of this prior distribution, see Verbeke and Lesaffre (1996). They show that, when the true distribution of the random effects is a mixture of normals, assuming that it is a single unmixed normal can lead to poor estimates of the random effects. The semiparametric approach we develop here is therefore very valuable in that it circumvents debate about the prior selection, as far as the random effects are concerned.

Another attraction of our approach is that it allows exact Bayesian inference, even in small sample sizes. This is accomplished through the use of the Gibbs sampler. We develop computational tools and demonstrate how the Gibbs sampler can be implemented for our model. We also show how to make Bayesian inference for all of the parameters in the model.

The nonparametric Bayesian approach for the random effects b_i is to specify a prior distribution on the space of all possible distribution functions. This prior is applied to the general prior distribution for the b_i. This can be accomplished with a Dirichlet process prior distribution. For random effects models, this means that the usual normal prior on the random effects is replaced with a nonparametric prior, followed by a Dirichlet process prior on that general distribution. The foundation of this technology is discussed in Ferguson (1973), where the Dirichlet process and its usefulness as a prior distribution are discussed. The practical application of such models, using the Gibbs sampler, has been pioneered by Doss (1994), MacEachern (1994), Escobar (1994), Bush and MacEachern (1996), Liu (1996), and Müller, Erkanli, and West (1996). Other important work in this area has been done by West et al. (1994), Escobar and West (1995), and MacEachern and Müller (1998).

The rest of this article is organized as follows. In Section 2, we describe the mixture of Dirichlet process (MDP) structure, which we propose for our model. In Section 3, we show how to apply the MDP structure to the random effects model. In Section 4, we illustrate our methodology with real data and, in addition, compare our approach to the fully parametric Bayesian model. In Section 5, we discuss our results and propose directions for future research.

2. Mixture of Dirichlet Process Models

The mixture of Dirichlet process model arises in cases of the following general situation.

Suppose an $n_i \times 1$ random vector x_i has a parametric distribution indexed by the $w \times 1$ vector θ_i, $i = 1, \ldots, n$. Then suppose the θ_i themselves have a prior distribution with known hyperparameters Ψ_0. Thus,

$$\begin{aligned} &\text{Stage 1: } [x_i \mid \theta_i] \sim D_{n_i}(h_1(\theta_i)), \\ &\text{Stage 2: } [\theta_i \mid \Psi_0] \sim D_w(h_2(\Psi_0)), \end{aligned} \tag{3}$$

where $D_s(\cdot)$ is a generic label for an s-dimensional parametric multivariate distribution and $h_1(\cdot)$ and $h_2(\cdot)$ are functions. The MDP model (Escobar, 1994; MacEachern, 1994) removes the assumption of a parametric prior at the second stage and replaces it with a general distribution G. The distribution G then in turn has a Dirichlet process prior (Ferguson, 1973), leading to

$$\begin{aligned} &\text{Stage 1: } [x_i \mid \theta_i] \sim D_{n_i}(h_1(\theta_i)), \\ &\text{Stage 2: } \theta_i \overset{\text{i.i.d.}}{\sim} G, \\ &\text{Stage 3: } [G \mid M, \Psi_0] \sim DP\left(M \cdot G_0(h_2(\Psi_0))\right), \end{aligned} \tag{4}$$

where G_0 is a w-dimensional parametric distribution and M is a positive scalar. The parameters of a Dirichlet process are $G_0(\cdot)$, a probability measure, and M, a positive scalar assigning mass to the real line. The parameter $G_0(\cdot)$, often called the base measure, is a distribution that approximates the true nonparametric shape of G. The scalar M reflects our prior belief about how similar the nonparametric distribution G is to the base measure $G_0(\cdot)$. There are two special cases in which the MDP model leads to the fully parametric case. As $M \to \infty$, $G \to G_0(\cdot)$, so that the base measure is the prior distribution for θ_i. Also, if $\theta_i \equiv \theta$ for all i, the same is true. For a more hierarchical modeling approach, it is possible to place prior distributions on (M, Ψ_0). In Section 3, we will place a prior on Ψ_0, but we will not do so for M. The specification in (4) results in a semiparametric specification in that a fully parametric distribution is given in Stage 1 and a nonparametric distribution is given in Stages 2 and 3.

The Polya urn representation of the Dirichlet process, described by Blackwell and MacQueen (1973), is useful for sampling purposes. It can be described as follows. The draw of θ_1 is always from the base measure. The draw of θ_2 is equal to θ_1 with probability p_1 and is from the base measure with probability $p_0 = 1 - p_1$. The draw of θ_3 is equal to θ_1 with probability p_1, equal to θ_2 with probability p_2, and is a draw from the base measure with probability $p_0 = 1 - (p_1 + p_2)$. The values of the p's change with each new draw. This process continues until θ_n is equal to each of the preceding θ's with probability $p_i, i \in 1 \cdots n-1$, and is a draw from the base measure with probability $p_0 = 1 - \Sigma_{i=1}^{n-1} p_i$. The values of $p_i, i = 0, \ldots, n-1$, are determined from the Dirichlet process parameters. In other words, the θ's are actually drawn from a mixture distribution where the mixing probabilities are determined by the Dirichlet process of Stage 3, thus giving rise to the MDP label. From this representation, it is clear that, if $\theta_i \equiv \theta$ for all i, then we draw θ from the base measure with probability 1, and thus the base measure is the prior.

The MDP model is simplified in practice by the Polya urn representation, using the fact that, marginally, the θ_i are distributed as the base measure along with the added property that $P(\theta_i = \theta_j, i \neq j) > 0$. Thus, the MDP structure results in what MacEachern (1994) calls a cluster structure among the θ_i's. This cluster structure partitions the n θ_i's into k sets or clusters, $0 < k \leq n$. All of the observations in a cluster share an identical value of θ, and subjects in different clusters have differing values of θ.

As derived in Escobar (1994), conditional on the other θ's, θ_i has the following mixture distribution:

$$p(\theta_i \mid x, \theta_{-i}) \propto \sum_{j \neq i} q_j \delta_{\theta_j} + M q_0 g_0(\theta_i) p(x_i \mid \theta_i), \tag{5}$$

where $x = (x_1, \ldots, x_n)$, $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$, and $p(x_i \mid \theta_i)$ is the sampling distribution of the x_i's. The values q_j and Mq_0 can be normalized to get the selection probabilities $p_i, i = 0, \ldots, n-1$, in the Polya urn scheme described above. In addition, δ_s is a degenerate distribution with point mass at s, and $g_0(\cdot)$ is the density corresponding to the probability measure $G_0(\cdot)$. Finally, $q_j = p(x_i \mid \theta_j)$, $j = 1, \ldots, i-1, i+1, \ldots, n$, and $q_0 = \int p(x_i \mid \theta) g_0(\theta) d\theta$.

