

Nonparametric Bayes local borrowing of information and clustering

BY DAVID B. DUNSON

Biostatistics Branch, National Institute of Environmental Health Sciences, P.O. Box

12233, RTP, NC 27709, USA

dunson1@niehs.nih.gov

SUMMARY

This article focuses on the problem of choosing a prior for a probability measure characterizing the joint distribution of multiple subject-specific parameters within a Bayesian hierarchical model. A local partition process prior is proposed, which has large support and induces dependent, local clustering. Subjects can be clustered together for a subset of their parameters, and one learns about similarities between subjects increasingly as parameters are added. The local partition process prior is constructed through a locally-weighted mixture of global and local components, resulting in a generalization of joint and independent Dirichlet process priors. Some basic properties of the process are described, including simple two-parameter expressions for marginal and conditional clustering probabilities. A slice sampler is developed which bypasses the need to approximate the countably infinite random measure in performing posterior computation. The methods are illustrated using simulation examples, and an application to hormone trajectory data from an epidemiologic study.

Some key words: Dirichlet process; Functional data; Local shrinkage; Meta analysis; Multi-task learning; Partition model; Slice sampling; Stick-breaking.

1. INTRODUCTION

1.1. Problem formulation

This article focuses on the problem of choosing a prior for a probability measure P characterizing the joint distribution of subject-specific parameters $\theta_i = \{\theta_{ij}\}_{j=1}^p$ within a hierarchical model. Letting $\theta_i \sim P$, with $\theta_{ij} \sim P_j$ and P_j denoting the j th marginal of P , we assume that P_j assigns probability one to the set of atoms $\{\Theta_{hj}\}_{h=1}^\infty$, for $j = 1, \dots, p$. Because P_j is discrete, subjects will be clustered into groups according to the atom they are allocated to. In particular, let $\psi_{ij} = h$ denote that $\theta_{ij} = \Theta_{hj}$, so that subject i is allocated to cluster h for component j . Then, $\psi_i = (\psi_{i1}, \dots, \psi_{ip})' \in \{1, 2, \dots, \infty\}^p$ defines the cluster membership for subject i for each of the p components.

Assume that P is a probability measure over a measurable Polish space (Ω, \mathcal{B}) , with Ω the sample space and \mathcal{B} the corresponding Borel σ -algebra. Note that $\theta_i \sim P$ implies that $\psi_i \sim Q = T(P)$, where $Q = T(P)$ is a probability measure on $\{1, \dots, \infty\}^p$. Then, a prior \mathcal{P} for P induces a prior \mathcal{Q} for Q through the mapping T . In choosing \mathcal{P} , it is appealing to induce local clustering of subjects, which allows subjects to be clustered together for a subset of the elements of θ_i . To clarify the distinction between local and global clustering, $\theta_{ij} = \theta_{i'j}$ if subjects i and i' are locally clustered for component j , while $\theta_i = \theta_{i'}$ if these subjects are globally clustered.

This article proposes a simple form for \mathcal{P} , which induces dependent, local clustering, while having weak support on the space of probability measures over (Ω, \mathcal{B}) . In addition, the induced prior $Q \sim \mathcal{Q}$ has the properties:

1. $\text{pr}\{Q(\psi) > \epsilon\} > \delta$ for all $\psi \in \{1, 2, \dots, \infty\}^p$ and some $\epsilon, \delta > 0$.
2. $\text{pr}\{Q(\psi_{ij} = \psi_{i'j}) \in A\} > \epsilon \forall$ Borel subsets $A \subset [0, 1]$, $\psi_l \sim Q$, $l = i, i'$, and some $\epsilon > 0$,
3. $\text{pr}\left\{\frac{Q(\psi_{ij}=\psi_{i'j}, \psi_{ij'}=\psi_{i'j'})}{Q(\psi_{ij'}=\psi_{i'j'})} \geq Q(\psi_{ij} = \psi_{i'j})\right\} = 1$, for all $j, j' \in \{1, \dots, p\}$, $j' \neq j$.

4. $\text{pr}\left\{\frac{Q(\psi_{ij}=\psi_{i'j},\psi_{ij'}=\psi_{i'j'})}{Q(\psi_{ij'}=\psi_{i'j'})} \in A\right\} > \epsilon$, for all Borel subsets $A \subset [Q(\psi_{ij} = \psi_{i'j}), 1]$, $j, j' \in \{1, \dots, p\}$, $j' \neq j$ and for some $\epsilon > 0$.

The first two conditions imply large support, while the second two relate to dependence in local clustering. Under condition 3, the probability of clustering two subjects for component j increases if these subjects are known to be clustered for another component. Condition 4 implies that any degree of positive dependence in local clustering is supported by the prior.

1.2. Motivating application and related literature

As a motivating application, suppose that interest focuses on borrowing of information in estimating multiple related functions. For example, these functions may consist of progesterone trajectories in early pregnancy for different women. Letting y_{it} denote the t th measurement in pregnancy i , suppose $y_{it} \sim N(\eta_i(s_{it}), \sigma^2)$, $i = 1, \dots, n$ and $t = 1, \dots, n_i$, where $s_i = (s_{i1}, \dots, s_{i,n_i})'$ are the measurement times and $\eta_i \in \mathcal{F}$ is the measurement error-corrected trajectory for woman i . Assuming \mathcal{F} corresponds to the linear span of the basis $b = \{b_j\}_{j=1}^p$, let

$$\begin{aligned} \eta_i(s) &= \sum_{j=1}^p \theta_{ij} b_j(s), \quad \forall s \in \mathcal{T}, \\ \theta_i &\sim P, \quad P \sim \mathcal{P}, \end{aligned} \tag{1}$$

where $\mathcal{T} \subset \mathfrak{R}^+$. Then, variability in the functions $\{\eta_i\}_{i=1}^n$ is controlled through the probability measure P characterizing the joint distribution of the subject-specific basis coefficients.

Typical functional data analysis methods require P to obey a parametric law, such as the Gaussian. As parametric models are often insufficiently flexible and do not accurately reflect available prior information, a number of authors have proposed assigning P a prior with large support on the space of probability measures over $(\mathfrak{R}^p, \mathcal{B})$. In the setting of wavelet-based models for functional data, Ray & Mallick (2006) proposed using a Dirichlet process prior (Ferguson, 1973, 1974) for P , following previous work on nonparametric Bayes modelling

of random effects distributions (Bush & MacEachern, 1996; Kleinman & Ibrahim, 1998; Ishwaran and Takahara, 2002). Bigelow & Dunson (2005, Technical Report, Department of Statistical Science, Duke University) applied a related approach in the setting of multivariate adaptive splines with unknown numbers and locations of knots. Dunson et al. (2008a) generalized this approach to allow nonparametric Bayes inference on changes in a response density over predictor clusters.

Due to the discreteness property, a Dirichlet process prior for P leads to global functional clustering. In particular, subjects i and i' will either have $\eta_i(s) = \eta_{i'}(s) \forall s$ or their functions will be constructed from independent draws of θ_i and $\theta_{i'}$ from the parametric base distribution P_0 . Global functional clustering is overly-restrictive, and it is common to observe functions that are quite similar overall but with distinct differences in local regions.

To allow local clustering and borrowing of information, an alternative to a joint Dirichlet process prior for P is to express $P = \otimes_{j=1}^p P_j$ as a product measure, with the P_j 's assigned independent Dirichlet process priors. However, the independence assumption seems inappropriate. Hoff (2006) proposed an alternative formulation, which used a Dirichlet process to define clusters of subjects that varied in both the elements of θ_i exhibiting a change relative to a baseline θ_0 and the magnitude of change for these elements. This approach allows local clustering, but only at the baseline value. Petrone et al. (2008) proposed a hybrid functional Dirichlet process that allows local allocation to clusters through a latent Gaussian process. Although the formulation is flexible, the use of a latent process to drive local clustering has some disadvantages in terms of interpretation, prior elicitation and posterior computation. Dunson et al. (2008b) proposed a matrix stick-breaking process that generalizes the Sethuraman (1994) stick-breaking representation of the Dirichlet process to allow dependent, local clustering. However, the matrix stick-breaking process fails to satisfy property 4 in forcing a high degree of dependence in local clustering.

This article proposes an alternative local partition process prior, which satisfies prop-

erties 1-4, while facilitating computation and theoretical work due to the simplicity of the formulation. Section 2 describes the proposed prior, derives some basic properties, and proposes a sparseness-favouring modification for high dimensional applications. Section 3 develops a slice sampler (Neal, 2003) for posterior computation, generalizing the approach proposed by Walker (2007) for Dirichlet process mixture models. Section 4 contains some simulation examples illustrating the approach. Section 5 applies the methodology to progesterone data in early pregnancy, and Section 6 contains a discussion.

