Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://amstat.tandfonline.com/loi/uasa20

Bayesian Inference on Changes in Response Densities Over Predictor Clusters

David B. Dunson, Amy H. Herring & Anna Maria Siega-Riz

To cite this article: David B. Dunson, Amy H. Herring & Anna Maria Siega-Riz (2008) Bayesian Inference on Changes in Response Densities Over Predictor Clusters, Journal of the American Statistical Association, 103:484, 1508-1517, DOI: 10.1198/016214508000001039

To link to this article: https://doi.org/10.1198/016214508000001039

Published online: 01 Jan 2012.

Submit your article to this journal

Article views: 251

View related articles

Citing articles: 18 View citing articles

Full Terms & Conditions of access and use can be found at
https://amstat.tandfonline.com/action/journalInformation?journalCode=uasa20

Bayesian Inference on Changes in Response Densities Over Predictor Clusters

David B. DUNSON, Amy H. HERRING, and Anna Maria SIEGA-RIZ

In epidemiology, it often is of interest to assess how individuals with different trajectories over time in an environmental exposure or biomarker differ with respect to a continuous response. For ease in interpretation and presentation of results, epidemiologists typically categorize predictors before analysis. To extend this approach to time-varying predictors, individuals can be clustered by their predictor trajectory, with the cluster index included as a predictor in a regression model for the response. This article develops a semiparametric Bayes approach that avoids assuming a prespecified number of clusters and allows the response to vary nonparametrically over predictor clusters. This methodology is motivated by interest in relating trajectories in weight gain during pregnancy to the distribution of birth weight adjusted for gestational age at delivery. In this setting, the proposed approach allows the tails of the birth weight density to vary flexibly over weight gain clusters.

KEY WORDS: Bivariate clustering; Dirichlet process; Functional predictor; Growth mixture model; Joint modeling; Latent class trajectory.

1. INTRODUCTION

1.1 Weight Gain During Pregnancy and Birth Weight

The National Research Council and Institute of Medicine (2007) recently published a report on the influence of pregnancy weight on maternal and child health. One goal of the report was to assess whether Institute of Medicine (IOM) pregnancy weight gain guidelines (Institute of Medicine 1990) need to be reexamined due to recent health trends and increasing obesity. Several studies that have examined the pattern of pregnancy weight gain in light of the IOM recommendations have found that most women do not follow the recommendations (Siega-Riz, Adair, and Hobel 1994; Carmichael, Abrams, and Selvin 1997; Dietz et al. 2006).

For clinical guidelines, simplicity is of paramount importance. For this reason, epidemiologists typically categorize continuous predictors before analysis. For example, in assessing health effects of body mass index (BMI), individuals are categorized into low weight, normal weight, overweight, and obese groups instead of including BMI as a continuous predictor in a flexible regression model for the health response. Simple tabular summaries of changes in the response between different BMI categories then can be presented, so that one set of recommendations can be made for all individuals in a category. For a time-varying predictor, such as pregnancy weight gain, predefining categories is not straightforward. But functional clustering methods can be used to group women into weight gain trajectory clusters, with the cluster IDs then providing a predictor of pregnancy outcomes and child health. Then clinicians can make recommendations based on simply examining a summary of changes in outcomes across weight gain clusters.

The goal of the present work was to develop an approach for assessing changes in infant birth weight across maternal weight gain clusters. For simplicity, we focus on data for $n = 1{,}888$ mothers with normal prepregnancy weight (pregravid body mass index, 19.8–26) and giving birth to singleton infants. In making weight gain recommendations to pregnant women, physicians are concerned primarily with inadequate or excessive weight gain. A women gaining too little weight may not provide adequate nutrition for the developing fetus, resulting in a small for gestational age baby and potential long-term developmental problems. At the other extreme, women gaining too much weight may provide too much nutrition to the fetus, resulting in a large for gestational age child who perhaps is at long-term risk of obesity. Thus the primary interest lies in how the tails of the birth weight density change across weight gain clusters. With this focus, we propose a Bayesian semiparametric approach, that allows a response density to change nonparametrically over predictor clusters.

1.2 Relevant Statistical Literature

There is an increasingly rich literature on model-based approaches for clustering through the use of finite mixture models. Finite mixture models for functional data are commonly referred to as latent class trajectory models. The approach of Jones, Nagin, and Roeder (2001), for example, characterizes functional data using a polynomial curve, with subjects having identical regression coefficients grouped into clusters. This approach can be implemented in SAS using PROC TRAJ (Jones and Nagin 2007). Muthén (2004) proposed an alternative approach that allows variability in trajectories among subjects in a trajectory cluster. James and Sugar (2003) instead characterized curves as linear combinations of basis functions, with subjects having similar basis coefficients clustered together. These methods can be used for joint modeling of functional predictors with a response by including the latent trajectory cluster index as a categorical predictor in the response model (Jones and Nagin 2007). The Bayesian information criterion (BIC) is commonly used to select the number of latent trajectory classes.

Considering the application of such an approach to pregnancy weight gain and birth weight data, clusters in the weight gain trajectories could be estimated, with the cluster IDs then included as categorical predictors in a linear regression model for the birth weight outcome. Gestational age at delivery and maternal covariates could be adjusted for within this model, while accommodating a heavy-tailed residual distribution to

David B. Dunson is Professor, Department of Statistical Science, Duke University, Durham, NC 27708-0251 (E-mail: *dunson@stat.duke.edu*). Amy H. Herring is Associate Professor, Department of Biostatistics, and Fellow, Carolina Population Center, The University of North Carolina, Chapel Hill, NC 27599. Anna Maria Siega-Riz is Associate Professor, Department of Epidemiology, Associate Professor, Department of Nutrition, and Fellow, Carolina Population Center, Ther University of North Carolina, Chapel Hill, NC 27599.
The authors thank David Umbach and David Kessler for helpful comments on a draft of the manuscript.

© 2008 American Statistical Association
Journal of the American Statistical Association
December 2008, Vol. 103, No. 484, Applications and Case Studies
DOI 10.1198/016214508000001039

better characterize the birth weight density. Such an approach would only allow the mean of the birth weight distribution to vary across weight gain clusters, however, and so it does not address our interest in studying changes in the tails of the distribution. The approach could be modified to allow the variance to change across the weight gain clusters as well, but the model still would not allow different-shaped distributions. Given that our primary interest is in assessing how pregnancy weight gain trajectories relate to risk of small and large infants, it seems critical to allow the birth weight densities to change flexibly over the trajectory clusters.

Bayesian functional clustering has recently been studied by Ray and Mallick (2006) and Heard, Holmes, and Stephens (2006). Related to the method of Ray and Mallick (2006), Bigelow and Dunson (2005) proposed expressing the curves using a spline basis, with subject-specific basis coefficients assigned a Dirichlet process (DP) prior (Ferguson 1973, 1974). This approach automatically groups the subjects into an unknown number of trajectory clusters, while allowing for uncertainty in the clustering process. For joint modeling, instead a DP prior can be assigned to the joint distribution of the basis coefficients and a location parameter in the response model. The result is a semiparametric Bayes version of the latent class trajectory methods cited earlier. Unfortunately, such an approach still assumes that the trajectory cluster-specific response distribution follows a parametric form, with only a location parameter varying.

A primary limitation of such methods is the restriction that predictor (trajectory) and response clusters are fundamentally the same entity, so that a subject assigned to a particular predictor cluster is automatically assigned to the corresponding response cluster. We propose an enriched formulation in which predictor and response clusters are separate, with a subject's allocation to a given predictor cluster informing, but not constraining, that subject's assignment to a response cluster. We develop this approach by starting with a semiparametric Bayes hierarchical model, which induces a dependent bivariate clustering structure. Posterior computation relies on a combined sequential updating and greedy search (SUGS) (Wang and Dunson 2007) and Gibbs sampling algorithm.

Section 2 proposes the nonparametric joint clustering model. Section 3 describes the algorithm for posterior computation and inferences. Section 4 applies the methods to the weight gain and birth weight data. Section 5 discusses the results.

