

Semiparametric Bayes Joint Modeling with Functional Predictors

David B. Dunson¹, Amy Herring^{2,3} and Anna Maria Siega-Riz^{3,4}

¹Biostatistics Branch, MD A3-03

National Institute of Environmental Health Sciences

P.O. Box 12233, RTP, NC 27709

dunson1@niehs.nih.gov

²Department of Biostatistics,

³Carolina Population Center,

⁴Department of Epidemiology and Department of Nutrition

The University of North Carolina at Chapel Hill

Summary We consider the problem of semiparametric Bayes joint modeling of predictors and a response variable, with a particular emphasis on functional predictors. Parametric models for the predictor and response are coupled through a joint distribution for subject-specific predictor and response coefficients. This joint distribution is assigned a flexible mixture prior, which allows the response distribution within predictor clusters to be unknown. Marginalizing out the random atoms and random weights, we obtain a useful closed form bivariate predictor rule. To avoid label ambiguity and accelerate computation, we propose a combined sequential updating and Gibbs sampling algorithm for posterior computation. The methods are applied to data on weight gain during pregnancy and birth weight.

Key Words: Bivariate clustering; Dirichlet process; Functional predictors; Growth mixture model; Joint modeling; Latent class trajectory; Prediction rule.

1. Introduction

This article focuses on flexible inferences on the relationship between a predictor and a response, with a particular emphasis on functional predictors. The goal is to limit parametric assumptions, allowing the distributions of the predictor and response to be unknown and to have a flexible dependence structure. In addition, for ease in interpretation and presentation of results, it is desirable to estimate predictor clusters along with response distributions specific to these clusters.

As motivation, we focus on a study relating the functional predictor: weight gain during pregnancy to the response: gestational age-adjusted birth weight. Clinicians and reproductive epidemiologists are interested in assessing how the birth weight distribution for full term pregnancies varies for women having different weight gain trajectories. In order to simplify clinical recommendations, it is useful to group women into different weight gain trajectory clusters, with the birth weight distribution possibly differing across clusters. The tails of the response distribution, corresponding to small and large for gestational age babies, are particularly important clinically. To capture changes in tail behavior, it is important to allow the response distribution to change flexibly across weight gain clusters.

Latent class trajectory models have been increasingly used in functional data analysis. For example, the approach of Jones, Nagin and Roeder (2001) 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) proposes an alternative approach, which allows variability in trajectories among subjects in a trajectory cluster. James and Sugar (2003) instead characterize 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). To select the number of latent trajectory

classes, the BIC is commonly used.

These approaches are quite appealing in dramatically reducing dimensionality in characterizing the relationship between an infinite dimensional function and a response, while facilitating inferences. Although a clear simplification of the data, it is nonetheless of interest to examine the trajectories in the different clusters and to assess how the response distribution varies across these trajectory clusters. That said, we have encountered an important pitfall in applying latent class trajectory models in applications relating time-varying predictors to pregnancy outcomes, such as gestational age and birth weight. The problem arises when one makes parametric assumptions about the trajectory cluster-specific response distributions. For example, for a continuous outcome, such as birth weight, a common strategy would be to assume that the response is normally distributed across subjects within a trajectory cluster, with the mean varying across clusters. As discussed in Section 2.1, we find that the performance is quite sensitive these assumptions, with violations leading to inflation of the number of functional clusters and misleading inferences about cluster-specific response densities.

To address this problem, we propose a Bayesian approach that allows the trajectory cluster-specific response densities to be unknown and varying flexibly across clusters. For recent articles on Bayesian functional clustering, refer to Ray and Mallick (2006) and Heard et al. (2006). Related to the method of Ray and Mallick (2006), Bigelow and Dunson (2007) proposed to express 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, one can instead assign a DP prior 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 above. Unfortunately, such an approach still assumes that the trajectory cluster-specific

response distribution follows a parametric form, with only a location parameter varying. Hence, the results are still quite sensitive to violations of these assumptions.

A primary limitation of the above latent class trajectory models 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 produces a dependent bivariate clustering structure upon marginalization. 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 model and derives the prediction rule. 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. Semiparametric Joint Modeling

2.1 Dirichlet process mixture model

For subject i ($i = 1, \dots, n$), let $f_i : \mathcal{T} \rightarrow \mathfrak{R}$ denote a functional predictor and let y_i denote a response variable. In our motivating application, \mathcal{T} is a Borel subset of \mathfrak{R}^+ . However, the methods apply to very general choices of \mathcal{T} , including continuous and discrete domains with or without bounds. Our interest focuses on assessing the relationship between f_i and y_i using the data $\{\mathbf{x}_i, y_i\}_{i=1}^n$, where $\mathbf{x}_i = (x_{i1}, \dots, x_{i, n_i})'$ and $x_{ij} = f_i(t_{ij}) + \epsilon_{ij}$, with $t_{ij} \in \mathcal{T}$ the time (or location) of measurement j on subject i , $j = 1, \dots, n_i$, and $\epsilon_{ij} \sim N(0, \tau_x^{-1})$.

We focus on the following semiparametric hierarchical model:

$$y_i \sim N(\mu_i, \tau_y^{-1}),$$

$$\begin{aligned}
x_{ij} &\sim \text{N}(f_i(t_{ij}), \tau_x^{-1}), \quad j = 1, \dots, n_i, \\
(f_i, \mu_i) &\sim Q,
\end{aligned} \tag{1}$$

for $i = 1, \dots, n$. Dependence between the functional predictor, $f_i \in \Omega$, and the response, $y_i \in \mathfrak{R}$, is characterized through the probability measure Q . We assume that Ω is the linear span of the basis functions $\{b_h\}_{m=1}^p$, so that any function $f_i \in \Omega$ has the basis expansion:

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

where $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{ip})'$ are subject-specific basis coefficients. Using this form, we have $f_i(t_{ij}) = \mathbf{b}'_{ij} \boldsymbol{\beta}_i$, where $\mathbf{b}_{ij} = (b_{ij1}, \dots, b_{ijp})'$ and $b_{ijm} = b_m(t_{ij})$.