2. DEPENDENT LOCAL PARTITION PROCESS

2.1. Formulation 1

Letting $\theta_i = \{\theta_{ij}\}_{j=1}^p \sim P$, with θ_{ij} a vector of dimension p_j and P a probability measure over (Ω, \mathcal{B}) , our focus is on specifying a prior \mathcal{P} for P that has support on the space of probability measures over (Ω, \mathcal{B}) , while satisfying properties 1-4 of §1.1. Let $\Theta_h = \{\Theta_{hj}\}_{j=1}^p \sim P_0$ denote species h , for $h = 1, \dots, \infty$, with P_0 a baseline probability measure over (Ω, \mathcal{B}) . A Dirichlet process prior for P , denoted $P \sim \text{DP}(\alpha P_0)$, can be expressed as $P = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_h}$, with $\pi = \{\pi_h\}_{h=1}^{\infty}$ probability weights sampled from a stick-breaking process (Sethuraman, 1994) and δ_{Θ} a probability measure concentrated at Θ . Hence, in sampling θ_i from the distribution implied by P , one allocates θ_i globally to one of the species, so that $\theta_i = \Theta_h$ for some $h \in \{1, \dots, \infty\}$.

In order to obtain a process that induces local clustering, one can instead allow the different components of θ_i to be allocated to different species, so that each individual is a hybrid (in the terminology of Petrone et al. 2008) of traits selected from different species. To accomplish this in a parsimonious manner that avoids the complication of a latent Gaussian process, let

$$\begin{aligned} \theta_{ij} &= \Theta_{\gamma_{ij}, j}, \quad i = 1, \dots, n, j = 1, \dots, p, \\ \gamma_{ij} &\sim z_{ij} \delta_{(0, \phi_{i0})} + (1 - z_{ij}) \delta_{(1, \phi_{ij})}, \quad j = 1, \dots, p, \end{aligned}$$

$$\begin{aligned}
z_{ij} &\sim \pi_j \delta_1 + (1 - \pi_j) \delta_0, \quad \pi_j \sim \text{Be}(1, \beta), \quad j = 1, \dots, p, \\
\phi_{ij} &\sim \sum_{h=1}^{\infty} \nu_{jh} \delta_h, \quad \nu_{jh} = \nu_{jh}^* \prod_{l < h} (1 - \nu_{jl}^*), \quad \nu_{jh}^* \sim \text{Be}(1, \alpha), \quad j = 0, 1, \dots, p, \quad (2)
\end{aligned}$$

where $\gamma_{ij} = (l, h)$ denotes that the j th component of the parameter vector for subject i is allocated to species (l, h) , for $l = 0, 1$ and $h = 1, \dots, \infty$. The l index is incorporated to allow two families of species, with $z_{ij} = 1$ ($z_{ij} = 0$) denoting allocation to a global (local) family for component j , $\pi_j = \text{pr}(z_{ij} = 1)$ is the probability of allocation to the global family for component j , ϕ_{i0} is the global species index for subject i , and ϕ_{ij} is the local species index for subject i and component j . The species $\Theta_\gamma = \{\Theta_{\gamma,j}\}_{j=1}^p \sim P_0$ independently for all $\gamma \in \{0, 1\} \times \{1, 2, \dots, \infty\}$. As shorthand for expression (2), we use $\theta_i \sim P$, $P \sim \text{LPP}_1(\alpha, \beta, P_0)$, with LPP_1 denoting version one of the local partition process prior (a second version is proposed in §1.2).

Let Θ_h denote the h th species within the set constructed from the union of all the species in families $l = 0$ and $l = 1$, with this union denoted by $\Theta = \bigcup_{l=0}^1 \Theta_{(l,-)}$, where $\Theta_{(l,-)}$ is the set of all species in family l , for $l = 0, 1$. Then, using the notation of §1.1, $\psi_{ij} = h$ denotes that $\theta_{ij} = \Theta_{hj}$, where $\Theta_h = \{\Theta_{hj}\}_{j=1}^p$ is the collection of traits for species h , and Θ_{jh} is the j th trait in this collection. The index h is arbitrary, but as a convention let $\Theta_h = \Theta_{(0,(h+1)/2)}$ for odd h and $\Theta_h = \Theta_{(1,h/2)}$ for even h , for $h = 1, \dots, \infty$. Under this convention, $\gamma_{ij} = (0, h)$ implies that $\psi_{ij} = 2h - 1$ and $\gamma_{ij} = (1, h)$ implies $\psi_{ij} = 2h$, $h = 1, \dots, \infty$.

To motivate the local partition process defined in (2), we first consider some important special cases. Note that in the limit as $\beta \rightarrow 0$, $\pi_j = 1$, $z_{ij} = 1$, $\gamma_{ij} = (0, \phi_{i0})$ and θ_i is set equal to the ϕ_{i0} th species from the global family. In this case, it is straightforward to show that one obtains $\theta_i \sim P$ with $P \sim \text{DP}(\alpha P_0)$ and subjects allocated globally to species. In addition, in the limit as $\beta \rightarrow \infty$, $\pi_j = 0$, $z_{ij} = 0$, $\gamma_{ij} = (1, \phi_{ij})$ and θ_{ij} is set equal to the j th component of the ϕ_{ij} th species from the local family. In this case, one obtains $\theta_{ij} \sim P_j$, with $P_j \sim \text{DP}(\alpha P_{0j})$ independently for $j = 1, \dots, p$, with P_{0j} the j th marginal of P_0 . Potentially,

dependence can be incorporated through the base measure P_0 , as in Cifarelli & Regazzini (1978), though this is a restrictive type of dependence.

Under prior (2) the marginal probability of local clustering of subjects i and i' for component j is as follows:

$$\textit{Proposition 1.} \quad \text{pr}(\theta_{ij} = \theta_{i'j}) = \left(\frac{1}{1+\alpha}\right)\left(\frac{1}{2+\beta}\right)\left(\beta + \frac{2}{1+\beta}\right) = \rho.$$

Figure 1 provides a heat plot of the local clustering probability as a function of the hyperparameters α and β . In the limit as $\alpha, \beta \rightarrow 0$, $\rho = 1$. In addition, ρ is monotone decreasing in α and β , with $\rho = 1/3$ for $\alpha = \beta = 1$, $\rho = 1/(1 + \alpha)$ in the limit as $\beta \rightarrow \infty$, and $\rho = 0$ in the limit as $\alpha \rightarrow \infty$. Hence, α and β are key hyperparameters controlling the probability of local clustering. To allow the data to inform about their values, gamma hyperpriors are chosen by letting $\alpha \sim \text{Ga}(a_\alpha, b_\alpha)$ and $\beta \sim \text{Ga}(a_\beta, b_\beta)$.

In addition, a simple closed form can be derived for the joint probability of local clustering of subjects i and i' for components j and j' .

$$\textit{Proposition 2.} \quad \text{pr}(\theta_{ij} = \theta_{i'j}, \theta_{ij'} = \theta_{i'j'}) = \left(\frac{1}{1+\alpha}\right)^2 \left(\frac{1}{2+\beta}\right)^2 \left\{ \left(\beta + \frac{2}{1+\beta}\right)^2 + \frac{4\alpha}{(1+\beta)^2} \right\}.$$

Note that the joint probability in Proposition 2 is strictly larger than the product of the marginal probabilities from Proposition 1, which implies that there is positive dependence in local clustering. The degree of positive dependence is monotonely decreasing with α and β , being controlled by the size of $4\alpha/(1 + \beta)^2$.

Let $\theta_{ij} \sim P_j$, with P_j the j th marginal from the joint prior, $P \sim \text{LPP}_1(\alpha, \beta, P_0)$. Then, P_j is a probability measure over $(\Omega_j, \mathcal{B}_j)$, where Ω_j is the sample space for the j th component of θ_i , and \mathcal{B}_j is the corresponding Borel σ -algebra. For any $B \in \mathcal{B}_j$, $P_j(B)$ is a random variable, with $P_j(B) \stackrel{D}{=} WX_1 + (1-W)X_2$, where $W \sim \text{Be}(1, \beta)$ and $X_l \sim \text{Be}(\alpha P_{0j}(B), \alpha\{1 - P_{0j}(B)\})$, for $l = 1, 2$, are independent random variables. Although the density of $P_j(B)$ does not have a simple form, the expectation and variance can be derived as

$$\text{E}\{P_j(B)\} = P_{0j}(B), \quad \text{V}\{P_j(B)\} = P_{0j}(B)\{1 - P_{0j}(B)\} \left(\frac{1}{1 + \alpha}\right) \left(\frac{1}{2 + \beta}\right) \left(\beta + \frac{2}{1 + \beta}\right),$$

so that the prior is centered on the base probability measure P_{0j} and the variance is simply $P_{0j}(B)\{1 - P_{0j}(B)\}\rho$, with ρ the probability of clustering shown in Proposition 1.