2. JOINT MODELING OF PREDICTOR CLUSTERS AND RESPONSE DENSITIES

2.1 Dirichlet Process Clustering

For ease in presentation, we focus on the pregnancy weight gain and infant birth weight application in describing the methodology, although the approach certainly can be used in other settings. Let f_i denote a smooth measurement error–corrected weight gain trajectory over time for woman i, for $i = 1, \ldots, n$, and let y_i denote the birth weight of her infant. In the measurement error component, we include not only true measurement errors in weighing the woman arising from scale and recording inaccuracies, but also short-term fluctuations in body weight that can occur during pregnancy due to water retention and other factors. We view f_i as a smooth function that is close (in a clinical/biological sense) to the true weight trajectory for woman i. Our goal is then to define clusters in the f_i's and assess how the birth weight density varies over these clusters.

The actual data collected for woman i are $\{\mathbf{x}_i, y_i\}_{i=1}^n$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{i,n_i})'$ and x_{ij} is the recorded weight at time t_{ij}, for $j = 1, \ldots, n_i$. The measurement times may vary somewhat for the different women under study. As a simple measurement model, we assume that

$$x_{ij} \sim \mathrm{N}(f_i(t_{ij}), \tau_x^{-1}), \qquad j = 1, \ldots, n_i, i = 1, \ldots, n, \quad (1)$$

where τ_x is a fixed measurement error precision (1/variance). In addition, we assume that each of the weight gain trajectories, f_i, can be characterized as a linear combination of prespecified basis functions, $\{b_h\}_{h=1}^p$, as

$$f_i(t) = \sum_{m=1}^{p} \beta_{im} b_m(t), \qquad \forall t \in \mathcal{T}, \quad (2)$$

where $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ip})'$ are subject-specific basis coefficients.

Thus variability in the weight gain trajectories is controlled by variability in the basis coefficients. For functional clustering, we assume that the basis coefficients have a discrete distribution by letting

$$\boldsymbol{\beta}_i \overset{iid}{\sim} P_1 = \sum_{h=1}^{K} \pi_h \delta_{\boldsymbol{\beta}_h^*}, \quad (3)$$

where K is an upper bound on the number of clusters, π_h is the probability weight on cluster h, δ_θ is a distribution concentrated at θ, and $\boldsymbol{\beta}_h^*$ are the basis coefficients characterizing the function in cluster h, for $h = 1, \ldots, K$. It is important to note that the number of clusters among the n women in the sample, denoted by k, is treated as unknown, and K provides just an upper bound on this number. By letting $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)' \sim \mathrm{Diri}(\alpha/K, \ldots, \alpha/K)$ and $\boldsymbol{\beta}_h^* \overset{iid}{\sim} P_{01}$, we obtain an accurate finite approximation to a DP prior, $P_1 \sim DP(\alpha P_{01})$, for the distribution of the basis coefficients (Ishwaran and Zarepour 2002). The accuracy of this approximation improves as α decreases, favoring a smaller value of k, and as K increases.

If we were interested only in clustering the weight gain trajectories, then models (1)–(3) would be sufficient, providing an accurate finite approximation to a DP mixture model. Such an approach is closely related to the work of Ray and Mallick (2006) and Bigelow and Dunson (2005). However, our goal is instead to assess changes in the density of birth weight across weight gain clusters, so that is becomes necessary to extend the specification to include y_i. If birth weight did not depend on pregnancy weight gain, then we could flexibly model the birth weight density using a DP mixture (DPM) of normals model,

$$f(y) = \int N(y; \mu, \sigma^2)\, dP(\mu), \quad (4)$$

where $P \sim DP(\lambda P_{02})$, with λ a precision parameter and P_{02} a parametric guess at the mixture distribution (e.g., normal) (Lo

1984; Escobar and West 1995). Again using the finite approximation to the DP of Ishwaran and Zarepour (2002), we then could obtain

$$f(y) = \sum_{l=1}^{L} \nu_l N(y; \mu_l^*, \sigma^2),$$

$$\boldsymbol{\nu} \sim \text{Diri}(\lambda/L, \ldots, \lambda/L),\ \mu_l^* \overset{iid}{\sim} P_{02}, \quad (5)$$

where L is an upper bound on the number of birth weight mixture components occupied by the n women in the sample.

2.2 Joint Modeling

In addressing our goal of allowing the birth weight density to vary flexibly over the pregnancy weight gain clusters, we need to link the functional clustering component, defined in (1)–(3), with the response component, defined in (5). A simple way to accomplish this is to choose a DP prior for the joint distribution of the basis coefficients, $\boldsymbol{\beta}_i$, and the response mean, μ_i, assuming that $y_i \sim \text{N}(\mu_i, \tau_y^{-1})$, with τ_y a precision parameter. Such an approach will automatically allocate the n women into k clusters, with each cluster having a unique weight gain trajectory and normal distribution for birth weight. Unfortunately, then only shifts in the mean birth weight would be allowed between weight gain clusters, which is counter to our interest in studying shifts in the left and right tails of the birth weight distribution.

Thus we instead propose a novel bivariate partitioning approach, which allows separate predictor and response clusters, while allowing dependent allocation to the two components. In particular, woman i's chance of falling into the lth birth weight cluster will depend on her pregnancy weight gain cluster. To formalize the dependence structure, we first let $\gamma_i \in \{1, \ldots, K\}$ represent the weight gain cluster index for woman i under models (1)–(3), so that $\boldsymbol{\beta}_i = \boldsymbol{\beta}_{\gamma_i}^*$ and $f_i(t) = \mathbf{b}(t)'\boldsymbol{\beta}_{\gamma_i}^*$, where $\mathbf{b}(t) = [b_1(t), \ldots, b_p(t)]'$. We then characterize the birth weight density for women having $\gamma_i = h$ as

$$f_h(y) = \int N(y; \mu, \sigma^2)\, dP_h(\mu), \quad (6)$$

where P_h is a mixture distribution specific to predictor (pregnancy weight gain) cluster h.

The joint DP approach mentioned earlier would let $P_h = \delta_{\mu_h^*}$, with the μ_h^* sampled independently for $h = 1, \ldots, K$, resulting in $f_h(y) = N(y; \mu_h^*, \sigma^2)$ as the birth weight density for women in weight gain cluster h. To instead allow the birth weight density to be a flexible mixture of normals within each weight gain cluster, while borrowing information, we let

$$P_h = (1 - \psi) P_{20}^* + \psi P_{2h}^*, \quad (7)$$

where $P_{20}^* = \sum_{l=1}^{L_0} \omega_{0l} \delta_{\mu_{0l}^*}$ is a global component shared by all predictor clusters, $0 \le \psi \le 1$ is a weight on the global component, and $P_{2h}^* = \sum_{l=1}^{L_0} \omega_{hl} \delta_{\mu_{hl}^*}$ is a local component specific to predictor cluster h, for $h = 1, \ldots, K$. We assume that $\boldsymbol{\omega}_h = (\omega_{h1}, \ldots, \omega_{h,L_0}) \sim \text{Diri}(\lambda/L_0, \ldots, \lambda/L_0)$ and $\mu_{hl}^* \overset{iid}{\sim} P_{02}$, $l = 1, \ldots, L_0$, $h = 0, 1, \ldots, K$, which implies that P_{2h}^* is assigned a finite approximation to a $DP(\lambda P_{02})$ prior, independently for $h = 0, 1, \ldots, K$. The $L_0 \to \infty$ limiting case of (7) is equivalent to the specification that Müller, Quintana, and Rosner (2004) used for borrowing information from related studies; however, in their setting, γ_i was a known study index instead of an unknown cluster index.

Our prior specification for the mixture distribution P_h implies further that

$$P_h = \sum_{l=1}^{L} \nu_{hl} \delta_{\mu_l^*}^*, \qquad \mu_l^* \overset{iid}{\sim} P_{02}, \quad (8)$$

where $L = (K+1)L_0$; μ_l^* is the lth element of $\boldsymbol{\mu}^* = (\boldsymbol{\mu}_0^{*\prime}, \boldsymbol{\mu}_1^{*\prime}, \ldots, \boldsymbol{\mu}_K^{*\prime})'$; $\nu_{hl} = (1 - \psi)\omega_{0l}$ for $l = 1, \ldots, L_0$; $\nu_{hl} = 0$ for $l \in \{L_0 + 1, \ldots, (h-1)L_0, hL_0 + 1, \ldots, L\}$; and $\nu_{hl} = \psi\omega_{hm}$ for $m = 1, \ldots, L_0$, $l = m + (h-1)L_0$. Thus there are at most L response (birth weight) clusters, with the first L_0 corresponding to global clusters that can occur in any of the K predictor clusters. Let $\phi_i = l$ if $\mu_i = \mu_l^*$ denote that woman i is allocated to the lth response cluster. Then L represents an upper bound on the number of clusters, because $\sum_{i=1}^{n} 1(\phi_i = l)$ often equals 0 for many $l \in \{1, \ldots, L\}$, particularly if L is large. Note that ψ is a key hyperparameter controlling the weight assigned to the global component and thus the degree of similarity in the birth weight distributions across the pregnancy weight gain clusters. To allow the data to inform more strongly, we assign a beta hyperprior to ψ.