To characterize heterogeneity among subjects while accommodating dependence between f_i and y_i , we let $\boldsymbol{\theta}_i = (\boldsymbol{\beta}'_i, \mu_i)' \sim P$, with P a probability measure on $(\mathfrak{R}^{p+1}, \mathcal{B})$, where \mathcal{B} is the Borel σ -algebra of subsets of \mathfrak{R}^{p+1} . Note that Q is induced through specification of P . Bigelow and Dunson (2007) proposed letting $P \sim DP(\alpha P_0)$, with $DP(\alpha P_0)$ denoting the Dirichlet process with precision parameter α and base measure P_0 . From Blackwell and MacQueen (1973), the DP prior implies the following prediction rule:

$$\boldsymbol{\theta}_1 \sim P_0, \quad (\boldsymbol{\theta}_i | \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{i-1}) \sim \left(\frac{\alpha}{\alpha + i - 1} \right) P_0 + \sum_{j=1}^{i-1} \left(\frac{1}{\alpha + i - 1} \right) \delta_{\boldsymbol{\theta}_j}, \tag{3}$$

where δ_θ is a measure concentrated at θ . This results in the n subjects being allocated to $k \leq n$ unique values $\boldsymbol{\theta}^* = \{\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*\}$. Subjects in cluster h have coefficients $\boldsymbol{\theta}_h^* = (\boldsymbol{\beta}_h^*, \mu_h^*)'$, with $f_h^*(t) = \mathbf{b}(t)' \boldsymbol{\beta}_h^*$ the cluster-specific functional predictor and μ_h^* the cluster-specific mean.

Although we obtain a highly flexible model in applying (1)-(2) with a DP prior on the distribution of the subject-specific parameters $\boldsymbol{\theta}_i$, there are drawbacks to this specification, which requires clusters in the functional predictor and response to be identical. To illustrate the problems that can arise, first suppose that the functional predictor is not informative about the response distribution. In this case, clustering will be driven by both variability in the functional trajectories and lack of fit in the Gaussian distribution for the response. If

the response is Gaussian, the model will tend to be inefficient in unnecessarily introducing a different mean for each functional cluster. If the response is non-Gaussian, the tendency will be to induce more clusters than are necessarily to explain the variability in the functional predictor. This leads to difficulty in interpretation. In Section 2.2, we propose an alternative that allows separate, but dependent, clustering in the functional predictor and response.

2.2 Joint partitioning

We start by considering the following hierarchical model for the subject-specific basis coefficients:

$$\beta_i \sim P_1 = \sum_{h=1}^H \pi_h \delta_{\beta_h^*}, \quad \boldsymbol{\pi} = (\pi_1, \dots, \pi_H)' \sim \text{Diri}\left(\frac{\alpha}{H}, \dots, \frac{\alpha}{H}\right), \quad \beta_h^* \stackrel{iid}{\sim} P_{01}, \quad (4)$$

which implies $P_1 \sim DP(\alpha P_{01})$ in the limit as $H \rightarrow \infty$ (Ishwaran and Zarepour, 2002). Note that this model can be equivalently expressed in hierarchical form as $\beta_i \sim \delta_{\beta_{\gamma_i}^*}$ with $\gamma_i \sim \sum_{h=1}^H \pi_h \delta_h$. Here, $\gamma_i \in \{1, \dots, H\}$ is an integer indexing the predictor cluster membership of subject i . In order to allow the response distribution to change flexibly across predictor clusters, we then let

$$\begin{aligned} (\mu_i | \gamma_i = h) &\sim P_{2h} = \sum_{l=1}^L \nu_{hl} \delta_{\mu_l^*}, \quad \mu_l^* \stackrel{iid}{\sim} P_{02}, \\ \boldsymbol{\nu}_h &= (\nu_{h1}, \dots, \nu_{hL})' \sim (1 - \psi) \delta_{\boldsymbol{\nu}_0^*} + \psi \delta_{\boldsymbol{\nu}_h^*}, \\ \boldsymbol{\nu}_h^* &\stackrel{iid}{\sim} \text{Diri}\left(\frac{\lambda}{L}, \dots, \frac{\lambda}{L}\right), \quad h = 0, 1, \dots, H, \end{aligned} \quad (5)$$

which implies that $P_{2h} \sim (1 - \psi)P_{20}^* + \psi P_{2h}^*$, with $P_{2h}^* \sim DP(\lambda P_{02})$ independently for $h = 0, 1, \dots, H$ in the limit as $L \rightarrow \infty$. As $L \rightarrow \infty$ the mixture distribution for the response model parameters is assigned a mixture of a global DP component, P_{20}^* , and a component specific to predictor cluster h , P_{2h}^* . For additional flexibility, we choose a beta hyperprior for ψ . This mixture form is related to that proposed by Müller, Quintana and Rosner (2004) for borrowing information from related studies. However, in their setting γ_i was a known study index instead of an unknown cluster index.

Note that (5) can be expressed in hierarchical form as $\mu_i \sim \mu_{\phi_i}^*$ with $(\phi_i | \gamma_i = h) \sim \sum_{l=1}^L \nu_{hl} \delta_l$, where $\phi_i \in \{1, \dots, L\}$ is a response cluster index. Clearly, the allocation of individuals to response clusters will depend in part on the predictor cluster allocation. In order to obtain insight into this clustering process, we use (4) and (5) as starting points in deriving a prediction rule, focusing on the limiting case as $H, L \rightarrow \infty$.

With this goal in mind, we start by introducing some additional notation. In particular, let $\boldsymbol{\beta}^{*(i-1)} = \{\boldsymbol{\beta}_h^*\}_{h=1}^{r_{i-1}}$ denote the r_{i-1} unique values of $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{i-1}$ in the order that they appear, and similarly let $\boldsymbol{\mu}^{*(i-1)} = \{\mu_l^*\}_{l=1}^{s_{i-1}}$ denote the s_{i-1} unique values of μ_1, \dots, μ_{i-1} in order of appearance. Let $\gamma_i = h$ if $\boldsymbol{\beta}_i = \boldsymbol{\beta}_h^*$ index membership of subject i in predictor cluster h , $\phi_i = l$ if $\mu_i = \mu_l^*$ index membership in response cluster l , and $\mathcal{S}_i = h$ denote that μ_i^* is an atom from P_{2h}^* . As will become clear, two subjects i and i' having $\gamma_i \neq \gamma_{i'}$ can have $\phi_i = \phi_{i'}$ if both subjects are assigned to the same atom within the global component P_{20}^* . Let $w_h^{(i-1)} = \sum_{j=1}^{i-1} 1_{(\gamma_j=h)}$ denote the number of subjects among $\{1, \dots, i-1\}$ in predictor cluster h , and $n_l^{(i-1)} = \sum_{j=1}^{i-1} 1_{(\phi_j=l)}$ denote the number of subjects among $\{1, \dots, i-1\}$ in response cluster l .