Theorem 1. Let $\theta_i \sim P$ with $P \sim \text{LPP}_1(\alpha, \beta, P_0)$, with hyperpriors chosen for α and β having support on \mathfrak{R}^+ . Then, properties 1-4 of §1.1 are satisfied.

2.2. Formulation 2

The prior proposed in (2) is appealing in leading to simple closed form expressions for the local clustering probabilities and the mean and variance of the random probability measure, while also satisfying properties 1-4 of §1.1. In addition, the prior is a two parameter generalization of joint and independent Dirichlet process priors, which are obtained in limiting special cases. The approach also results in straightforward computation using minor adaptations of Markov chain Monte Carlo algorithms developed for Dirichlet process mixture models. However, in applications involving large p , it is appealing for computational reasons to consider sparse alternatives that avoid the incorporation of a separate stick-breaking process for each component.

As a simplification of (2), which may be particularly useful in applications having large p , replace the last line of expression (2) with the following:

$$\phi_{ij} \sim \sum_{h=1}^{\infty} \nu_h \delta_h, \quad \nu_h = \nu_h^* \prod_{l < h} (1 - \nu_l^*), \quad \nu_h^* \sim \text{Be}(1, \alpha), \quad j = 0, 1, \dots, p, \quad (3)$$

where the top lines of (2) are unchanged. The hierarchical prior in this modified case is denoted as $\theta_i \sim P$, $P \sim \text{LPP}_2(\alpha, \beta, P_0)$. For the $\text{LPP}_2(\alpha, \beta, P_0)$ prior, in the limit as $\beta \rightarrow 0$, $\theta_i \sim P$ with $P \sim \text{DP}(\alpha P_0)$, so that the joint Dirichlet process is obtained as a limiting special case. However, in the limit as $\beta \rightarrow \infty$, instead of independent Dirichlet process priors for P_j , $j = 1, \dots, p$, one obtains the following dependent Dirichlet process (MacEachern, 1999, 2001) formulation:

$$P_j = \sum_{h=1}^{\infty} \nu_h \delta_{\Theta_{hj}}, \quad j = 1, \dots, p, \quad \Theta_h = \{\Theta_{hj}\}_{j=1}^p \sim P_0, \quad (4)$$

with $P_j \sim \text{DP}(\alpha P_{0j})$ marginally. Expression (4) corresponds to the fixed-p formulation of the dependent Dirichlet process, which has been applied by De Iorio et al. (2004) to model collections of densities and by Gelfand et al. (2005) in spatial data analysis.

Under the modified local partition process, the local clustering probability $\text{pr}(\theta_{ij} = \theta_{ij'})$ is identical to the expression shown in Proposition 1. However, the joint clustering probability is modified as shown in Proposition 3.

$$\text{Proposition 3. } \text{pr}(\theta_{ij} = \theta_{i'j}, \theta_{ij'} = \theta_{i'j'}) =$$

$$\left(\frac{1}{1+\alpha}\right)\left(\frac{1}{2+\beta}\right)^2\left(\beta + \frac{2}{1+\beta}\right)\left\{\frac{(6+\alpha)\beta}{(2+\alpha)(3+\alpha)} + \frac{2}{1+\beta}\right\}.$$

Note that this joint probability is strictly larger than the product of the marginal clustering probabilities shown in Proposition 1, implying positive dependence in local clustering. Varying the values of α and β widely, one obtains similar values for the conditional clustering probabilities under Propositions 2 and 3, with the values under Proposition 3 slightly higher. Under Proposition 3 $\text{pr}(\theta_{ij} = \theta_{i'j} \mid \theta_{ij'} = \theta_{i'j'})$ is constrained to fall in a subset of the interval $[\text{pr}(\theta_{ij} = \theta_{i'j}), 1]$, so property 4 is not satisfied. Because this subset comprises almost the entire interval, with only values very close to the left boundary excluded, this constraint is not important from a practical perspective.

3. POSTERIOR COMPUTATION

3.1. *Slice Sampler*

Assume that $y_i \sim g(\theta_i, \tau)$, where y_i is an $n_i \times 1$ vector of measurements for subject i , $\theta_i \sim P$ is a vector of parameters specific to subject i , and τ is a vector of population parameters. Initially letting $P \sim \text{LPP}_1(\alpha, \beta, P_0)$, we propose to use an adaptation of the slice sampling approach of Walker (2007) to implement posterior computation. This slice sampler avoids the need to rely on a truncation approximation, as in Ishwaran & James

(2001). Although we could also implement a retrospective sampler, modifying the approach of Papaspiliopoulos & Roberts (2008), the slice sampler is considerably simpler to implement.

In addition to the latent variables defined in (2), we introduce latent variables $u_i = \{u_{ij}\}_{j=0}^p$, for $i = 1, \dots, n$, where the complete data joint likelihood of y , u and z is

$$\prod_{i=1}^n \left\{ g(y_i; \Theta_{\gamma_i}, \tau) 1(u_{i0} < \nu_{0\phi_{i0}}) \prod_{j=1}^p 1(u_{ij} < \nu_{j\phi_{ij}}) \pi_j^{z_{ij}} (1 - \pi_j)^{1-z_{ij}} \right\}, \quad (5)$$

where the u_{ij} s are constrained to fall in $(0, 1)$. We can now define the full conditional posterior distributions to be used in a Gibbs sampler as follows:

Step 1. For the latent u_{ij} , the conditional is $\text{Unif}(0, \nu_{j\phi_{ij}})$, for $j = 0, 1, \dots, p$.

Step 2. For the latent z_{ij} , the conditional distribution is $\text{Ber}(p_{ij})$, with

$$p_{ij} = \frac{\pi_j g(y_i; \Theta_{\gamma_i(z_{ij}=1)}, \tau)}{\pi_j g(y_i; \Theta_{\gamma_i(z_{ij}=1)}, \tau) + (1 - \pi_j) g(y_i; \Theta_{\gamma_i(z_{ij}=0)}, \tau)},$$

where $\gamma_i(z_{ij} = l)$ denotes the current value of γ_i with γ_{ij} replaced with $(0, \phi_{i0})$ for $l = 1$ and $(1, \phi_{ij})$ for $l = 0$.

Step 3. For the stick-breaking variable, ν_{jh}^* , the conditional density is proportional to

$$(1 - \nu_{jh}^*)^{\alpha-1} \prod_{i=1}^n 1\left(\nu_{j\phi_{ij}}^* \prod_{l < \phi_{ij}} (1 - \nu_{jl}^*) > u_{ij}\right).$$

Letting $\phi_j^* = \max\{\phi_{ij}, i = 1, \dots, n\}$, the conditional distribution of ν_{jh}^* corresponds to the $\text{Be}(1, \alpha)$ prior for $h > \phi_j^*$, while for $h \leq \phi_j^*$ the posterior is a $\text{Be}(1, \alpha)$ distribution truncated to fall in the (a_{jh}, b_{jh}) interval with

$$\begin{aligned} a_{jh} &= \max \left\{ \frac{u_{ij}}{\prod_{l < h} (1 - \nu_{jl}^*)}, i : \phi_{ij} = h \right\}, \\ b_{jh} &= 1 - \max \left\{ \frac{u_{ij}}{\nu_{j\phi_{ij}}^* \prod_{l < \phi_{ij}, l \neq h} (1 - \nu_{jl}^*)}, i : \phi_{ij} > h, \right\}. \end{aligned}$$

Step 4. The conditional probability of $\phi_{ij} = h$ is proportional to $1(h \in A_{ij})g(y_i; \Theta_{\gamma_i(\phi_{ij}=h)}, \tau)$,

where $A_{ij} = \{h : \nu_{jh} > u_{ij}\}$ is a finite subset of $\{1, 2, \dots, \infty\}$ obtained by first sampling

ν_{jh}^* , for $h = 1, \dots, \phi_j^*$, with ϕ_j^* the smallest value satisfying

$$\sum_{h=1}^{\phi_j^*} \nu_{jh}^* \prod_{l < h} (1 - \nu_{jl}^*) > 1 - u_j^*,$$

where $u_j^* = \min\{u_{ij}, i = 1, \dots, n\}$.