2.3 Properties of Clustering Process

Our interest lies in assessing changes in the birth weight density across pregnancy weight gain clusters and not in birth weight clustering. Thus allowing different birth weight clusters in this application is simply a convenient tool for allowing flexible changes in the density. It remains of interest to characterize the process for allocating women to response components in a predictor cluster-dependent manner, however. This process is not intuitive from the model specification given in Section 2.2 without further derivations.

Introducing some additional notation, let $\boldsymbol{\beta}^{*(i-1)} = \{\boldsymbol{\beta}_h^*\}_{h=1}^{r_{i-1}}$ denote the r_{i-1} unique values of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{i-1}$ in the order in which they appear, and, similarly, let $\boldsymbol{\mu}^{*(i-1)} = \{\mu_l^*\}_{l=1}^{s_{i-1}}$ denote the s_{i-1} unique values of $\mu_1, \ldots, \mu_{i-1}$ in their order of appearance. Let $\gamma_i = h$ if $\boldsymbol{\beta}_i = \boldsymbol{\beta}_h^*$ index membership of subject i in predictor cluster h, let $\phi_i = l$ if $\mu_i = \mu_l^*$ index membership in response cluster l, and let $\mathcal{S}_l = h$ denote that μ_l^* is an component from P_{2h}^*. Two subjects, i and i', having $\gamma_i \ne \gamma_{i'}$ can have $\phi_i = \phi_{i'}$ if both subjects are assigned to the same cluster within the global component, P_{20}^*. Let $w_h^{(i-1)} = \sum_{j=1}^{i-1} 1_{(\gamma_i = h)}$ denote the number of subjects among $\{1, \ldots, i-1\}$ in predictor cluster h, and let $n_l^{(i-1)} = \sum_{j=1}^{i-1} 1_{(\phi_i = l)}$ denote the number of subjects among $\{1, \ldots, i-1\}$ in response cluster l.

Theorem 1. Assuming (3) and (8), we obtain the conditional distribution

$$\left(\boldsymbol{\beta}_i \mid \boldsymbol{\beta}^{*(i-1)}, \boldsymbol{\gamma}^{(i-1)}\right)$$

$$\sim \left(\frac{\alpha(1 - r_{i-1}/K)}{\alpha + i - 1}\right) P_{01} + \sum_{h=1}^{r_{i-1}} \left(\frac{\alpha/K + w_h^{(i-1)}}{\alpha + i - 1}\right) \delta_{\boldsymbol{\beta}_h^*}, \quad (9)$$

noting that $\boldsymbol{\beta}_i$ is conditionally independent of $\mathbf{S}^{(i-1)}$, $\boldsymbol{\mu}^{*(i-1)}$, and $\boldsymbol{\phi}^{(i-1)}$. In addition,

$$\begin{aligned}
&\left(\mu_i \mid \mathbf{S}^{(i-1)}, \boldsymbol{\mu}^{*(i-1)}, \boldsymbol{\phi}^{(i-1)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\beta}^{*(i)}\right) \\
&\quad \sim \left\{ \frac{(1-\psi)\lambda(L - \sum_{l=1}^{s_{i-1}} 1_{(\mathcal{S}_l=0)})/L}{\lambda + \sum_{l=1}^{s_{i-1}} 1_{(\mathcal{S}_l=0)} n_l^{(i-1)}} \right. \\
&\qquad \left. + \frac{\psi\lambda(L - \sum_{l=1}^{s_{i-1}} 1_{\mathcal{S}_l=\gamma_i})/L}{\lambda + \sum_{l=1}^{s_{i-1}} 1_{(\mathcal{S}_l=\gamma_i)} n_l^{(i-1)}} \right\} P_{02} \\
&\qquad + \sum_{l=1}^{s_{i-1}} \left\{ \frac{(1-\psi) 1_{(\mathcal{S}_l=0)} (\lambda/L + n_l^{(i-1)})}{\lambda + \sum_{t=1}^{s_{i-1}} 1_{(\mathcal{S}_t=0)} n_t^{(i-1)}} \right. \\
&\qquad\qquad \left. + \frac{\psi 1_{(\mathcal{S}_l=\gamma_i)} (\lambda/L + n_t^{(i-1)})}{\lambda + \sum_{t=1}^{s_{i-1}} 1_{(\mathcal{S}_l=\gamma_i)} n_t^{(i-1)}} \right\} \delta_{\mu_l^*}. \qquad (10)
\end{aligned}$$

Expression (9) is a straightforward consequence of model (3) and properties of the Dirichlet distribution. Closely related forms are well known. Expression (10) follows from expression (8) using a similar approach to that used in deriving (9), taking care to keep track of the component from which each atom is drawn.

Note that in the limiting case as $K \to \infty$, expression (9) simplifies to the Polya urn scheme of Blackwell and MacQueen (1973), whereas for $L \to \infty$, (10) becomes

$$\begin{aligned}
&\left(\mu_i \mid \boldsymbol{S}^{(i-1)}, \boldsymbol{\mu}^{*(i-1)}, \boldsymbol{\phi}^{(i-1)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\beta}^{*(i)}\right) \\
&\quad \sim \left\{ \frac{(1-\psi)\lambda}{\lambda + \sum_{l=1}^{s_{i-1}} 1_{(\mathcal{S}_l=0)} n_l^{(i-1)}} \right. \\
&\qquad \left. + \frac{\psi\lambda}{\lambda + \sum_{l=1}^{s_{i-1}} 1_{(\mathcal{S}_l=\gamma_i)} n_l^{(i-1)}} \right\} P_{02} \\
&\qquad + \sum_{l=1}^{s_{i-1}} \left\{ \frac{(1-\psi) 1_{(\mathcal{S}_l=0)} n_l^{(i-1)}}{\lambda + \sum_{t=1}^{s_{i-1}} 1_{(\mathcal{S}_t=0)} n_t^{(i-1)}} \right. \\
&\qquad\qquad \left. + \frac{\psi 1_{(\mathcal{S}_l=\gamma_i)} + n_t^{(i-1)}}{\lambda + \sum_{t=1}^{s_{i-1}} 1_{(\mathcal{S}_l=\gamma_i)} n_t^{(i-1)}} \right\} \delta_{\mu_l^*}. \qquad (11)
\end{aligned}$$

The conditional distributions (9) and (10) provide insight into the joint clustering process for predictors and responses. Due to exchangeability of the subjects, we can obtain a simple form for the conditional and unconditional probabilities of allocating subjects i and j to the same response component given their predictor allocation. The conditional prior probability of response clustering for i and j is

$$\Pr(\mu_i = \mu_j \mid \gamma_i, \gamma_j) = \frac{\lambda/L + 1}{\lambda + 1} \left\{ \psi^2 1_{(\gamma_i=\gamma_j)} + (1-\psi)^2 \right\}, \qquad (12)$$

and the marginal prior probability is

$$\Pr(\mu_i = \mu_j) = \frac{\lambda/L + 1}{\lambda + 1} \left\{ \frac{\psi^2(\alpha/K + 1)}{\alpha + 1} + (1-\psi)^2 \right\}. \qquad (13)$$

Thus in the limit as $\psi \to 0$, $\Pr(\mu_i = \mu_j) = (\lambda/L + 1)/(\lambda + 1)$ and predictor and response clustering occur independently. In the limit as $\psi \to 1$, $\Pr(\mu_i = \mu_j) = 0$ if subjects i and j are not in same predictor cluster and $\Pr(\mu_i = \mu_j) = (\lambda/L + 1)/(1 + \lambda)$ otherwise. In the limiting case as $\psi \to 1$ and $\lambda \to 0$, the same clusters of subjects are obtained for the predictor and response component. Taking $\psi \to 1, \lambda \to 0, K \to \infty$, we obtain the joint DP approach described in the beginning of Section 2.1.

