Published in final edited form as:
Biometrika. 2009 ; 96(2): 249–262. doi:10.1093/biomet/asp021.

Nonparametric Bayes local partition models for random effects

DAVID B. DUNSON
Department of Statistical Science, Box 90251, Duke University, Durham, North Carolina 27708, U.S.A

DAVID B. DUNSON: dunson@stat.duke.edu

Summary

This paper focuses on the problem of choosing a prior for an unknown random effects distribution within a Bayesian hierarchical model. The goal is to obtain a sparse representation by allowing a combination of global and local borrowing of information. A local partition process prior is proposed, which 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. Some basic properties 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.

Keywords

Dirichlet process; Functional data; Local shrinkage; Meta-analysis; Multi-task learning; Partition model; Slice sampling; Stick-breaking

1. Introduction

Random effects models are commonly used in the analysis of longitudinal and functional data. Suppose that data on subject i ($i = 1, \ldots, n$) consist of n_i measurements, $y_i = (y_{i1}, \ldots, y_{in_i})^{\mathrm{T}}$, collected at times $t_i = (t_{i1}, \ldots, t_{in_i})^{\mathrm{T}}$, and let

$$\begin{aligned} y_{is} &= \eta_i(t_{is}) + \varepsilon_{is}, \quad \varepsilon_{is} \sim N(0, \sigma^2), \\ \eta_i(t) &= \sum_{j=1}^{p} \theta_{ij} b_j(t), \quad \theta_i \sim P, \end{aligned} \tag{1}$$

where η_i is a function for subject i, $b = \{b_j\}_{j=1}^{p}$ are basis functions, σ^2 is the measurement error variance, $\theta_i = (\theta_{i1}, \ldots, \theta_{ip})^{\mathrm{T}}$ are basis coefficients for subject i and heterogeneity in the curves is characterized through the random effects distribution P. A potential concern is sensitivity to the choice of random effects distribution. As described by Heard et al. (2006), one strategy is to use a latent class model with $P = \sum_{h=1}^{k} \pi_h \delta_{\Theta_h}$, where δ_{Θ} denotes a degenerate distribution with all its mass on Θ, π_h is the probability that $\theta_i = \Theta_h$ and k is the number of functional clusters. Dirichlet process priors (Ferguson, 1973, 1974) allow $k = \infty$ and increasing numbers of clusters with sample size, and have been widely used for random effects distributions (Bush & MacEachern, 1996; Müller & Rosner, 1997).

Although much of the recent literature on mixture models has focused on clustering as a primary goal, such models are useful even if there is no interest in clustering. For example, in the analysis of longitudinal data, one may be interested in inferences on the average

trajectory over time, in estimating variability in the trajectories across subjects, or in inferences on fixed effects coefficients, with the random effects considered a nuisance. In each of these cases, it is desirable to allow the random effects distribution, P, to be unknown. In many applications, p is moderate to large, so that it is necessary to obtain a sparse approach for characterizing the unknown P due to the curse of dimensionality.

Sparsity can be formalized through favouring a small number of components. This can be accomplished through a Dirichlet process prior, $P \sim \mathrm{DP}(\alpha P_0)$, where α is a concentration parameter and P_0 is a base measure. The number of unique values of $\{\theta_i\}_{i=1}^n$ increases at a rate proportional to $\alpha \log n$, so that the number of occupied components tends to be small relative to n. However, a drawback of the Dirichlet process is the assumption of global partitioning. In particular, $\theta_{ij} = \Theta_{hj}$ implies that $\theta_{ij'} = \Theta_{hj'}$ for all $j' \neq j$. Hence, two subjects in the same cluster for any of their random effects are forced to be in the same cluster for all of their random effects. In functional data analysis, it is common for two subjects to have similar trajectories across certain regions while having local deviations. Under the Dirichlet process, either such subjects will be inappropriately clustered together, obscuring local differences, or they will be allocated to separate clusters.

This paper proposes a class of local partition process priors for parsimonious modelling of unknown random effects distributions. The general structure can be characterized as follows:

$$\theta_{ij} = \Theta_{\psi_{ij}\, j}, \quad \Theta_h \sim P_0, \quad \psi_i = (\psi_{i1}, \ldots, \psi_{ip})^{\mathrm{T}} \sim Q, \quad (2)$$

where $\psi_{ij} \in \{1, \ldots, \infty\}$ denotes the cluster index for the jth random effect from subject i, $\theta_i = \Theta_{\psi_i} = (\Theta_{\psi_{i1}1}, \ldots, \Theta_{\psi_{ip}p})^{\mathrm{T}}$, the elements of $\Theta = \{\Theta_h\}_{h=1}^{\infty}$ are independent and identically distributed, and Q is a probability distribution over $\{1, 2, \ldots, \infty\}^p$.

In specifying Q, it is convenient to let $\Theta = \Theta_0 \cup \Theta_1$, with $\Theta_0 = \{\Theta_{0h}\}_{h=1}^{\infty}$ global coefficient vectors and $\Theta_1 = \{\Theta_{1h}\}_{h=1}^{\infty}$ local coefficient vectors. As a convention, let Θ_0 denote the odd-numbered elements of Θ, with Θ_1 the even-numbered elements. To induce $\psi_i \sim Q$, let

$$\psi_{ij} = z_{ij}(2\phi_{i0} - 1) + (1 - z_{ij})2\phi_{ij}, \quad z_i \sim G, \quad \phi_i = (\phi_{i0}, \phi_{i1}, \ldots, \phi_{ip})^{\mathrm{T}} \sim H, \quad (3)$$

where $z_{ij} = 1$ denotes allocation of θ_{ij} to a global component, $z_{ij} = 0$ denotes allocation to a local component, $\phi_{i0} \in \{1, 2, \ldots, \infty\}$ is a global cluster index for subject i and $\phi_{ij} \in \{1, 2, \ldots, \infty\}$ is a local cluster index for subject i and element j. From (2)-(3), the ith subject's random effects vector is equivalent to $\Theta_{0\phi_{i0}}$ for those elements having $z_{ij} = 1$, while the remaining elements are selected from a variety of local coefficient vectors. This specification allows a combination of global and local borrowing of information. Proofs are included in an Appendix.

