

# Posterior Simulation across Nonparametric Models for Functional Clustering

Jamie L. Bigelow<sup>1</sup> and David B. Dunson<sup>1,2</sup>

<sup>1</sup>Department of Statistical Science

Duke University

<sup>2</sup>National Institute of Environmental Health Sciences

Biostatistics Branch, MD A3-03

P.O. Box 12233

Research Triangle Park, NC 27709

USA

February 7, 2008

SUMMARY. By choosing a species sampling random probability measure for the distribution of the basis coefficients, a general class of nonparametric Bayesian methods for clustering of functional data is developed. Allowing the basis functions to be unknown, one faces the problem of posterior simulation over a high-dimensional space of semi-parametric models. To address this problem, we propose a novel Metropolis-Hastings algorithm for moving between models, with a nested generalized collapsed Gibbs sampler for updating the model parameters. Focusing on Dirichlet process priors for the distribution of the basis coefficients in multivariate linear spline models, we apply the approach to the problem of clustering of hormone trajectories. This approach allows the number of clusters and the shape of the trajectories within each cluster to be

---

*email:* [jbigelow@duke.edu](mailto:jbigelow@duke.edu)

unknown.

KEY WORDS: Dirichlet process; Functional data analysis; Latent class; Model averaging; Nonparametric Bayes; RJMCMC; Species sampling

## 1. Introduction

Semiparametric Bayes models have been increasingly used in applications in recent years, due largely to substantial improvements in the simplicity and efficiency of computational algorithms. For example, it has become common to use Dirichlet process (DP) priors (Ferguson, 1973; Ferguson, 1974) to allow for uncertainty in specifying distributions within a hierarchical Bayesian model. Such DP mixture (DPM) models (Antoniak, 1974; Escobar and West, 1995) can be implemented routinely using a variety of Markov chain Monte Carlo (MCMC) algorithms (MacEachern, 1994; MacEachern and Müller, 1998; Neal, 2000; Jain and Neal, 2004; Papaspiliopoulos and Roberts, 2006). For some recent applications of DPMs, refer to Kottas et al., 2002; Laws and O’Hagan, 2002; Griffin and Steel, 2004; Dunson et al., pear; Sha et al., 2006; Xue et al., 2006.

Although there is an increasingly rich literature on semiparametric Bayes methods and applications, few approaches are available for accommodating uncertainty in specifying such models. An important article in this area is Basu and Chib, 2003, who propose an approach for calculating marginal likelihoods and Bayes factors for DPM models. This approach is useful when the focus is on comparing a small number of competing models, some of which have DP components. Cai and Dunson, 2005 instead develop a method for selection of fixed and random effects in a linear mixed model having a nonparametric prior on the random effects distributions.

Our focus is on developing Bayesian methods for functional clustering. A useful

approach for characterizing functional data is to express each function as a linear combination of basis functions, with the basis coefficients then assigned a prior distribution. In the literature, this prior distribution is often chosen to be Gaussian. For related approaches, refer to Brumback and Rice, 1998; Rice and Wu, 2001; Ke and Wang, 2001; James and Sugar, 2003. Assuming a common set of basis functions, subjects with identical or similar basis coefficients can be clustered together (James and Sugar (2003), Ma et al. (2005)).

James and Hastie (2001) apply linear discriminant analysis to divided functional data into classes, where the number (and sometimes nature) of classes is pre-determined. de la Cruz and Quintana (2005) use Bayesian methods and Marshall and Barón (2000) develop a mixed effects model for classification of hormone trajectories into pre-defined groups. With the goal of identifying Olympic athletes who use growth hormone injections, Brown et al. (2001) develop a Bayesian method which defines trajectory classes based on a training dataset with known classification. Muthén and Shedden (1999) use an EM algorithm to identify trajectories in young adult drinking behaviour that are likely to lead to alcohol dependence. Heard et al. (2006) apply Bayesian hierarchical clustering to cluster curves in a gene regulation application. Ray and Mallick (2006) use a DP prior in a wavelet model for functional clustering, exploiting the discreteness property of the DP.

These methods assume that the basis functions are prespecified. As noted by many authors, the assumption of a fixed basis is often insufficiently flexible, motivating a literature on adaptive regression splines (DiMatteo et al., 2001; Kass et al., 2003). It is particularly important to use adaptive methods that allow the number and locations of knots to be unknown in multivariate settings, because the number of prespecified knots that need to be included for sufficient flexibility increases dramatically with number

of predictors. Holmes and Mallick (2001) propose Bayesian regression with adaptive multivariate linear splines (see also Hansen and Kooperberg, 2002; Wood et al., 2002; Holmes and Mallick, 2002). Bigelow and Dunson (2007) generalize this approach to hierarchical functional data assuming a normal distribution on the basis coefficients.

Motivated by applications to clustering of hormone trajectories in the menstrual cycle, we replace the normality assumption with a species sampling (SS) prior (Pitman, 1995; Pitman, 1996; Ishwaran and James, 2003a) on the distribution of the basis coefficients. The SS formulation encompasses a very broad class of random probability measures, including the DP. If the basis functions were known, such an approach would be straightforward to implement following Ishwaran and James (2003a), being equivalent to placing a SS prior on the random effects distribution in a linear mixed model. The DP special case has been considered by many authors (Bush and MacEachern, 1996; Mukhopadhyay and Gelfand, 1997; Kleinman and Ibrahim, 1998; Ishwaran and Takahara, 2002). However, when the basis functions are unknown, we are faced with the problem of posterior simulation over a high-dimensional list of competing semiparametric Bayes models.

To address this problem, we propose a novel Metropolis-Hastings transition probability for moving between models, with a nested Gibbs sampler to update parameters within a model. This general algorithm is applied to the problem of posterior simulation over species sampling models for functional data, taking advantage of properties of the SS prior to construct an efficient algorithm.

Section 2 describes the space of models under consideration. Section 3 proposes the Metropolis-Hastings with nested Gibbs algorithm. Section 4 outlines the steps of the algorithm in the functional data application. Section 5 contains a simulation example, Section 6 applies the approach to progesterone data, and Section 7 discusses

the results.

## 2. Species Sampling Priors for Functional Clustering

### 2.1 Models for samples of curves

Data for subject  $i$  ( $i = 1, \dots, n$ ) consist of  $\mathbf{y}_i = (y_{i1}, \dots, y_{i,n_i})$ , a vector of  $n_i$  error-prone observations of an unknown function  $f_i : \mathcal{X} \rightarrow \mathfrak{R}$ , along with an  $n_i \times p$  matrix of covariates,  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{i,n_i})'$ . In particular, the  $j$ th observation for subject  $i$ ,  $y_{ij}$ , has corresponding covariates  $\mathbf{x}_{ij} \in \mathcal{X}$ , where  $\mathcal{X} \subset \mathfrak{R}^p$ . We assume the following measurement structure:

$$y_{ij} = f_i(\mathbf{x}_{ij}) + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \tau^{-1}), \quad (1)$$

where  $\epsilon_{ij}$  is a normally distributed measurement error. Our focus is on flexible methods for characterizing the collection of functions  $\{f_1, \dots, f_n\}$ , allowing clustering of different subjects and borrowing of information across subjects. Although we focus on the form in (1), it is straightforward to apply the methods when (i) a parametric adjustment for additional covariates,  $\mathbf{z}'_i \boldsymbol{\beta}$ , is added to the mean; (ii)  $f_i(\cdot)$  is expressed in additive form as a sum of unknown 1-dimensional functions for each predictor; or (iii) model (1) is embedded in a larger hierarchical model.