To demonstrate the MDP model, consider the germinal example of Escobar (1994) and Escobar and West (1995). Suppose that x_i has the univariate normal distribution with unknown mean θ_i

and known variance σ_x^2. In this case, we have $n_i = 1, i = 1, \ldots, n$. Also assume that each θ_i has the univariate normal distribution. Then (3) becomes

$$\text{Stage 1: } [x_i \mid \theta_i, \sigma_x] \sim \mathrm{N}(\theta_i, \sigma_x^2),$$
$$\text{Stage 2: } [\theta_i \mid \mu, \sigma_\theta] \sim \mathrm{N}(\mu, \sigma_\theta^2).$$

The MDP model removes the assumption of normality at the second stage, resulting in

$$\text{Stage 1: } [x_i \mid \theta_i, \sigma_x] \sim \mathrm{N}(\theta_i, \sigma_x^2),$$
$$\text{Stage 2: } \theta_i \sim G,$$
$$\text{Stage 3: } [G \mid M, \Psi_0] \sim DP\left(M \cdot G_0(h_2(\Psi_0))\right).$$

Suppose $G_0 = \mathrm{N}(\mu, \sigma_\theta^2)$ so that $\Psi_0 = (\mu, \sigma_\theta^2)$. In this case, the unnormalized selection probability q_j is equal to $p(x_i \mid \theta_j) = \phi(x_i \mid \theta_j, \sigma_x^2)$, where $\phi(\cdot \mid \mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2. With probability proportional to q_j, $\theta_i \sim \delta_{\theta_j}$, which means that $\theta_i = \theta_j$ with probability 1. The unnormalized selection probability q_0 is given by

$$q_0 = \int p(x_i \mid \theta, \sigma_x^2) g_0(\theta \mid \Psi_0) d\theta \equiv \int \phi(x_i \mid \theta, \sigma_x^2) \phi(\theta \mid \mu, \sigma_\theta^2) d\theta.$$

With probability proportional to $M \times q_0$,

$$p(\theta_i \mid x_i) \propto g_0(\theta \mid \Psi_0) p(x_i \mid \theta) = \mathrm{N}(\theta \mid \mu, \sigma_\theta^2) \mathrm{N}(x_i \mid \theta, \sigma_x^2),$$

where $\mathrm{N}(s \mid \mu, \sigma^2)$ indicates that s has a normal distribution with mean μ and variance σ^2. Then

$$[\theta_i \mid x_i] \sim \mathrm{N}\left(\left[\left(\sigma_\theta^2 + \sigma_x^2\right)^{-1} \sigma_\theta^2 \sigma_x^2\right]\left(\frac{\mu}{\sigma_\theta^2} + \frac{x_i}{\sigma_x^2}\right), \left(\sigma_\theta^2 + \sigma_x^2\right)^{-1} \sigma_\theta^2 \sigma_x^2\right).$$

In the example above, selecting G_0 to be normal when the sampling distribution of the data is normal emulates the conjugate relationship between sampling distribution and prior in the usual Bayesian hierarchy. In the MDP case, the sampling distribution is conjugate to the base measure. MDP models with base measures and sampling distributions that are conjugate in this fashion are called conjugate MDP models by MacEachern (1994). The computational advantages of the conjugate MDP model are clear from the example. First, q_0 has a closed form. Second, the distribution of θ_i corresponding to q_0 is from the same exponential family as the base measure. As a result, Gibbs sampling in the conjugate model described above can proceed in a relatively straightforward fashion, as described in Section 3.

When conjugacy is not assumed, the integral needed for q_0 typically has no closed-form solution. Because this integral must be evaluated n times within each Gibbs cycle, the cost in time of numerical integration is compounded, as is the cost in accuracy of approximations. In addition, the corresponding distribution of θ_i is from some nonstandard family, so that sampling θ_i is more difficult in these situations as well. Work by MacEachern and Müller (1998) and Walker and Damien (1997) describes a technique whereby nonconjugate MDP models can be fit without numerical integration or approximation, but it is not discussed here.

3. Dirichlet Process Priors in Random Effects Models

In this section, we describe how the MDP model can be applied to the longitudinal random effects model. We show the full conditional distributions and how Gibbs sampling can be implemented for both the conjugate and a nonconjugate case. Section 3.1 describes in more detail a special case of a model presented in West et al. (1994). Section 3.2 presents an extension of that work to a nonconjugate MDP model.

3.1 *Multivariate Normal Base Measure*

If we assume the sampling distribution for y to be normal, then a normal base measure for the random effects completes a conjugate MDP model. Let the distribution of the vector of outcomes y_i for subject i be

$$[y_i \mid \beta, b_i, \sigma^2] \sim \mathrm{N}_{n_i}(X_i\beta + Z_i b_i, \sigma^2 I_{n_i}).$$

Letting $\tau = \sigma^{-2}$, the prior specifications are

$$\tau \sim \text{gamma}\left(\frac{\alpha_0}{2}, \frac{\lambda_0}{2}\right),$$

$$\begin{aligned}\beta &\sim \mathrm{N}_p(\mu_0, \Sigma_0),\\ b_i &\sim G,\\ G &\sim DP(M \cdot \mathrm{N}_v(0, D)).\end{aligned} \tag{6}$$

Model (6) implies that there are p 'fixed' effects and v random effects.

When G is a fully parametric prior, the joint posterior can be easily found, as in Wilks et al. (1993). Then, following usual algebraic routes, the full conditionals for β and τ are

$$[\beta \mid b, \tau, y] \sim \mathrm{N}_p(Ta, T), \tag{7}$$

where $T = (\tau \Sigma_{i=1}^n X_i^{\mathrm{T}} X_i + \Sigma_0^{-1})^{-1}$, $a = \Sigma_{i=1}^n X_i^{\mathrm{T}}(y_i - Z_i b_i) + \Sigma_0^{-1}\mu_0$, and

$$[\tau \mid \beta, b, y] \sim \text{gamma}\left(\frac{n+\alpha_0}{2}, \frac{\sum_i r_i^{\mathrm{T}} r_i + \lambda_0}{2}\right), \tag{8}$$

where $r_i = y_i - X_i\beta - Z_i b_i$. These full conditionals are unchanged by the MDP model and can be inserted as displayed above.

From the discussion of MDP models in Section 2, we find, as in West et al. (1994),

$$\begin{aligned}p(b_i \mid \beta, \tau, y, b_{-i}) \propto &\sum_{j \neq i} \phi(y_i \mid X_i\beta + Z_i b_j, \tau I_{n_i}) \cdot \delta_{b_j}\\ &+ \left\{ M \int \phi(y_i \mid X_i\beta + Z_i b_i, \tau I_{n_i}) \phi(b_i \mid 0, D) db_i \right\}\\ &\times \phi(b_i \mid 0, D) p(y_i \mid b_i, \beta, \tau, y_j),\end{aligned}$$

where b_{-i} denotes the random effects for the subjects excluding subject i. Also, as in Section 2, δ_s is a degenerate distribution with point mass at s. After some algebra, we get

$$\begin{aligned}p(b_i \mid \beta, \tau, y, b_{-i}) \propto &\left(\sum_{j \neq i} \tau^{n_i/2} \exp\left[\frac{-\tau}{2} (y_i - X_i\beta - Z_i b_j)^{\mathrm{T}} (y_i - X_i\beta - Z_i b_j) \right] \cdot \delta_{b_j} \right)\\ &+ M |Q_i|^{1/2} |D|^{-1/2} \tau^{n_i/2}\\ &\times \phi(b_i \mid 0, D) p(y_i \mid b_i, \beta, \tau),\end{aligned} \tag{9}$$

where $Q_i = (D^{-1} + \tau Z_i^{\mathrm{T}} Z_i)^{-1}$ and $U_i = (\tau Z_i Q_i Z_i^{\mathrm{T}} - I)$. In the above specification, each summand is separated into two elements. The first element is a mixing probability, and the second is a distribution to be mixed. So with probability proportional to