2.4 Chinese Restaurants for Families

A Chinese restaurant metaphor is often used as an aid in understanding DP clustering. Under the Chinese restaurant process (CRP) (Aldous 1985; Pitman 1996), customers arrive sequentially at a restaurant with infinitely many tables, each of which can seat infinitely many individuals. The first customer is seated at the first table, and as additional customers arrive, they are assigned either to an occupied table or to an empty table. For the ith customer, the probability of assignment to an empty table is $\alpha/(\alpha + i - 1)$, whereas the probability of assignment to the jth occupied table is $n_{ij}/(\alpha + i - 1)$, with n_{ij} denoting the number of customers at table j when individual i arrives. The CRP allocates customers to tables according to the $DP(\alpha P_0)$ Polya urn scheme, with the individuals at table j sharing a dish $\boldsymbol{\theta}_j^*$ sampled from P_0.

This metaphor can be generalized to describe the process of Sections 2.1 and 2.2 in the limiting case as $K, L \to \infty$. Suppose that families arrive at a Chinese restaurant sequentially and are assigned to tables by a typical CRP with parameter α, with parents at table j sharing a dish $\boldsymbol{\beta}_j^*$ sampled from P_{01}. Now suppose that there are bags containing infinitely many games at the front of the restaurant and at each of the tables. When a family arrives, the children choose a game to share from either the bag at the front (with probability $1 - \psi$) or the bag at their table (with probability ψ). The first group of siblings to sample from a bag is given the first game in the bag, whereas the ith group of siblings to sample from a bag is given a new game with probability $\lambda/(\lambda + i - 1)$ and otherwise join a game already drawn from the bag with probability $m_{ij}/(\lambda + i - 1)$, with m_{ij} the number of earlier children assigned to the same game. Games are generated independently from P_{02}.

In our metaphor, the predictors are represented by the parents, with the predictor clusters the tables and the dishes the parameter values specific to a predictor cluster. In addition, the responses are represented by the children, with children assigned to the same game belonging to the same response cluster.

3. POSTERIOR COMPUTATION

For ease in interpretation, to more rapidly analyze the large data set that we are faced with, and to address the label-switching problem (Stephens 2000), we avoided a full Markov chain Monte Carlo (MCMC) analysis of the model proposed in Section 2. Instead, we rely on a combined sequential updating with greedy search (SUGS) (Wang and Dunson 2007) and Gibbs sampling algorithm. The SUGS algorithm sequentially updates the prior by adding subjects one at a time, selecting the cluster that maximizes the conditional posterior model probability. This results in a greedy stepwise search for a good partition of subjects into clusters, while requiring only a single cycle of computation for each subject.

We propose using SUGS to choose the partition of subjects into functional predictor clusters and Gibbs sampling for posterior computation of the predictor cluster-specific response

densities. We start by introducing an infinite sequence of predictor clusters, $\{\boldsymbol{\beta}_h^*\}_{h=1}^{\infty}$. The posterior distribution of the parameters in predictor cluster h, conditionally on the data for subjects $1, \ldots, i$ and on predictor cluster allocation $\boldsymbol{\gamma}^{(i)} = (\gamma_1, \ldots, \gamma_{i-1})'$, is

$$\left(\boldsymbol{\beta}_h^* \mid \mathbf{X}^{(i)}, \boldsymbol{\gamma}^{(i)}\right) \propto P_{01}(\boldsymbol{\beta}_h^*) \prod_{j=1}^{i} \left\{ \prod_{l=1}^{n_i} N(x_{jl}; \mathbf{b}_{jl}' \boldsymbol{\beta}_h^*, \tau_x^{-1}) \right\}^{1_{(\gamma_j = h)}}, \quad (14)$$

which is simply P_{01} updated with the likelihood for those subjects in predictor cluster h among subjects $\{1, \ldots, i\}$. Using the updating equations given in the Appendix, (14) is available in closed form for conjugate P_{01}.

Due to the well-known label ambiguity problem (Stephens 2000), the index values in $\boldsymbol{\gamma}$ are inherently arbitrary, and the important information is simply which subjects are grouped together. For example, $\gamma = 1$ could be a cluster of women with rapid weight gain, and $\gamma = 2$ could be a cluster with no weight gain in the first trimester followed by slow weight gain. The label $\gamma = 1$ for the rapid weight gain cluster is arbitrary, and an identical marginal likelihood could be obtained by switching the labels to let $\gamma = 2$ for the women in the rapid weight gain cluster and $\gamma = 1$ for the women in the slow weight gain cluster. Thus, without restriction, we let $\gamma_1 = 1$ to assign subject $i = 1$ to the first predictor cluster. The SUGS algorithm proceeds as follows:

1. Let $\gamma_1 = 1$ and calculate $(\boldsymbol{\beta}_1^* \mid \mathbf{x}_1, \gamma_1 = 1)$ using (14).
2. For $i = 2, \ldots, n$, letting $m_{i-1}^{\gamma} = \max\{\gamma_1, \ldots, \gamma_{i-1}\}$, perform the following steps:

   a. Choose $\gamma_i \in \{1, \ldots, m_{i-1}^{\gamma} + 1\}$ to maximize the conditional posterior probability of $\gamma_i = h$,

$$\Pr\left(\gamma_i = h \mid \mathbf{X}^{(i)}, \mathbf{y}^{(i-1)}, \boldsymbol{\gamma}^{(i-1)}, \boldsymbol{\phi}^{(i-1)}, \tau_x\right) = u_{ih} \int \prod_{j=1}^{n_i} N(x_{ij}; \mathbf{b}_{ij}' \boldsymbol{\beta}_h^*, \tau_x^{-1}) \times d\left(\boldsymbol{\beta}_h^* \mid \mathbf{X}^{(i-1)}, \boldsymbol{\gamma}^{(i-1)}\right) \Bigg/ \left\{ \sum_{l=1}^{m_{i-1}^{\gamma}+1} u_{il} \int \prod_{j=1}^{n_i} N(x_{ij}; \mathbf{b}_{ij}' \boldsymbol{\beta}_l^*, \tau_x^{-1}) \times d\left(\boldsymbol{\beta}_l^* \mid \mathbf{X}^{(i-1)}, \boldsymbol{\gamma}^{(i-1)}\right) \right\},$$

   for $h = 1, \ldots, r_{i-1} + 1$, $u_{ih} = (\alpha/K + w_h^{(i-1)})/(\alpha + i - 1)$, for $h = 1, \ldots, r_{i-1}$, and $u_{ih} = \alpha(r_{i-1}/K - 1)/(\alpha + i - 1)$ for $h = r_{i-1} + 1$.

   b. Compute $(\boldsymbol{\beta}_{\gamma_i}^* \mid \mathbf{X}^{(i-1)}, \boldsymbol{\gamma}^{(i-1)})$ using the data $\mathbf{x}_i$ for subject i.

Note that the conditional probability of allocating subject i to predictor cluster h in step 2a does not depend on unknowns in the response component model, so the allocation of subjects to predictor clusters can proceed through this fast sequential updating algorithm in a first stage. No approximations are needed in the sequential updating for conjugate P_{01}. The procedure selects a single partition, $\boldsymbol{\gamma}$, instead of model averaging across possible partitions. Model averaging is appealing for prediction but does not address our interest in inference on changes in the response distribution as the functional predictor changes or in obtaining results that are easily interpretable to a subject matter audience.

In the second stage, we apply a Gibbs sampling algorithm with the following steps:

1. Update the allocation of subjects to response clusters, $\boldsymbol{\phi}$, by sampling from the multinomial full conditional posterior distribution of ϕ_i given $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\gamma}$, ψ, $\boldsymbol{\nu}^*$, $\boldsymbol{\mu}^*$, and τ_y, for $i = 1, \ldots, n$.
2. Update $\boldsymbol{\nu}_h^*$, for $h = 0, 1, \ldots, K$, by sampling from the conjugate Dirichlet conditional posterior distribution.
3. Update the cluster-specific means $\boldsymbol{\mu}^*$ from their normal full conditional posterior distribution.
4. Update the precision parameter τ_y from its gamma full conditional.
5. Update ψ from its beta full conditional posterior distribution.