Let  $\boldsymbol{\mu}_M = \{\mu_{M1}, \dots, \mu_{Mk_M}\}$  denote a set of  $k_M$  basis functions, with  $\mu_{Mh} : \mathcal{X} \rightarrow \mathfrak{R}$ , for  $h = 1, \dots, k_M$ , and  $M \in \mathcal{M}$ , with  $\mathcal{M}$  denoting the model space. Then, under model  $M$ , we let

$$\begin{aligned} f_i(\mathbf{x}_{ij}) &= \sum_{h=1}^{k_M} b_{Mih} \mu_{Mh}(\mathbf{x}_{ij}) = \mathbf{H}_{Mi} \mathbf{b}_{Mi} \\ \mathbf{b}_{Mi} &\sim G_M, \end{aligned} \quad (2)$$

where  $\mathbf{b}_{Mi} = (b_{Mi1}, \dots, b_{Mik_M})'$  are basis coefficients, and  $G_M$  is a probability measure characterizing the distribution of these basis coefficients. When the basis functions to

be included are known, so  $M$  is known, expression (2) is a typical linear hierarchical model. In this case, curves for different subjects,  $i$  and  $i'$ , are clustered together,  $f_i = f_{i'}$ , if and only if the basis coefficients are equal,  $\mathbf{b}_{Mi} = \mathbf{b}_{Mi'}$  (for typical families of basis functions). In the setting of wavelet-based functional models, Ray and Mallick (2006) incorporated a DP prior on the coefficients to allow clustering.

## 2.2 Species sampling priors

Treating  $G_M$  as unknown, we let  $G_M \sim \mathcal{P}_M$ , where  $G_M$  is a probability measure on  $(\mathfrak{R}^{k_M}, \mathcal{B})$  and  $\mathcal{P}_M$  is a probability measure on the space of probability measures on  $(\mathfrak{R}^{k_M}, \mathcal{B})$ , with  $\mathcal{B}$  the Borel  $\sigma$ -algebra of subsets of  $\mathfrak{R}^{k_M}$ . Hence, it remains to specify  $\mathcal{P}_M$ , with one convenient choice corresponding to the Dirichlet process (DP). As a broader class of random probability measures, we focus on the species sampling models (SSMs) described by Ishwaran and James (2003a), which characterize  $\mathcal{P}_M$  as follows:

$$\mathcal{P}_M(\cdot) = \sum_h p_{Mh} \delta_{\Theta_{Mh}}(\cdot) + \left(1 - \sum_h p_{Mh}\right) G_{M0}(\cdot), \quad (3)$$

where  $0 < p_{Mh} < 1$  are random probability weights with  $\sum_h p_{Mh} \leq 1$ ,  $\delta_{\Theta}$  is a probability measure concentrated at  $\Theta$ ,  $\Theta_{Mh} \stackrel{iid}{\sim} G_{M0}$ , and  $G_{M0}$  is a nonatomic distribution over  $(\mathfrak{R}^{k_M}, \mathcal{B})$ . The Dirichlet process prior with precision  $\alpha$  and base distribution  $G_{M0}$  is obtained in the special case in which  $p_{Mh} = V_{Mh} \prod_{l < h} (1 - V_{Ml})$  and  $V_{Mh} \stackrel{iid}{\sim} \text{beta}(1, \alpha)$ , for  $h = 1, \dots, \infty$ , so that  $\sum_h p_{Mh} = 1$  a.s. and the second term in (3) drops out. Some other special cases include the Pitman-Yor process (Pitman and Yor (1997)), finite dimensional Dirichlet priors (Ishwaran and Zarepour (2002b), Ishwaran and Zarepour (2002a)), and stick-breaking measures (Ishwaran and James (2001)).

Assuming that  $\mathbf{b}_{Mi} \stackrel{iid}{\sim} G_M$ , with  $G_M \sim \mathcal{P}_M$  and  $\mathcal{P}_M$  following expression (3), one

obtains the following prediction rule upon marginalizing out  $G_M$  (Pitman (1996)):

$$P\{\mathbf{b}_{M1} \in \cdot\} = G_{M0}(\cdot), \quad P\{\mathbf{b}_{M,i+1} \in \cdot \mid \mathbf{b}_{M1}, \dots, \mathbf{b}_{Mi}\} = w_{0i}G_{M0}(\cdot) + \sum_{h=1}^{r(i)} w_{hi}\delta_{\Theta_{Mh}^*}(\cdot), \quad (4)$$

where  $w_{0i}, w_{hi}$  are non-negative measurable functions of  $\{\mathbf{b}_{Mh}, h = 1, \dots, i\}$ , and the atoms  $\{\Theta_{M1}^*, \dots, \Theta_{Mr(i)}^*\}$  correspond to the unique elements of  $\{\mathbf{b}_{M1}, \dots, \mathbf{b}_{Mi}\}$  in the order that they appear. The elements of  $\{w_{hi}, h = 0, \dots, r(i)\}$  sum to one and are assumed to satisfy the conditions described by Hansen and Pitman (2000) to ensure exchangeability of  $\mathbf{b}_{M1}, \mathbf{b}_{M2}, \dots$ . In the special case in which  $w_{0i} = \alpha/(\alpha + i)$  and  $w_{hi} = n_{hi}/(\alpha + i)$ , with  $n_{hi} = \sum_{h=1}^i \mathbf{1}(\mathbf{b}_{Mh} = \Theta_h^*)$  denoting the number of previous subjects assigned to cluster  $h$ , expression (4) reduces to the Dirichlet process prediction rule or Pólya urn scheme of Blackwell and MacQueen (1973).

Expression (4) can be used to characterize the joint marginal law of the basis coefficients for the different subjects under model  $M$ . However, from Pitman (1996), the joint distribution of  $\mathbf{b}_M = \{\mathbf{b}_{M1}, \dots, \mathbf{b}_{Mn}\}$  can also be characterized in terms of the unique values and an exchangeable partition probability function (EPPF) governing the random allocation of subjects to these values. Let  $\mathbf{p}_M = \{\mathcal{I}_{M1}, \dots, \mathcal{I}_{M, n(\mathbf{p})}\}$  denote the partition of  $\{1, \dots, n\}$ , where  $\mathcal{I}_{Mh} = \{l : \mathbf{b}_{Ml} = \Theta_{Mh}^*\}$  and  $n(\mathbf{p}_M)$  is the number of unique values. Then, following Pitman (1996) and Ishwaran and James (2003a),

$$P\{\mathbf{p}_M, \Theta_{Mh}^* \in B_h, h = 1, \dots, n(\mathbf{p}_M)\} = \pi(\mathbf{p}_M) \prod_{h=1}^{n(\mathbf{p}_M)} G_{M0}(B_h), \quad (5)$$

where  $\pi(\mathbf{p}_M)$  is the EPPF, which depends only on the cardinalities of the sets  $\mathcal{I}_{Mh}$ .

Note that each unique value,  $\Theta_{Mh}^*$ , corresponds to a functional cluster, as subjects allocated to this unique value will have identical basis coefficients and hence identical curves. The unique values are generated independently from  $G_{M0}$ , while the EPPF,  $\pi(\mathbf{p}_M)$ , induces a prior on the number and relative sizes of the functional clusters.

This implies that all SSMs draw the unique values by iid sampling from  $G_{M0}$ , so that different specifications of the functions,  $w_{0i}, w_{hi}$ , only impact the prior on the assignment of individuals to clusters. Although general characterizations of the EPPF are very difficult, it is straightforward to sample from  $\pi(\mathbf{p}_M)$  using (4).