$$\tau^{n_i/2} \exp\left(\frac{-\tau}{2} (y_i - X_i\beta - Z_i b_j)^{\mathrm{T}} (y_i - X_i\beta - Z_i b_j) \right),$$

we select from distribution δ_{b_j}, which means that we set $b_i = b_j$. Also, with probability proportional to

$$M |Q_i|^{1/2} |D|^{-1/2} \tau^{\frac{n_i}{2}} \int \exp\left\{ \frac{\tau}{2} \left[(y_i - X_i\beta)^{\mathrm{T}} U_i (y_i - X_i\beta) \right] \right\} db_i,$$

we select from

$$p(b_i \mid \beta, \tau, y_i) \propto \phi(b_i \mid 0, D) p(y_i \mid b_i, \beta, \tau, y_j),$$

meaning we sample b_i from its full conditional,

$$[b_i \mid \beta, \tau, y_i] \sim \mathrm{N}_v\left(\tau Q_i Z_i^{\mathrm{T}} (y_i - X_i\beta), Q_i \right).$$

This results in a mixture distribution where one piece is a normal distribution and all of the others are point masses.

There is some plausible intuition behind the above mixture scheme. If subject i has a relatively large residual using subject j's random effect b_j, then b_j is relatively less likely to be chosen as the random effect for subject i. Conversely, if subject i has a relatively small residual using subject j's random effect, then the random effect b_j is relatively more likely to be chosen as the random effect for subject i. On the other hand, the greater the residuals for subject i, assuming its random effect parameters are all 0, the greater the probability that subject i gets a new value as its random effect.

The Gibbs sampler for $p(\beta, b, \tau \mid D, y)$ can be described as follows:

(0) Select starting values $b^{(0)}$ and $\tau^{(0)}$. Set $i = 0$.
(1) Sample $\beta^{(i+1)}$ from $[\beta \mid b^{(i)}, \tau^{(i)}, y]$ according to equation (7).
(2) Sample $\tau^{(i+1)}$ from $[\tau \mid \beta^{(i+1)}, b^{(i)}, y]$ according to equation (8).
(3.1) Sample $b_1^{(i+1)}$ from $[b_1 \mid \beta^{(i+1)}, b_{-1}^{(i)}, \tau^{(i+1)}, y]$ according to equation (9).
$\vdots$
(3.n) Sample $b_n^{(i+1)}$ from $[b_n \mid \beta^{(i+1)}, b_{-n}^{(i+1)}, \tau^{(i+1)}, y]$ according to equation (9).
(4) Set $i = i + 1$. Return to step (1).

Typically, the covariance matrix D in the base measure of the Dirichlet process in model (6) is unknown and therefore a suitable prior distribution must be specified for it. Note that, once this has been accomplished, the base measure is no longer marginally normal.

For convenience, suppose

$$D^{-1} \sim \text{Wishart}(d_0, c_0 R_0),$$

where $d_0 \geq v$, $c_0 > 0$, and R_0 is a $v \times v$ positive definite matrix. Then *a priori*

$$p\left(D^{-1} \mid d_0, c_0, R_0\right) \propto \left|D^{-1}\right|^{\frac{d_0 - v - 1}{2}} \exp\left\{\frac{-1}{2}\text{tr}\left((c_0 R_0)^{-1} D^{-1}\right)\right\},$$

where $\text{tr}(\cdot)$ denotes trace.

After choosing random effects for each subject, the subjects will be grouped into clusters in which the subjects have equal b_i's. That is, after selecting a new b_i for each subject i in the sample, there will be some number $k, 0 < k \leq n$, of unique values among the b_i's. Denote these unique values by $\gamma_l, l = 1 \cdots k$. Additionally, let l represent the set of subjects with common random effect γ_l. Note that knowing the random effects is equivalent to knowing k, all of the γ's, and the cluster memberships l. Then for the purposes of calculating the full conditionals of D^{-1}, the γ_l are k independent observations from $\text{N}_v(0, D)$. Thus, as in West et al. (1994),

$$\begin{aligned} p\left(D^{-1} \mid b, \beta, y, \tau\right) &= p\left(D^{-1} \mid \gamma, l, \beta, y, \tau\right) \\ &\propto \left|D^{-1}\right|^{\frac{d_0 + k - v - 1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left((c_0 R_0)^{-1} D^{-1} - \frac{1}{2}\sum_{l=1}^{k} \gamma_l^{\text{T}} D^{-1} \gamma_l\right)\right\}, \end{aligned}$$

so that

$$\left[D^{-1} \mid b, \beta, y, \tau\right] \sim \text{Wishart}\left(d_0 + k, \left((c_0 R_0)^{-1} + \sum_{1}^{k} \gamma_l \gamma_l^{\text{T}}\right)^{-1}\right). \tag{10}$$

The Gibbs sampling scheme above must be modified by an insertion of a step to sample $D^{(i)}$ from distribution (10).

One additional piece of the model is recommended by Bush and MacEachern (1996) as an aid to convergence for the Gibbs sampler. To speed mixing over the entire parameter space, they suggest moving around the γ's after determining how the b_i's are grouped. The conditional density of γ_l is

$$p(\gamma_l \mid \beta, \tau, b, D, y) \propto \phi(\gamma_l \mid 0, D) \prod_{i \in l} p(y_i \mid \beta, b, D, \tau),$$

which implies that

$$[\gamma_l \mid \beta, \tau, b, y] \sim \text{N}\left(\tau Q_l \sum_{i \in l} Z_i^{\text{T}}(y_i - X_i\beta), Q_l\right), \tag{11}$$

where $Q_l = (D^{-1} + \tau \Sigma_{i \in l} Z_i^{\text{T}} Z_i)^{-1}$.

The Gibbs sampling scheme is modified by inserting the following steps between steps (3.n) and (4).

(3a.1) Sample $\gamma_1^{(i+1)}$ from $[\gamma_1 \mid \beta^{(i+1)}, \tau^{(i+1)}, D^{(i)}, y]$ according to equation (11).
$\vdots$
(3a.k) Sample $\gamma_k^{(i+1)}$ from $[\gamma_k \mid \beta^{(i+1)}, \tau^{(i+1)}, D^{(i)}, y]$ according to equation (11).

Note that selecting a new value for γ_l changes the b_i's for subjects in cluster l.

Then the final Gibbs sampling scheme is

(0) Select starting values $b^{(0)}, \tau^{(0)}$, and $D^{(0)}$. Set $i = 0$.
(1) Sample $\beta^{(i+1)}$ from $[\beta \mid b^{(i)}, \tau^{(i)}, D^{(i)}, y]$ according to equation (7).
(2) Sample $\tau^{(i+1)}$ from $[\tau \mid \beta^{(i+1)}, b^{(i)}, D^{(i)}, y]$ according to equation (8).
(3.1) Sample $b_1^{(i+1)}$ from $[b_1 \mid \beta^{(i+1)}, b_{-1}^{(i)}, D^{(i)}, \tau^{(i+1)}, y]$ according to equation (9).
$\vdots$
(3.n) Sample $b_n^{(i+1)}$ from $[b_n \mid \beta^{(i+1)}, b_{-n}^{(i+1)}, D^{(i)}, \tau^{(i+1)}, y]$ according to equation (9).
(3a.1) Sample $\gamma_1^{(i+1)}$ from $[\gamma_1 \mid \beta^{(i+1)}, D^{(i)}, \tau^{(i+1)}, y]$ according to equation (11).
$\vdots$
(3a.k) Sample $\gamma_k^{(i+1)}$ from $[\gamma_k \mid \beta^{(i+1)}, D^{(i)}, \tau^{(i+1)}, y]$ according to equation (11).
(4) Sample ${D^{-1}}^{(i+1)}$ from $[D^{-1} \mid \beta^{(i+1)}, b^{(i+1)}, \tau^{(i+1)}, y]$ according to equation (10).
(5) Set $i = i + 1$. Return to step (1).