Each of these steps is simple to implement rapidly, involving direct sampling from standard distributions. In cases that we have considered, the algorithm is quite efficient, with rapid apparent convergence and good rates of mixing. Note that, unlike for the predictor component, we are model averaging across partitions $\boldsymbol{\phi}$ of subjects into response clusters. This results in more realistic measures of uncertainty in estimates of the conditional response densities with no difficulties in interpretation, as we illustrate in Section 4.

4. WEIGHT GAIN DURING PREGNANCY AND BIRTH WEIGHT

4.1 Description of Data and Scientific Problem

There is considerable interest in assessing the relationship between maternal weight gain during pregnancy and health outcomes. Inadequate or excessive weight gain during critical stages of pregnancy may indicate or even contribute to problems for both the developing child and the mother. For normal weight women, the IOM recommends gaining 4–6 lbs in the first trimester and 1 lb/week during the second to third trimesters, a recommendation motivated by the desire to avoid low birth weight.

Using data from the Pregnancy, Infection, and Nutrition (PIN) study (Savitz et al. 1999), our goal is to relate trajectories in weight gain during pregnancy to the distribution of gestational age at delivery-adjusted birth weight. The PIN study enrolled women during the second trimester of pregnancy. The woman's height was measured, she was asked her weight before pregnancy at the first prenatal visit or the time of recruitment, and weight data were collected from each clinic visit during pregnancy. A total of $n = 1{,}888$ women had a baseline BMI in the [19.8, 26] range, corresponding to "normal weight" in the IOM guidelines. The number of weight measurements per woman ranged from 1 to 29, with a mean of 10.9, discarding weights that were clearly misrecorded.

To characterize the weight gain trajectories, we used a cubic spline model with knots at the trimesters, motivated by recommendations for weight gain, which are trimester-specific,

$$f_i(t) = \beta_{i1}t + \beta_{i2}t^2 + \beta_{i3}t^3 + \beta_{i4}(t-13)_+^3 + \beta_{i5}(t-26)_+^3,$$

where t is time of gestation in weeks, with $t = 0$ at the time of the last menstrual period. As upper bounds on the number of clusters per component, we let $K = 5$ and $L = 4$. This value of K was motivated by interpretability and computational efficiency. Repeating the analysis for $K = 25$, we obtained essentially the same five common trajectory clusters but with several additional outlying trajectory clusters containing <3% of the subjects. The approach of Jones and Nagin (2007) selected $K = 5$ as the number of trajectory clusters based on the BIC. The value of $L = 4$ was motivated by the observation that a mixture of four normal densities is sufficiently flexible to capture a broad variety of densities. In addition, very similar results were obtained for larger values of L.

To complete a Bayesian specification, we choose P_{01} as $N(\boldsymbol{\beta}_i; \boldsymbol{\beta}_0, \Psi\tau_x^{-1})G(\tau_x; a_x, b_x)$ and P_{02} as $N(\mu_i; \kappa\tau_y^{-1})\times G(\tau_y; a_y, b_y)$, with $G(a, b)$ the gamma distribution with mean a/b and variance a/b^2. We use a default specification with $\boldsymbol{\beta}_0 = (\mathbf{b}'\mathbf{b})^{-1}\mathbf{b}'\mathbf{x}$ (global least squares estimate of the basis coefficients), $\Psi = n(\mathbf{b}'\mathbf{b})^{-1}$ (a unit information prior covariance), $\mu_0 = \overline{y}$ (global mean of y), $\kappa = 1$ (unit information), $\alpha = \lambda = 1$ (a widely used default value for the DP precision), $a_x = 2$, $b_x = 2\,\mathrm{var}(x_{ij} - \mathbf{b}_{ij}'\boldsymbol{\beta}_0)$, $a_y = 2$, and $b_y = 2\,\mathrm{var}(y_i)$. In addition, we choose a uniform(0, 1) hyperprior for ψ. Results were robust to local changes in the hyperparameter values, and we obtained very similar results for a wide variety of alternative specifications. For example, we tried normalizing the data, and then letting $\boldsymbol{\beta}_0 = \mathbf{0}$, $\Psi =$ identity, $\mu_0 = 0$, and $a_x = b_x = a_y = b_y = 1$. In addition, we modified the hyperparameters in the initial specification by a factor of 2.

4.2 Analysis and Results

We applied the combined sequential updating and Gibbs sampling algorithm proposed in Section 3, taking 50,000 Gibbs samples. To assess convergence and mixing of the Gibbs sampler, we monitored values of the birth weight density within different pregnancy weight gain clusters and for various birth weight values, including values in the tails and center of the distribution. Based on examination of trace plots and application of standard diagnostic tests, we found no evidence of lack of convergence, and mixing was good. Note that the meaning of the labels for the response (birth weight) clusters will vary over the Gibbs iterations, but this does not present a problem for estimating posterior summaries of the pregnancy weight gain cluster-specific birth weight densities.

The 1,888 women were clustered into five weight gain trajectory clusters, which we order according to total weight gain, with the number (%) in each cluster: (lowest weight gain, magenta, 98 (5.2%); blue, 617 32.7%; black, 774 (41.0%); cyan, 175 (9.3%); and green, 224 (11.9%). Figure 1 shows the estimated weight gain trajectory clusters and corresponding pointwise 95% credible intervals. These trajectories are scientifically reasonable, with most women having a slower rate of weight gain in the first trimester (or even small amounts of weight loss, common in women with nausea and vomiting associated with

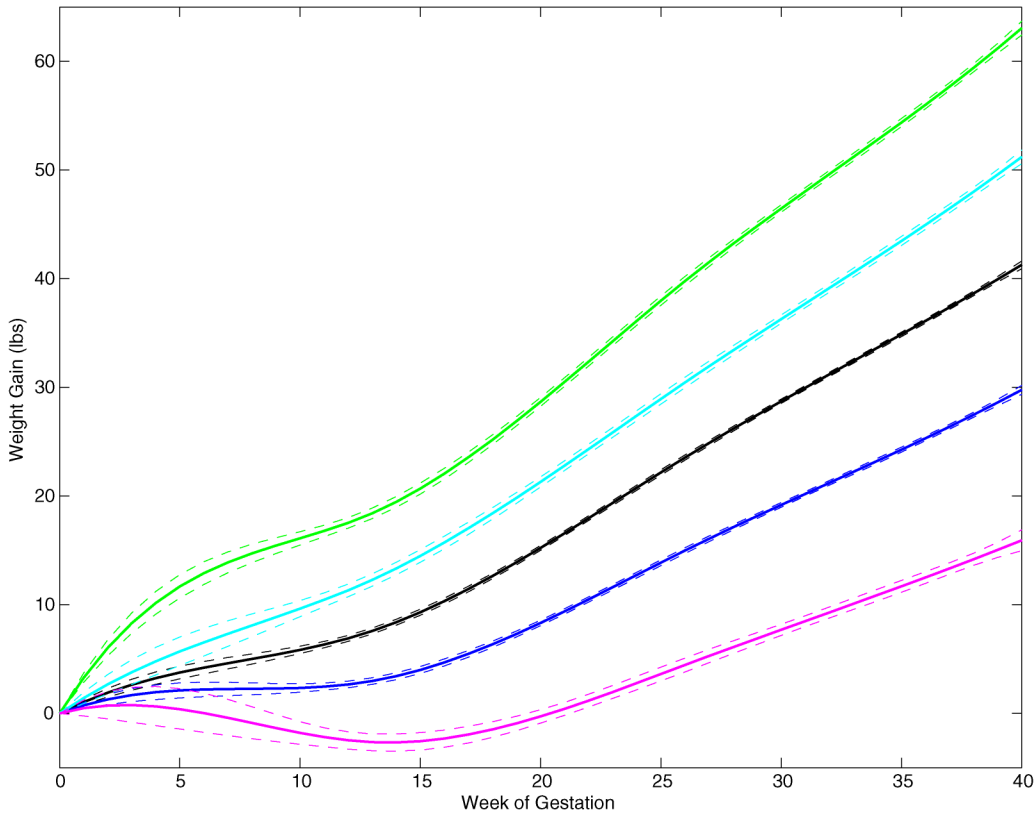

Figure 1. Estimated weight gain trajectory clusters (——) and 95% credible intervals (- - - -). The proportions of women in each cluster were blue, 32.7%; black, 41.0%; green, 11.9%; magenta, 5.2%; and cyan, 9.3%.

pregnancy-related hormonal changes), followed by an approximately linear rate of gain during the second and third trimesters. The blue cluster, which closely corresponds to the IOM recommendations and has an average gain of 30 pounds, contains approximately 1/3 of the women. Most of the remaining women are in the black cluster, with an average gain of just over 40 pounds, exceeding the IOM recommendation. Twenty percent of the women have even higher weight gain, with women in the cyan and green clusters gaining an average of 50 and 60 pounds, respectively, whereas 5% are in the lowest weight gain cluster (magenta), corresponding to an average weight gain around 15 pounds. Repeating the analysis for a random reordering of women in the data set, we obtained similar results. To illustrate the extent to which the real data mimic these estimated trajectories, we plot the raw data for 25 randomly selected women in each cluster in Figure 2.