Theorem 1. Assuming (4) and (5), we obtain the prediction rule

$$(\boldsymbol{\beta}_i | \boldsymbol{\beta}^{*(i-1)}, \boldsymbol{\gamma}^{(i-1)}) \sim \left(\frac{\alpha(1 - r_{i-1}/K)}{\alpha + i - 1} \right) P_{01} + \sum_{h=1}^{r_{i-1}} \left(\frac{\alpha/K + w_h^{(i-1)}}{\alpha + i - 1} \right) \delta_{\boldsymbol{\beta}_h^*} \quad (6)$$

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

$$\begin{aligned} & (\mu_i | \mathbf{S}^{(i-1)}, \boldsymbol{\mu}^{*(i-1)}, \boldsymbol{\phi}^{(i-1)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\beta}^{*(i)}) \\ & \sim \left\{ \frac{(1 - \psi)\lambda(L - \sum_{l=1}^{s_{i-1}} 1_{(\mathcal{S}_l=0)})/L}{\lambda + \sum_{l=1}^{s_{i-1}} 1_{(\mathcal{S}_l=0)} n_l^{(i-1)}} + \frac{\psi\lambda(L - \sum_{l=1}^{s_{i-1}} 1_{(\mathcal{S}_l=\gamma_i)})/L}{\lambda + \sum_{l=1}^{s_{i-1}} 1_{(\mathcal{S}_l=\gamma_i)} n_l^{(i-1)}} \right\} P_{02} \\ & + \sum_{l=1}^{s_{i-1}} \left\{ \frac{(1 - \psi)1_{(\mathcal{S}_l=0)}(\lambda/L + n_l^{(i-1)})}{\lambda + \sum_{t=1}^{s_{i-1}} 1_{(\mathcal{S}_t=0)} n_t^{(i-1)}} + \frac{\psi 1_{(\mathcal{S}_l=\gamma_i)}(\lambda/L + n_t^{(i-1)})}{\lambda + \sum_{t=1}^{s_{i-1}} 1_{(\mathcal{S}_t=\gamma_i)} n_t^{(i-1)}} \right\} \delta_{\mu_l^*}. \quad (7) \end{aligned}$$

Expression (6) is a straightforward consequence of model (4) and properties of the Dirichlet distribution. Closely related forms are well known. Expression (7) follows from expression

(5) using a similar approach to that used in deriving (6), taking care to keep track of the component from which each atom is drawn.

Note that in the limiting case as $K \rightarrow \infty$, expression (6) simplifies to the Blackwell and MacQueen (1973) Polya urn scheme, while for $L \rightarrow \infty$ (7) becomes

$$\begin{aligned}
& (\mu_i | \mathbf{S}^{(i-1)}, \boldsymbol{\mu}^{*(i-1)}, \boldsymbol{\phi}^{(i-1)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\beta}^{*(i)}) \\
& \sim \left\{ \frac{(1-\psi)\lambda}{\lambda + \sum_{l=1}^{s_{i-1}} \mathbf{1}_{(S_l=0)} n_l^{(i-1)}} + \frac{\psi\lambda}{\lambda + \sum_{l=1}^{s_{i-1}} \mathbf{1}_{(S_l=\gamma_i)} n_l^{(i-1)}} \right\} P_{02} \\
& + \sum_{l=1}^{s_{i-1}} \left\{ \frac{(1-\psi)\mathbf{1}_{(S_l=0)} n_l^{(i-1)}}{\lambda + \sum_{t=1}^{s_{i-1}} \mathbf{1}_{(S_t=0)} n_t^{(i-1)}} + \frac{\psi\mathbf{1}_{(S_l=\gamma_i)} + n_t^{(i-1)}}{\lambda + \sum_{t=1}^{s_{i-1}} \mathbf{1}_{(S_t=\gamma_i)} n_t^{(i-1)}} \right\} \delta_{\mu_i^*}. \quad (8)
\end{aligned}$$

The prediction rule (6)-(7) provides insight into the joint clustering process for predictors and response. 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

$$\Pr(\mu_i = \mu_j | \gamma_i, \gamma_j) = \frac{\lambda/L + 1}{\lambda + 1} \{ \psi^2 \mathbf{1}_{(\gamma_i=\gamma_j)} + (1-\psi)^2 \}, \quad (9)$$

while the marginal prior probability is

$$\Pr(\mu_i = \mu_j) = \frac{\lambda/L + 1}{\lambda + 1} \left\{ \frac{\psi^2(\alpha/K + 1)}{\alpha + 1} + (1-\psi)^2 \right\}. \quad (10)$$

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

By allowing within-subject dependence in the allocation to predictor and response clusters, the conditional distribution of the response given predictors is allowed to vary flexibly.

As noted in Section 1, typical latent class joint models assume a parametric form (e.g., Gaussian) for the conditional distribution of the response given the predictor cluster. In contrast, our proposed specification characterizes the conditional response distribution within a predictor cluster as a (potentially infinite) mixture of parametric models, with the mixture distribution changing flexibly across the predictor clusters.

2.3 Chinese restaurants for families

A Chinese restaurant metaphor is often used as an aid in understanding the DP prediction rule and alternative prediction rules arising from a species sampling sequence. 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 either assigned to an occupied table or to an empty table. For the i th customer, the probability of assignment to an empty table is $\alpha/(\alpha + i - 1)$, while the probability of assignment to the j th 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 $DP(\alpha P_0)$ prediction rule in (3), with the individuals at table j sharing a dish θ_j^* sampled from P_0 .