3.2 *A Modified MDP Model*

Samples of the b_i from a modified distribution can be obtained by specifying the base measure as a multivariate normal scale mixture. The multivariate t distribution can be obtained as a scale mixture of the multivariate normal distribution as follows. If

$$[x \mid \eta] \sim \mathrm{N}_p(\mu, \eta^{-1}\Sigma), \tag{12}$$

$$\eta \sim \mathrm{gamma}(s/2, s/2), \tag{13}$$

then the marginal distribution of x is $St_p(s, \mu, \Sigma)$, where $St_p(s, \mu, \Sigma)$ is a p-dimensional Student-t distribution with s degrees of freedom, mean μ, and dispersion matrix Σ. From a sampling perspective, sampling from

$$\eta \sim \mathrm{gamma}(s/2, s/2)$$

and then from

$$[x \mid \eta] \sim \mathrm{N}_q(\mu, \eta^{-1}\Sigma)$$

is equivalent to sampling from $St_p(s, \mu, \Sigma)$. Thus, marginally, the distribution of x is multivariate t, with η being integrated out. This representation is often used in the Gibbs sampling literature (see Wakefield et al. (1995) and references therein).

The representation (12) and (13) of the multivariate t inserted into a MDP model allows a closed-form solution to q_0 and yields samples of the b_i from a modified distribution. The specifications for the modified MDP model are as follows:

$$\begin{aligned}
[y_i \mid \beta, b_i, \sigma^2] &\sim \mathrm{N}_{n_i}(X_i\beta + Z_i b_i, \tau I_{n_i}),\\
\tau &\sim \mathrm{gamma}\left(\frac{\alpha_0}{2}, \frac{\lambda_0}{2}\right),\\
\beta &\sim \mathrm{N}_p(\mu_0, \Sigma_0),\\
b_i &\sim G,\\
G &\sim DP(M \cdot \mathrm{N}_v(0, \eta^{-1}D)),
\end{aligned} \tag{14}$$

and

$$\eta \sim \mathrm{gamma}(\rho/2, \rho/2). \tag{15}$$

We note that, with a common η for all i, we generate dependent samples for the b_i's. To obtain independent samples for the b_i's, we must specify a separate η_i for each b_i and take the η_i's to be i.i.d. gamma variates for $i = 1, \ldots, n$.

Once again, it may be helpful to examine the parametric random effects case where $G = \mathrm{N}(0, \eta^{-1}D)$. Here

$$p(\beta, \tau, b, \eta \mid y) \propto \exp\left\{\frac{-\tau}{2}\left[\sum_{i=1}^{n}(y_i - X_i\beta - Z_i b_i)^{\mathrm{T}}(y_i - X_i\beta - Z_i b_i) + \lambda_0\right] - \frac{1}{2}(\beta - \mu_0)^{\mathrm{T}}\Sigma_0^{-1}(\beta - \mu_0)\right\}$$

$$\cdot \exp\left\{\frac{-\eta}{2}\left[\sum_{i=1}^{n} b_i^{\mathrm{T}} D^{-1} b_i + \rho\right]\right\} \eta^{[(nv+\rho)/2]-1} \tau^{[(\alpha_0+\Sigma n_i)/2]-1}. \tag{16}$$

Note that we use a somewhat simpler generation of the multivariate t than is usually found in applications of the Gibbs sampler.

The full conditionals for β and τ in the parametric multivariate t model are the same as in the parametric normal model, but

$$[b_i \mid \eta, \beta, \tau, y] \sim \mathrm{N}\left(\tau Q_i^{\eta} Z_i^{\mathrm{T}}(y_i - X_i\beta), Q_i^{\eta}\right),$$

where $Q_i^{\eta} = (\tau Z_i^{\mathrm{T}} Z_i + \eta D^{-1})^{-1}$ and

$$[\eta \mid b, \beta, \tau, y] \sim \text{gamma}\left(\frac{\rho + nv}{2}, \frac{\rho + \sum_i b_i^{\mathrm{T}} D^{-1} b_i}{2}\right).$$

For the modified MDP model in (14) and (15), the full conditional distributions for β and τ are the same as in the normal MDP case.

From the discussion of MDP models in Section 2, we find

$$\begin{aligned} p(b_i \mid \beta, \tau, y, b_{-i}, \eta) \propto & \sum_{j \neq i} \phi(y_i \mid X_i\beta + Z_i b_j, \tau I_{n_i}) \cdot \delta_{b_j} \\ & + \left\{ M \int \phi(y_i \mid X_i\beta + Z_i b_i, \tau I_{n_i}) \phi(b_i \mid 0, \eta^{-1} D) db_i \right\} \\ & \times \phi(b_i \mid 0, \eta^{-1} D) p(y_i \mid b_i, \beta, \tau). \end{aligned}$$

After some algebraic reduction, we get

$$\begin{aligned} p(b_i \mid \beta, \tau, y, b_{-i}) \propto & \left(\sum_{j \neq i} \tau^{n_i/2} \exp\left[\frac{-\tau}{2}(y_i - X_i\beta - Z_i b_j)^{\mathrm{T}}(y_i - X_i\beta - Z_i b_j)\right] \cdot \delta_{b_j} \right) \\ & + M |Q_i^{\eta}|^{1/2} |D|^{-1/2} \tau^{n_i/2} \eta^{\rho/2} \exp\left\{\frac{\tau}{2}\left[(y_i - X_i\beta)^{\mathrm{T}} U_i^{\eta} (y_i - X_i\beta)\right]\right\} \\ & \times \phi(b_i \mid 0, D) p(y_i \mid b_i, \beta, \tau, y_j), \end{aligned} \tag{17}$$

where $U_i^{\eta} = (\tau Z_i Q_i^{\eta} Z_i^{\mathrm{T}} - I)$ and Q_i^{η} is as defined in the parametric multivariate t model. So with probability proportional to

$$\tau^{n_i/2} \exp\left(\frac{-\tau}{2}(y_i - X_i\beta - Z_i b_j)^{\mathrm{T}}(y_i - X_i\beta - Z_i b_j)\right),$$

$b_i = b_j$. And with probability proportional to

$$M |Q_i|^{1/2} |D|^{-1/2} \tau^{n_i/2} \eta^{v/2} \exp\left\{\frac{\tau}{2}\left[(y_i - X_i\beta)^{\mathrm{T}} U_i^{\eta} (y_i - X_i\beta)\right]\right\},$$

$$[b_i \mid \beta, \tau, \eta, y] \sim \mathrm{N}\left(\tau Q_i^{\eta} Z_i^{\mathrm{T}}(y_i - X_i\beta), Q_i^{\eta}\right).$$