Figure 3 shows the estimated birth weight densities and cumulative distribution functions (adjusted linearly for gestational age at delivery) for women in each weight gain trajectory cluster. Women in the lowest weight gain cluster (magenta), who had a dip in weight in early pregnancy, had significantly lighter infants than women in the other clusters. The women in the cluster corresponding to the IOM recommendation (blue) had significantly lighter infants compared with the women in the three higher weight gain clusters. Although the estimated birth weight distributions are unimodal, there is clear evidence of nonnormality, with heavy tails and a tendency toward a left skew. Figure 4 plots estimated posterior densities for different percentiles of the birth weight distribution specific to each trajectory cluster. Interestingly, although there were significantly more low birth weight infants in the lowest weight gain trajectory cluster, the difference between the cluster most closely corresponding to the IOM recommendations and the three higher weight gain clusters was more apparent for normal and high birth weight infants. Table 1 provides posterior probabilities for quantile-specific contrasts. Using a standard cutpoint for low birth weight infants (<2,500 g), women in the lowest weight gain cluster had significantly more low birth weight infants (17.2%) than women in the higher weight gain clusters (9.3%, 6.8%, 6.6%, and 6.8%), which is of considerable concern given that low birth weight infants are at increased risk for morbidity and mortality.

Applying the approach of Jones and Nagin (2007) in SAS, we obtained similar trajectories to those shown in Figure 1, though the estimates of Jones and Nagin were less smooth early and late in gestation due to edge effects. As in our analysis, the approach of Jones and Nagin concluded that women in the two lowest weight gain trajectory clusters had infants with significantly lower birth weight, adjusting for gestational age at delivery. But by assuming normality of the birth weight distribution for women within a trajectory cluster, with only the means varying across trajectory clusters, the approach of Jones and Nagin does not allow separate inferences on different quantiles of the birth weight distribution. Such inferences are of primary importance in assessing effects of maternal weight gain on risk of small and large for gestational age infants.

A simple alternative is to use frequentist kernel smoothing to obtain estimated birth weight densities within each trajectory class produced by the approach of Jones and Nagin. Such an analysis produces similar but less smooth estimates as those shown in Figure 3. Our proposed approach has the advantage of borrowing information across the latent classes in

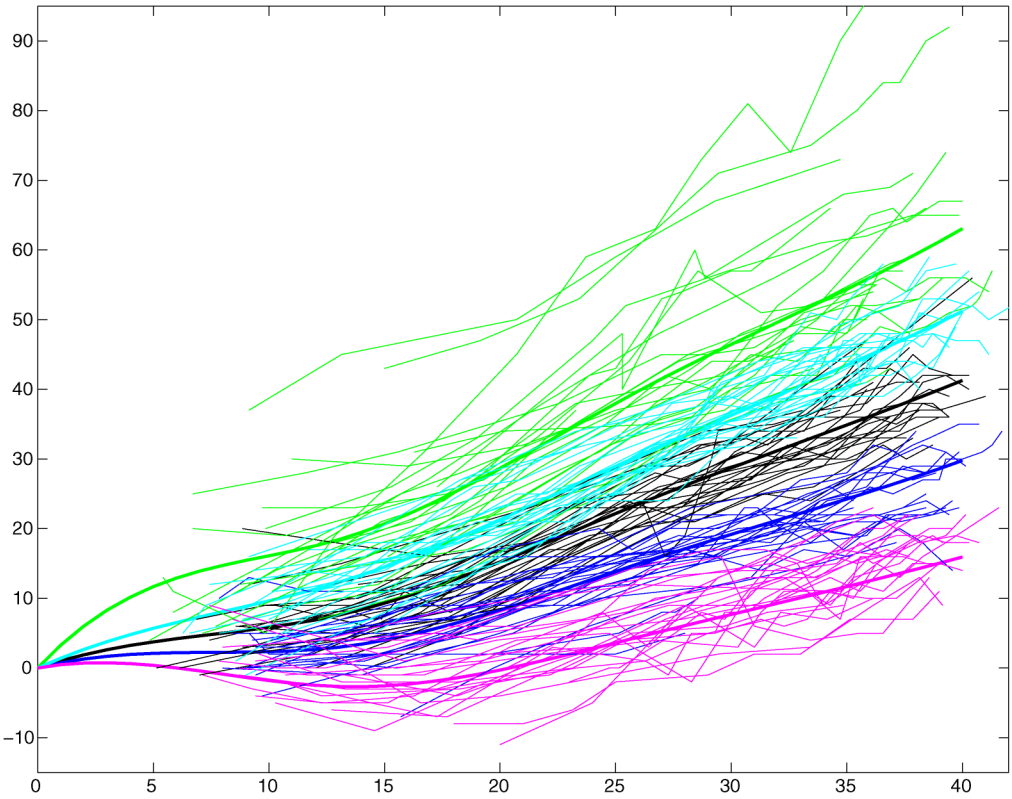

Figure 2. Raw data for 25 randomly selected women in each identified trajectory cluster.

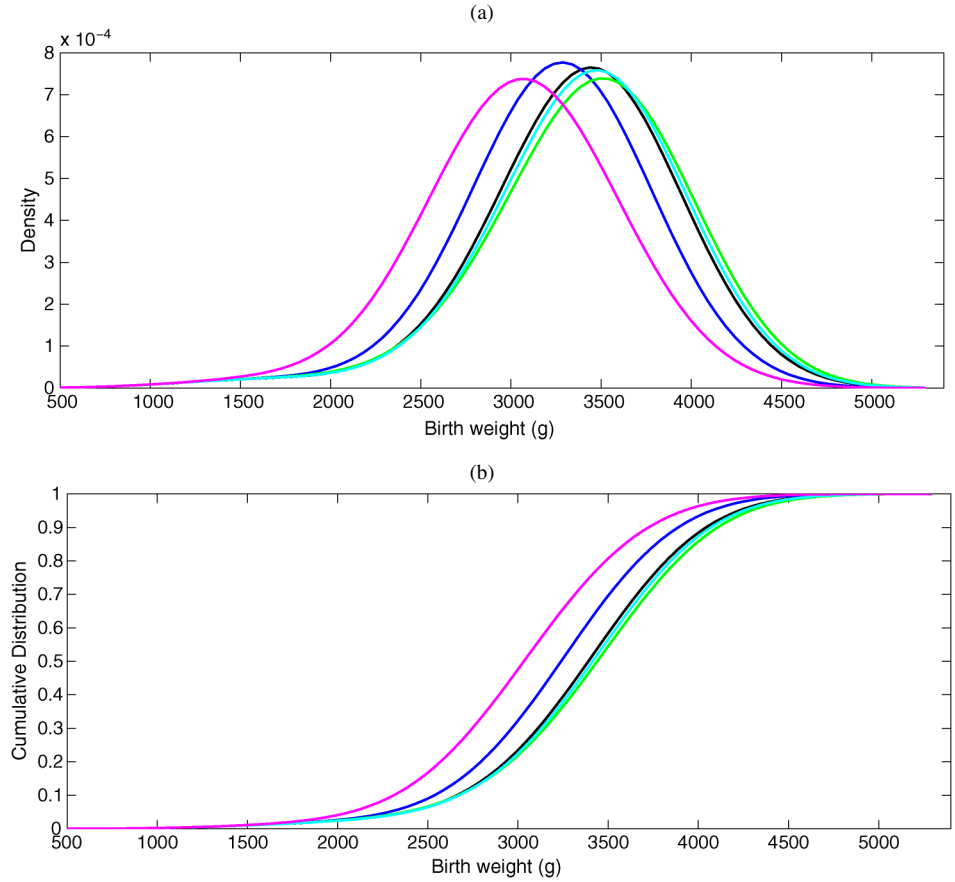

Figure 3. Estimated density (a) and cumulative distribution function (b) for birth weight at 40 weeks for women in each weight gain trajectory cluster.

performing density estimation. In addition, our approach automatically produces measures of uncertainty, while also allowing calculation of posterior probabilities for quantile contrasts.