2. Dependent local partition process

2·1. Formulation 1

To choose G and H in (3), let

$$z_{ij} \sim \mathrm{Ber}(\nu_j), \quad \nu_j \sim \mathrm{Be}(1, \beta) \quad (j = 1, \ldots, p),$$
$$\phi_{ij} \sim \sum_{h=1}^{\infty} \pi_{jh} \delta_h, \quad \pi_{jh} = \pi^*_{jh} \prod_{l<h} (1 - \pi^*_{jl}), \quad \pi^*_{jh} \sim \mathrm{Be}(1, \alpha) \quad (j = 0, 1, \ldots, p), \quad (4)$$

where ν_j is the probability of allocation to the global component for the jth random effect, β is a hyperparameter controlling the overall weight on the local components and the elements

of ϕ_i are assigned independent stick-breaking priors, with the hyperparameter α controlling how rapidly the weights π_{jh} decrease towards zero as the index h increases. The stick-breaking priors are chosen following the Sethuraman (1994) representation of the Dirichlet process for simplicity, though extensions to more general stick-breaking priors (Ishwaran & James, 2001) are straightforward. Let $P \sim \text{LPP}_1(\alpha, \beta, P_0)$ denote the prior on the unknown random effects distribution induced through (2)-(4).

The prior $P \sim \text{LPP}_1(\alpha, \beta, P_0)$ induces borrowing of information across subjects through both global and local clustering. In particular, if subjects i and i' are assigned to the same global cluster, then $\phi_{i0} = \phi_{i'0}$, and it becomes more likely that these subjects will have identical values for multiple elements of their random effects vectors, particularly if β is not large. For small β and high-dimensional random effects vectors, subjects in the same global clusters will have identical values for most of their basis coefficients, while having occasional local deviations. This structure addresses the limitations of Dirichlet process priors mentioned in § 1. In a recent application using Dirichlet process priors for the distribution of a high-dimensional latent factor vector underlying gene expression measurements, we noticed a tendency of the Dirichlet process to induce a large number of clusters, as even individuals that were quite similar overall had important local differences. In contrast, the local partition process induced a small number of global and local clusters, leading to substantial gains in efficiency.

One of the appealing characteristics of the Dirichlet process prior is the availability of simple expressions for the prior probability of clustering marginalizing out P. In particular, $\theta_i \sim P$ with $P \sim \text{DP}(\alpha P_0)$ implies that $\text{pr}(\theta_i = \theta_{i'}) = 1/(1 + \alpha)$. This illustrates the role of the precision parameter α in controlling the tendency of individuals to be clustered together. As described in Proposition 1, a similarly simple expression is obtained for the marginal prior probability of $\text{pr}(\theta_{ij} = \theta_{i'j})$ under $P \sim \text{LPP}_1(\alpha, \beta, P_0)$.

Proposition 1—*Assuming $\theta_i \sim P$ with $P \sim \text{LPP}_1(\alpha, \beta, P_0)$,*

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

Hence, α and β are key hyperparameters controlling the clustering probability ρ with ρ monotone decreasing in α and β. In the limit as $\alpha, \beta \to 0$, $\rho = 1$ and all individuals are automatically grouped into a single cluster with $\theta_i = \theta$ for all i. To allow the data to provide information about appropriate values for α, β, let $\alpha \sim \text{Ga}(a_\alpha, b_\alpha)$ and $\beta \sim \text{Ga}(a_\beta, b_\beta)$.

If subjects i and i' are in the same cluster for the j'th random effect, so that $\theta_{ij'} = \theta_{i'j'}$, one has information that these subjects are similar. Ideally, such information would be propagated to other random effects, increasing the probability of $\theta_{ij} = \theta_{i'j}$ for $j \neq j'$. Proposition 2 demonstrates that the prior $P \sim \text{LPP}_1(\alpha, \beta, P_0)$ automatically induces the desired dependence in local clustering. In addition, the dependence structure has a simple form, which provides insight into the flexibility of the approach and role of the hyperparameters.

Proposition 2—*Assuming $\theta_i \sim P$ with $P \sim \text{LPP}_1(\alpha, \beta, P_0)$,*

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

The joint probability in Proposition 2 is strictly larger than the product of the marginal probabilities from Proposition 1, which implies that $\text{pr}(\theta_{ij} = \theta_{i'j})$ increases given knowledge that $\theta_{ij'} = \theta_{i'j'}$. The degree of positive dependence in local clustering is monotonely increasing with α and decreasing with β, being controlled by the size of $4\alpha/(1+\beta)^2$.

As $P \sim \text{LPP}_1(\alpha, \beta, P_0)$ provides a prior for the unknown random effects distribution P, it is important to assess basic properties, such as the mean and variance. Following common convention, the notation P is used to denote both a random probability measure and the corresponding distribution. It is assumed that P_0 is a probability measure over a measurable Polish space $(\Omega, \mathcal{B})$, with Ω the sample space and $\mathcal{B}$ the corresponding Borel σ-algebra. Let $\theta_{ij} \sim P_j$, with P_j the jth 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 jth 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 $P_j(B) \stackrel{D}{=} W X_1 + (1-W)X_2$, 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 centred on the base probability measure P_{0j} and the variance is $P_{0j}(B)\{1 - P_{0j}(B)\}\rho$, with ρ the probability of clustering shown in Proposition 1.