Ishwaran and James (2003a) provides a number of alternative Monte Carlo algorithms for posterior computation in species sampling mixture models. These algorithms can be used directly when the basis functions are known. In the DP mixture model special case, the collapsed Gibbs sampler of MacEachern (1994) has been widely used due to simplicity of implementation and efficiency in most cases. One of the algorithms proposed by Ishwaran and James (2003a) generalized the collapsed Gibbs sampler from DP to SS mixture models. In considering extensions to unknown  $M$ , we focus on this algorithm.

Let  $\Theta_M^* = \{\Theta_{M1}^*, \dots, \Theta_{Mr_M}^*\}$  denote the  $r_M$  unique values of  $\mathbf{b}_M$ , and let  $\mathbf{S}_M = (S_{M1}, \dots, S_{Mn})'$ , with  $S_{Mi} = h$  if  $\mathbf{b}_{Mi} = \Theta_{Mh}^*$ . The generalized collapsed Gibbs sampler alternates between (i) updating the allocation of subjects to clusters,  $\mathbf{S}_M$ , and number of clusters,  $r_M$ , by sampling from the conditional posterior distribution of  $S_{Mi}$ , for  $i = 1, \dots, n$ ; and (ii) updating the unique value,  $\Theta_{Mh}^*$ , for  $h = 1, \dots, r_M$ , through sampling from the conditional posterior, which is proportional to the prior,  $G_{M0}(\Theta_{Mh}^*)$ , multiplied by the likelihood contribution for those subjects allocated to cluster  $h$ .

### 2.3 *Uncertainty in basis functions*

In many cases, the assumption of a known set of basis functions is insufficiently flexible, and there is uncertainty in  $M$ . For example, we are interested in developing models for reproductive hormone data in menstrual cycles incorporating information on timing relative to the cycle start and relative to ovulation, in addition to other covariates. Potentially, a multivariate linear spline basis can be used. However, the

number and location of the knots will be unknown. Although one can potentially choose a dense grid of knots, this would involve a large number of basis functions, leading to an increased computational burden and problems with overfitting.

For this reason, a number of authors have considered the basis functions to be unknown (Biller (2000), Holmes and Mallick (2001), Hansen and Kooperberg (2002), Wood et al. (2002), Holmes and Mallick (2002), Holmes and Mallick (2003), Lindstrom (2002), among others). However, previous adaptive methods either assume fixed basis coefficients for each set of basis functions or model the distribution of the basis coefficients parametrically. RJMCMC (Green (1995)) is the standard approach for posterior simulation over the model space in this setting.

In order to allow for uncertainty in basis function selection, while treating the distribution of the basis coefficients as unknown using the species sampling approach of Section 2.2, we propose a different approach. To complete a Bayesian specification, we need to choose (1) a model space  $\mathcal{M}$  of different possibilities for the basis functions, and (2) a prior over  $\mathcal{M}$ . In applications, focus on multivariate linear spline basis functions, with the prior probability of model  $M$  proportional to a prior on the number of basis functions,  $p(k_M) = \binom{T}{k_M-1}^{-1} K^{-1}$ , where  $K$  is the maximum number possible and  $T$  is very large. However, our approach applies to arbitrary choices of  $\mathcal{M}$  and  $p(M)$ .

### 3. Metropolis-Hastings with Nested Gibbs Algorithm

#### 3.1 Background

To construct an MCMC algorithm for posterior simulation over  $\mathcal{M}$ , we propose a Metropolis-Hastings with nested Gibbs algorithm. The Metropolis-Hastings step moves between models and the Gibbs algorithm conducts posterior simulation within a model. We focus first on the Metropolis-Hastings step, which involves specification of (1)  $S(M', M)$ , the probability of proposing a transition from the current model  $M$

to  $M'$ ; and (2)  $Q(M', M)$ , the probability of accepting the proposal.

The choice of  $S(M', M)$  can follow along standard lines, as in Holmes and Mallick (2001), so we focus on  $Q(M', M)$ . Following Hastings (1970), this probability is

$$Q(M', M) = \min \left\{ 1, \frac{p(\mathbf{y} | M')p(M')S(M, M')}{p(\mathbf{y} | M)p(M)S(M', M)} \right\}, \quad (6)$$

where  $p(\mathbf{y} | M)$  is the marginal likelihood under model  $M$ . Through careful specification of the prior on the model space,  $R = \frac{p(M')S(M, M')}{p(M)S(M', M)}$  can be treated as a known constant (Holmes and Mallick, 2000, Denison et al., 2002). Unfortunately,  $p(\mathbf{y} | M)$  does not have a closed form and the Bayes factor  $p(\mathbf{y} | M)/p(\mathbf{y} | M')$  is intractable, so (6) is also intractable.

We consider alternative acceptance probabilities to (6), which can be calculated easily.

### 3.2 Proposed Metropolis-Hastings transition probability

Although  $p(\mathbf{y} | M)$  is unavailable, we can often find a tractable conditional likelihood of the form  $p(\mathbf{y} | M, \mathcal{V})$ , where  $\mathcal{V}$  is a set of model parameters. When  $\mathcal{V}$  are any parameters common to both  $M$  and  $M'$ , expression (6) can be written:

$$Q(M', M) = \min \left[ 1, \frac{p(\mathbf{y} | M', \mathcal{V})p(\mathcal{V} | M')p(\mathcal{V} | M, \mathbf{y})}{p(\mathbf{y} | M, \mathcal{V})p(\mathcal{V} | M)p(\mathcal{V} | M', \mathbf{y})} \times R \right] \quad (7)$$

using the candidate's formula  $p(\mathbf{y} | M, \mathcal{V})p(\mathcal{V} | M)/p(\mathcal{V} | M, \mathbf{y})$  for the marginal  $p(M | \mathbf{y})$ . The ratio  $\frac{p(\mathcal{V} | M')p(\mathcal{V} | M, \mathbf{y})}{p(\mathcal{V} | M)p(\mathcal{V} | M', \mathbf{y})}$  may be difficult to calculate. A common approach (Holmes and Mallick, 2001; Denison et al., 2002) is to instead use the acceptance probability:

$$Q_2(M', M) = \min \left[ 1, \frac{p(\mathbf{y} | M', \mathcal{V})}{p(\mathbf{y} | M, \mathcal{V})} \times R \right] \quad (8)$$

*Theorem 1.* The acceptance probability  $Q_2$  leads to samples from  $p(M | \mathcal{V}, \mathbf{y})$  if  $p(\mathcal{V} | M)$  is constant across all  $M \in \mathcal{M}$ .

*Proof.* Without loss of generality, for two models  $M$  and  $M'$  in  $\mathcal{M}$ :

$$Q_2(M', M) = \frac{p(\mathbf{y}|M', \mathcal{V})}{p(\mathbf{y}|M, \mathcal{V})} \times R = \frac{p(\mathbf{y}|M', \mathcal{V})p(M')S(M, M')}{p(\mathbf{y}|M, \mathcal{V})p(M)S(M', M)}$$

$$Q_2(M, M') = 1$$

Using the detailed balance equations to find the stationary distribution,  $\pi(M)$  we get:

$$\pi(M)S(M', M) \frac{p(\mathbf{y}|M', \mathcal{V})p(M')S(M, M')}{p(\mathbf{y}|M, \mathcal{V})p(M)S(M', M)} = \pi(M')S(M, M') \quad \forall M \in \mathcal{M}$$

$$\frac{\pi(M)}{p(\mathbf{y}|M, \mathcal{V})p(M)} = \frac{\pi(M')}{p(\mathbf{y}|M', \mathcal{V})p(M')} \quad (9)$$

Solving these equations for  $\pi(M)$ , we get the following limiting distribution:

$$\pi(M) = \frac{p(M|\mathbf{y}, \mathcal{V})}{p(\mathcal{V}|M)} \quad \forall M \in \mathcal{M}$$

If  $p(\mathcal{V}|M)$  is constant across all  $M \in \mathcal{M}$ , then  $\pi(M) = p(M|\mathbf{y}, \mathcal{V})$ .  $\square$

There is a more general valid acceptance probability, where the priors on  $\mathcal{V}$  over the model space need not be equal, but the ratio of the priors under two models must be computed.