5. DISCUSSION

In this article we have proposed a method for assessing changes in a response density across clusters in a functional predictor. We applied this approach to a rich data set containing detailed records of weight gain during pregnancy, gestational age at delivery, and birth weight for a large (by typical reproductive epidemiology standards) sample of woman. We found that women in a low pregnancy weight gain cluster had significantly more low birth weight infants, adjusting for gestational age at delivery. In addition, we found that there was a highly significant increase in the risk of large for gestational age infants among women gaining more weight than the current clinical recommendations. These results are of substantial clinical and public health interest, given that most women gain more weight than recommended.

Motivated by label ambiguity problems and difficulties in performing efficient computation, we allocated subjects into predictor clusters using a fast sequential updating and greedy search procedure, using Gibbs sampling to obtain samples from the conditional posterior for the predictor cluster-specific response densities. Our view is that clustering of the predictor data is useful as a dimensionality reduction technique and as an aid in interpreting the complex relationship between a high-dimensional predictor and a response with an unknown distribution. We generally do not recommend interpreting such clusters as corresponding to biologically or clinically distinct groups of individuals, however, because the formation of clusters is necessarily dependent on parametric assumptions.

Following Quintana (2006), we view partitioning as a model selection problem. From this viewpoint, our sequential updating and greedy search algorithm is a type of stepwise selection procedure for rapidly partitioning the subjects into clusters. Our proposed joint clustering prior should be useful in other applications in which allocation of a subject to a cluster for one task predicts cluster allocation for another distinct task. In addition, even when clustering is not of interest, the method can be used to generate flexible joint partition models. Joint partitioning extends current Bayesian partition models (Barry and Hartigan 1992; Quintana and Iglesias 2003; Holmes, Denison, Ray, and Mallick 2005; Quintana 2006) to allow flexible joint modeling of data from different sources. Related problems have been considered in the machine learning literature from a different perspective (Barnard et al. 2003).

APPENDIX: POSTERIOR UPDATING

A.1 Predictor Component

Assuming that $(\boldsymbol{\beta} \mid \tau_x) \sim \mathrm{N}_p(\boldsymbol{\beta}_0, \Psi\tau_x^{-1})$ and $\tau_x \sim G(a_x, b_x)$ and updating with the likelihood $\prod_{j=1}^{n_i} N(x_{ij}; \mathbf{b}'_{ij}\boldsymbol{\beta}, \tau_x^{-1})$, we ob-

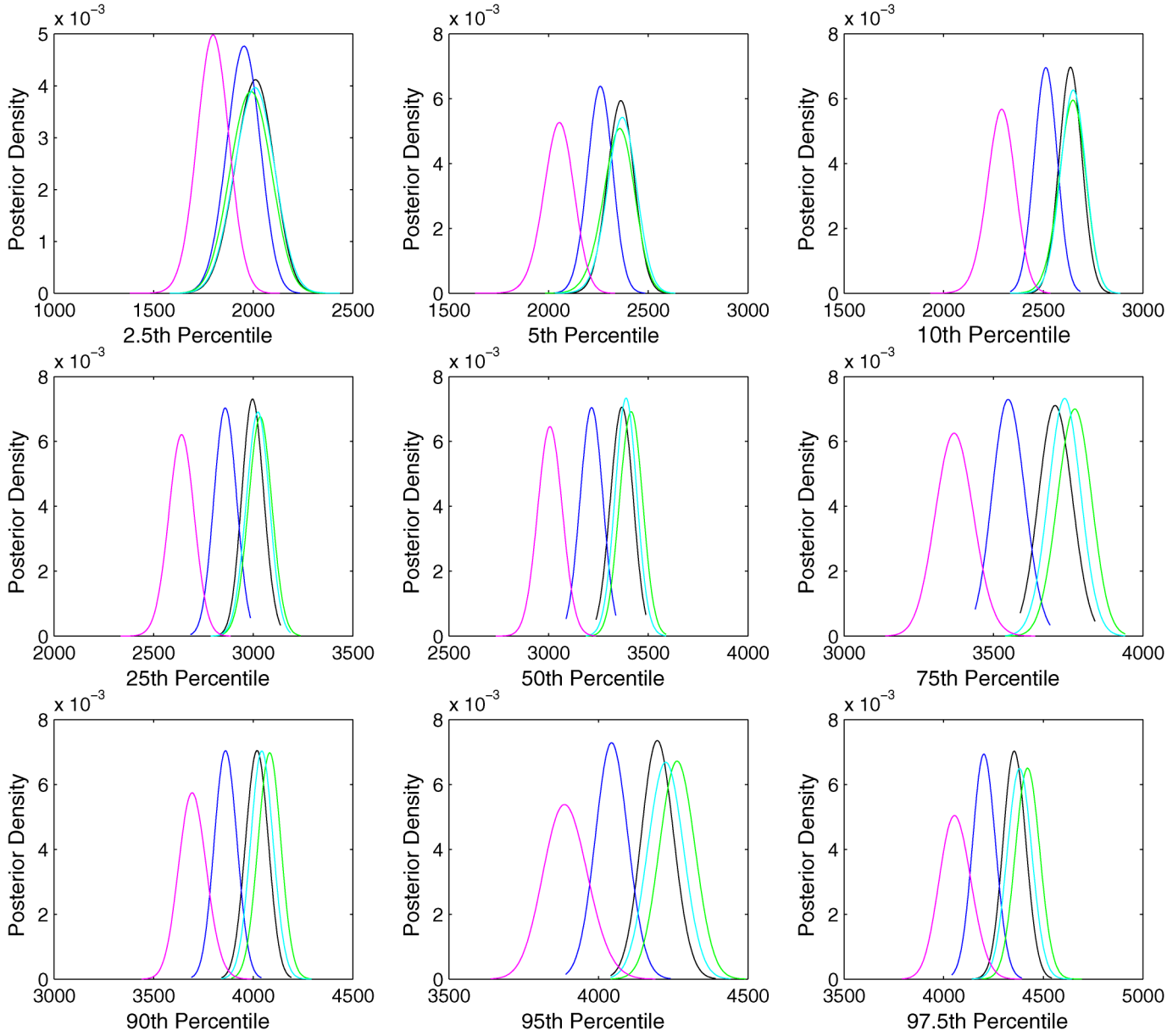

Figure 4. Estimated posterior densities of selected percentiles of the birth weight distribution within each weight gain trajectory cluster.

tain $(\boldsymbol{\beta} \mid \tau_x, \mathbf{x}_i) \sim \mathrm{N}_p(\widehat{\boldsymbol{\beta}}, \widehat{\Psi}\tau_x^{-1})$ and $(\tau_x \mid \mathbf{x}_i) \sim G(\widehat{a}_x, \widehat{b}_x)$, where $\widehat{\Psi} = (\Psi^{-1} + \sum_{j=1}^{n_i} \mathbf{b}_{ij}\mathbf{b}_{ij}')^{-1}$, $\widehat{\boldsymbol{\beta}} = \widehat{\Psi}(\Psi^{-1}\boldsymbol{\beta}_0 + \sum_{j=1}^{n_i} \mathbf{b}_{ij}x_{ij})$, $\widehat{a}_x = a_x + n_i/2$, and

$$\widehat{b}_x = b_x + \frac{1}{2}\left\{\boldsymbol{\beta}_0'\Psi^{-1}\boldsymbol{\beta}_0 + \sum_{j=1}^{n_i} x_{ij}^2 - \widehat{\boldsymbol{\beta}}'\widehat{\Psi}^{-1}\widehat{\boldsymbol{\beta}}\right\}.$$

In addition, the marginal likelihood of $\mathbf{x}_i$ obtained in integrating out $(\boldsymbol{\beta}, \tau_x)$ is

$$f(\mathbf{x}_i) = \frac{\Gamma((\nu + n_i)/2)}{(\pi\nu)^{n_i/2}|\sigma^2\boldsymbol{\Sigma}|^{1/2}\Gamma(\nu/2)} \times \left(1 + \frac{(\mathbf{x}_i - \boldsymbol{\mu}_x)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_x)}{\sigma^2\nu}\right)^{-(\nu+n_i)/2},$$