This metaphor can be generalized to describe the process of Section 2.2 in the limiting case as $K, L \rightarrow \infty$. Suppose 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 β_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, either from 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, while the i th group of siblings to sample from a bag is given a

new game with probability $\lambda/(\lambda + i - 1)$ and otherwise joins 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

3.1 *Gibbs sampling*

For Dirichlet process mixture models, posterior computation typically relies on Gibbs sampling, either marginalizing out the random measure using the Pólya urn scheme (Escobar, 1994; Bush and MacEachern, 1996) or approximating the measure by truncation (Ishwaran and James, 2001). The efficient Pólya urn Gibbs sampler proposed by Bush and MacEachern (1996) can be generalized by relying on the bivariate prediction rule in (6)-(7). However, we encountered two problems in implementing Gibbs sampling algorithms in our motivating application. The first is that Gibbs sampling is extremely computationally intensive, with computation increasing substantially as the number of subjects and observations per subject increases. The second problem is label switching in which the meaning of the clusters varies across the samples. We first addressed this problem using the Stephens (2000) post-processing approach. However, this added considerably to the already substantial computational burden. In order for the method to be practical to implement routinely, dramatically faster alternatives are needed.

3.2 *Combined sequential updating and Gibbs sampling*

For DPMs a variety of alternatives to MCMC have been proposed, including iid sequential importance sampling (SIS) (MacEachern, Clyde and Liu, 1999; Quintana and Newton,

2000; Ishwaran and James, 2003), partial predictive recursion (PPR) (Newton and Zhang, 1999; Tao et al., 1999), variational Bayes (VB) (Blei and Jordan, 2006) and sequential updating with greedy search (SUGS) (Wang and Dunson, 2007). The SIS approaches are still computationally expensive, so will not be considered further. Both PPR and VB rely on approximations to the posterior having unknown accuracy, and there is evidence that VB results in underestimation of posterior uncertainty for mixture models (Wang and Titterton, 2005). In addition, we have observed VB results for DP mixture models to be highly sensitive to starting values due to the use of an EM algorithm in approximating the multimodal posterior.

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 only requiring a single cycle of computation for each subject. For conjugate priors, the exact joint posterior distribution under the selected partition model is available in closed form. An unappealing property of the SUGS algorithm, which is shared by most stepwise selection procedures and by PPR, is order dependence. However, order dependence can be removed by repeating the procedure for random reorderings and choosing the partition resulting in the highest predictive likelihood (Wang and Dunson, 2007).

We propose a combined SUGS and Gibbs sampling algorithm, with sequential updating used to choose the partition of subjects into functional predictor clusters and Gibbs sampling used 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 $1, \dots, i$ and on predictor cluster allocation $\boldsymbol{\gamma}^{(i)} = (\gamma_1, \dots, \gamma_{i-1})'$, is

$$(\boldsymbol{\beta}_h^* | \mathbf{X}^{(i)}, \boldsymbol{\gamma}^{(i)}) \propto P_{01}(\boldsymbol{\beta}_h^*) \prod_{j=1}^i \left\{ \prod_{l=1}^{n_i} N(x_{jl}; \mathbf{b}'_{jl} \boldsymbol{\beta}_h^*, \tau_x^{-1}) \right\}^{1_{(\gamma_j=h)}}, \quad (11)$$

which is simply P_{01} updated with the likelihood for those subjects in predictor cluster h among subjects $\{1, \dots, i\}$. Using the updating equations in the Appendix, (11) is available in closed form for conjugate P_{01} .

Due to the well known label ambiguity problem (Stephens, 2000), the index values in γ 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 one could obtain an identical marginal likelihood 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. Hence, 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 $(\beta_1^* | \mathbf{x}_1, \gamma_1 = 1)$ using (11).
2. For $i = 2, \dots, n$,
 - (a) Choose $\gamma_i \in \{1, \dots, m_{i-1}^\gamma + 1\}$ to maximize the conditional probability of $\gamma_i = h$:

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

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

- (b) Update $(\beta_{\gamma_i}^* | \mathbf{X}^{(i-1)}, \boldsymbol{\gamma}^{(i-1)})$ using the data \mathbf{x}_i for subject i .

Note that the conditional probability of allocating subject i to predictor cluster h in step 2 (a) does not depend on unknowns in the response component model, so the allocation of subjects to predictor clusters can proceed via 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, γ , 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 steps:

1. Update the allocation of subjects to response clusters, ϕ , by sampling from the multinomial full conditional posterior distribution of ϕ_i given \mathbf{y} , \mathbf{X} , γ , ψ , ν^* , μ^* and τ_y , for $i = 1, \dots, n$.
2. Update ν_h^* , for $h = 0, 1, \dots, K$, by sampling from the conjugate Dirichlet conditional posterior distribution.
3. Update the cluster-specific means μ^* 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 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 ϕ of subjects into response clusters. This results in more realistic measures of uncertainty in estimates of the conditional response densities, without leading to any 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. Insufficient or excessive weight gain during critical stages of pregnancy may indicate or even contribute to problems of both the developing child and the mother. For normal weight women, the Institute of Medicine (IOM) recommends gaining 4-6 lbs in the first trimester and 1lb/week during the 2nd to 3rd trimester, 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 the pregnancy. The woman’s height was measured, she was asked her weight prior to pregnancy at the first prenatal visit or the time of recruitment, and weight was collected from each clinic visit during pregnancy. There were $n = 1,888$ women having 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 less than 3% of the subjects. The Jones and Nagin (2007) approach selected $K = 5$ as the number of trajectory clusters based on the BIC criteria. The value of $L = 4$ was motivated by the observation that a

mixture of 4 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 $N(\mu_i; \kappa\tau_y^{-1})G(\tau_y; a_y, b_y)$, with $G(a, b)$ the gamma distribution having 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), $\mu_0 = \bar{y}$ (global mean of y), $\kappa = 1$ (unit information), $\alpha = \lambda = 1$ (a widely used default value for the DP precision), $a_x = 2$, $b_x = 2\text{var}(x_{ij} - \mathbf{b}'_{ij}\boldsymbol{\beta}_0)$, $a_y = 2$ and $b_y = 2\text{var}(y_i)$. In addition, we choose a uniform(0,1) hyperprior for ψ . Results were robust to local changes in the hyperparameter values.

4.2 Analysis and results

We applied the combined sequential updating and Gibbs sampling algorithm proposed in Section 3, taking 50,000 Gibbs samples. The 1,888 women were clustered into 5 weight gain trajectory clusters, with the following number in each cluster: (1, blue) 617 (32.7%), (2, black) 774 (41.0%), (3, green) 224 (11.9%), (4, magenta) 98 (5.2%), (5, cyan) 175 (9.3%), with the clusters listed in order of appearance.

