

Bayesian isotonic density regression

Lianming Wang and David B. Dunson

*Biostatistics Branch, MD A3-03
National Institute of Environmental Health Sciences
U.S. National Institutes of Health
P.O. Box 12233, RTP, NC 27709
wangl3@niehs.nih.gov*

In estimating and performing inferences on conditional response distributions given predictors, stochastic ordering constraints can be incorporated to express prior knowledge and improve efficiency. This article proposes a nonparametric Bayes approach for modeling an uncountable collection of stochastically ordered distributions indexed by a continuous predictor. Theory is developed to allow priors to be chosen with large support through a dependent Dirichlet process (DDP) specification. Choosing monotone splines for functional atoms within the DDP representation, an efficient MCMC algorithm is developed for posterior computation. Methods are developed for hypothesis testing and graphical presentation of results. The approach is evaluated through a simulation study and is applied to an epidemiologic study.

Key Words: Conditional density estimation; Dependent Dirichlet process; Hypothesis test; Nonparametric Bayes; Quantile regression; Stochastic ordering.

1. Introduction

In studying the relationship between a continuous predictor x and a response variable y , it is common to have prior knowledge of stochastic ordering in the conditional distribution of y given x . For example, as age increases during adolescence, the distribution of height conditional on age increases stochastically. Also, in environmental toxicology and epidemiology studies, it is often reasonable to assume that the severity of an adverse health outcome is stochastically non-decreasing with dose of a potentially-adverse exposure. In such cases, incorporation of stochastic ordering constraints can improve efficiency in estimating conditional distributions and in testing for local or global changes as x increases. The focus of this article is on addressing these problems using nonparametric Bayesian methods.

In the Bayesian literature on incorporation of ordering constraints, most of the focus has been on parametric models. As noted by Gelfand, Smith and Lee (1992), constraints are typically straightforward to include in parametric models through defining a prior with appropriate support. To avoid parametric assumptions on the mean, a number of authors have focused on models of the form $y_i = f(x_i) + \epsilon_i$, where $f(\cdot)$ is an unknown non-decreasing function and ϵ_i 's are iid random errors. Lavine and Mockus (1995) allowed both $f(\cdot)$ and the residual distribution to be unknown through modeling a scaled version of $f(\cdot)$ as an unknown distribution function assigned a Dirichlet process (DP) prior (Ferguson, 1973; 1974). Neelon and Dunson (2004) proposed an alternative approach which characterized $f(\cdot)$ using linear splines having flat regions corresponding to no change. Holmes and Heard (2003) instead used a piecewise constant model for $f(\cdot)$ having unknown changepoints.

To our knowledge, previous Bayesian isotonic regression methods focus on flexible modeling of the mean or median, while restricting the residual density to be constant. In many applications, the assumption of a constant residual density is overly restrictive. For example, we are interested in studying changes in the distribution of gestational age at delivery with predictors, such as dose of DDE. As dose of an adverse exposure increases, we anticipate more premature deliveries and stochastic decreases in the distribution of gestational age at delivery. However, we cannot characterize such effects in a biologically plausible manner through a mean or median regression model with a constant residual distribution. Because the right tail of the distribution, which corresponds to the gestational age at full term deliveries, tends to be robust to exposures, we need to allow the different quantiles of the distribution of y to vary differentially with x .

Although this can potentially be accomplished through an isotonic quantile regression method, such as that proposed by Abrevaya (2005), our goal is not to characterize a single selected quantile, but instead to estimate and perform inferences on the entire conditional distribution. If x were a simple binary group indicator, this could be accomplished through the nonparametric Bayes methods proposed by Arjas and Gasbarra (1996) and Gelfand and Kottas (2000). The Gelfand and Kottas approach, which characterizes stochastically ordered distributions through a product of DP priors, can also be applied in multiple group cases. Hoff (2003b) considered the more general problem of Bayesian modeling of finite collections of distributions subject to a partial stochastic ordering. His approach relied on the theory of Hoff (2003a), which considered modeling of convex collections of probability measures. Dunson and Peddada (2007) proposed a restricted dependent Dirichlet process

prior (rDDP) for the same problem, developing efficient methods for posterior computation and hypothesis testing.

We extend these methods to the case in which x is continuous, considering priors π_P for $P = (P_x, x \in \mathcal{X})$, an uncountable collection of stochastically ordered probability measures. Letting P_x denote the measure specific to predictor value x , we focus on the constraint $P_x \preceq P_{x'}$, for all any two points x and x' in \mathcal{X} . The notation $P_x \preceq P_{x'}$ denotes that $P_x(a, \infty) \leq P_{x'}(a, \infty)$ for any $a \in \mathfrak{R}$. Our first goal will be to choose a prior π_P with full support on the space of stochastically ordered collections P . We provide theory on the existence of such a prior and propose a formulation based on a dependent Dirichlet process (DDP) (MacEachern, 1999; 2001). This formulation facilitates the development of methodology for local and global hypothesis testing and results in straightforward posterior computation. Using our proposed prior for a collection of predictor-dependent mixture distributions, we develop methods for estimation and inferences on stochastically ordered conditional densities.