Also,

$$p(\eta \mid b, \beta, \tau, y) \propto \eta^{[(kv+\rho)/2]-1} \exp\left\{\frac{-\eta}{2}\left[\sum_{l=1}^{k} \gamma_l^{\mathrm{T}} D^{-1} \gamma_l + \rho\right]\right\},$$

which leads to

$$[\eta \mid b, \beta, \tau, y] \sim \text{gamma}\left(\frac{\rho + kv}{2}, \frac{\rho + \sum_1^k \gamma_l^{\mathrm{T}} D^{-1} \gamma_l}{2}\right). \tag{18}$$

As in the normal case, it is desirable to introduce a Wishart prior on D^{-1}. Suppose $D^{-1} \sim \text{Wishart}(d_0, c_0 R_0)$. Then

$$\begin{aligned} p(D^{-1} \mid b, \beta, \tau, \eta, y) &= p(D^{-1} \mid \gamma, l, \beta, \tau, \eta, y) \\ &\propto |D^{-1}|^{k/2} \exp\left(\frac{-1}{2}\sum_{l}^{k} \eta \gamma_l^{\mathrm{T}} D^{-1} \gamma_l\right) \\ &\quad \times |D^{-1}|^{(d_0 - v - 1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left((c_0 R_0)^{-1} D^{-1}\right)\right), \end{aligned}$$

so that

$$\left[D^{-1} \mid b, \eta, \beta, y, \tau\right] \sim \text{Wishart}\left(d_0 + k, \left((c_0 R_0)^{-1} + \sum_{l}^{k} \eta \gamma_l \gamma_l^{\mathrm{T}}\right)^{-1}\right). \tag{19}$$

As in the normal MDP case, we will want to move around the γ_l. For the modified MDP model,

$$p(\gamma_l \mid \beta, b, \tau, D, \eta, y) \propto \phi(\gamma_l \mid 0, \eta^{-1} D) \prod_{i \in l} p(y_i \mid \beta, b, \tau, D),$$

so that

$$\gamma_l \mid \beta, b, \tau, D, y \sim \mathrm{N}\left(\tau Q_l^{\eta}\left(\sum_{i \in l} Z_i (y_i - X_i \beta), Q_l^{\eta}\right)\right), \tag{20}$$

where $Q_l^{\eta} = (\eta D^{-1} + \tau \Sigma_{i \in l} Z_i^{\mathrm{T}} Z_i)^{-1}$.

So the proposed Gibbs sampling scheme is as follows:

(0) Select starting values $b^{(0)}, \eta^{(0)}, \tau^{(0)}$, and $D^{(0)}$. Set $i = 0$.
(1) Sample $\beta^{(i+1)}$ from $[\beta \mid b^{(i)}, \tau^{(i)}, D^{(i)}, \eta^{(i)}, y]$ according to equation (7).
(2) Sample $\tau^{(i+1)}$ from $[\tau \mid \beta^{(i+1)}, b^{(i)}, D^{(i)}, \eta^{(i)}, y]$ according to equation (8).
(3.1) Sample $b_1^{(i+1)}$ from $[b_1 \mid \beta^{(i+1)}, b_{-1}^{(i)}, D^{(i)}, \tau^{(i+1)}, \eta^{(i)}, y]$ according to equation (17).
$\vdots$
(3.n) Sample $b_n^{(i+1)}$ from $[b_n \mid \beta^{(i+1)}, b_{-n}^{(i+1)}, D^{(i)}, \tau^{(i+1)}, \eta^{(i)}, y]$ according to equation (17).
(3a.1) Sample $\gamma_1^{(i+1)}$ from $[\gamma_1 \mid \beta^{(i+1)}, D^{(i)}, \tau^{(i+1)}, \eta^{(i)}, y]$ according to equation (20).
$\vdots$
(3a.k) Sample $\gamma_k^{(i+1)}$ from $[\gamma_k \mid \beta^{(i+1)}, D^{(i)}, \tau^{(i+1)}, \eta^{(i)}, y]$ according to equation (20).
(4) Sample $\eta^{(i+1)}$ from $[\eta \mid \beta^{(i+1)}, b^{(i+1)}, D^{(i)}, \tau^{(i+1)}, y]$ according to equation (18).
(5) Sample ${D^{-1}}^{(i+1)}$ from $[D^{-1} \mid \beta^{(i+1)}, b^{(i+1)}, \tau^{(i+1)}, \eta^{(i+1)}, y]$ according to equation (19).
(6) Set $i = i + 1$. Return to step (1).

4. Application

This section contains an analysis of a longitudinal repeated measurements data set. The data come from an AIDS patients study performed by the AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Disease. The data are public and come from the study labeled ACTG 116B/117. For a full description of the study, see Kahn et al. (1992).

The study was intended to test the effectiveness of switching from zidovudine (AZT) to didanosine (ddI), in light of the fact that the effectiveness of AZT is known to diminish with time. The design called for subjects who had been taking AZT for at least 4 months prior to enrollment in the study. After enrolling in the study, subjects were randomized to receive a continuation of AZT, 500 milligrams per day of ddI, or 750 milligrams per day of ddI. As part of the data collection, CD4 cell counts were measured at baseline and at 2, 8, 12, 16, and 24 weeks after enrollment in the study.

For the purposes of simplifying the example, only the AZT group and the 750-mg ddI group are included in this analysis. Additionally, subjects were further required to have been taking AZT for 6 months in order to be included. Finally, at least two CD4 count measurements were required for inclusion. These restrictions left a total of 151 individuals with 659 measurements between them.

We fit a model to the log CD4 cell counts, including an intercept (x_0), time in weeks (x_1), a treatment group indicator (x_2), and a time by treatment interaction (x_3) as fixed covariates; an intercept(z_0) and time in weeks (z_1) were included as covariates for the random effects. The time was centered around its mean in both the fixed and random covariate matrix. This was done to reduce correlation among the parameters and is recommended in Gibbs sampling for linear models to decrease autocorrelation among the samples. (See Gilks and Roberts (1996) for a further discussion.)

The data from the 500-mg ddI group, which was not used in the analysis, was used to elicit parameters for the prior distribution of the covariance matrix D in the base measure. This elicitation technique is analogous to using data from a previous experiment to specify prior distributions. (For more information about elicitation procedures, see Ibrahim and Laud (1994), Laud and Ibrahim (1995), or Laud and Ibrahim (1996).) We emphasize that the data used to generate the priors are not part of the data used in the analysis. In any case, this elicitation is for illustrative purposes.

Prior parameters for the Wishart prior on D^{-1} were obtained in the following manner. First, d_0 was chosen to be 10 and $c_0 = 1$. Then $d_0^{-1}R_0$ was taken to have the same diagonal elements as $\hat{D}^{-1}$, where $\hat{D}$ is the maximum likelihood estimate of D for the 500-mg ddI group. The off-diagonal elements of R_0 were taken to be 0. Thus,

$$R_0 = \begin{pmatrix} .0926 & 0 \\ 0 & 113.9 \end{pmatrix}.$$

Though it might be desirable in some circumstances to have a flatter prior on the covariance matrix in the base measure, this can lead to nonconvergence in parametric random effects models (see Cowles, Carlin, and Connett (1996) for an example).