Nonparametric Bayes procedures allow uncertainty in parametric specifications through priors with large support. The proposed prior, $P \sim \text{LPP}_1(\alpha, \beta, P_0)$, has weak support on the space of probability measures over $(\Omega, \mathcal{B})$. In addition, as formalized in Theorem 1, the proposed process induces a highly flexible prior on the joint distribution of the local cluster indices, $\psi_i = (\psi_{i1}, \ldots, \psi_{ip})^{\text{T}} \sim Q$, with Q a probability measure on $\{1, 2, \ldots, \infty\}^p$.

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 $\Re^+$. Then, the following properties are satisfied*:

Property 1. $\text{pr}\{Q(\psi) > \varepsilon\} > \delta$ *for all* $\psi \in \{1, 2, \ldots, \infty\}^p$ *and some* $\varepsilon, \delta > 0$;

Property 2. $\text{pr}\{Q(\psi_{ij} = \psi_{i'j}) \in A\} > \varepsilon$ *for all Borel subsets* $A \subset [0, 1]$, $\psi_l \sim Q$, $l = i, i'$, *and some* $\varepsilon > 0$;

Property 3. $\text{pr}\{\dfrac{Q(\psi_{ij}=\psi_{i'j}, \psi_{ij'}=\psi_{i'j'})}{Q(\psi_{ij'}=\psi_{i'j'})} \geqslant Q(\psi_{ij}=\psi_{i'j})\} = 1$, *for all* $j, j' \in \{1, \ldots, p\}$, $j' \neq j$; *and*

Property 4. $\text{pr}\{\dfrac{Q(\psi_{ij}=\psi_{i'j}, \psi_{ij'}=\psi_{i'j'})}{Q(\psi_{ij'}=\psi_{i'j'})} \in A\} > \varepsilon$, *for all Borel subsets* $A \subset [Q(\psi_{ij} = \psi_{i'j}), 1]$, $j, j' \in \{1, \ldots, p\}$, $j' \neq j$ *and for some* $\varepsilon > 0$.

2·2. Formulation 2

In applications involving large p, it is appealing for computational reasons to consider sparse alternatives to $P \sim \text{LPP}_1(\alpha, \beta, P_0)$ that avoid the incorporation of a separate stick-breaking process for each component. As a simplification of (4), replace the second line with

$$\phi_{ij} \sim \sum_{h=1}^{\infty} \pi_h \delta_h, \quad \pi_h = \pi_h^* \prod_{l<h} (1 - \pi_l^*), \quad \pi_h^* \sim \text{Be}(1, \alpha) \ \ (j = 0, 1, \ldots, p).$$

This modified case is denoted as $P \sim \text{LPP}_2(\alpha, \beta, P_0)$. This specification assumes a common prior distribution for each of the elements of ϕ_i, resulting in a more parsimonious model. Under $P \sim \text{LPP}_2(\alpha, \beta, P_0)$, the local clustering probability $\text{pr}(\theta_{ij} = \theta_{i'j})$ is identical to the expression shown in Proposition 1, so the hyperparameters α and β have a similar role under both specifications. The main question is whether one maintains sufficient flexibility in characterizing dependence in local partitioning, which can be addressed through modifying Proposition 2 as follows.

Proposition 3—*Assuming $\theta_i \sim P$ with $P \sim \text{LPP}_2(\alpha, \beta, P_0)$, then*

$$\text{pr}(\theta_{ij}=\theta_{i'j}, \theta_{ij'}=\theta_{i'j'}) = \left(\frac{1}{1+\alpha}\right)\left(\frac{1}{2+\beta}\right)^2\left(\beta+\frac{2}{1+\beta}\right)\left\{\frac{(6+\alpha)\beta}{(2+\alpha)(3+\alpha)}+\frac{2}{1+\beta}\right\}.$$

As in Proposition 2, the joint probability is strictly larger than the product of the marginal clustering probabilities, implying positive dependence in local clustering. To investigate the flexibility in the degree of positive dependence, one can vary the values of α and β widely. Ideally, as α and β are varied, any degree of positive dependence could be obtained, ranging between independence of $\theta_{ij} = \theta_{i'j}$ and $\theta_{ij'} = \theta_{i'j'}$ to perfect dependence, with $\theta_{ij'} = \theta_{i'j'}$ implying $\theta_{ij} = \theta_{i'j}$. Under Property 4 in Theorem 1, the prior $P \sim \text{LPP}_1(\alpha, \beta, P_0)$ allows any degree of positive dependence. In contrast, it follows from Proposition 3 that the prior $P \sim \text{LPP}_2(\alpha, \beta, P_0)$ induces a value for $\text{pr}(\theta_{ij} = \theta_{i'j} \mid \theta_{ij'} = \theta_{i'j'})$ that is constrained to fall in a subset of the interval $[\text{pr}(\theta_{ij} = \theta_{i'j}), 1]$, so that 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.

2·3. Special cases and connections

A variety of priors that have been proposed previously arise as special cases of the local partition process formulation in (2) and (3). Considering the case in which $G = \delta_1$, so that $z_{ij} = 1$ for all i, j, one obtains $\psi_{ij} = \psi_i$, with $\psi_i \sim H^*$, where H^* is induced from H. By choosing H^* to have a stick-breaking form, one can obtain any prior in the class proposed by Ishwaran & James (2001), which includes the Dirichlet process and two-parameter Poisson Dirichlet process (Pitman & Yor, 1997) as special cases. For the priors $P \sim \text{LPP}_1(\alpha, \beta, P_0)$ and $P \sim \text{LPP}_2(\alpha, \beta, P_0)$, $G = \delta_1$ corresponds to the limiting case $\beta \to 0$ in which we obtain $P \sim \text{DP}(\alpha P_0)$.