Step 5. The conditional distribution of $\Theta_{(l,h)}$ is proportional to

$$g_0(\Theta_{(l,h)}) \prod_{i=1}^n g(y_i; \Theta_{\gamma_i}, \tau),$$

assuming the probability measure P_0 has density g_0 with respect to Lesbesgue measure.

Step 6. For the probability π_j , we have $\text{Be}(1 + \sum_i z_{ij}, \beta + \sum_i (1 - z_{ij}))$.

Step 7. For the parameters τ , the conditional distribution is proportion to $f(\tau) \prod_{i=1}^n g(y_i; \Theta_{\gamma_i}, \tau)$.

Step 8. For the hyperparameter, β , the conditional distribution is

$$\text{Ga}\left(a_\beta + p, b_\beta - \sum_{j=1}^p \log(1 - \pi_j)\right).$$

Step 9. For the hyperparameter, α , the conditional distribution is

$$\text{Ga}\left(a_\alpha + \sum_{j=0}^p \phi_j^*, b_\alpha - \sum_{j=0}^p \sum_{h=1}^{\phi_j^*} \log(1 - \nu_{jh}^*)\right).$$

Each of the above sampling steps is simple to implement and we have observed good rates of mixing and apparent convergence in several applications to simulated and real data. Steps 5 and 7 simplify when conjugate priors are chosen. For example, when $g(y_i; \Theta_{\gamma_i}, \tau)$ is the likelihood for a Gaussian linear model and P_0 is chosen to obey a Gaussian law, the conditional distribution in step 5 is Gaussian.

3.2. Formulation 2

The slice sampler of §3.1 can be easily modified for formulation 2 of the local partition process. We start with a modification of the complete data joint likelihood in (5) to drop the first subscript on each of the ν 's. Then the only change in steps 1 and 2 of the Gibbs sampler is to drop the first subscript on the ν 's. In step 3, the conditional density for the stick-breaking variable, ν_h^* , is proportional to

$$(1 - \nu_h^*)^{\alpha-1} \prod_{i=1}^n \prod_{j=0}^p 1\left(\nu_{\phi_{ij}}^* \prod_{l < \phi_{ij}} (1 - \nu_l^*) > u_{ij}\right),$$

Letting $\phi^* = \max\{\phi_{ij}, i = 1, \dots, n, j = 0, 1, \dots, p\}$, the conditional distribution of ν_h^* is $\text{Be}(1, \alpha)$ for $h > \phi^*$, while for $h \leq \phi^*$ the posterior is a $\text{Be}(1, \alpha)$ truncated to (a_h, b_h) with

$$\begin{aligned} a_h &= \max \left\{ \frac{u_{ij}}{\prod_{l < h} (1 - \nu_l^*)}, i, j : \phi_{ij} = h, i = 1, \dots, n, j = 0, 1, \dots, p \right\}, \\ b_h &= 1 - \max \left\{ \frac{u_{ij}}{\nu_{\phi_{ij}}^* \prod_{l < \phi_{ij}, l \neq h} (1 - \nu_l^*)}, i, j : \phi_{ij} > h, i = 1, \dots, n, j = 0, 1, \dots, p \right\}. \end{aligned}$$

Step 4 is modified to let $A_{ij} = \{h : \nu_h > u_{ij}\}$, where ν_h^* is sampled for $h = 1, \dots, \phi^*$, with ϕ^* the smallest value satisfying $\sum_{h=1}^{\phi^*} \nu_h^* \prod_{l < h} (1 - \nu_l^*) > 1 - \min\{u_{ij}, i = 1, \dots, n, j = 0, 1, \dots, p\}$.

Steps 5-8 are unchanged, and in step 9 the conditional distribution of α is

$$\text{Ga}\left(a_\alpha + \phi^*, b_\alpha - \sum_{h=1}^{\phi^*} \log(1 - \nu_h)\right).$$

Note that the major computational advantage relative to the algorithm of §3.1 for formulation 1 occurs in step 3, since one can update a single vector of stick-breaking random variables instead of $(p+1)$ vectors. As these vectors can be high-dimensional, this can convey a considerable computational savings, particularly in cases in which p is moderate to large.

4. SIMULATION EXAMPLE

To illustrate the approach and assess the performance, we first analyzed functional data simulated under the model described in §1.2, with $\mathcal{T} = [0, 1]$ and the basis functions corresponding to $b_1(s) = 1$, $b_{j+1}(s) = \exp(-\psi \|s - \xi_j\|^2)$, $j = 1, \dots, p-1$, with ξ_1, \dots, ξ_{p-1}

equally-spaced kernel locations, $\psi = 25$ and $p = 20$. We assumed there was a single species, $\Theta_1 = (\Theta_{11}, \dots, \Theta_{1p})'$, in the global family, and a single species, $\Theta_2 = (\Theta_{21}, \dots, \Theta_{2p})'$, in the local family. We sampled the elements in Θ_1 and Θ_2 independently from a mixture of a point mass at 0 (with probability 0.5) and a standard normal density. The point mass allowed a different subset of basis functions to be important for each of the two species. Then, assuming there were $n = 16$ individual functions, we let $\theta_{ij} = z_{ij}\Theta_{1j} + (1 - z_{ij})\Theta_{2j}$, for $j = 1, \dots, p$, with z_{ij} sampled independently from Bernoulli(0.5). One can generate a very rich variety of curves through locally selecting from among a small number of vectors of unique basis coefficients. In generating the data, we assumed 40 equally-spaced observations along each curve with a measurement error variance of $\sigma^2 = 0.2$.

We initially implemented the original formulation of the local partition process using the slice sampler proposed in §3.1. A Bayesian specification was completed with $\text{Ga}(1, 1)$ hyperpriors for α and β , a $\text{Ga}(0.1, 0.1)$ hyperprior for σ^{-2} , and P_0 chosen to correspond to the multivariate Gaussian distribution with mean zero and identity covariance. The slice sampler was run for 25,000 iterations, with the first 5,000 samples discarded as a burn-in and every 20th sample collected to thin the chain. The sampler appeared to converge rapidly and to mix efficiently based on examination of trace plots of α, β, σ^2 and function values at a variety of locations and for a variety of individuals. Due to label switching issues, it is not reliable to monitor the elements of Θ_h in assessing convergence and mixing.

Figure 2 shows the data, true curves and estimates for each subject. We clearly do a good job in estimating each of the functions, with all the estimates very close to the truth except for the subject in the (2, 3) panel. However, even for this subject the shape of the function was estimated well. As summaries of overall performance, we estimated the average absolute bias across all locations on all curves (0.192), the mean square error (0.060), the maximum absolute bias (0.506) and the average width of the 99% pointwise credible interval (1.01). The estimated posterior mean of α was 0.17, with a 95% credible interval of [0.10, 0.29]. The

small value of α suggests the individuals are assigned to few local and global species. Indeed, we find that the posterior means of ϕ_j^* are less than 2 for $j = 0, 1, \dots, p$, with $\hat{\phi}_0^* = 1.62$ and $\hat{\phi}_j^* \in [0.99, 1.30]$, for $j = 1, \dots, p$. The estimated posterior mean of β was 1.23, with a 95% credible interval of $[0.51, 2.42]$, suggesting the probability of allocation to local species is close to 0.5.

For comparison, we also implemented a Dirichlet process mixture model using the slice sampler and the same hyperparameter values chosen above (for those common to both models). We ran the slice sampler for the same number of iterations as previously, and obtained similar convergence and mixing results. However, the Dirichlet process analysis induced only minimal borrowing of information across the different functions, with posterior probability of allocating the 16 functions to 13 clusters greater than 0.99. The resulting function estimates were still reasonable, but the performance was not nearly as good as for the local partition process. In particular, the average absolute bias was increased by 26%, the mean square error by 60%, and the maximum absolute bias by 57%. The average credible interval widths were similar, but the true functions fell well outside the 99% intervals in 2/16 of the cases. This was the first simulation scenario we tried, but we found a general improvement for the local partition process over the Dirichlet process in a broad variety of scenarios, with the gain very substantial in many cases. The gains are most notable when data for each function are sparse. For example, we obtained notably larger gains for the local partition process when repeating our simulation for 10 or 20 observations per individual.

We also implemented the slice sampler for the second formulation of the local partition process. As anticipated, this resulted in a substantial saving in computational time compared with the first formulation. In addition, the results were better in terms of function estimation performance, with average absolute bias = 0.168, mean square error = 0.046, maximum absolute bias = 0.520, and an average 99% credible interval width of 0.793. It does not appear that the more concentrated posterior distributions reflected lack of proper account

for uncertainty in that the true functions were all entirely enclosed in the 99% pointwise intervals. Hence, the narrower intervals more likely reflect an improvement in efficiency due to the incorporation of fewer parameters in the second formulation.