*Theorem 2.* If the ratio  $\frac{p(\mathcal{V}|M')}{p(\mathcal{V}|M)}$  is known for any  $\{M, M'\}$  in  $\mathcal{M}$ , then the following acceptance probability satisfies the detailed balance equations for  $\pi(M) = p(M|\mathbf{y}, \mathcal{V})$ :

$$Q_3(M', M) = \min \left\{ \left[ 1, \frac{p(\mathbf{y}|M', \mathcal{V})p(\mathcal{V}|M')}{p(\mathbf{y}|M, \mathcal{V})p(\mathcal{V}|M)} \times R \right] \right\} \quad (10)$$

$\square$

Theorem 2 is easily verified by modifying the proof of Theorem 1. The acceptance probability in  $Q_3$  has the advantage over  $Q_2$  that it allows conditioning on parameters that have different priors for the different models in  $\mathcal{M}$ . Although we are motivated by the nonparametric functional clustering application, this acceptance probability represents a novel Metropolis-Hastings algorithm for posterior simulation over model spaces, which should have broad utility.

#### 4. Posterior Simulation over SSMs for Functional Data

Assume that the data are generated from the hierarchical functional model described in (1) and (2), with  $G_M \sim \mathcal{P}_M$  and  $\mathcal{P}_M$  a species sampling prior having

$$\begin{aligned} G_{M0} &= N_{k_M}(\boldsymbol{\beta}_M, \tau_M^{-1} \boldsymbol{\Delta}_M^{-1}) \\ \boldsymbol{\beta}_M &\sim N_{k_M}(\mathbf{0}, \tau_M^{-1} \lambda_M^{-1} \mathbf{I}_{k_M}) \\ \pi(\tau_M, \lambda_M, \boldsymbol{\delta}_M) &\propto \tau_M^{a_\tau - 1} \exp(-b_\tau \tau_M) \lambda_M^{a_\lambda - 1} \exp(-b_\lambda \lambda_M) \prod_{l=1}^{k_M} (\delta_{Ml}^{a_\delta - 1} \exp(-b_\delta \delta_{Ml})) \end{aligned} \quad (11)$$

where  $\boldsymbol{\Delta}_M = \text{diag}(\boldsymbol{\delta}_M)$  and  $a_\tau$ ,  $b_\tau$ ,  $a_\lambda$ ,  $b_\lambda$ ,  $a_\delta$  and  $b_\delta$  are pre-specified hyperparameters constant across models. In sensitivity analyses, we found posterior inferences robust to the choice of these hyperparameters.

Conditionally on  $\boldsymbol{\delta}_M$ ,  $\lambda_M$ , and  $\mathbf{S}_M$ , the allocation of subjects to unique values of the basis coefficients, it follows from (1), (2), (5) and (11) after some algebra that

$$p(\mathbf{y}|M, \boldsymbol{\delta}_M, \lambda_M, \mathbf{S}_M) = C(\lambda_M, k_M) |\mathbf{R}_M|^{-\frac{1}{2}} (b_\tau + \frac{A_M}{2})^{-(\frac{N}{2} + a_\tau)} \prod_{l=1}^{k_M} \delta_{Ml}^{r_M/2} \prod_{j=1}^{r_M} |\mathbf{U}_{Mj}|^{-\frac{1}{2}} \quad (12)$$

where this conditional likelihood under model  $M$  is a closed form expression of

$$\begin{aligned} \mathbf{U}_{Mj} &= \sum_{i \in \mathcal{I}_j} (\Delta_M + \mathbf{H}'_{Mj} \mathbf{H}_{Mj}) \text{ for } j = 1, \dots, r_M \\ \mathbf{R}_M &= \lambda_M \mathbf{I}_{k_M} + r_M \boldsymbol{\Delta}_M - \boldsymbol{\Delta}_M \left( \sum_{j=1}^r \mathbf{U}_{Mj}^{-1} \right) \boldsymbol{\Delta}_M \\ A_M &= \mathbf{y}' \mathbf{y} - \sum_{j=1}^r \left\langle \sum_{i \in \mathcal{I}_{Mj}} \mathbf{H}'_{Mi} \mathbf{y}_i, \mathbf{U}_{Mj}^{-1} \right\rangle - \left\langle \boldsymbol{\Delta}_M \sum_{j=1}^{r_M} \mathbf{U}_{Mj}^{-1} \mathbf{H}'_{Mi} \mathbf{y}_i, \mathbf{R}_M^{-1} \right\rangle \\ C(\lambda_M, k_M) &= \frac{b_\tau^{a_\tau} \lambda_M^{\frac{k_M}{2}} \Gamma(\frac{N}{2} + a_\tau)}{\Gamma(a_\tau) (2\pi)^{\frac{N}{2}}} \end{aligned}$$

with  $\langle A, B \rangle$  denoting the quadratic form  $A'BA$ .

Our Metropolis with nested Gibbs sampler alternates between the following steps: (1) propose a move from model  $M$  to  $M'$  according to  $S(M', M)$ ; (2) accept this move with probability  $Q_3(M', M)$ ; and (3) update the unknowns within the current model, including the configuration,  $\mathbf{S}_M$ , the unique values,  $\Theta_M^*$ , the parameters within  $G_{M0}$ , and the residual precision,  $\tau_M$ , using the generalized collapsed Gibbs sampler (Ishwaran and James (2003a)). In calculating  $Q_3(M', M)$ , we marginalize out parameters that are not shared by  $M$  and  $M'$ , avoiding the need to explicitly consider a change in dimension. In typical implementations, there is at most one unshared parameter, so such marginalization involves approximation of a one-dimensional integral. When a proposal to move from  $M$  to  $M'$  is accepted, we initialize the (at most one) unshared parameter to a reasonable starting value to begin the nested Gibbs chain.

Calculate of  $Q_3(M', M)$  in this case requires some additional explanation. For the conditions of Theorem 2 to hold,  $\mathcal{V}$  must be parameters that are defined under both models. Hence, we let  $\mathcal{V} = \{\mathbf{S}, \boldsymbol{\delta}_{M \cap M'}, \lambda\}$ , where  $\boldsymbol{\delta}_{M \cap M'}$  corresponds to the subset of  $\boldsymbol{\delta}$  corresponding to basis functions common to both  $M$  and  $M'$ . Then, we have

$$Q_3(M', M) = \min \left\{ 1, \frac{p(\mathbf{y}|M', \mathbf{S}_M, \boldsymbol{\delta}_{M \cap M'}, \lambda_M)p(\mathbf{S}_M, \boldsymbol{\delta}_{M \cap M'}, \lambda_M|M')}{p(\mathbf{y}|M, \mathbf{S}_M, \boldsymbol{\delta}_{M \cap M'}, \lambda_M)p(\mathbf{S}_M, \boldsymbol{\delta}_{M \cap M'}, \lambda_M|M)} \times R \right\}. \quad (13)$$

Assuming  $M$  and  $M'$  differ by at most one basis function, the conditional likelihoods are either available from (12) or are available up to a one-dimensional integral, which can easily be evaluated numerically or through a Laplace approximation. We find that this Laplace approximation is highly accurate in the cases we have considered.