Another important special case corresponds to $G = \delta_0$, so that $z_{ij} = 0$ for all i, j, which occurs in the limit as $\beta \to \infty$ for the priors $P \sim \text{LPP}_1(\alpha, \beta, P_0)$ and $P \sim \text{LPP}_2(\alpha, \beta, P_0)$. In the $P \sim \text{LPP}_1(\alpha, \infty, P_0)$ case, one obtains $\theta_{ij} \sim P_j$, with $P_j \sim \text{DP}(\alpha P_{0j})$ independently for $j = 1, \ldots, p$, with P_{0j} the jth 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. In the $P \sim \text{LPP}_2(\alpha, \infty, P_0)$ case, instead of independent Dirichlet process priors for P_j ($j = 1, \ldots, p$), one obtains the following dependent Dirichlet process (MacEachern, 1999),

$$P_j = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_{hj}} \quad (j = 1, \ldots, p), \qquad \Theta_h = \{\Theta_{hj}\}_{j=1}^{p} \sim P_0, \quad (5)$$

with $P_j \sim \text{DP}(\alpha P_{0j})$ marginally. Expression (5) corresponds to the fixed-π formulation of the dependent Dirichlet process (De Iorio et al., 2004).

Previous nonparametric Bayes methods for discrete random probability measures consider a single cluster index ψ_i, while local partition processes provide a methodology for multivariate $\psi_i = (\psi_{i1}, \ldots, \psi_{ip})^{\mathrm{T}}$. A competing formulation relies on a matrix stick-breaking process (Dunson et al., 2008), which induces dependent local clustering through including stick-breaking random variables for every subject and predictor. This leads to substantial computational difficulties for large n or p, while requiring $\mathrm{pr}(\theta_{ij} = \theta_{i'j} \mid \theta_{ij'} = \theta_{i'j'}) > 0{\cdot}5$.

3. Posterior computation

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 $P \sim \mathrm{LPP}_1(\alpha, \beta, P_0)$, we propose to use an adaptation of the slice sampling approach of Walker (2007) to implement posterior computation. This slice sampler avoids the need for a finite approximation, and is simpler to implement than the retrospective sampler of Papaspiliopoulos & Roberts (2008).

In addition to the latent variables defined in (4), we introduce latent variables $u_i = \{u_{ij}\}_{j=0}^{p}$ ($i = 1, \ldots, n$), where the complete data joint likelihood of y, u and z is

$$\prod_{i=1}^{n} \left\{ g(y_i; \Theta_{\psi_i}, \tau) 1(u_{i0} < \pi_{0\phi_{i0}}) \prod_{j=1}^{p} 1(u_{ij} < \pi_{j\phi_{ij}}) \nu_j^{z_{ij}} (1-\nu_j)^{1-z_{ij}} \right\}, \quad (6)$$

where the u_{ij}s are constrained to fall in (0, 1). The sampler proceeds as follows.

Step 1. For the latent u_{ij}, the conditional is $\mathrm{Unif}(0, \pi_{j\phi_{ij}})$ ($j = 0, 1, \ldots, p$).

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

$$p_{ij} = \frac{\nu_j g(y_i; \theta_{i(z_{ij}=1)}, \tau)}{\nu_j g(y_i; \theta_{i(z_{ij}=1)}, \tau) + (1-\nu_j) g(y_i; \theta_{i(z_{ij}=0)}, \tau)},$$

where $\theta_{i(z_{ij}=l)}$ denotes the current value of θ_i with the jth element set equal to $\Theta_{0\phi_{i0}j}$ for $l = 1$ and $\Theta_{1\phi_{ij}j}$ for $l = 0$.

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

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

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

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

For all h such that $\sum_{i=1}^{n} 1(\phi_{ij}=h)=0$, we have $a_{jh} = 0$. In addition, $b_{jh} = 1$ for $h=\phi_j^*$.

Step 4. The conditional probability of $\phi_{ij} = h$ is proportional to $1(h \in A_{ij})g(y_i; \theta_{h(\phi_{ij}=h)}, \tau)$, where $A_{ij} = \{h : \pi_{jh} > u_{ij}\}$ is a finite subset of $\{1, 2, \ldots, \infty\}$ obtained by first sampling π_{jh}^*, for $h = 1, \ldots, \tilde{\phi}_j$, with $\tilde{\phi}_j$ the smallest value satisfying

$$\sum_{h=1}^{\tilde{\phi}_j} \pi_{jh}^* \prod_{l<h} (1 - \pi_{jl}^*) \geqslant 1 - u_j^*,$$

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

Step 5. The conditional distribution of Θ_{lh} is proportional to

$$g_0(\Theta_{lh}) \prod_{i=1}^{n} g(y_i; \theta_i, \tau),$$

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

Step 6. For the probability ν_j, the conditional density is $\mathrm{Be}\{1 + \Sigma_i z_{ij}, \beta + \Sigma_i (1 - z_{ij})\}$.

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

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

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

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

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

Each of these steps is simple to implement, and good rates of mixing and convergence have been observed in several applications. Steps 5 and 7 simplify when conjugate priors are chosen. For example, when $g(y_i; \theta_i, \tau)$ 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.