Figure 1 shows the estimated weight gain trajectory clusters along with pointwise 95% credible intervals. These trajectories are quite reasonable, with most having a slower rate of gain in the first trimester and an approximately linear rate of gain during the second to third trimester. The blue cluster, which closely corresponds to the IOM recommendations, contains approximately one third of the women. Most of the remaining women are in the black cluster, which corresponds to a moderate increase in the rate of gain over the blue cluster. Twenty percent of women have an even higher rate of weight gain, falling in the cyan and green clusters, while 5% are in the magenta cluster corresponding to an unusually low rate of weight gain. Repeating the analysis for a random reordering of the women in the

data set, we obtained similar results. To illustrate the extent to which the real data mimic these 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. From this figure, the women in the magenta cluster had significantly lighter babies than women in the other clusters, which is consistent with the lower weight gain for these women. In addition, the women in the blue cluster had significantly lighter babies than women in the three higher weight gain clusters. Although the estimated birth weight distributions are unimodal, there was clear evidence of non-normality with heavy tails and a tendency towards 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 babies in the magenta cluster, the difference between the blue cluster and the higher weight gain clusters was more apparent for normal and high birth weight babies. Table 1 provides posterior probabilities for quantile-specific contrasts.

Applying the Jones and Nagin (2007) approach in SAS, we obtained similar trajectories to those shown in Figure 1, though the Jones and Nagin estimates were less smooth early and late in gestation due to edge effects. As in our analysis, the Jones and Nagin approach concluded that women in the two lowest weight gain trajectory clusters had babies with significantly lower birth weight adjusting for gestational age at delivery. However, by assuming normality of the birth weight distribution for women within a trajectory cluster, with only the means varying across trajectory clusters, the Jones and Nagin approach 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 babies.

A simple alternative is to use frequentist kernel smoothing to obtain estimated birth weight densities within each trajectory class produced by the Jones and Nagin approach. Such an analysis produces similar, but less smooth, estimates to those shown in Figure 3. Our proposed approach has the advantage of borrowing information across the latent classes in performing density estimation. In addition, our approach automatically produces measures of uncertainty, while also allowing calculation of posterior probabilities for quantile contrasts.

5. Discussion

This article proposes a method for semiparametric joint modeling of a functional predictor with a response variable. Motivated by label ambiguity problems and difficulties in performing efficient computation, we allocate subjects into predictor clusters using a fast sequential updating and greedy search procedure, with Gibbs sampling used 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 interpretation of the complex relationship between a high dimensional predictor and a response having an unknown distribution. However, we do not in general recommend interpreting such clusters as corresponding to biologically or clinically distinct groups of individuals, because the formation of clusters is necessarily dependent on parametric assumptions. In addition, there is typically substantial uncertainty involved in clustering, as the number of partitions is enormous and it is difficult to search efficiently over the space of all possible partitions. For this reason and because there are typically many competing partitions that are difficult to distinguish based on the data alone, it is important to avoid over-interpretation.

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 to rapidly partition the subjects into clusters based on their predictor data. Our proposed prediction rule and the associated 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 et al., 2005; Quintana 2006) to allow flexible joint modeling of data from different sources. Related problems have been considered in the machine learning literature, but from a different perspective (Barnard et al., 2003).

ACKNOWLEDGMENTS

The authors thank David Umbach and David Kessler for helpful comments on a draft of the paper.

Appendix A: Posterior Updating

A.1 Predictor component

Assuming $(\boldsymbol{\beta} | \tau_x) \sim N_p(\boldsymbol{\beta}_0, \Psi\tau_x^{-1})$ 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 obtain $(\boldsymbol{\beta} | \tau_x, \mathbf{x}_i) \sim N_p(\hat{\boldsymbol{\beta}}, \hat{\Psi}\tau_x^{-1})$ and $(\tau_x | \mathbf{x}_i) \sim G(\hat{a}_x, \hat{b}_x)$, where $\hat{\Psi} = (\Psi^{-1} + \sum_{j=1}^{n_i} \mathbf{b}_{ij}\mathbf{b}'_{ij})^{-1}$, $\hat{\boldsymbol{\beta}} = \hat{\Psi}(\Psi^{-1}\boldsymbol{\beta}_0 + \sum_{j=1}^{n_i} \mathbf{b}_{ij}x_{ij})$, $\hat{a}_x = a_x + n_i/2$ and

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

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

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

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

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

A.2 Response component

Assuming $(\mu | \tau_y) \sim N(\mu_0, \kappa\tau_y^{-1})$ and $\tau_y \sim G(a_y, b_y)$ and updating with the likelihood $N(y_i; \mu, \tau_y^{-1})$, we obtain $(\mu | \tau_y, y_i) \sim N(\hat{\mu}, \hat{\kappa}\tau_y^{-1})$ and $(\tau_y | y_i) \sim G(\hat{a}_y, \hat{b}_y)$, where $\hat{\kappa} = (\kappa^{-1} + 1)^{-1}$, $\hat{\mu} = \hat{\kappa}(\kappa^{-1}\mu_0 + y_i)$, $\hat{a}_y = a_y + 1/2$ and $\hat{b}_y = b_y + 1/2(\kappa^{-1}\mu_0^2 + y_i^2 - \hat{\kappa}^{-1}\hat{\mu}^2)$. 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 $\sigma^2 = 2b_y(1 + \kappa)/\nu$.

References

- Aldous, D.J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XII* (edited by P.L. Hennequin). Springer Lecture Notes in Mathematics, Vol. 1117.
- 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. *Annals of Statistics*, 20, 260-279.
- Bigelow, J.L. and Dunson, D.B. (2007). Posterior simulation across nonparametric models for functional clustering. *Journal of the Royal Statistical Society B*, revision submitted.
- Blackwell, D. and MacQueen, J. (1973). Ferguson distributions via Pólya urn schemes. *Annals of Statistics*, 1, 353-355.
- Blei, D. and Jordan, M. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1, 121-144.
- Bush, C.A. and MacEachern, S.N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika*, 83, 275-285.