Calculate of  $Q_3(M', M)$  also requires specification of the prior ratio:

$$\frac{p(\mathbf{S}_M, \boldsymbol{\delta}_{M \cap M'}, \lambda_M|M')}{p(\mathbf{S}_M, \boldsymbol{\delta}_{M \cap M'}, \lambda_M|M)} = \frac{p(\mathbf{S}_M|M')p(\boldsymbol{\delta}_{M \cap M'}|M')p(\lambda_M|M')}{p(\mathbf{S}_M|M)p(\boldsymbol{\delta}_{M \cap M'}|M)p(\lambda_M|M)} = \frac{p(\mathbf{S}_M|M')}{p(\mathbf{S}_M|M)},$$

so that it remains to calculate the ratio of prior probabilities of the cluster allocation  $\mathbf{S}_M$  under models  $M$  and  $M'$ . Fortunately, it follows directly from (5) that  $\frac{p(\mathbf{S}_M|M')}{p(\mathbf{S}_M|M)} = 1$ .

## 5. Simulations

### 5.1 *Simulated data*

To illustrate the method we focus on the Dirichlet process special case, using trajectories simulated from four clusters, with 25 subjects per cluster and 10 observations per subject. There were two covariates. One can be thought of as time and consisted of randomly generated continuous values from 0 to 5. The other was a time-dependent covariate, which increased linearly with time. Each subject's trajectory is generated from a linear combination of this continuous covariate and one of four smooth functions of time. These four smooth functions, along with the data points generated for each cluster, are shown in Figure 1. Within-subject dependence was induced by adding a subject-specific random effect and a random residual error to each observation.

[Figure 1 about here.]

Escobar (1994) demonstrates that high values of  $\alpha$  yield large numbers of clusters. For interpretability, we wish to have only a small number of clusters. Based on the guidelines given in Escobar (1994), we chose  $\alpha = 0.5$ . The hyperparameters for the gamma priors on the precision parameters were all set to 0.05, yielding somewhat vague priors centred at 1. Sensitivity analyses conducted for a wide range of alternative priors for the precision parameters showed results to be robust, with only the posterior summary of  $\lambda$  changing appreciably.

We ran the algorithm in Matlab for 20,000 iterations after a 5,000 iteration burn-in period. Sample collection took approximately 10 hours. At each iteration, we collected the cluster membership for each subject, the individual cluster trajectories, and the population mean trajectory.

To identify underlying groups from our set of 20,000 sampled cluster structures, we use a hierarchical clustering algorithm. Following Medvedovic and Sivaganesan

(2002), who searched for clusters of co-expressed genes in microarray data, we define the total distance between two vector observations to be the proportion of MCMC samples at which two observations were put in different clusters. Observations separated by small distances are often grouped together and thus are more likely to truly belong to the same cluster. Heard et al. (2006) also conduct Bayesian hierarchical clustering of curves, but their method does not allow uncertainty in the basis functions characterizing the curves.

To cluster the trajectories, we require that any observations separated by less than some threshold distance be grouped together. This threshold is the minimum distance allowed between two observations in separate groups. A high threshold yields few groups, and a low threshold will result in many groups. The choice of the threshold may be largely driven by the application. Hierarchical clustering lends itself readily to the creation of a cluster tree, a visual representation of how tight the clusters are and how the choice of threshold affects the number of clusters.

## 5.2 *Simulation results*

Figure 2 shows the top of the hierarchical tree. The four clusters are correctly distinguished. We notice that observations generated from the same cluster sometimes have low probability of being grouped together. However, observations from different clusters were virtually never grouped together.

[Figure 2 about here.]

We looked at the resulting groups when the minimum between-group distance was 60%, meaning that observations were grouped together if they appeared in the same cluster more than 40% of the time. Ten groups resulted. Each of the four underlying clusters was split into one large group, ranging in size from 15-24 observations, and

one to three smaller groups ranging in size from 1-6 observations.

Figure 3 shows the fitted curve for each of the four clusters as well as the fitted curve for each of the ten classes. The model fit the population trajectories closely. Examination of the raw data indicates that the model correctly identified subjects in each of the four groups with unusual random effects.

[Figure 3 about here.]

We assessed convergence by examining the stability of the distributions of the number of basis functions in the model, number of clusters, and the precision parameters  $\tau$  and  $\lambda$ . Traceplots of these parameters provided no evidence against convergence.

## 6. Progesterone example

### 6.1 *Data*

We analyze hormone data from the North Carolina Early Pregnancy Study (EPS; Wilcox et al., 1988; Baird et al., 1997). Women in this study provided daily urine specimens. We consider daily urinary progesterone measurements from the onset of menstruation up until two days after ovulation. Implantation of a conceptus is known to affect progesterone, and this region is designed to precede implantation. We randomly selected one menstrual cycle from each of 172 women to use in our analysis. If a woman had both conception and non-conception cycles available, we randomly selected first the cycle type and then the cycle. Conception cycles made up 65 (38%) of the 172 cycles.

We used Bayesian multivariate linear splines to model the trajectories, allowing for unknown numbers and locations of knots and for a nonparametric distribution of the random woman-specific basis coefficients around the population value. Our goal is to identify trajectory clusters and to describe any clusters that are predictive of

high or low conception probability. Covariates include day since menstruation and day relative to ovulation, both of which may be related to hormone trajectories (van Zonneveld et al., 2003). Pre-specified basis functions were deemed insufficiently flexible for characterizing the data, because it is not at all clear where to place the knots in the 2-d predictor space. In addition, by averaging across models with different numbers and locations of knots we obtain smooth trajectories and better account for our true uncertainty. We ran the algorithm for 50,000 iterations after a 10,000 iteration burn-in period.

## 6.2 *Progesterone results*

Analysis of the EPS data yielded several interesting results. Unlike the more ideal simulated data case, there were no completely disjoint classes to examine. As in the simulation, we selected a minimum between-group distance of 60%. This yielded one very large group with 133 observations and eight small groups with 1-14 observations each. Figure 4 shows the estimated progesterone trajectories for the groups. In general, the trajectory in the dominant class appears to contain mid-range values and be relatively flat. The other classes tend to have sharp decreases at the start or end of the window, or to have exceptionally high or low values. These smaller clusters likely represent less common hormone patterns, and these trajectories may be clinically important.

[Figure 4 about here.]

Perhaps more informative is Figure 5, which displays the raw data for each of the nine groups. The method has appropriately grouped similar observations together. In addition, the presence of sets of very similar curves and very different shapes for each cluster support the decision to relax the parametric assumption that all women are

normally distributed around a trajectory described by some mean set of coefficients and instead describe it through a DP.

[Figure 5 about here.]

### 6.3 *Sensitivity Analysis*

Adequate mixing and convergence of cluster allocation is sometimes of concern in DP models, as the algorithm has a tendency to become trapped in a local mode (Jain and Neal, 2004). To combat this, we collected a large number of samples. In addition, we found that parallel chains from different initial cluster allocations yielded nearly identical results, and the number of clusters appeared to mix adequately across iterations. This led us to conclude that the mixing was adequate to avoid local mode problems. We also examined the implications of using different values of  $\alpha$ . Large values of  $\alpha$  tended to increase the average number of clusters at each iteration. In the post-processing of the samples, this meant that distances between observations tended to be larger. The same final groups were found, although requiring a larger threshold between-group distance. Small values of  $\alpha$  had the expected opposite effect.