Relatively flat priors were chosen for all other parameters. In particular, $\mu_0 = (0\ 0\ 0\ 0)^{\mathrm{T}}$, $\Sigma_0 = 10{,}000I_4$, $\alpha_0 = 2$, and $\lambda_0 = .0002$. A range of values for the parameter M was chosen to reflect small, moderate, and large departures from normality for the distribution of the random effects. A value of $M = 27$ reflects a small departure from normality, with the average number of clusters $\bar{k} \approx 50$. A value of $M = 5$ reflects a moderate departure from normality, with $\bar{k} \approx 25$. A value of $M = 1.25$ reflects a large departure from normality, with $\bar{k} \approx 12$.

Table 1 shows posterior medians and highest posterior density (HPD) regions for parameters in the multivariate normal parametric and MDP models with the three values of M. Table 2 reproduces these results for multivariate t parametric and modified base measure MDP models with the same values of M. Gibbs samplers were run for 22,000 iterations, with the first 500 being discarded as a burn-in period. In addition, due to high autocorrelation, every 10th iteration was used and the rest discarded, making a total sample size of 2150. Convergence of the Gibbs sampler was assessed via the Geweke (1992) method, using the CODA (Best, Cowles, and Vines, 1995) suite of diagnostics in S-plus. Most of the parameters had Geweke statistics within ± 1.96, indicating convergence is plausible. A more appropriate measure might be the technique of Raftery and Lewis (1992), which is designed for testing the convergence of estimates of percentiles. However, results from this technique depend on the amount of autocorrelation in the iterates. Thus, in this case, it indicated that infeasibly large runs of the Gibbs sampler would be necessary for convergence.

The values in Tables 1 and 2 show an unusual result. Except for β_2, the coefficient for the treatment indicator, the medians of the posterior distributions of the regression coefficients $\beta = (\beta_0\ \beta_1\ \beta_2\ \beta_3)^{\mathrm{T}}$ remain approximately the same. The reason for this deviation is shown in the density estimates presented in Figure 1. We see from this that, while there is a piece of the density that is near 0 for all values of M, as M decreases, the distribution of β_2 takes on a bimodal appearance with a great deal of weight centered around .5. While the HPD region includes zero for all the values of M used here, it is reasonable to suppose that, for a suitably small M, the analysis would indicate that a treatment difference of zero was unlikely.

This is a clear example of why it is important to correctly model the distribution of the random effects; very different results may be obtained as a result of using the MDP model rather than the fully parametric model. Thus, the model we have presented is certainly valuable in this case and may be useful in others as well. The 750-mg ddI group has been coded as 1 in this example and the AZT group as 0, so these results suggest that, in fact, switching to 750 mg of ddI may be better than staying on AZT, a fact not in evidence using a fully parametric analysis. Note that the time covariate has been centered so that this represents the group difference at the middle of the study.

Table 1

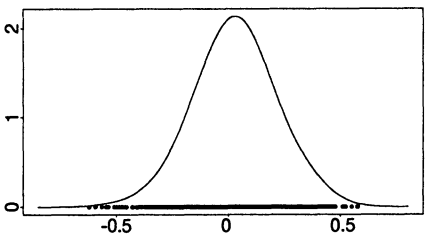
Two and one-half, 50, and 97.5 percentiles for various parameters from normal MDP models and the parametric normal model. β_0 is the intercept, β_1 is the slope over time, β_2 is the treatment effect, β_3 is the slope for the treatment–time interaction, and σ^2 is the error variance. $b_{0,i}$ is the intercept for subject i, and $b_{1,i}$ is its slope over time. $D(0,0)$ is the variance of the intercepts in the base measure, $D(1,1)$ is the variance of the slopes, and $D(0,1)$ is their covariance. $\bar{k}$ is the average number of clusters observed in the course of sampling.

	Normal MDP model			Usual RE
Parameter	$M = 1.25$	$M = 5$	$M = 27$	$M = 10^7$
β_0	(2.67, 3.52, 4.30)	(2.98, 3.63, 4.50)	(3.22, 3.74, 4.28)	(3.40, 3.67, 3.94)
β_1	(−.0488, −.0282, −.00815)	(−.0441, −.0280, −.0111)	(−.0394, −.0273, −.0148)	(−.0374, −.0276, −.0178)
β_2	(−.295, .446, .623)	(−.294, .375, .500)	(−.383, .0803, .529)	(−.331, .0293, .388)
β_3	(−.00216, .00907, .0208)	(−.00213, .00948, .0220)	(−.00365, .00992, .0229)	(−.00327, .0103, .0240)
σ^2	(.165, .188, .214)	(.158, .181, .207)	(.153, .175, .201)	(.147, .168, .192)
$b_{0,148}$	(−1.39, −.495, .534)	(−1.53, −.559, .213)	(−1.38, −.709, −.00444)	(−1.15, −.628, −.0739)
$b_{1,148}$	(−.0349, −.00786, .0226)	(−.0392, −.00963, .0261)	(−.0461, −.00852, .0329)	(−.0483, −.00486, .040)
$b_{0,149}$	(.184, .838, 1.95)	(.0804, .803, 1.57)	(.333, .887, 1.42)	(.586, 1.01, 1.42)
$b_{1,149}$	(−.0337, .00580, .0386)	(−.0324, .00424, .0362)	(−.0325, .00176, .0323)	(−.0281, .00456, .0367)
$D(0,0)$	(.839, 1.53, 3.22)	(.799, 1.37, 2.40)	(.795, 1.20, 1.86)	(.863, 1.07, 1.34)
$D(0,1)$	(−.019, .00127, .0214)	(−.0146, .000639, .0163)	(−.00959, .000822, .0103)	(−.00557, .000668, .00735)
$D(1,1)$	(.000380, .000732, .00166)	(.000375, .000684, .00173)	(.000344, .000589, .00104)	(.000364, .000576, .000903)
$\bar{k}$	12.0	21.2	48.0	151

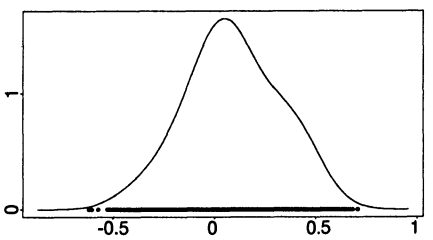
Table 2

Two and one-half, 50, and 97.5 percentiles for various parameters from the modified MDP models and the parametric multivariate t model. Parameters have the same meaning as in Table 1.