To modify the algorithm for formulation 2, drop the first subscript on each of the πs in (6) and in Steps 1 and 2. In Step 3, the conditional density for π_h^* is proportional to

$$(1 - \pi_h^*)^{\alpha - 1} \prod_{i=1}^{n} \prod_{j=0}^{p} 1\left\{\pi_{\phi_{ij}}^* \prod_{l<\phi_{ij}} (1 - \pi_l^*) > u_{ij}\right\}.$$

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

$$a_h = \max\left\{\frac{u_{ij}}{\prod_{l<h}(1-\pi_l^*)}, i, j{:}\phi_{ij}=h, i=1,\ldots,n, j=0,1,\ldots,p\right\},$$
$$b_h = 1-\max\left\{\frac{u_{ij}}{\pi^*_{\phi_{ij}}\prod_{l<\phi_{ij},l\neq h}(1-\pi_l^*)}, i, j{:}\phi_{ij}>h, i=1,\ldots,n, j=0,1,\ldots,p\right\}.$$

Step 4 is modified to let $A_{ij} = \{h : \pi_h > u_{ij}\}$, where π_h^* is sampled for $h = 1, \ldots, \tilde{\phi}$, with $\tilde{\phi}$ the smallest value satisfying $\sum_{h=1}^{\tilde{\phi}} \pi_h^* \prod_{l<h}(1-\pi_l^*) \geqslant 1 - \min\{u_{ij}, i=1,\ldots,n, j=0,1,\ldots,p\}$. Steps 5–8 are unchanged, and in Step 9 the conditional distribution of α is

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

The major computational advantage relative to formulation 1 occurs in Step 3.

4. Simulation example

To illustrate the approach, a simulation example was considered following (1)-(4), with $t_{is} \in [0, 1]$ and basis functions $b_1(s) = 1$, $b_{j+1}(s) = \exp(-\psi\|s - \xi_j\|^2)$ ($j = 1, \ldots, p-1$), with $\xi_1, \ldots, \xi_{p-1}$ equally spaced kernel locations, $\psi = 25$ and $p = 20$. To generate $n = 16$ curves, let $\theta_{ij} = z_{ij}\Theta_{0j} + (1 - z_{ij})\Theta_{1j}$ for $i = 1, \ldots, 16$ and $j = 1, \ldots, 20$, with $z_{ij} \sim \mathrm{Ber}(0{\cdot}5)$ and the elements of Θ_0 and Θ_1 sampled independently from a mixture of a point mass at zero, having probability 0·5, and a standard normal density. The point mass allowed exclusion of basis functions. There were 40 equally spaced observations along each curve with $\sigma^2 = 0{\cdot}2$.

The data were analyzed using $P \sim \mathrm{LPP}_1(\alpha, \beta, P_0)$, $P \sim \mathrm{LPP}_2(\alpha, \beta, P_0)$ and $P \sim \mathrm{DP}(\alpha P_0)$, with Ga(1, 1) hyperpriors for α and β, a Ga(0·1, 0·1) hyperprior for σ^{-2}, and P_0 chosen to correspond to the multivariate Gaussian distribution with zero-mean 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 twentieth sample collected to thin the chain. In each case, 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, it is not reliable to monitor the elements of Θ_h in assessing convergence and mixing.

Figure 1 shows the data, true curves and estimates under $P \sim \mathrm{LPP}_1(\alpha, \beta, P_0)$. The estimates are close to the truth, with average absolute bias = 0·192, mean square error = 0·060, maximum absolute bias = 0·506 and an average 99% pointwise credible interval width of 1·01. The estimated posterior mean of σ^2 was 0·21, with a 95% credible interval of [0·18, 0·26]. The estimate for α was 0·17, with 95% credible interval [0·10, 0·29]. The small value of α suggests the individuals are assigned to few local and global clusters. For component j ($j = 0, 1, \ldots, p$), all subjects are allocated to the first $\phi_j^* = \max\{\phi_{ij}, i=1,\ldots,n\}$ clusters. The posterior means of ϕ_j^* are less than 2 for $j = 0, 1, \ldots, p$, with $\widehat{\phi_0^*} = 1 \cdot 62$ and $\widehat{\phi_j^*} \in [0 \cdot 99, 1 \cdot 30]$, for $j = 1, \ldots, p$. The probability of allocation to the local component for a random element of a subject's random effects vector is $1/(1 + \beta)$. Hence, β provides a measure of the relative importance of the local and global components, with both components receiving equal weight for $\beta = 1$ and the global component dominant for $\beta < 1$. For the simulation, we obtained $\hat{\beta} = 1{\cdot}23$, with a 95% credible interval of [0·51, 2·42], so that the two components receive close to equal weight.

For the analysis under $P \sim \mathrm{LPP}_2(\alpha, \beta, P_0)$, there was a substantial reduction in computational time and the results were improved, 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. The estimated posterior mean of β was 1·56, with a 95% credible interval of [0·64, 3·25]· In addition, $\hat{\sigma^2} = 0{\cdot}20$, with a 95% credible interval of [0·18, 0·23]. It does not appear that the more concentrated posterior distributions reflected a 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.

For the Dirichlet process, there was 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. The average absolute bias was increased by 26%, the mean square error by 60% and the maximum absolute bias by 57% relative to the first formulation. The average credible interval widths were similar, but the true functions fell well outside the 99% intervals in two out of 16 of the cases. In addition, the estimated residual variance was $\hat{\sigma^2} = 0{\cdot}27$, with 99% credible interval [0·23, 0·31] not including the true value of 0·20. This overestimation of the residual variance suggests a poor job at recovering the signal. This was the first simulation scenario considered, but there was 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. Gains are most notable when data for each function are sparse.