## 7. **Discussion**

We have proposed a method for posterior simulation across a high-dimensional space of models for curves characterized in terms of basis functions with an unknown distribution on the basis coefficients. Although Ishwaran and James (2003b) considered mixture models where kernels have different forms, including different numbers of parameters, our approach is (to our knowledge) the first Bayesian method to treat both the basis functions and the distribution of the basis coefficients as unknown. This is done in a very general manner, allowing arbitrary classes of basis functions and a species sampling prior on the basis coefficients, though we focus on multivariate linear splines and the Dirichlet process special case in the examples.

Posterior simulation across the model space relied on a new Metropolis-Hastings transition probability, which can be implemented efficiently in species sampling models due to the property of independence between the allocation to unique values and the prior on the unique values. We found the algorithm to be efficient at exploring model spaces in our simulations. In between steps for moving between models, parameters were updated using a nested Gibbs sampler, relying on the generalized collapsed Gibbs algorithm of Ishwaran and James (2003a). Although our focus was on functional data, the algorithm can be applied directly for model selection and averaging in large lists of competing nonparametric models, adding to the (as of yet under-developed) literature on calculating Bayes factors for comparing nonparametric models (refer to Basu and Chib (2003)). For example, potential applications include variable selection in semiparametric hierarchical models, simultaneous variable selection and clustering in multi-task learning and meta analysis applications, and selection between non-nested hierarchical models with nonparametric priors on the random effects distributions.

#### ACKNOWLEDGEMENTS

This research was supported by the Intramural Research Program of the NIH, and NIEHS. We would like to thank Allen Wilcox, Donna Baird and Clare Weinberg for generously providing the data and for their helpful comments on the approach. Thanks also to the Associate Editor and two reviewers for their extremely useful comments.

#### REFERENCES

Antoniak, C. (1974). Mixtures of Dirichlet processes with application to nonparametric problems. *The Annals of Statistics* **2**, 1152–1174.

- Baird, D., Wilcox, A., Weinberg, C., Kamel, F., McConnaughey, D., Musey, P. and Collins, D. (1997). Preimplantation hormonal differences between the conception and non-conception menstrual cycles of 32 normal women. *Human Reproduction* **12**, 2607–2613.
- Basu, S. and Chib, S. (2003). Marginal likelihood and Bayes factors for Dirichlet process mixture models. *Journal of the American Statistical Association* **98**, 224–235.
- Bigelow, J. and Dunson, D. (2007). Bayesian adaptive regression splines for hierarchical data. *Biometrics* **63**, 724–732.
- Biller, C. (2000). Adaptive Bayesian regression splines in semiparametric generalized linear models. *Journal of Computational and Graphical Statistics* **9**, 122–140.
- Blackwell, D. and MacQueen, J. (1973). Ferguson distributions via pólya urn schemes. *The Annals of Statistics* **1**, 353–355.
- Brown, P., Kenward, M. and Bassett, E. (2001). Bayesian discrimination with longitudinal data. *Biostatistics* **2**, 417–432.
- Brumback, B. and Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association* **93**, 961–976.
- Bush, C. and MacEachern, S. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83**, 275–285.
- Cai, B. and Dunson, D. (2005). Variable selection in nonparametric random effects models. Technical report, Department of Statistical Science, Duke University.
- de la Cruz, R. and Quintana, F. (2005). A model-based approach to Bayesian classification with applications to predicting pregnancy outcomes from longitudinal  $\beta$ -hCG profiles. <http://www.mat.puc.cl/quintana/trbp.pdf>.
- Denison, D., Holmes, C., Mallick, B. and Smith, A. (2002). *Bayesian methods for*

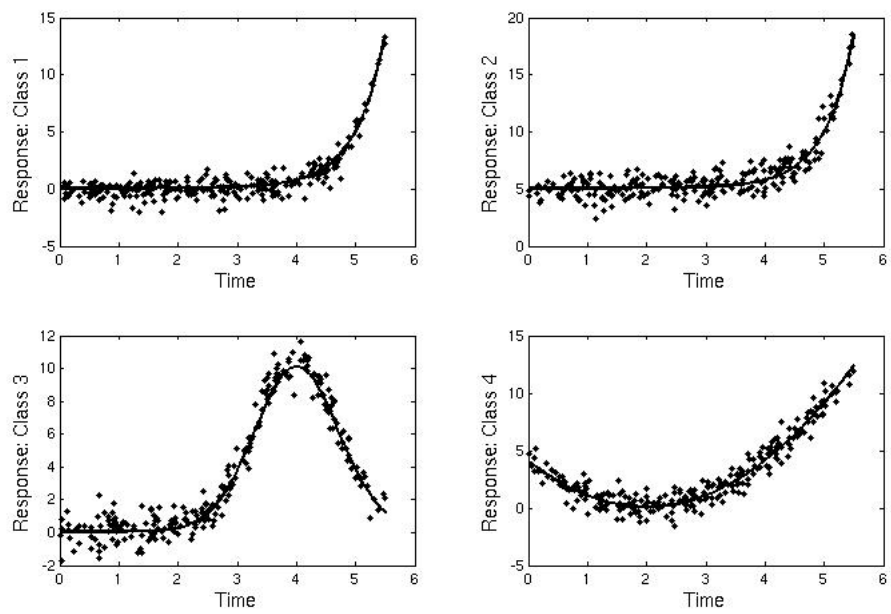
- nonlinear classification and regression*. John Wiley and Sons, Chichester, West Sussex, England.
- DiMatteo, I., Genovese, C. and Kass, R. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88**, 1055–1073.
- Dunson, D. B., Herring, A. H. and Mulheri-Engel, S. A. (to appear). Bayesian selection and clustering of polymorphisms in functionally-related genes. *Journal of the American Statistical Association* .
- Escobar, M. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89**, 268–277.
- Escobar, M. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.
- Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.
- Ferguson, T. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics* **2**, 615–629.
- Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.
- Griffin, J. and Steel, M. (2004). Semiparametric bayesian inference for stochastic frontier models. *Journal of Econometrics* **123**, 121–152.
- Hansen, B. and Pitman, J. (2000). Prediction rules for exchangeable sequences related to species sampling. *Statistics and Probability Letters* **46**, 251–256.
- Hansen, M. and Kooperberg, C. (2002). Spline adaptation in extended linear models. *Statistical Science* **17**, 2–20.
- Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

- Heard, N., Holmes, C. and Stephens, D. (2006). A quantitative study of gene regulation involved in the immune response of anopheline mosquitoes: An application of bayesian hierarchical clustering of curves. *Journal of the American Statistical Association* **101**, 18–29.
- Holmes, C. and Mallick, B. (2000). Bayesian wavelet networks for nonparametric regression. *IEEE Transactions on Neural Networks* **11**, 27–35.
- Holmes, C. and Mallick, B. (2001). Bayesian regression with multivariate linear splines. *Journal of the Royal Statistical Society, Series B* **63**, 3–17.
- Holmes, C. and Mallick, B. (2003). Generalized nonlinear modeling with multivariate free-knot regression splines. *Journal of the American Statistical Association* **98**, 352–368.
- Holmes, C.C. Denison, D. and Mallick, B. (2002). Accounting for model uncertainty in seemingly unrelated regressions. *Journal of Computational and Graphical Statistics* **11**, 533–551.
- Ishwaran, H. and James, L. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173.
- Ishwaran, H. and James, L. (2003a). Generalized weighted chinese restaurant processes for species sampling mixture models. *Statistica Sinica* **13**, 1211–1235.
- Ishwaran, H. and James, L. (2003b). Some further developments for stick-breaking priors: finite and infinite clustering and classification. *Sankhya Series A* **65**, 577–592.
- Ishwaran, H. and Takahara, G. (2002). Independent and identically distributed monte carlo algorithms for semiparametric linear mixed models. *Journal of the American Statistical Association* **97**, 1154–1166.
- Ishwaran, H. and Zarepour, M. (2002a). Dirichlet prior sieves in finite normal mixtures.