5. HORMONE CURVE APPLICATION

5.1. *Background and motivation*

To further illustrate the approach, we considered an application to post-ovulatory progesterone data collected in early pregnancy for $n = 165$ women in the Wilcox et al. (1988) Early Pregnancy Study. In particular, the progesterone metabolite, PdG, was measured in urine based on daily samples. Letting y_{it} denote the measurement of $\log(\text{PdG})$ for woman i t days after the estimated day of ovulation, we assume $y_{it} \sim t_{\kappa}(\eta_i(s_{it}), \sigma^2)$, where $t_{\kappa}(\mu, \sigma^2)$ denotes the t-density, with mean μ , degrees of freedom κ and scale σ^2 . In addition, η_i is a measurement error-corrected smooth trajectory in $\log(\text{PdG})$ for woman i , and s_{it} denotes the timing of the measurement relative to ovulation. We generalize the model in (1) to allow t-distributed measurement errors, because it is well known that the assay error distribution has heavy tails and contains outliers. The data contain between $n_i = 4$ and $n_i = 40$ observations per woman, with an average of $\bar{n} = 23.1$. In general, the data are collected daily in the morning, but there are gaps, with most of these gaps corresponding to censoring in which the woman stopped collecting urine prior to day 40.

The goal of our analysis is to obtain a flexible, yet parsimonious representation of the progesterone trajectories. Bigelow & Dunson (2008) previously analyzed these data using a Dirichlet process prior for the distribution of the woman-specific basis coefficients in a spline model assuming normally-distributed measurement errors. In their analysis, they estimated 31 PdG trajectory clusters, with 18 of these being singletons containing just one woman. Given that many of their reported trajectory clusters have similar shapes with only local deviations, it is our expectation that replacing the Dirichlet process prior on the basis

coefficients with a local partition process prior can produce a more parsimonious and flexible representation of the data.

In analyzing the data, we used the same approach applied in §4 for the simulated data after standardizing time to the $[0, 1]$ interval and adapting the approach to accommodate t-distributed measurement errors. This adaptation was straightforward by expressing the t-distribution as a scale mixture of normals by letting $y_{it} \sim N(\eta_i(s_{it}), \xi_i^{-1}\sigma^2)$, with $\xi_i \sim \text{Ga}(\kappa/2, \kappa/2)$ and with the degrees of freedom κ assigned a $\text{Ga}(1, 1)$ hyperprior to favour very heavy tails to correspond with our prior knowledge. Under this structure, the full conditional posterior distribution for ξ_i is gamma, while we update κ using Metropolis-Hastings.

5.2. Analysis and results

We focus initially on the first formulation of the local partition process. As in the simulated examples, we obtained good rates of mixing and apparent convergence. Figure 3 plots the data, posterior mean trajectories and 99% pointwise credible intervals for 16 randomly selected women. It is clear from this plot and from examination of plots for the other women in the study (not shown) that we obtain an excellent fit to the data. In addition, the estimated value of the hyper-parameter α was $\hat{\alpha} = 0.41$, with a 95% credible interval of $[0.25, 0.63]$. This value suggests that there are few unique global or local vectors of basis coefficients that are occupied by the subjects, so that we obtain a sparse representation of the data. Indeed, the posterior probability of $\phi_0^* = 3$ was 0.97, suggesting that all subjects were allocated to one of three global species (unique vector of basis coefficients) with high probability. The number of local species occupied by the subjects also tended to be small, having an estimated average value of 2.91.

In addition, the estimated value of the hyper-parameter β was 1.05, with a 95% credible interval of $[0.63, 1.59]$. Recall that values close to 0 for β provide support for a joint Dirichlet process prior on the distribution of the basis coefficients, while large values provide support

for independent Dirichlet process priors for the different coefficients. A value close to one instead provides evidence for a local partition process balanced between the two extremes, with a 50-50 chance of a randomly selected basis coefficient being drawn from a global versus local species. Hence, there is clear evidence in the data favoring our proposed approach over the Dirichlet process. The estimated degrees of freedom in the t-distribution was 2.05, with a 95% credible interval of [2.01, 2.11], suggesting very heavy tails. Commenting further on the trajectories estimated in Figure 3, we note that many of the trajectories increase rapidly after ovulation and then flatten out within a week or two, while other trajectories increase initially and then decrease. It is likely that the increasing trajectories correspond to healthy pregnancies, while the trajectories that peak shortly after implantation (typically 7-10 days after ovulation) and then decline correspond to early pregnancy losses.

Repeating the analysis for a joint Dirichlet process prior, the posterior mean of α was 3.60, with a 95% credible interval of [1.84, 6.32]. The α parameter was slow-mixing in the Markov chain Monte Carlo implementation, with high autocorrelation. However, because traceplots of the subject-specific basis coefficients and function estimates were well behaved, the mixing problems with α did not appear to impact our inferences. The posterior mean number of clusters was 21.4, with a 95% interval of [15, 37], reflecting large posterior uncertainty in clustering and a substantially larger number of clusters occupied than in the local partition process analysis. Figure 4 shows the estimated trajectories for the same women plotted in Figure 3. In most cases, the estimates are similar to those obtained for the LPP₁ analysis. However, there are several exceptions in which the Dirichlet process approach produced an estimate inconsistent with the data. For example, consider the subplots in the (4,2) and (4,3) coordinates. We noted similarly poor performance for many of the other women in the study, particularly those having relatively few observations. This likely reflects a tendency of the joint Dirichlet process to overly-favour clustering together of subjects unless there is abundant data available to suggest this clustering is not supported. Hence, in sparse

data situations, one anticipates dramatic gains for the local partition process.

Finally, we implemented the sparse variant of the local partition process. The estimate of α was 0.38 with a 95% interval of $[0.10, 0.95]$, the estimate of β was 1.87 with a 95% interval of $[1.10, 2.92]$, and the estimated number of occupied species was very similar to that for the LPP_1 . The estimates for the 16 example women are shown in Figure 5. It is clear the estimates are very similar to those for the LPP_1 . However, there was a substantial decrease in the time required per iteration of the slice sampler. The first formulation took 112 seconds per 100 iterations in Matlab on a MacBook Pro laptop, while the Dirichlet process implementation took 30 seconds and the sparse local partition process took 69 seconds. We repeated each of the analyses for a variety of hyperparameter values, with the variance multiplied by 2 and divided by 2 for α, β, P_0 and κ . There were no noticeable differences in the results.

6. DISCUSSION

This article has proposed a generalization of the Dirichlet process, which allows dependent local clustering and borrowing of information. Our particular emphasis was on functional data analysis applications in which each function is characterized as a linear combination of basis functions, and the goal is to flexibly borrow information to more efficiently estimate the individual functions. In order to favour more borrowing of information across the individual functions and obtain a more parsimonious representation of the data, one can allow the basis coefficients for an individual to be locally selected from a small number of vectors of unique basis coefficients. The vectors of unique basis coefficients can be viewed as representing underlying commonalities across the different functions, resulting in a discrete nonparametric analog of functional principal components. This type of idea seems very promising as a tool for generating sparse characterizations of complex multivariate and functional data in a variety of settings. Our proposed local partition process provides a sim-

ple, yet flexible approach for characterizing the local selection process. The slice sampling implementation we have proposed is quite simple and efficient to implement, while allowing posterior computation for the infinite-dimensional nonparametric process instead of a finite approximation.

ACKNOWLEDGMENT

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

APPENDIX

Proofs

Proof of Proposition 1.

In order to derive the marginal probability of local clustering, $\text{pr}(\theta_{ij} = \theta_{i'j})$, let

$$\begin{aligned} \text{pr}(\theta_{ij} = \theta_{i'j}) &= \int \sum_{l=0}^1 \sum_{m=0}^1 \text{pr}(\theta_{ij} = \theta_{i'j} \mid z_{ij} = l, z_{i'j} = m) \text{pr}(z_{ij} = l, z_{i'j} = m \mid \pi_j) df(\pi_j) \\ &= \int \frac{1}{1+\alpha} \{ \pi_j^2 + (1-\pi_j)^2 \} \frac{1}{B(1, \beta)} (1-\pi_j)^{\beta-1} d\pi_j \\ &= \frac{1}{1+\alpha} \left\{ \left(\frac{1}{1+\beta} \right) \left(\frac{2}{2+\beta} \right) + \frac{\beta}{2+\beta} \right\}, \end{aligned}$$

with Proposition 1 following directly, where $f(\pi_j)$ denotes the $\text{Be}(1, \beta)$ prior distribution for π_j and $B(a, b)$ denotes the beta(a, b) function.