There has been a great deal of recent interest in developing methods for dependent collections of probability measures. Much of this work relies on generalizations of the Sethuraman (1994) characterization of the Dirichlet process (DP) to incorporate dependence through the weights and/or the atoms. We follow a number of authors in building on the DDP proposed by MacEachern (1999, 2001). For example, De Iorio et al. (2004) used the DDP to create ANOVA-type dependence in random measures, while Gelfand et al. (2005) consider a spatial application. Griffin and Steel (2006) developed order-based DDPs (π DDPs), in which both the weights and atoms are associated with a stochastic process. Alternative formulations based on predictor-dependent mixtures of Dirichlet processes have been proposed by Dunson et al. (2007) and Dunson and Park (2006). These methods do not allow stochastic ordering constraints.

Section 2 provides a representation theorem and proposes a Dirichlet process formulation. Section 3 discusses Bayesian inference and develops methods for hypothesis testing. Section 4 describes a simple and efficient Gibbs sampler for posterior computation. Section 5 and 6 evaluate and illustrate the proposed method with a simulation study and an epidemiologic application. Section 7 concludes with some remarks. Proofs and some details of posterior computation are included in Appendixes.

2. Stochastically ordered probability measures

2.1 Theoretical results

In this section, we establish some theoretical results about an uncountable collection of stochastically ordered probability measures indexed by a continuous predictor. We will develop a generalization of Choquet's theorem (Hoff, 2003a), which states that each point in a convex set C of probability measures can be written as a mixture over the extreme points of C , that is

$$P = T(Q) = \int_{exC} P^* dQ(P^*). \quad (2.1)$$

The mixing measure Q is a probability measure on exC , the set of the extreme points of C . The map T in (2.1) is called the Barycenter map. Since Q is not constrained, inference on P can be simplified via estimation of the unconstrained measure Q through $P = T(Q)$.

In Choquet's theorem, an element P of C is a single probability measure, while in this paper we consider that P is a collection of measures subject to stochastic ordering. We can not use Choquet's theorem directly and must extend (2.1) for a convex set with elements being collections of measures. Hoff (2003a) developed an extension of Choquet's theorem for a finite collection of measures subject to a partial stochastic ordering. We further extend Hoff (2003a) to the case in which P is an uncountable collection of measures subject to simple stochastic ordering.

Suppose that \mathcal{U} and \mathcal{X} are two subsets of the real line \mathfrak{R} and thus are separable metric spaces. Let $\mathcal{B}(\mathcal{U})$ be the Borel subsets of \mathcal{U} and \mathcal{P} be the set of probability measures on $\mathcal{B}(\mathcal{U})$. \mathcal{U} and \mathcal{X} are used as the sample spaces of response variable u and continuous predictor x , respectively. We focus on nondecreasing stochastic ordering on $(P_x, x \in \mathcal{X})$ in this part, which means $P_x \preceq P_{x'}$ for any two points $x, x' \in \mathcal{X}$ with $x \leq x'$. The other case of stochastic ordering, in which $P_x \preceq P_{x'}$ for $x \geq x'$, can be dealt with similarly.

Define

$$\mathcal{P}_{\mathcal{X}} = \{(P_x, x \in \mathcal{X}) : P_x \in \mathcal{P}, \forall x \in \mathcal{X}\}$$

and

$$C_{\mathcal{X}} = \{(P_x, x \in \mathcal{X}) \in \mathcal{P}_{\mathcal{X}} : P_x \preceq P_{x'}, \text{ for any } x \leq x', x, x' \in \mathcal{X}\}, \quad (2.2)$$

where $\mathcal{P}_{\mathcal{X}}$ is the set of uncountable collections of probability measures on $\mathcal{B}(\mathcal{U})$ indexed by $x \in \mathcal{X}$, while $C_{\mathcal{X}}$ is the subset of $\mathcal{P}_{\mathcal{X}}$ corresponding to collections satisfying nondecreasing stochastic ordering.

Lemma 1. $C_{\mathcal{X}}$ is a closed convex set.

The closure of a convex set ensures the existence of extreme points. Extreme points are defined as those points in the convex set that cannot be written as a convex combination of any two other points in the convex set. Extreme points are important since any point in the convex set can be represented by a mixture of extreme points as in (2.1). Let $s \in \mathcal{R}_{\mathcal{X}}$ be a function with $\mathcal{R}_{\mathcal{X}}$ a function space and $S_{\mathcal{X}} \subset \mathcal{R}_{\mathcal{X}}$, the subset corresponding to non-decreasing functions having $s(x) \leq s(x')$ for all $x < x'$.