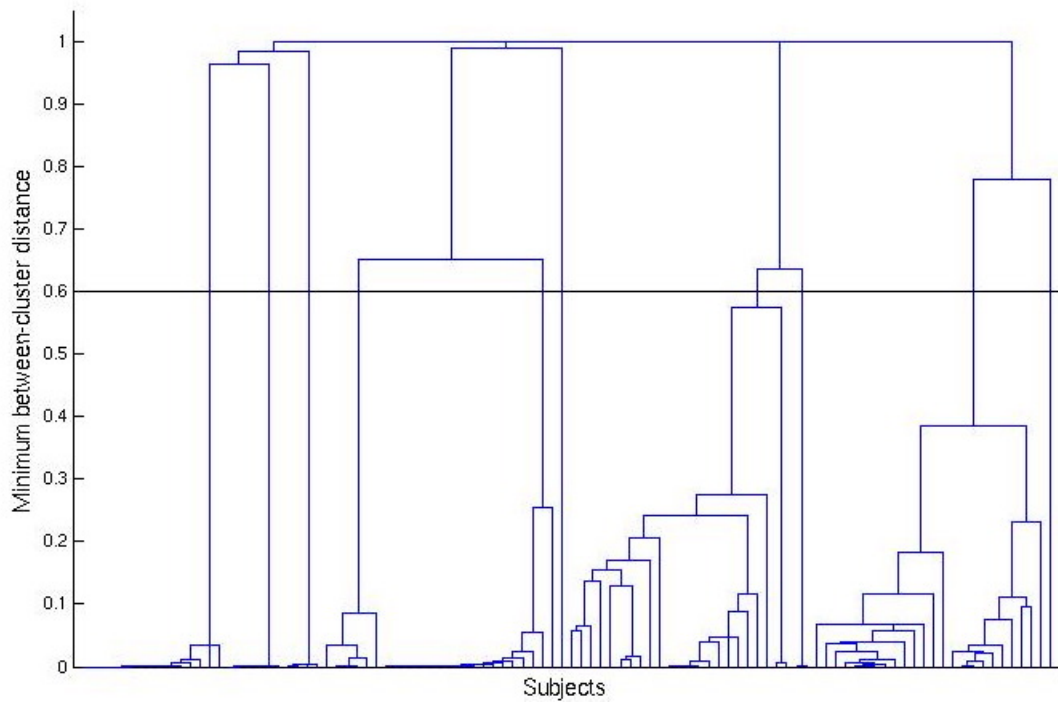
- Statistica Sinica* **12**, 941–963.
- Ishwaran, H. and Zarepour, M. (2002b). Exact and approximate sum-representations for the dirichlet process. *Canadian Journal of Statistics* **30**, 269–283.
- Jain, S. and Neal, R. (2004). A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *Journal of Computational and Graphical Statistics* **13**, 158–182.
- James, G. and Hastie, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society, Series B* **63**, 533–550.
- James, G. and Sugar, C. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association* **98**, 397–408.
- Kass, R., Ventura, V. and Cai, C. (2003). Statistical smoothing of neuronal data. *NETWORK: Computation in Neural Systems* **14**, 5–15.
- Ke, C. and Wang, Y. (2001). Semiparametric nonlinear mixed-effects models and their application (with discussion). *Journal of the American Statistical Association* **96**, 1272–1298.
- Kleinman, K. and Ibrahim, J. (1998). A semi-parametric Bayesian approach to the random effects model. *Biometrics* **54**, 921–938.
- Kottas, A., Branco, M. D. and Gelfand, A. E. (2002). A nonparametric bayesian modeling approach for cytogenetic dosimetry. *Biometrics* **58**, 593–600.
- Laws, D. J. and O’Hagan, A. (2002). A hierarchical bayes model for multilocation auditing. *Journal of the Royal Statistical Society, Series D* **51**, 431–450.
- Lindstrom, M. (2002). Bayesian estimation of free-knot splines using reversible jumps. *Computational Statistics & Data Analysis* **41**, 255–269.
- Ma, P., Castillo-Davis, C., Zhong, W. and Liu, J. (2005). Curve clustering to discover patterns in time-course gene expression data. *Working paper available*

- at: <http://ilabs.inquiry.wiuc.edu/ilab/fallbiosem/documents/2380/home/ma-et-al-2005.pdf>.
- MacEachern, S. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics- Simulation and Computation* **23**, 727–741.
- MacEachern, S. and Müller, P. (1998). Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 223–238.
- Marshall, G. and Barón, A. (2000). Linear discriminant models for unbalanced longitudinal data. *Statistics in Medicine* **19**, 1961–1981.
- Medvedovic, M. and Sivaganesan, S. (2002). Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics* **18**, 1194–1206.
- Mukhopadhyay, S. and Gelfand, A. (1997). Dirichlet process mixed generalized linear models. *Journal of the American Statistical Association* **92**, 633–639.
- Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* **55**, 463–469.
- Neal, R. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265.
- Papaspiliopoulos, O. and Roberts, G. (2006). Retrospective mcmc for dirichlet process hierarchical models. *Biometrika* **in press**.
- Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory Related Fields* **102**, 145–158.
- Pitman, J. (1996). Some developments of the blackwell-macqueen urn scheme. In Ferguson, T., Shapley, L. and MacQueen, J., editors, *Statistics, Probability and Game Theory*, pages 245–267. IMS Lecture Notes-Monograph Series.
- Pitman, J. and Yor, M. (1997). The two-parameter poisson-dirichlet distribution de-