- Escobar, M.D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, 89, 268-277.
- Escobar, M.D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 577-588.
- Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1, 209-230.
- Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *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.
- Ishwaran, H. and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96, 161-173.
- Ishwaran, H. and James, L.F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statistica Sinica*, 13, 1211-1235.
- Ishwaran, H. and Takahara, G. (2002). Independent and identically distributed Monte Carlo algorithms for semiparametric linear mixed models. *Journal of the American Statistical Association*, 97, 1154-1166.
- Ishwaran, H. and Zarepour, M. (2003). 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. *Annals of Statistics*, 12, 351-357.
- MacEachern, S.N., Clyde, M. and Liu, J.S. (1999). Sequential importance sampling for nonparametric Bayes models: The next generation. *Canadian Journal of Statistics*, 27, 251-267.
- Müller, P., Quintana, F. and Rosner, G. (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society B*, 66, 735-749.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modelling and related techniques for longitudinal data. In D. Kaplan (ed.), *Handbook of Quantitative Methodology for the Social Sciences*, 345-368, Newbury Park, CA: Sage Publications.
- Newton, M.A. and Zhang, Y. (1999). A recursive algorithm for nonparametric analysis with missing data. *Biometrika*, 86, 15-26.
- Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory* (edited by T.S. Ferguson, L.S. Shapley, and J.B. MacQueen), 245-267. IMS Lecture Notes-Monograph Series, vol 30.

- 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 B*, 65, 557-574.
- Quintana, F.A. and Newton, M.A. (2000). Computational aspects of nonparametric Bayesian analysis with applications to the modeling of multiple binary sequences. *Journal of Computational and Graphical Statistics*, 9, 711-737.
- Ray, S. and Mallick, B. (2006). Functional clustering by Bayesian wavelet methods. *Journal of the Royal Statistical Society B*, 68, 305-332.
- Savitz, D.A., Dole, N., Williams, J., Thorp, J.M., McDonald, T., Carter, A.C., Eucker, B. (1999). Determinants of participation in an epidemiological study of preterm delivery. *Paediatric and Perinatal Epidemiology*, 13, 114-125.
- Stephens, (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society B*, 62, 795-809.
- Tao, H., Palta, M., Yandell, B.S. and Newton, M.A. (1999). An estimation method for the semiparametric mixed effects model. *Biometrics*, 55, 102-110.
- Wang, L. and Dunson, D.B. (2007). Fast posterior computation in Dirichlet process mixture models. *in preparation*.
- Wang, B. and Titterton, D.M. (2005). Inadequacy of interval estimates corresponding to variational Bayesian approximations. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics* (eds R.G. Cowell and Z. Ghahramani), 373-380, Society for Artificial Intelligence and Statistics.

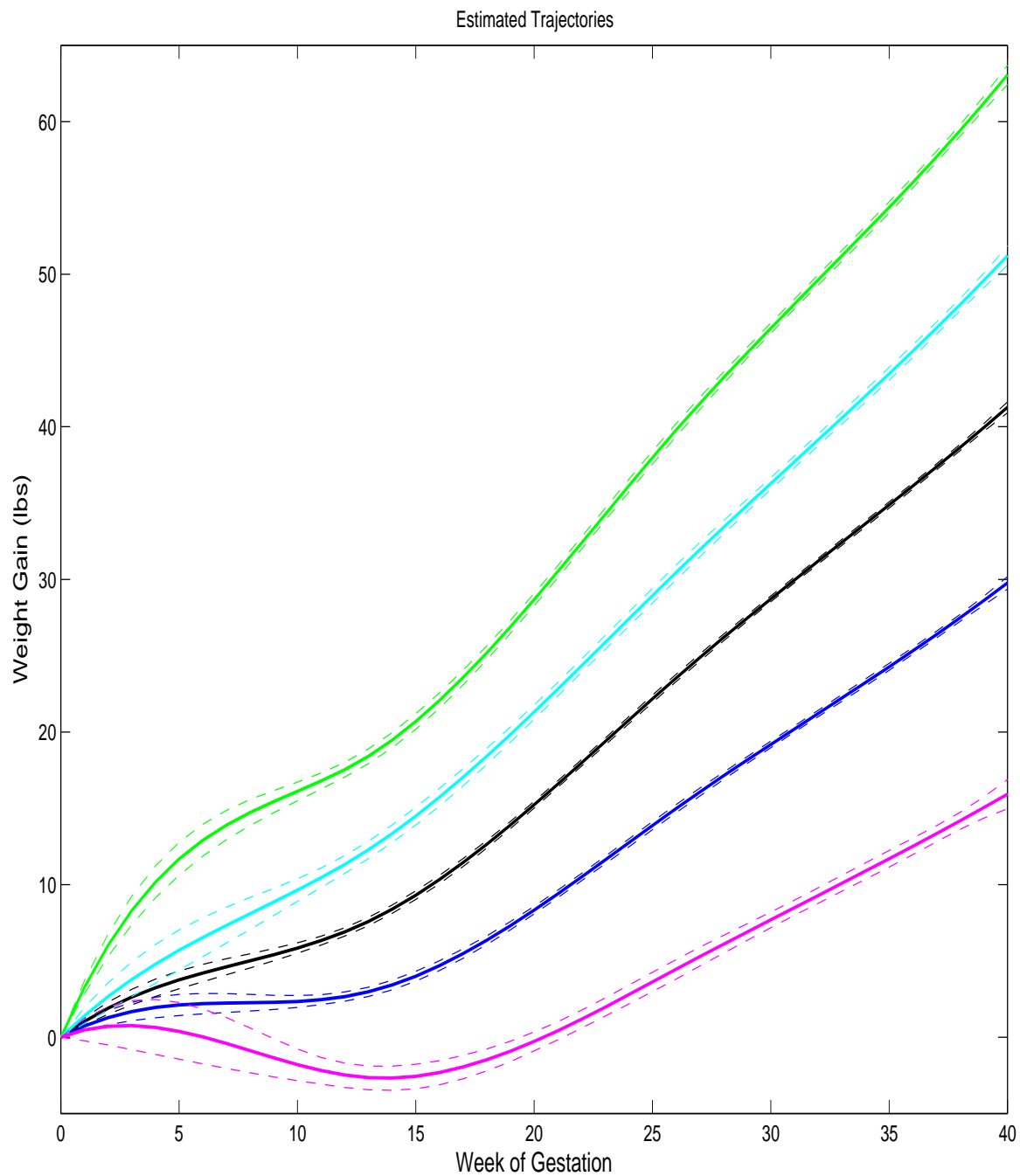


Figure 1. Estimated weight gain trajectory clusters (solid lines) and 95% credible intervals (dashed lines). The proportions of women in each cluster were: (1, blue) 32.7%, (2, black) 41.0%, (3, green) 11.9%, (4, magenta) 5.2%, and (5, cyan) 9.3%.