Proof of Proposition 2.

The joint probability of local clustering follows along similar lines,

$$\begin{aligned} &\text{pr}(\theta_{ij} = \theta_{i'j}, \theta_{ij'} = \theta_{i'j'}) \\ &= \int \frac{1}{1+\alpha} \left[\frac{1}{1+\alpha} \{ \pi_j^2 (1-\pi_{j'})^2 + (1-\pi_j)^2 \pi_{j'}^2 + (1-\pi_j)^2 (1-\pi_{j'})^2 \} + \pi_j^2 \pi_{j'}^2 \right] \\ &\quad \times \frac{1}{B(1, \beta)^2} (1-\pi_j)^{\beta-1} (1-\pi_{j'})^{\beta-1} d\pi_j d\pi_{j'} \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1}{1+\alpha}\right)^2 \left\{ \frac{4\beta}{(1+\beta)(2+\beta)^2} + \left(\frac{\beta}{2+\beta}\right)^2 \right\} + \left(\frac{1}{1+\alpha}\right) \left(\frac{1}{1+\beta}\right)^2 \left(\frac{2}{2+\beta}\right)^2 \\
&= \left(\frac{1}{1+\alpha}\right)^2 \left(\frac{1}{2+\beta}\right)^2 \left\{ \left(\beta + \frac{2}{1+\beta}\right)^2 + \frac{4\alpha}{(1+\beta)^2} \right\}.
\end{aligned}$$

Proof of Proposition 3.

Express the joint probability of local clustering as

$$\begin{aligned}
\text{pr}(\theta_{ij} = \theta_{i'j}, \theta_{ij'} = \theta_{i'j'}) &= \int \sum_{l_1=0}^1 \sum_{l_2=0}^1 \text{pr}(\theta_{ij} = \theta_{i'j} | z_{ij} = l_1, z_{i'j} = l_2) \text{pr}(z_{ij} = l_1, z_{i'j} = l_2 | \pi_j) df(\pi_j) \\
&\quad \times \int \sum_{l_3=0}^1 \sum_{l_4=0}^1 \text{pr}(\theta_{ij'} = \theta_{i'j'} | \theta_{ij} = \theta_{i'j}, z_{ij'} = l_3, z_{i'j'} = l_4) \text{pr}(z_{ij'} = l_3, z_{i'j'} = l_4 | \pi_{j'}) df(\pi_{j'}) \\
&= \left[\frac{1}{1+\alpha} \int \{\pi_j^2 + (1-\pi_j)^2\} df(\pi_j) \right] \left[\frac{6+\alpha}{(2+\alpha)(3+\alpha)} \int (1-\pi_{j'})^2 df(\pi_{j'}) + \int \pi_{j'} df(\pi_{j'}) \right] \\
&= \left(\frac{1}{1+\alpha}\right) \left(\frac{1}{2+\beta}\right)^2 \left(\beta + \frac{2}{1+\beta}\right) \left\{ \frac{(6+\alpha)\beta}{(2+\alpha)(3+\alpha)} + \frac{2}{1+\beta} \right\}.
\end{aligned}$$

Proof of Theorem 1.

To demonstrate property 1, first note that

$$Q(\psi) = \left\{ \prod_{j:\psi_j \in I_o} \pi_j \nu_{0\psi_j}^* \prod_{l < \psi_j} (1 - \nu_{0l}^*) \right\} \left\{ \prod_{j:\psi_j \in I_e} (1 - \pi_j) \nu_{j\psi_j}^* \prod_{l < \psi_j} (1 - \nu_{jl}^*) \right\},$$

where $\psi \in \{1, 2, \dots, \infty\}^p$, $I_o = \{1, 3, 5, \dots\}$, and $I_e = \{2, 4, 6, \dots\}$. Let $\bar{Q}(\psi)$ denote the median of the distribution of $Q(\psi)$ for fixed ψ . Note that $\bar{Q}(\psi)$ is expressed as a product of finitely many random variables in $(0, 1)$, so that $\text{pr}\{Q(\psi) > 0\} = 1$. It follows that $\bar{Q}(\psi) > 0$ and hence $\text{pr}\{Q(\psi) > \epsilon_\psi\} = 0.5$, with $\epsilon_\psi = \bar{Q}(\psi) > 0$. Letting $\epsilon = \min[\epsilon_\psi, \psi \in \{1, 2, \dots, \infty\}^p]$, we have $\text{pr}\{Q(\psi) > \epsilon\} \geq 0.5$.

To demonstrate property 2, first note that

$$\text{pr}\{Q(\theta_{ij} = \theta_{i'j}) \in A\} = \int 1\{\rho(\alpha, \beta) \in A\} df(\alpha, \beta),$$

where $\rho(\alpha, \beta)$ is defined in proposition 1, with the dependence on α and β now explicit, and $f(\alpha, \beta)$ is the prior density for α, β on $(0, \infty) \times (0, \infty)$. For any point $a \in A$, there exists

a corresponding region $d(a) \subset (0, \infty) \times (0, \infty)$ such that $\rho(\alpha, \beta) = a$ for all $(\alpha, \beta) \in d(a)$. Letting $d(A) = \cup_{a \in A} d(a)$, $\text{pr}\{Q(\theta_{ij} = \theta_{i'j}) \in A\} = \int_{d(A)} f(\alpha, \beta) d\alpha d\beta$. For all Borel subsets $A \subset (0, 1)$, the area of $d(A)$ is greater than 0, so $\text{pr}\{Q(\theta_{ij} = \theta_{i'j}) \in A\} > 0$ as long as $f(\alpha, \beta) > 0$ for all $(\alpha, \beta) \in (0, \infty) \times (0, \infty)$, which holds for independent gamma priors on α and β .

Property 3 follows from propositions 1 and 2. To demonstrate property 4, note that

$$\text{pr}\left\{\frac{Q(\theta_{ij} = \theta_{i'j}, \theta_{ij'} = \theta_{i'j'})}{Q(\theta_{ij'} = \theta_{i'j'})} \in A\right\} = \int 1\{\rho_2(\alpha, \beta) \in A\} f(\alpha, \beta) d\alpha d\beta,$$

where $\rho_2(\alpha, \beta) = \text{pr}(\theta_{ij} = \theta_{i'j} | \theta_{ij'} = \theta_{i'j'}, \alpha, \beta)$. Hence, given that $f(\alpha, \beta) > 0$ for all $(\alpha, \beta) \in (0, \infty) \times (0, \infty)$, it suffices to show that there exists a region $d_2(a) \subset (0, \infty) \times (0, \infty)$ of (α, β) values that result in $\rho_2(\alpha, \beta) = a$, with $d_2(a) \neq \emptyset$ for all $a \in (0, 1)$. This condition follows directly if there exists an (α, β) solution to the equations $\rho(\alpha, \beta) = a, \rho_2(\alpha, \beta) = b$, for every point (a, b) in $0 < a < b < 1$, which is easily verified to hold.

REFERENCES

- BIGELOW, J.L. & DUNSON, D.B. (2008). Bayesian semiparametric joint models for functional predictors. *J. Am. Statist. Assoc.*, under invited revision.
- BUSH, C.A. & MACEACHERN, S.N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika*, **83**, 275-85.
- CIFARELLI, D.M. & REGAZZINI, E. (1978). Nonparametric statistical problems under partial exchangeability: The use of associative means (in Italian). *Annali dell'Instituto di Matematica Finanziara dell'Universitá di Torino, Serie III*, 12, 1-36.
- DE IORIO, M., MÜLLER, P., ROSNER, G. L. & MACEACHERN, S. N. (2004). An ANOVA model for dependent random measures. *J. Am. Statist. Assoc.*, **99**, 205-15.

- DUNSON, D.B., HERRING, A.H. & SIEGARIZ, A.M. (2008a). Bayesian inference on changes in response densities over predictor clusters. *J. Am. Statist. Assoc.*, to appear.
- DUNSON, D.B., XUE, Y. & CARIN, L. (2008b). The matrix stick-breaking process: Flexible Bayes meta analysis. *J. Am. Statist. Assoc.*, to appear.
- FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209-30.
- FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.*, **2**, 615-29.
- GELFAND, A.E., KOTTAS, A. & MACEACHERN, S.N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J. Am. Statist. Assoc.*, **100**, 1021-35.
- HOFF, P.D. (2006). Model-based subspace clustering. *Bayesian Analysis*, **1**, 321-44.
- ISHWARAN, H. & JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Am. Statist. Assoc.* **96**, 161-73.
- ISHWARAN, H. & TAKAHARA, G. (2002). Independent and identically distributed Monte Carlo algorithms for semiparametric linear mixed models. *J. Am. Statist. Assoc.*, **97**, 1154-66.
- KLEINMAN, K.P. & IBRAHIM, J.G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics*, **54**, 921-38.
- MACEACHERN, S. N. (1999). Dependent Nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association.