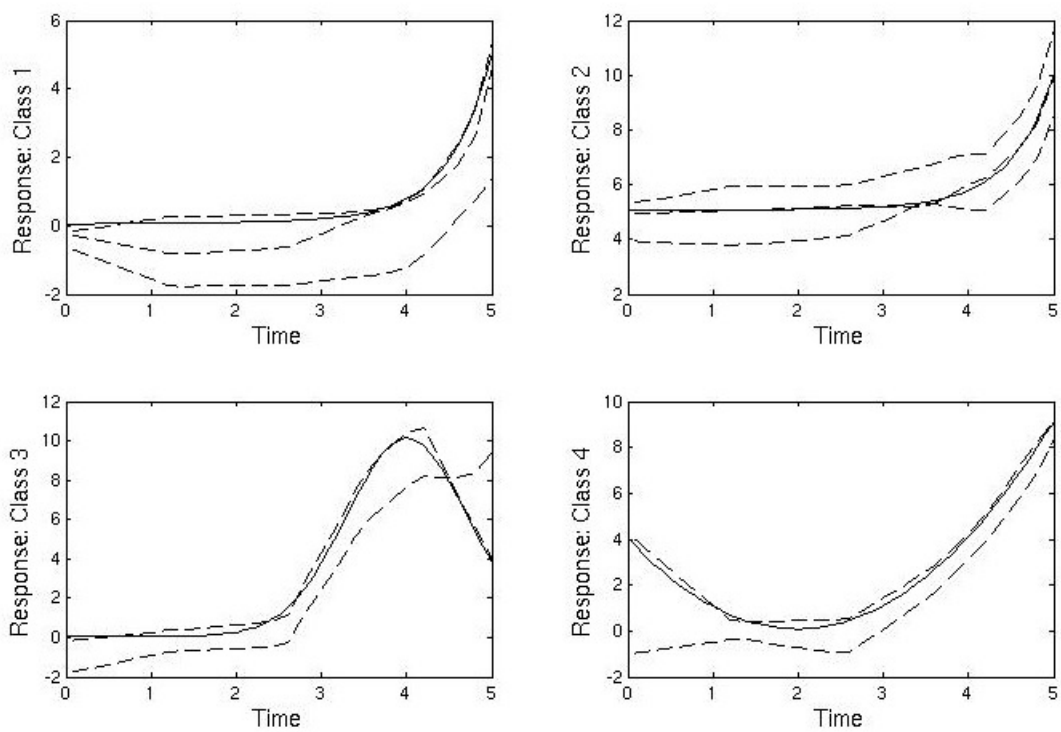
- rived from a stable subordinator. *Annals of Probability* **25**, 855–900.
- Ray, S. and Mallick, B. (2006). Functional clustering by bayesian wavelet methods. *Journal of the Royal Statistical Society B* **68**, 305–332.
- Rice, J. and Wu, C. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253–259.
- Sha, N., Tadesse, M. and Vannucci, M. (2006). Variable selection in clustering via dirichlet process mixture models. *Biometrika* **in press**.
- van Zonneveld, P., Scheffer, G., Broekmans, F., Blankenstein, M., de Jong, F., Looman, C., Habbema, J. and te Velde, E. (2003). Do cycle disturbances explain the age-related decline of female fertility? Cycle characteristics of women aged over 40 years compared with a reference population of young women. *Human Reproduction* **18**, 495–501.
- Wilcox, A., Weinberg, C., O’Connor, J., Baird, D., Schlatterer, J., Canfield, R., Armstrong, E. and Nisula, B. (1988). Incidence of early loss of pregnancy. *New England Journal of Medicine* **319**, 189–194.
- Wood, S., Jiang, W. and Tanner, M. (2002). Bayesian mixture of splines for spatially adaptive nonparametric regression. *Biometrika* **89**, 513–528.
- Xue, Y., Liao, X., Carin, L. and Krishnapuram, B. (2006). Multi-task learning for classification with Dirichlet process priors. *under revision* .



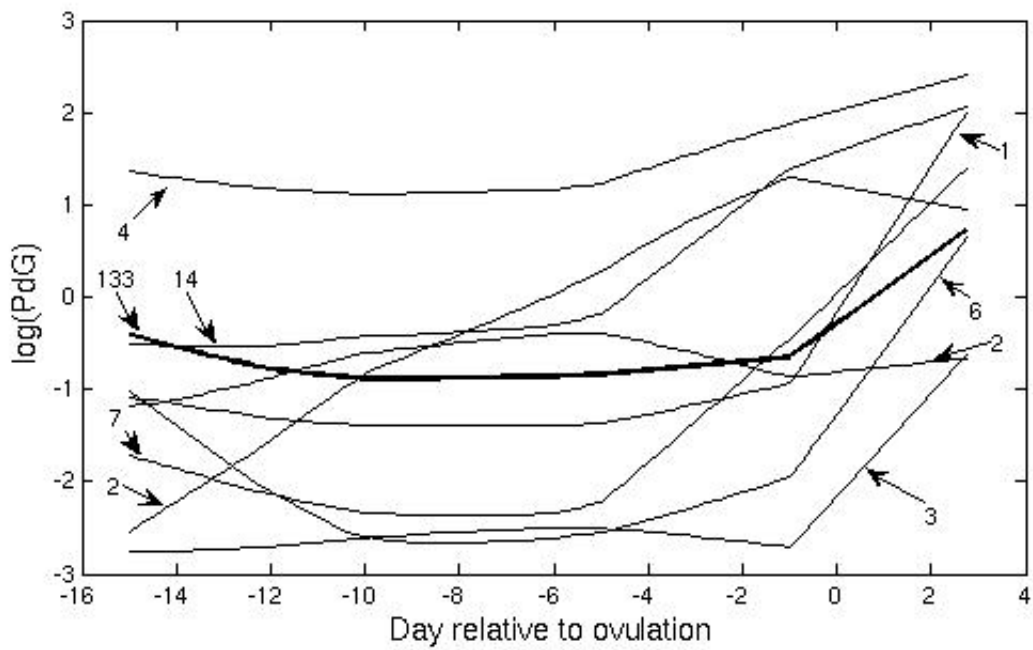
**Figure 1.** Underlying population curves (lines) and data (points) for each of the four clusters. Each cluster had 250 observations from 25 subjects



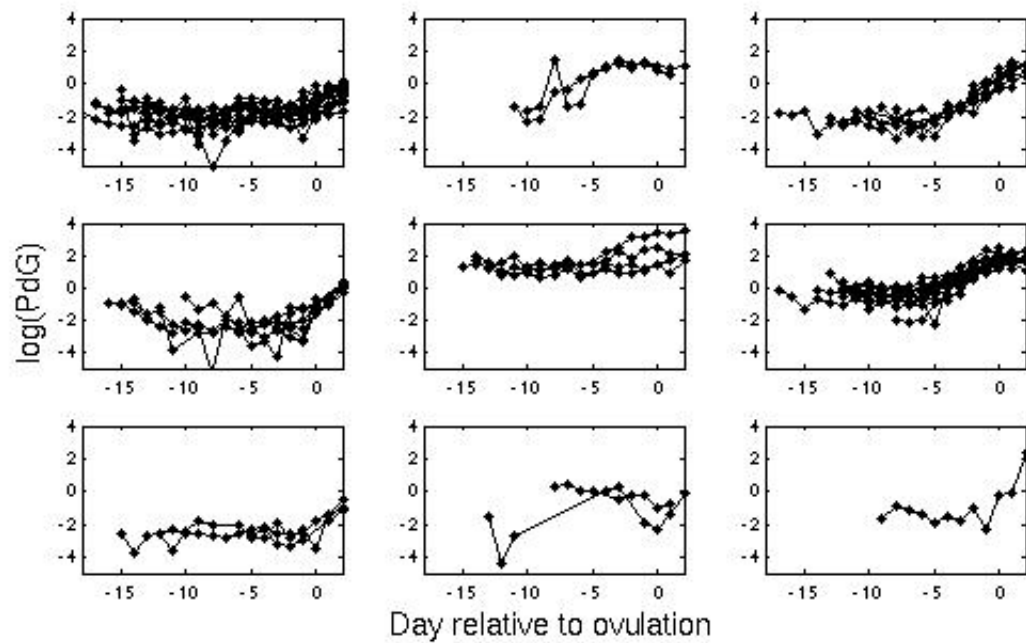
**Figure 2.** Hierarchical cluster tree. The between-group distance is given on the y-axis, where the distance is the posterior probability that the two classes are distinct. The four true clusters are the four classes at the top of the tree, and more classes are seen as the minimum distance decreases. The solid horizontal line is the threshold at which our classes were created.



**Figure 3.** Four true clusters (solid lines) and the ten classes (dashed lines)



**Figure 4.** The dark line is the estimated trajectory for the large group. The lighter lines are the estimated trajectories for the smaller groups. Arrows indicate the number of cycles in each cluster.



**Figure 5.** These are data from all nine groups. There are twenty trajectories presented for the dominant cluster (first plot), and the other eight groups are entirely displayed.