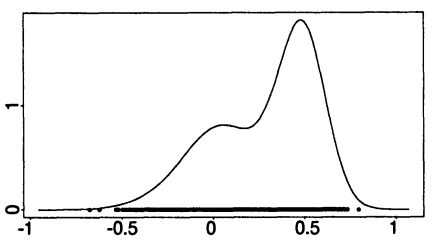
	Normal MDP model			Usual RE
Parameter	$M = 1.25$	$M = 5$	$M = 27$	$M = 10^7$
β_0	(2.75, 3.54, 4.71)	(2.90, 3.61, 4.36)	(3.36, 3.78, 4.27)	(3.43, 3.68, 3.92)
β_1	(−.0497, −.0281, −.00902)	(−.0459, −.0284, −.0121)	(−.0394, −.0276, −.0156)	(−.0369, −.0276, −.0181)
β_2	(−.300, .450, .628)	(−.337, .310, .623)	(−.422, .0593, .527)	(−.298, .0308, .380)
β_3	(−.00259, .00916, .0212)	(−.00268, .00955, .0222)	(−.00276, .0101, .0233)	(−.00293, .0104, .0235)
σ^2	(.165, .188, .217)	(.160, .181, .206)	(.154, .176, .201)	(.149, .169, .192)
$b_{0,148}$	(−1.79, −.495, .514)	(−1.48, −.576, .299)	(−1.38, −.764, −.0777)	(−1.14, −.622, −.0899)
$b_{1,148}$	(−.0355, −.00838, .0228)	(−.0385, −.00776, .0271)	(−.0412, −.00819, .0328)	(−.0498, −.00415, .0382)
$b_{0,149}$	(−.0149, .825, 1.64)	(.248, .859, 1.59)	(.342, .865, 1.36)	(.562, .995, 1.39)
$b_{1,149}$	(−.0332, .00618, .0385)	(−.0291, .00380, .0376)	(−.0302, .00150, .0334)	(−.0277, .00417, .0363)
$D(0,0)$	(.811, 1.60, 3.42)	(.822, 1.50, 2.92)	(.837, 1.45, 2.62)	(.835, 1.36, 2.18)
$D(0,1)$	(−.0195, .00157, .0232)	(−.0159, .00107, .0173)	(−.0109, .000933, .0138)	(−.00762, .000993, .00986)
$D(1,1)$	(.000383, .000744, .00161)	(.000372, .000701, .00147)	(.000368, .000646, .00122)	(.000383, .000655, .00113)
$\bar{k}$	11.9	21.5	48.1	151



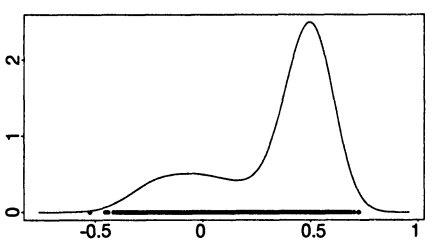
a) $M = 10^7$



b) $M = 27$



c) $M = 5$



d) $M = 1.25$

Figure 1. Posterior density of β_2 under various values of M for the normal MDP model with vague prior on β.

From the posterior distribution of the selected random effects, different facts are revealed. Not only are the HPD regions different, but the median values are very different as well. This is to be expected, as the random effects are directly affected by the relaxation of the normal assumption. Some experimentation with smaller data sets shows that subjects' random effects tend to be shrunken from OLS regression estimates for a subject toward the nearest largest mode in the distribution of the random effects. The shrinkage is more pronounced in subjects where the

deviations of the single measurements from a straight line are large. This behavior seems intuitive in light of the mixing distribution shown in equation (9). The elements of the matrix D are not easy to interpret in the MDP model since they have a very complex role in the marginal posterior distribution of the b_i.

Another conclusion may be drawn from the evidence in Tables 1 and 2. The main difference between the two tables is that the HPD regions for the random effects are wider under the modified base measure. The similarity of the results in Tables 1 and 2 indicates that, for these data, the results are not sensitive to the choice of one of these two base measures.

We now consider a more informative prior specification for β. For (β_0, β_1), we take the prior mean and covariance matrix to be the maximum likelihood estimates of these parameters using the data from the 500-mg ddI group. Since the previous study does not involve the treatments in the current study, little *a priori* information is available regarding (β_2, β_3). We therefore specify vague prior parameters for these coefficients. Specifically, we take the prior mean of (β_2, β_3) to be $(0, 0)$. This is equivalent to assuming *a priori* that the treatment makes no difference. Moreover, the prior covariance matrix for (β_2, β_3) is taken as a scalar multiple of the prior covariance matrix for (β_0, β_1). In particular, we take the multiple to be nine, which implies prior standard deviations for (β_2, β_3) that are three times as large as those of (β_0, β_1). Finally, we take the prior covariances between (β_0, β_1) and (β_2, β_3) to be zero since the 500-mg ddI group contains no information regarding these parameters. Thus, we are led to

$$\mu_0 = \begin{pmatrix} 3.7 \\ -.0205 \\ 0 \\ 0 \end{pmatrix}, \qquad \Sigma_0 = \begin{pmatrix} .0168 & .000031 & 0 & 0 \\ .000031 & .000028 & 0 & 0 \\ 0 & 0 & .15 & .000279 \\ 0 & 0 & .000279 & .000252 \end{pmatrix}.$$

Other parameters are unchanged from the previous discussion. Results from this model are presented in Table 3.

Table 3 shows results in keeping with the trends described above. The main result of the sharper prior on β is an attenuation of the width of the HPD region for all parameters. This is true for the covariate of the treatment indicator as well. The bimodal posterior distribution observed in the normal MDP model with flat priors is still present in Figure 2, but the mode near .5 is less pronounced. This is to be expected since the prior places a great deal of added weight around zero.

5. Conclusion

In this article, we have applied a general technique for Bayesian nonparametrics to an important class of models, the random effects model for longitudinal data. Our technique involved specifying a nonparametric prior for the distribution of the random effects and a Dirichlet process prior on the space of prior distributions for that nonparametric prior. Moreover, we also present the use of a modified MDP model. The resulting model was then fit with a Gibbs sampler. The proposed procedure represents a new application of the MDP model to longitudinal data.

The important contributions of this article revolve around the semiparametric model for the random effects. The implementation of the MDP model for the general random effects model has not been laid out in the detail presented here. In addition, the interpretation of the model may be of some use to applied biostatisticians, who often use the random effects model. There has been little direct focus on this model and little data analysis of the semiparametric random effects model. The modified MDP model is a new generalization, as is the computational imputation of this model. Also, the application to and discussion of an interesting data set is helpful for understanding the importance and utility of this model. Finally, we demonstrate how to make Bayesian inference for all of the parameters in our model.

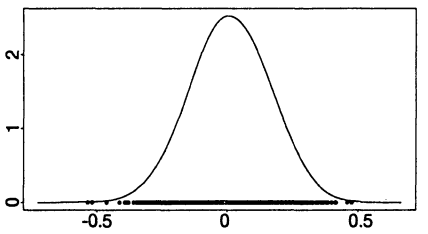
For our example, the Gibbs sampler takes about 1 hour to run for 1000 iterations on a Sparcstation 4. A potential drawback of the Gibbs sampler for this problem is that more iterations are needed for small values of M. However, reparameterization techniques, such as those suggested in Gilks and Roberts (1996) and Gelfand, Sahu, and Carlin (1995), may speed up convergence.

Future work suggested by this article includes allowing a different η_i for each i in the modified base measure and perhaps more complicated base measures, such as multivariate t or uniform base measures. More complex models, similar to those of Tsiatis et al. (1995), Mori et al. (1992), DeGruttola and Tu (1994), or Wu and Carroll (1988), would be a natural extension of this work. The MDP model for the random effects would be particularly useful in these kinds of models since they depend heavily on the random effects, which are greatly affected by the MDP model. Another field where the MDP model may prove useful is in generalized linear mixed models, where inferences may be sensitive to the distributional assumptions on the random effects.

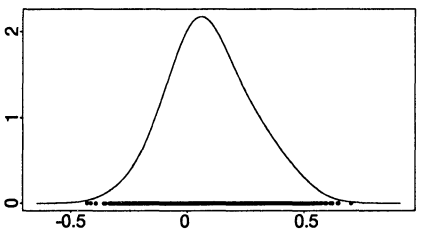
Table 3

Two and one-half, 50, and 97.5 percentiles for various parameters from the normal MDP models and the parametric normal model, with a sharp prior distribution on the β's. Parameters have the same meaning as in Table 1.