5. Hormone curve application

5·1. Background and motivation

The approach is applied to post-ovulatory progesterone data collected in early pregnancy for $n = 165$ women (Wilcox et al., 1988). The progesterone metabolite, PdG, was measured in urine. Letting y_{is} denote the sth measurement of log(PdG) for woman i occurring t_{is} days after the estimated day of ovulation, we assume $y_{is} \sim t_\kappa\{\eta_i(t_{is}), \sigma^2\}$, where η_i is a smooth trajectory in log(PdG) for woman i and $t_\kappa(\mu, \sigma^2)$ denotes the t-density, with mean μ, degrees of freedom κ and scale σ^2. Expression (1) is generalized to allow t-distributed measurement errors. The data contain between $n_i = 4$ and $n_i = 40$ observations per woman, with an average of $\bar{n} = 23{\cdot}1$. 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. Censoring occurred if menses resumed following an early pregnancy loss or the six-month collection period ended. Given the plausible missingness mechanisms, it is reasonable to assume that missingess is conditionally independent of the missing PdG values given the observed PdG values, implying data are missing at random.

The goal is to obtain a flexible, yet parsimonious representation of the progesterone trajectories. Bigelow & Dunson (2009) 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. 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 representation.

In analyzing the data, we used the 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 $y_{it} \sim N\left\{\eta_i(s_{it}), \xi_i^{-1}\sigma^2\right\}$, with $\xi_i \sim \text{Ga}(\kappa/2, \kappa/2)$ and with the degrees of freedom κ assigned a 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 κ is updated using Metropolis–Hastings steps.

5·2. Analysis and results

We focus initially on the first formulation of the local partition process. Using multiple chains with widely distributed starting points, the slice sampler was observed to exhibit good rates of convergence and mixing. Figure 2 shows the results under $P \sim \text{DP}(\alpha P_0)$ and $P \sim \text{LPP}_1(\alpha, \beta, P_0)$ for five women. It is clear from this plot, and from examination of plots for the other women in the study, that the local partition process method produces a good fit. The estimated value of the hyperparameter α was $\hat{\alpha} = 0{\cdot}41$, with a 95% credible interval of [0·25, 0·63]. This value suggests that few global and local clusters are needed, so that a sparse representation of the data is obtained. Indeed, the posterior probability of $\phi_0^* = 3$ was 0·97, suggesting that subjects were allocated to three global clusters. The number of local clusters also tended to be small, having an estimated average value of 2·91.

The estimated value of the hyperparameter β was 1·05, with a 95% credible interval of [0·63, 1·59]. Recall that values of β close to zero 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 component. Hence, there is clear evidence in the data favouring 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. This is as expected, since most of the values are tightly distributed about a smooth trajectory, but there are extreme outlying values in the dataset. Commenting further on the estimated hormone curves, 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 a week or more after ovulation and then decline correspond to early pregnancy losses.

Repeating the analysis for $P \sim \text{DP}(\alpha P_0)$, 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 than in the local partition process analysis. In most cases, the estimates are similar to those obtained for $P \sim \text{LPP}_1(\alpha, \beta, P_0)$. However, as illustrated in Fig. 2, there are multiple exceptions in which the Dirichlet process approach produced an estimate inconsistent with the data. We noted particularly poor performance for women having relatively few observations. This likely reflects a tendency of the Dirichlet process to overly favour clustering together of subjects unless there are abundant data available to suggest that 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], and the estimate of β was 1·87 with a 95%

interval of [1·10, 2·92]. The larger estimate of β suggests that $P \sim \mathrm{LPP}_2(\alpha, \beta, P_0)$ places more weight on the local component than $P \sim \mathrm{LPP}_1(\alpha, \beta, P_0)$. This is likely due to the fact that there is less of a penalty to be paid for introducing additional local clusters under $P \sim \mathrm{LPP}_2(\alpha, \beta, P_0)$ due to the incorporation of common stick-breaking weights. The estimates for the overall number of clusters and the log(PdG) curves were very similar to those for the $\mathrm{LPP}_1(\alpha, \beta, P_0)$. 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 paper proposes a generalization of the Dirichlet process, which allows dependent local clustering and borrowing of information. The emphasis is 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 analogue 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. The proposed local partition process provides a simple, yet flexible approach for characterizing the local selection process. The slice sampling implementation is quite simple and efficient to implement, while allowing posterior computation for the infinite-dimensional nonparametric process instead of a finite approximation.

Appendix

Proof of Proposition 1

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

$$\begin{aligned}\mathrm{pr}(\theta_{ij}=\theta_{i'j}) &= \int \sum_{l=0}^{1}\sum_{m=0}^{1} \mathrm{pr}(\theta_{ij}=\theta_{i'j}|z_{ij}=l, z_{i'j}=m)\mathrm{pr}(z_{ij}=l, z_{i'j}=m|\nu_j) d f(\nu_j)\\ &= \int \tfrac{1}{1+\alpha}\{\nu_j^2+(1-\nu_j)^2\}\tfrac{1}{B(1,\beta)}(1-\nu_j)^{\beta-1} d\nu_j\\ &= \tfrac{1}{1+\alpha}\left\{\left(\tfrac{1}{1+\beta}\right)\left(\tfrac{2}{2+\beta}\right)+\tfrac{\beta}{2+\beta}\right\},\end{aligned}$$

with Proposition 1 following directly, where $f(\pi_j)$ denotes the Be(1, β) 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:

$$\begin{aligned}\mathrm{pr}(\theta_{ij} &= \theta_{i'j}, \theta_{ij'}=\theta_{i'j'})\\ &= \int \tfrac{1}{1+\alpha}\left[\tfrac{1}{1+\alpha}\left\{\nu_j^2(1-\nu_{j'})^2+(1-\nu_j)^2\nu_{j'}^2+(1-\nu_j)^2(1-\nu_{j'})^2\right\}+\nu_j^2\nu_{j'}^2\right]\\ &\quad\times\tfrac{1}{B(1,\beta)^2}(1-\nu_j)^{\beta-1}(1-\nu_{j'})^{\beta-1}d\nu_j d\nu_{j'}\\ &= \left(\tfrac{1}{1+\alpha}\right)^2\left\{\tfrac{4\beta}{(1+\beta)(2+\beta)^2}+\left(\tfrac{\beta}{2+\beta}\right)^2\right\}+\left(\tfrac{1}{1+\alpha}\right)\left(\tfrac{1}{1+\beta}\right)^2\left(\tfrac{2}{2+\beta}\right)^2\\ &= \left(\tfrac{1}{1+\alpha}\right)^2\left(\tfrac{1}{2+\beta}\right)^2\left\{\left(\beta+\tfrac{2}{1+\beta}\right)^2+\tfrac{4\alpha}{(1+\beta)^2}\right\}.\end{aligned}$$

Proof of Proposition 3

Express the joint probability of local clustering as

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

Proof of Theorem 1

To demonstrate property 1, first let

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

where $\psi\in\{1, 2, \ldots, \infty\}^p$, $I_o=\{1, 3, 5, \ldots\}$ and $I_e=\{2, 4, 6, \ldots\}$. 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 $\mathrm{pr}\{Q(\psi)>0\}=1$. It follows that $\bar{Q}(\psi)>0$ and hence $\mathrm{pr}\{Q(\psi)>\varepsilon_\psi\}=0{\cdot}5$, with $\varepsilon_\psi=\bar{Q}(\psi)>0$. Letting $\varepsilon=\min[\varepsilon_\psi, \psi\in\{1, 2, \ldots, \infty\}^p]$, we have $\mathrm{pr}\{Q(\psi)>\varepsilon\}\geqslant 0{\cdot}5$.

To demonstrate Property 2, first let

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

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 $d(A)=\cup_{a\in A}d(a)$, $\mathrm{pr}\{Q(\theta_{ij}=\theta_{i'j})\in A\}=\int_{d(A)}f(\alpha, \beta)d\alpha\, d\beta$. For all Borel subsets $A\subset(0, 1)$, the area of $d(A)$ is greater than 0, so $\mathrm{pr}\{Q(\theta_{ij}=\theta_{i'j})=A\}>0$ 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, first let

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

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

References

Bigelow JL, Dunson DB. Bayesian semiparametric joint models for functional predictors. J Am Statist Assoc. 2009; 104:26–36.

Bush CA, MacEachern SN. A semiparametric Bayesian model for randomised block designs. Biometrika. 1996; 83:275–85.

Cifarelli DM, Regazzini E. Nonparametric statistical problems under partial exchangeability: The use of associative means (in italian). Annali del'Instituto di Matematica Finianziara dell'Universitá di Torino, Serie III. 1978; 12:1–36.

De Iorio M, Müller P, Rosner G, MacEachern SN. An ANOVA model for dependent random measures. J Am Statist Assoc. 2004; 99:205–15.

Dunson DB, Xue Y, Carin L. The matrix stick-breaking process: Flexible Bayes meta analysis. J Am Statist Assoc. 2008; 103:317–27.

Ferguson TS. A Bayesian analysis of some nonparametric problems. Ann Statist. 1973; 1:209–30.

Ferguson TS. Prior distributions on spaces of probability measures. Ann Statist. 1974; 2:615–29.

Heard NA, Holmes CC, Stephens DA. A quantitative study of gene regulation involved in the immune response of anopheline mosquitoes: An application of Bayesian hierarchical clustering of curves. J Am Statist Assoc. 2006; 101:18–29.

Ishwaran H, James LF. Gibbs sampling methods for stick-breaking priors. J Am Statist Assoc. 2001; 96:161–73.

MacEachern, SN. Proc Sect Bayesian Statist Sci. Alexandria, VA: American Statistical Association; 1999. Dependent nonparametric processes; p. 50-55.

Müller P, Rosner GL. A Bayesian population model with hierarchical mixture priors applied to blood count data. J Am Statist Assoc. 1997; 92:1279–92.

Papaspiliopoulos O, Roberts GO. Retrospective MCMC for Dirichlet process hierarchical models. Biometrika. 2008; 95:169–86.

Pitman J, Yor M. The two parameter Poisson–Dirichlet distribution derived from a stable subordinator. Ann Prob. 1997; 25:855–900.

Sethuraman J. A constructive definition of Dirichlet priors. Statist Sinica. 1994; 4:639–50.

Walker SG. Sampling the Dirichlet mixture model with slices. Comm Statist B. 2007; 36:45–54.