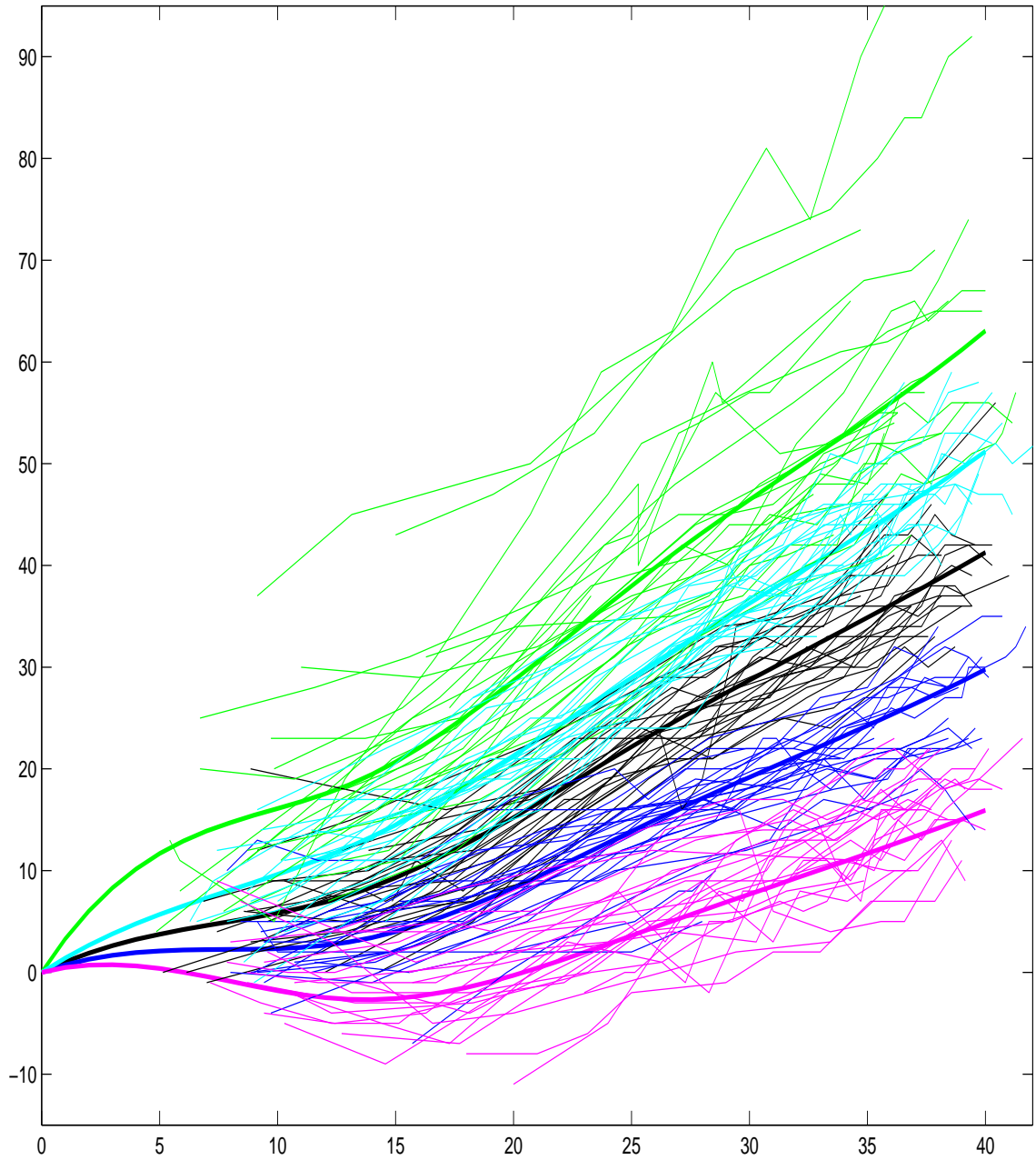


Figure 2. Raw data for 25 randomly selected women in each identified trajectory cluster.