- MACEACHERN, S. N. (2001), Decision Theoretic Aspects of Dependent Nonparametric Processes. In *Bayesian Methods With Applications to Science, Policy, and Official Statistics*, Ed. E. George, Creta: ISBA, pp551-60.
- NEAL, R.M. (2003). Slice sampling. *Ann. Statist.*, **31**, 705-41.
- PAPASPILIOPOULOS, O. & ROBERTS, G. (2008). Retrospective MCMC for Dirichlet process hierarchical models. *Biometrika*, to appear.
- PETRONE, S., GUINDANI, M. & GELFAND, A.E. (2008). Hybrid Dirichlet processes for functional data. *J. R. Statist. Soc. B*, under invited revision.
- RAY, S. & MALLICK, B. (2006). Functional clustering by Bayesian wavelet methods. *J.R. Statist. Soc. B* **68**, 305-32.
- SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica*, **4**, 639-50.
- WALKER, S.G. (2007). Sampling the Dirichlet mixture model with slices. *Comm. Statist. Sim. Comput.*, **36**, 45-54.
- WILCOX, A., WEINBERG, C.R., O'CONNOR, J., BAIRD, D., SCHLATTERER, J., CANFIELD, R., ARMSTRONG, E. & NISULA, B. (1988). Incidence of early loss of pregnancy. *N. Engl. J. Med.*, **319**, 189-94.

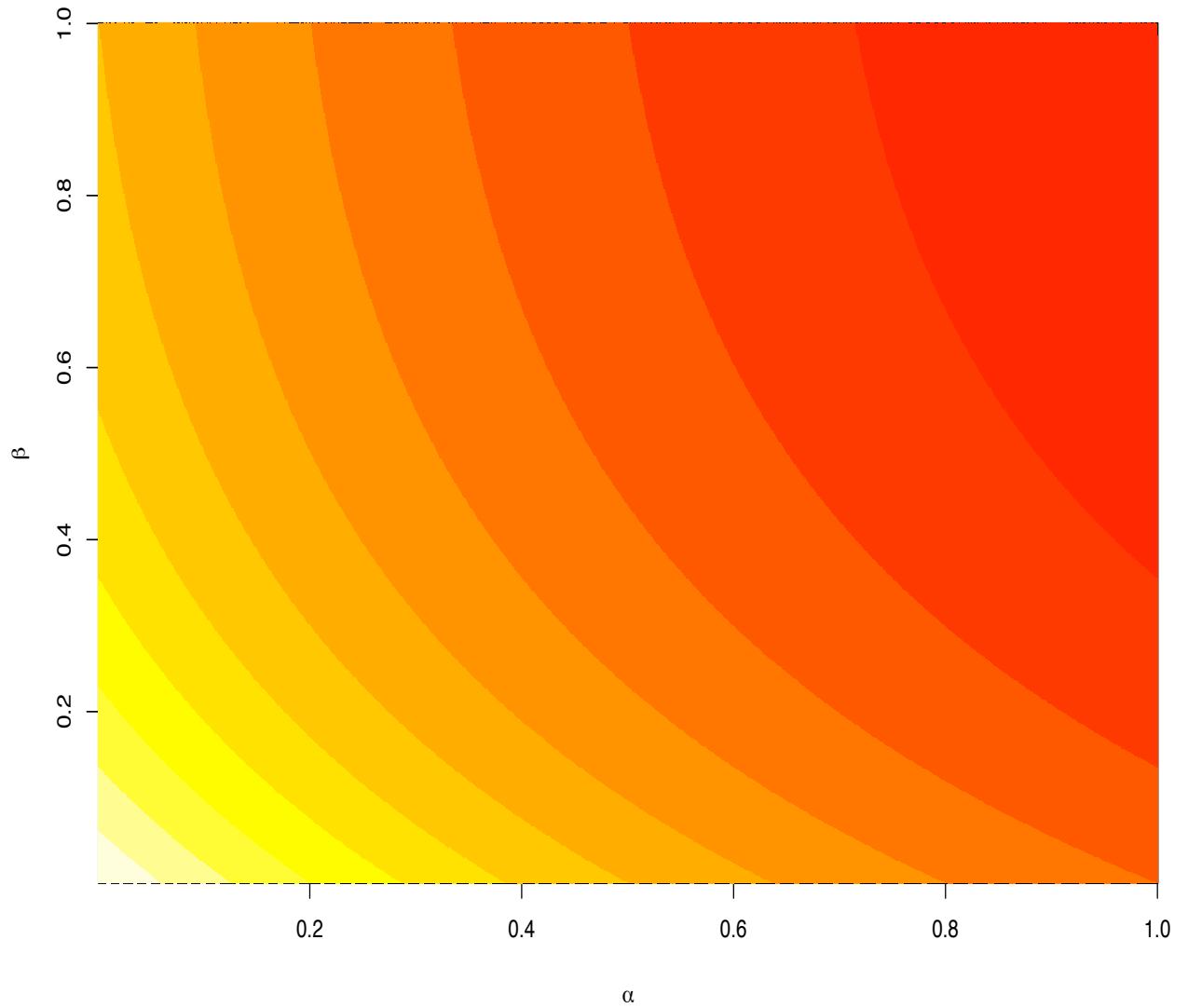


Fig. 1. Heat plot of the marginal probability of local clustering, $\text{pr}(\theta_{ij} = \theta_{i'j'})$, as a function of the hyperparameters α and β . Values range from $1/3$ to 1 as colors change from red to yellow to white.

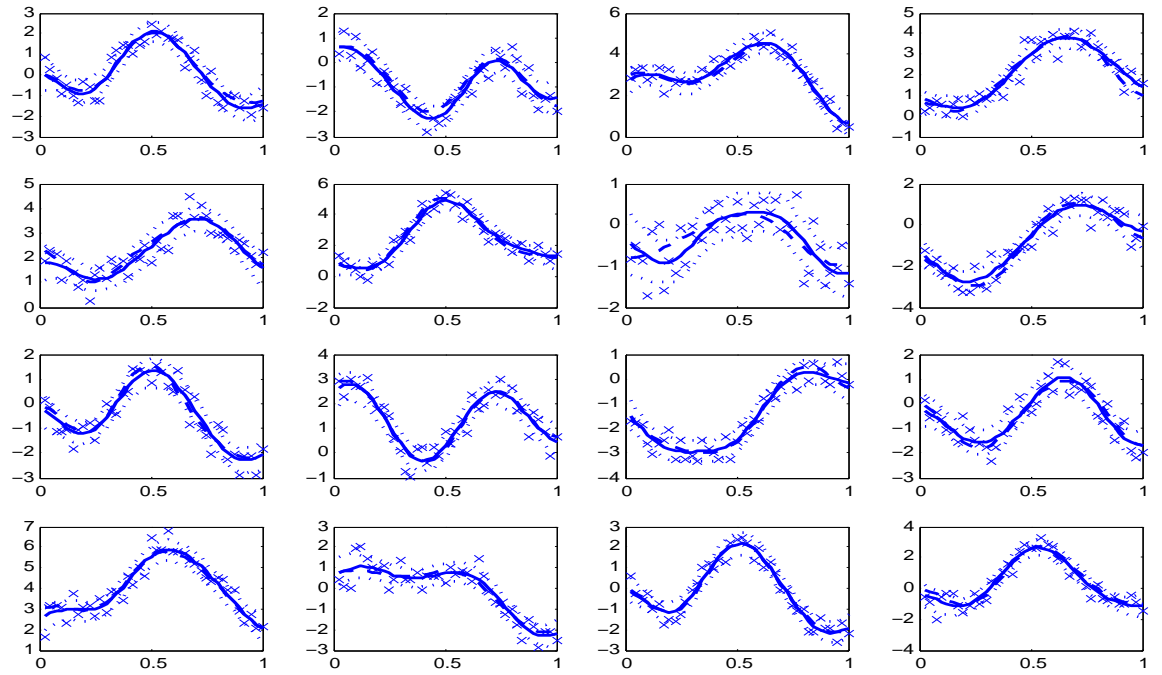


Fig. 2. Data and results for the simulation example. Each panel corresponds to one of the 16 subjects in the study, data points are marked with \times , the true functions are represented with dashed lines, the posterior means with solid lines, and 99% pointwise credible intervals with dotted lines.