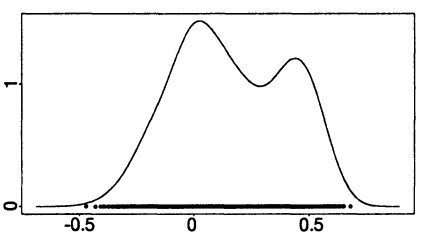
	Normal MDP model			Usual RE
Parameter	$M = 1.25$	$M = 5$	$M = 27$	$M = 10^7$
β_0	(3.45, 3.69, 3.93)	(3.45, 3.70, 3.92)	(3.48, 3.71, 3.94)	(3.51, 3.68, 3.85)
β_1	(−.0312, −.0225, −.0131)	(−.0313, −.0226, −.0138)	(−.0316, −.0233, −.0152)	(−.0309, −.0239, −.0169)
β_2	(−.308, .185, .587)	(−.266, .138, .565)	(−.239, .0827, .477)	(−.274, .00914, .286)
β_3	(−.00383, .00717.0179)	(−.00389, .00694, .0176)	(−.00472, .00675, .0178)	(−.00473, .00610, .0175)
σ^2	(.164, .189, .216)	(.159, .181, .208)	(.154, .174, .199)	(.148, .168, .192)
$b_{0,148}$	(−1.17, −.717, −.0419)	(−1.19, −.685, −.0718)	(−1.20, −.681, −.149)	(−1.15, −.631, −.139)
$b_{1,148}$	(−.0368, −.0117, .0154)	(−.0405, −.0115, .0262)	(−.0455, −.0116, .0319)	(−.0514, −.00501, .0383)
$b_{0,149}$	(.308, .802, 1.49)	(.313, .856, 1.44)	(.431, .917, 1.38)	(.616, 1.01, 1.42)
$b_{1,149}$	(−.0347, −.00274, .0256)	(−.0326, −.00131, .0304)	(−.0323, .000174, .0311)	(−.0277, .00398, .0371)
$D(0,0)$	(.867, 1.50, 3.04)	(.781, 1.34, 2.40)	(.903, 1.19, 1.85)	(.847, 1.07, 1.35)
$D(0,1)$	(−.0177, .00158, .0201)	(−.0147, .000878, .0151)	(−.00963, .000582, .0103)	(−.00586, .000761, .00733)
$D(1,1)$	(.000377, .000709, .00148)	(.000369, .000674, .00131)	(.000352, .000589, .00103)	(.000360, .000571, .000898)
$\bar{k}$	11.9	21.8	48.3	151



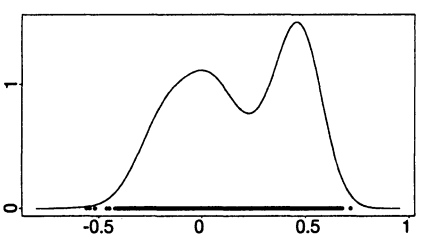
a) $M = 10^7$



b) $M = 27$



c) $M = 5$



d) $M = 1.25$

Figure 2. Posterior density of β_2 under various values of M for the normal MDP model with sharp prior on β.

Acknowledgement

This work was supported in part by NIH grant MH 17119 and NIH grant CA 70101-01.

Résumé

Dans les modèles longitudinaux à effets aléatoires, les effets aléatoires sont classiquement distribués selon une loi normale à la fois dans les modèles classiques et bayésiens. Nous proposons un modèle

bayésien, qui autorise des effets aléatoires avec des distributions a priori de Dirichlet pour la distribution des effets aléatoires; les calculs sont réalisables à l'aide de l'échantillonneur de Gibbs. Un exemple utilisant des marqueurs dans une étude sur le Sida est fourni pour illustrer la méthode.

References

Best, N. G., Cowles, M. K., and Vines, S. K. (1995). *CODA: Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output.* Version 0.3. Cambridge: MRC Biostatistics Unit.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Polya urn schemes. *Annals of Statistics* **1,** 353–355.

Bush, C. A. and MacEachern, S. N. (1996). A semi-parametric Bayesian model for randomized block designs. *Biometrika* **83,** 275–285.

Cowles, M. K., Carlin, B. P., and Connett, J. E. (1996). Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of the American Statistical Association* **91,** 86–98.

DeGruttola, V. and Tu, X. M. (1994). Modeling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics* **50,** 1003–1014.

Doss, H. (1994). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. *Annals of Statistics* **22,** 1763–1786.

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89,** 268–277.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90,** 578–588

Ferguson, T. S. (1973). A Bayesian analysis of some non-parametric problems. *The Annals of Statistics* **1,** 209–230.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika* **82,** 479–488.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics,* Volume 4, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds), 169–193. Oxford: Clarendon Press.

Gilks, W. R. and Roberts, G. O. (1996). Strategies for improving MCMC. In *Markov Chain Monte Carlo in Practice,* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (eds), 89–114. London: Chapman and Hall.

Ibrahim, J. G. and Laud, P. W. (1994). A predictive approach to the analysis of designed experiments. *Journal of the American Statistical Association* **89,** 309–319.

John. (1971). *Statistical Design and Analysis.* New York: MacMillan.

Kahn, J. O., Lagakos, S. W., Richman, D. D., et al. (1992). A controlled trial comparing zidovudine with didanosine in human immunodeficiency virus infection. *New England Journal of Medicine* **327,** 581–587.

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

Laud, P. W. and Ibrahim, J. G. (1995). Predictive model selection. *Journal of the Royal Statistical Society, Series B* **57,** 247–262.

Laud, P. W. and Ibrahim, J. G. (1996). Predictive specification of prior model probabilities in variable selection. *Biometrika* **83,** 267–274.

Liu, J. (1996). Nonparametric hierarchical Bayes via sequential imputation. *Annals of Statistics* **24,** 911–930.

MacEachern, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics* **23,** 727–741.

MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics,* in press.

Mori, M., Woodworth, G. G., and Woolson, R. F. (1992). Application of empirical Bayes inference to estimation of rate of change in the presence of informative right censoring. *Statistics in Medicine* **11,** 621–631.

Müller, P., Erkanli, A., and West, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* **83,** 67–79.

Raftery, A. L. and Lewis, S. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics,* Volume 4, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds), 763–773. Oxford: Clarendon Press.

Tsiatis, A. A., DeGruttola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90,** 27–37.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* **91,** 217–221.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., and Gelfand, A. E. (1994). Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. *Applied Statistics* **43,** 201–221.

Walker, S. G. and Damien, P. (1997). *Sampling a Dirichlet process mixture model.* Technical Report, University of Michigan Business School, Ann Arbor.

West, M., Müller, P., and Escobar, M. D. (1994). Hierarchical priors and mixture models, with applications in regression and density estimation. In *Aspects of Uncertainty: A Tribute to D. V. Lindley,* A. F. M. Smith and P. R. Freeman (eds). London: Wiley.

Wilks, W. R., Wang, C. C., Yvonnet, B., and Coursaget, P. (1993). Random effects models for longitudinal data using Gibbs sampling. *Biometrics* **49,** 441–453.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44,** 175–188.

Received August 1996; *revised July* 1997; *accepted September* 1997.