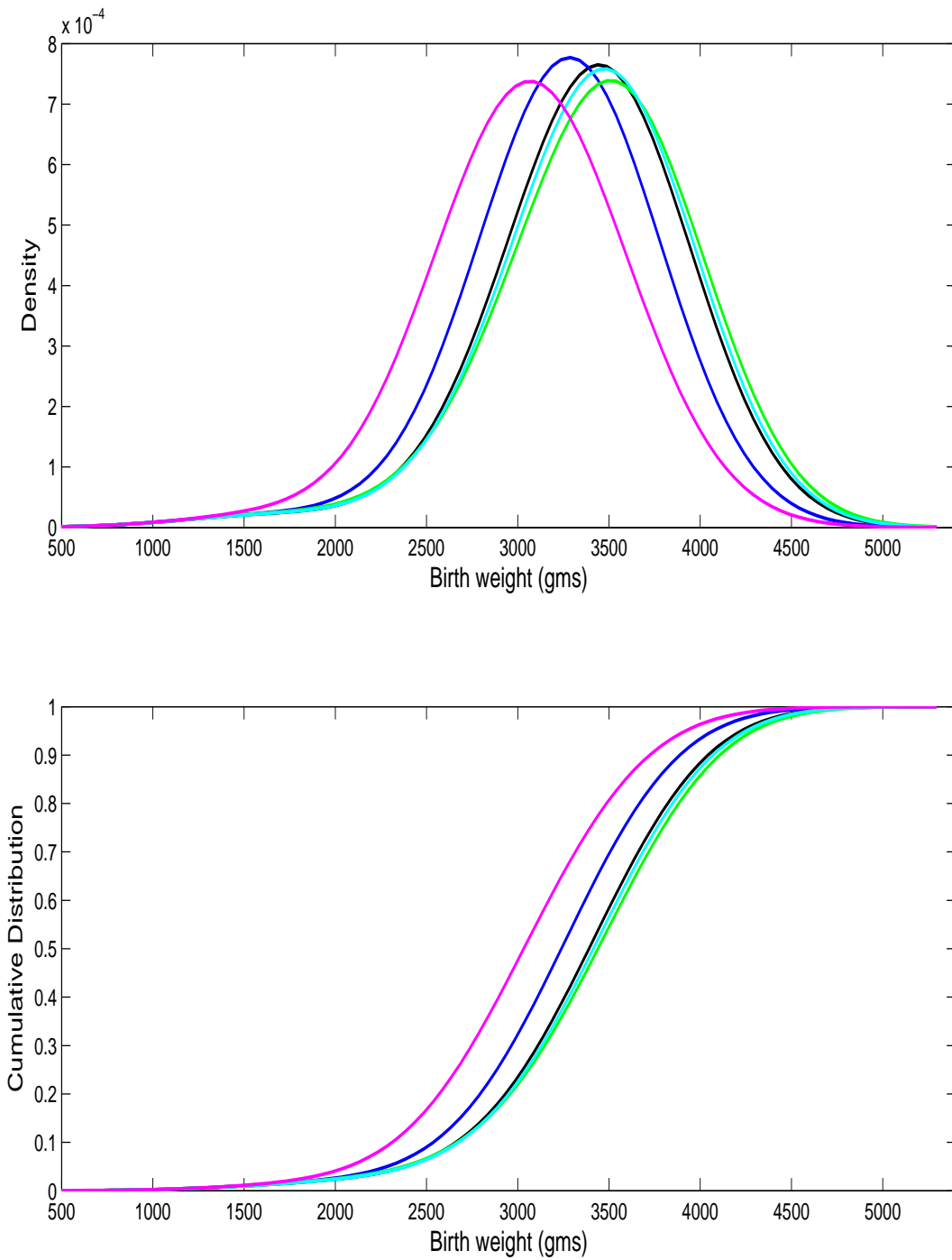


Figure 3. Estimated density and cumulative distribution function for birth weight at 40 weeks for women in each weight gain trajectory cluster.

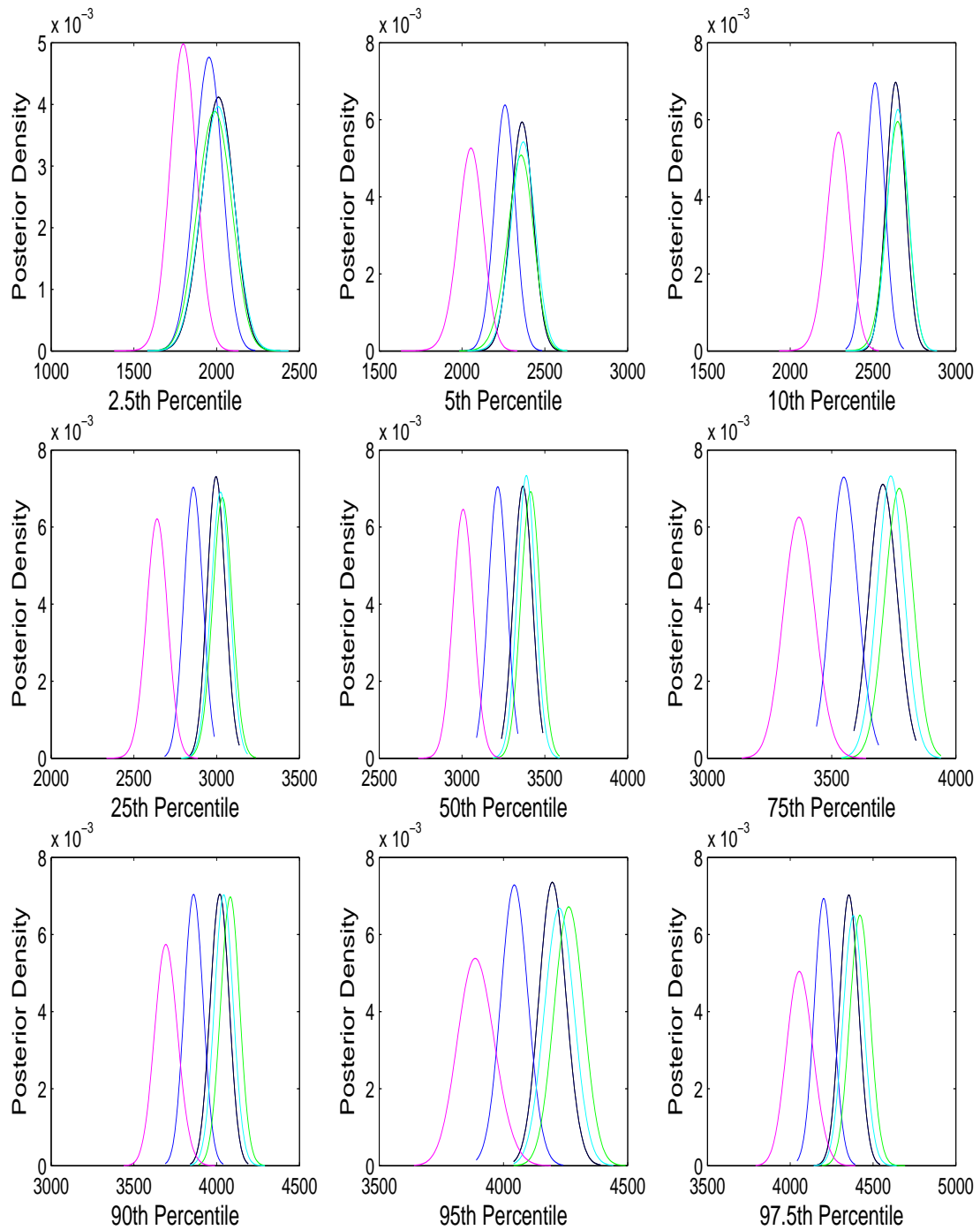


Figure 4. Estimated posterior densities of selected percentiles of the birth weight distribution within each weight gain trajectory cluster.

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

Contrast	Percentile of distribution								
	2.5th	5th	10th	25th	50th	75th	90th	95th	97.5th
magenta < blue	0.998	> 0.999	1.000	1.000	1.000	0.999	0.995	0.988	0.976
blue < black	0.815	0.992	> 0.999	1.000	1.000	1.000	1.000	1.000	> 0.999
black < cyan	0.185	0.309	0.370	0.505	0.444	0.579	0.497	0.518	0.501
black < green	0.127	0.271	0.394	0.648	0.732	0.859	0.875	0.901	0.875
cyan < green	0.141	0.231	0.292	0.365	0.507	0.607	0.650	0.644	0.650