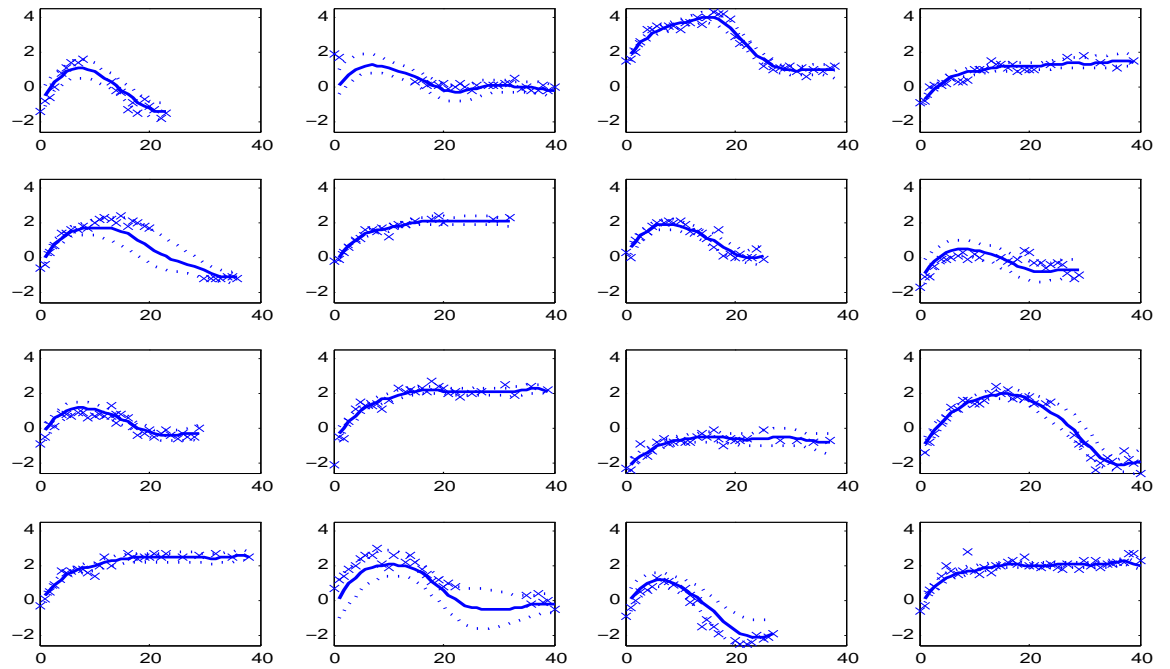


Fig. 3. Log(PdG) data and LPP₁-based function estimates for 16 women randomly selected from the women in the Early Pregnancy Study. The data points are marked with \times , the posterior means are solid lines, and 95% pointwise credible intervals are dotted lines. The x-axis scale is time in days starting at the estimated day of ovulation.

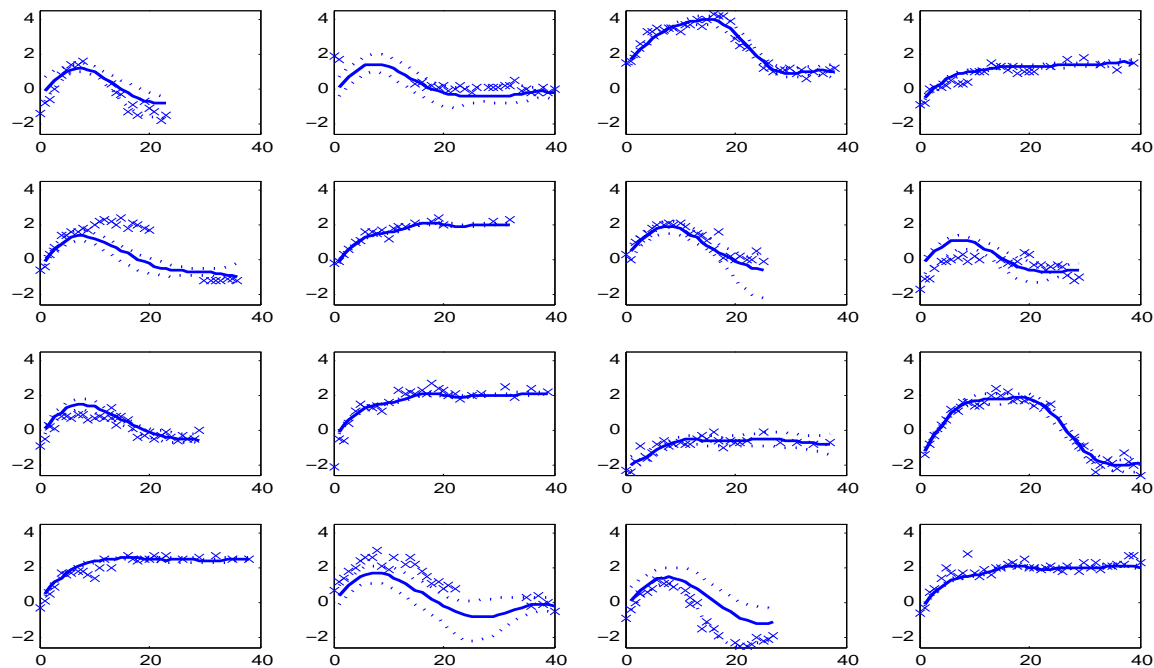


Fig. 4. $\text{Log}(\text{PdG})$ data and DP-based function estimates for the same 16 women considered in Fig. 3. The posterior means are solid lines, and 95% pointwise credible intervals are dotted lines. The x-axis scale is time in days starting at the estimated day of ovulation.

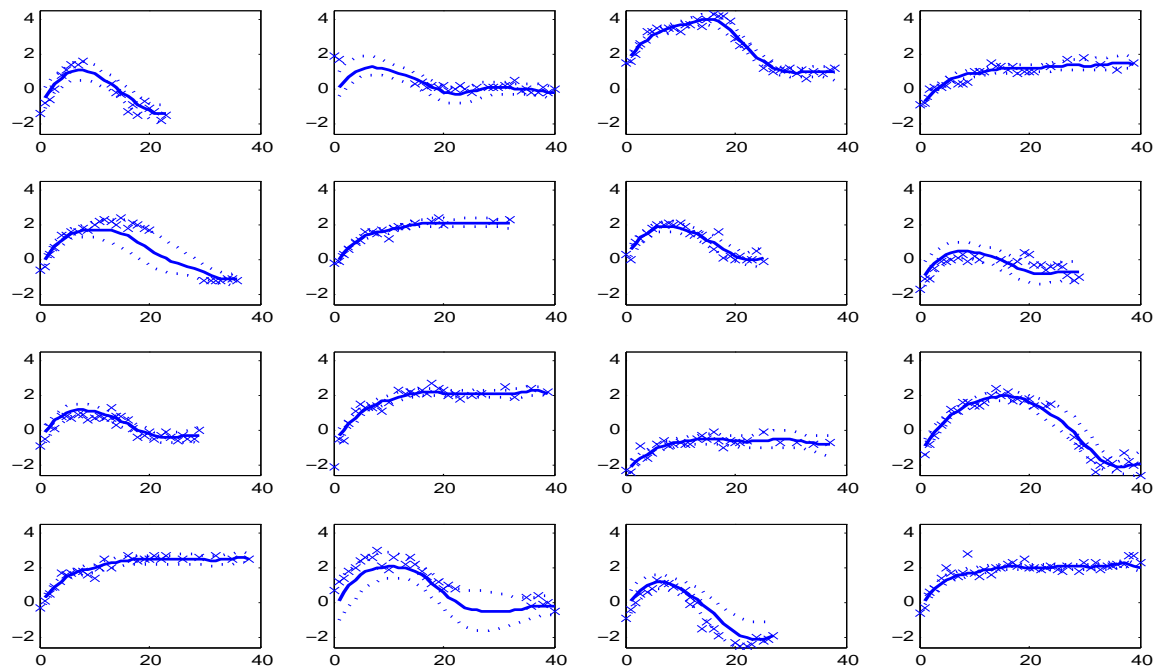


Fig. 5. Log(PdG) data and LPP₂-based function estimates for the same 16 women considered in Figs. 3-4. The posterior means are solid lines, and 95% pointwise credible intervals are dotted lines. The x-axis scale is time in days starting at the estimated day of ovulation.