Table 1. Posterior probabilities of orderings in selected quantiles of the birth weight distribution between weight gain trajectory clusters

	Percentile of distribution								
Contrast	2.5th	5th	10th	25th	50th	75th	90th	95th	97.5th
Magenta < blue	.998	>.999	1.000	1.000	1.000	.999	.995	.988	.976
Blue < black	.815	.992	>.999	1.000	1.000	1.000	1.000	1.000	>.999
Black < cyan	.185	.309	.370	.505	.444	.579	.497	.518	.501
Black < green	.127	.271	.394	.648	.732	.859	.875	.901	.875
Cyan < green	.141	.231	.292	.365	.507	.607	.650	.644	.650

where $\nu = 2a_x$, $\mathbf{\Sigma} = (\mathbf{I}_{n_i} - \mathbf{b}_i \widehat{\Psi} \mathbf{b}_i')^{-1}$, $\boldsymbol{\mu}_x = \mathbf{\Sigma}(\mathbf{b}_i \widehat{\Psi} \Psi^{-1} \boldsymbol{\beta}_0)$, $\mathbf{b}_i = (\mathbf{b}_{i1}, \ldots, \mathbf{b}_{in_i})'$, and

$$\sigma^2 = \frac{1}{\nu}\big(2b_x + \boldsymbol{\beta}_0'(\Psi^{-1} - \Psi^{-1}\widehat{\Psi}\Psi^{-1})\boldsymbol{\beta}_0 - \boldsymbol{\mu}_x' \mathbf{\Sigma}^{-1} \boldsymbol{\mu}_x\big).$$

A.2 Response Component

Assuming that $(\mu \mid \tau_y) \sim \mathrm{N}(\mu_0, \kappa\tau_y^{-1})$ and $\tau_y \sim G(a_y, b_y)$ and updating with the likelihood $N(y_i; \mu, \tau_y^{-1})$, we obtain $(\mu \mid \tau_y, y_i) \sim \mathrm{N}(\widehat{\mu}, \widehat{\kappa}\tau_y^{-1})$ and $(\tau_y \mid y_i) \sim G(\widehat{a}_y, \widehat{b}_y)$, where $\widehat{\kappa} = (\kappa^{-1} + 1)^{-1}$, $\widehat{\mu} = \widehat{\kappa}(\kappa^{-1}\mu_0 + y_i)$, $\widehat{a}_y = a_y + 1/2$, and $\widehat{b}_y = b_y + 1/2(\kappa^{-1}\mu_0^2 + y_i^2 - \widehat{\kappa}^{-1}\widehat{\mu}^2)$. In addition, the marginal likelihood of y_i obtained in integrating out (μ, τ_y) is

$$f(y_i) = \frac{\Gamma((\nu+1)/2)}{(\pi\nu)^{1/2}\sigma\Gamma(\nu/2)}\left(1 + \frac{(y_i - \mu_y)^2}{\sigma^2\nu}\right)^{-(\nu+1)/2},$$

where $\nu = 2a_y$, $\mu_y = \mu_0$, and $\sigma^2 = 2b_y(1+\kappa)/\nu$.

[Received June 2006. Revised January 2008.]

REFERENCES

Aldous, D. J. (1985), "Exchangeability and Related Topics," in *École d'Été de Probabilités de Saint-Flour XII*, ed. P. L. Hennequin, **???**: Springer pp. **???**–**???**.

Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D. M., and Jordan, M. I. (2003), "Matching Words and Pictures," *Journal of Machine Learning Research*, 3, 1107–1135.

Barry, D., and Hartigan, J. A. (1992), "Product Partition Models for Change Point Problems," *The Annals of Statistics*, 20, 260–279.

Bigelow, J. L., and Dunson, D. B. (2005), "Semiparametric Classification in Hierarchical Functional Data Analysis," Discussion Paper 2005-18, Duke University, Dept. of Statistical Science.

Blackwell, D., and MacQueen, J. (1973), "Ferguson Distributions via Pólya Urn Schemes," *The Annals of Statistics*, 1, 353–355.

Carmichael, S., Abrams, B. A., and Selvin, S. (1997), "The Pattern of Maternal Weight Gain in Women With Good Pregnancy Outcomes," *American Journal of Public Health*, 87, 1984–1988.

Dietz, P. M., Callaghan, W. M., Cogswell, M. E., Morrow, B., Ferre, C., and Schieve, L. A. (2006), "Combined Effects of Prepregnancy Body Mass Index and Weight Gain During Pregnancy on the Risk of Preterm Delivery," *Epidemiology*, 17, 170–177.

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

——— (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–629.

Heard, N. A., Holmes, C. C., and Stephens, D. A. (2006), "A Quantitative Study of Gene Regulation Involved in the Immune Response of *Anopheline* Mosquitoes: An Application of Bayesian Hierarchical Clustering of Curves," *Journal of the American Statistical Association*, 101, 18–29.

Holmes, C. C., Denison, D. G. T., Ray, S., and Mallick, B. K. (2005), "Bayesian Prediction via Partitioning," *Journal of Computational and Graphical Statistics*, 14, 811–830.

Institute of Medicine (1990), *Nutrition During Pregnancy*, Washington, DC: National Academies Press.

Ishwaran, H., and Zarepour, M. (2002), "Dirichlet Prior Sieves in Finite Normal Mixtures," *Statistica Sinica*, 12, 941–963.

James, G., and Sugar, C. (2003), "Clustering for Sparsely Sampled Functional Data," *Journal of the American Statistical Association*, 98, 397–408.

Jones, B. L., and Nagin, D. S. (2007), "Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them," *Sociological Methods & Research*, 35, 542–571.

Jones, B. L., Nagin, D. S., and Roeder, K. (2001), "A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories," *Sociological Methods & Research*, 29, 374–393.

Lo, A. Y. (1984), "On a Class of Bayesian Nonparametric Estimates: I, Density Estimates," *The Annals of Statistics*, 12, 351–357.

Müller, P., Quintana, F., and Rosner, G. (2004), "A Method for Combining Inference Across Related Nonparametric Bayesian Models," *Journal of the Royal Statistical Society*, Ser. B, 66, 735–749.

Muthén, B. (2004), "Latent Variable Analysis: Growth Mixture Modelling and Related Techniques for Longitudinal Data," in *Handbook of Quantitative Methodology for the Social Sciences*, ed. D. Kaplan, Newbury Park, CA: Sage, pp. 345–368.

National Research Council and Institute of Medicine (2007), *Influence of Pregnancy Weight on Maternal and Child Health: Workshop Report, Committee on the Impact of Pregnancy Weight on Maternal and Child Health*, Washington, DC: National Academies Press.

Pitman, J. (1996), "Some Developments of the Blackwell–MacQueen Urn Scheme," in *Statistics, Probability and Game Theory*, eds. T. S. Ferguson, L. S. Shapley, and J. B. MacQueen, Heyward, CA: Institute of Mathematical Statistics, pp. 245–267.

Quintana, F. A. (2006), "A Predictive View of Bayesian Clustering," *Journal of Statistical Planning and Inference*, 13, 2407–2429.

Quintana, F. A., and Iglesias, P. L. (2003), "Bayesian Clustering and Product Partition Models," *Journal of the Royal Statistical Society*, Ser. B, 65, 557–574.

Ray, S., and Mallick, B. (2006), "Functional Clustering by Bayesian Wavelet Methods," *Journal of the Royal Statistical Society*, Ser. B, 68, 305–332.

Savitz, D. A., Dole, N., Williams, J., Thorp, J. M., McDonald, T., Carter, A. C., and Eucker, B. (1999), "Determinants of Participation in an Epidemiological Study of Preterm Delivery," *Paediatric and Perinatal Epidemiology*, 13, 114–125.

Siega Riz, A. M., Adair, L. S., and Hobel, C. J. (1994), "Institute of Medicine Maternal Weight Gain Recommendations and Pregnancy Outcome in a Predominantly Hispanic Population," *Obstetrics and Gynecology*, 84, 565–573.

Stephens, M. (2000), "Dealing With Label Switching in Mixture Models," *Journal of the Royal Statistical Society*, Ser. B, 62, 795–809.

Wang, L., and Dunson, D. B. (2007), "Fast Posterior Computation in Dirichlet Process Mixture Models," discussion paper, ISDS, Duke University.