Wilcox A, Weinberg CR, O'Connor J, Baird D, Schlatterer J, Canfield R, Armstrong E, Nisula B. Incidence of early loss of pregnancy. New Engl J Med. 1988; 319:189–94. [PubMed: 3393170]

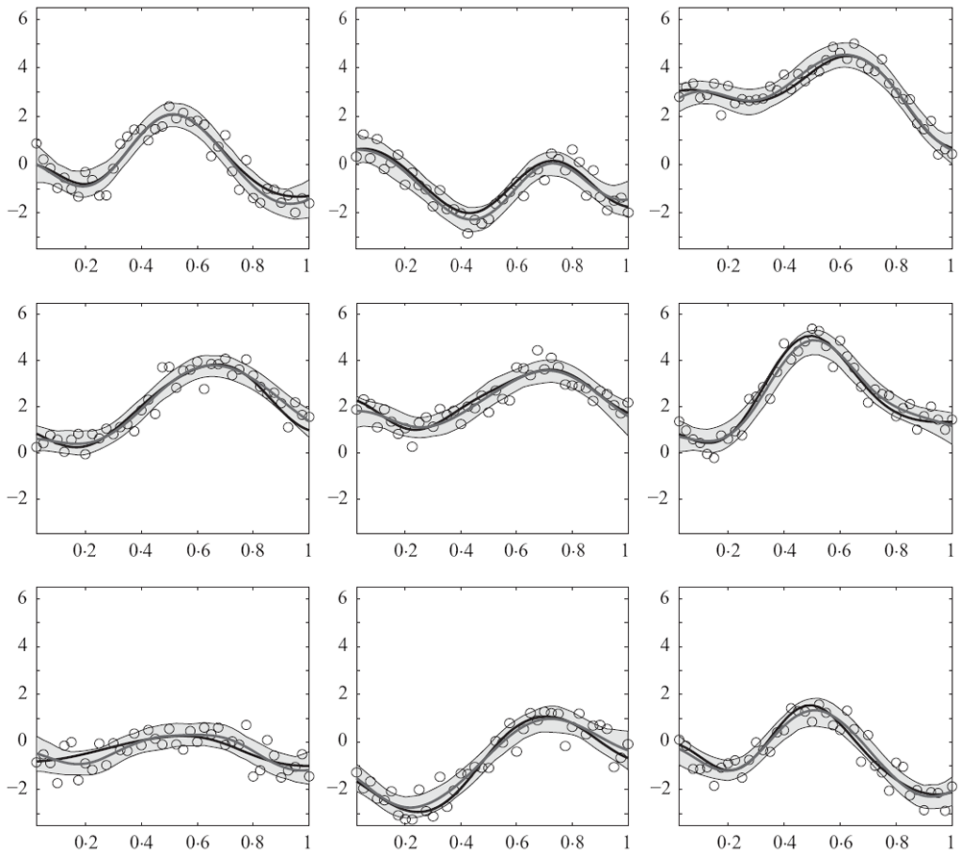

Fig. 1.
Data and results for 9 of the 16 subjects in the simulation example. Each panel corresponds to one subject in the study, the true functions are represented with black lines, the posterior means with grey lines and 99% pointwise credible intervals are shaded.

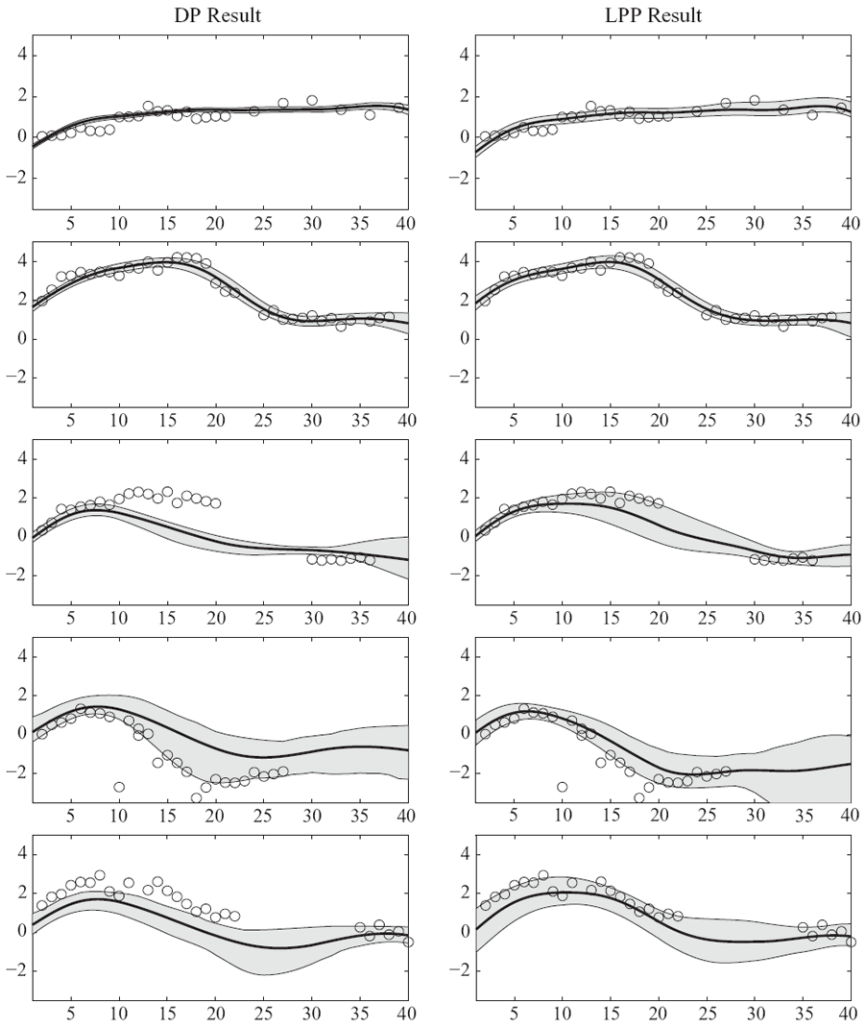

Fig. 2.

Log PdG function estimates for selected women in the Early Pregnancy Study. The posterior means are solid lines and 95% pointwise credible intervals are shown in grey. The *x*-axis scale is time in days starting at the estimated day of ovulation.