Lemma 2. The set of extreme points $exC_{\mathcal{X}}$ of $C_{\mathcal{X}}$ is

$$exC_{\mathcal{X}} = \{(\delta_{s(x)}, x \in \mathcal{X}) : s \in S_{\mathcal{X}}\}.$$

From Lemma 2, any extreme point of $C_{\mathcal{X}}$ can be indexed by a nondecreasing function s , and any nondecreasing function corresponds to an extreme point of $C_{\mathcal{X}}$. This one to one relationship suggests that studies of $exC_{\mathcal{X}}$ can be conducted through studying the more familiar space $S_{\mathcal{X}}$.

Let $T(Q) = \int_{S_{\mathcal{X}}} (\delta_{s(x)}, x \in \mathcal{X}) dQ(s)$, where Q is a probability measure on $(S_{\mathcal{X}}, \sigma_{S_{\mathcal{X}}})$ and $\sigma_{S_{\mathcal{X}}}$ is a σ algebra on $S_{\mathcal{X}}$ satisfying some conditions shown in the Appendix A.

Lemma 3. $T(Q) \in C_{\mathcal{X}}$ for any probability measure Q on $(S_{\mathcal{X}}, \sigma_{S_{\mathcal{X}}})$.

We establish the following generalization of Choquet's theorem,

Theorem 1. For any $P \in C_{\mathcal{X}}$, there exists a mixing measure Q such that

$$P = T(Q) = \int_{S_{\mathcal{X}}} (\delta_{s(x)}, x \in \mathcal{X}) dQ(s), \quad (2.3)$$

where Q is a probability measure on $(S_{\mathcal{X}}, \sigma_{S_{\mathcal{X}}})$.

Formula (2.3) is similar to formula (2.1) but differs in the integration space. A reparameterization is applied in (2.3) based on the fact that any extreme point can be indexed by a $s \in S_{\mathcal{X}}$. The above results yield in the following Proposition directly,

Proposition. A collection of measures $P = (P_x, x \in \mathcal{X})$ is in $C_{\mathcal{X}}$ defined in (2.2) if and only if there exists a probability measure Q on $(S_{\mathcal{X}}, \sigma_{S_{\mathcal{X}}})$ such that $Q(s(x) \in \cdot) = P_x(\cdot)$ for all $x \in \mathcal{X}$.

The integral representation (2.3) in Theorem 1 provides a feasible way to specify a prior for an uncountable collection of measures P subject to a stochastic ordering constraint. Letting π_Q be a prior for the mixing measure Q , a prior π_P for $P \in C_{\mathcal{X}}$ can be induced through $P = T(Q)$. Thus, the problem of finding a prior for constrained $P \in C_{\mathcal{X}}$ is simplified to finding a prior for Q on $(S_{\mathcal{X}}, \sigma_{S_{\mathcal{X}}})$.

Lemma 4. The Barycenter map T defined above is continuous.

The continuity of T is important in ensuring that the induced prior π_P has full support on $C_{\mathcal{X}}$. In other words, π_P has a positive probability on any open subset of $C_{\mathcal{X}}$. The argument is as follows: letting A be an open subset of $C_{\mathcal{X}}$, $\pi_P(A) = \Pr(T(Q) \in A) = \pi_Q(T^{-1}A) > 0$ since $T^{-1}A$ is an open set due to the continuity of T .

2.2 Prior construction

We now construct a prior for $P \in C_{\mathcal{X}}$ based on the theory established in Section 2.1. Suppose that Q is a probability measure that is assigned a $DP(\alpha H)$ prior, where α is a precision parameter and H is a base probability measure on $(S_{\mathcal{X}}, \sigma_{S_{\mathcal{X}}})$. By the stick-breaking representation of the DP (Sethuraman, 1994), we have

$$Q = \sum_{h=1}^{\infty} p_h \delta_{\Theta_h}, \quad p_h = V_h \prod_{l < h} (1 - V_l), \quad (2.4)$$

where $V_h \stackrel{i.i.d.}{\sim} \text{Beta}(1, \alpha)$, independently from $\Theta_h \stackrel{i.i.d.}{\sim} H$ for $h = 1, \dots, \infty$.

Let Q_x be the marginal of Q at $x \in \mathcal{X}$. Then for any Borel set $B \in \mathcal{B}(\mathcal{U})$, we have

$$Q_x(B) = Q(B') = \sum_{h=1}^{\infty} p_h \delta_{\Theta_h}(B') = \sum_{h=1}^{\infty} p_h 1_{\{\Theta_h \in B'\}} = \sum_{h=1}^{\infty} p_h 1_{\{\Theta_h(x) \in B\}} = \sum_{h=1}^{\infty} p_h \delta_{\Theta_h(x)}(B),$$

where $B' = \{s \in S_{\mathcal{X}} : s(x) \in B\}$. This stick-breaking representation of Q_x suggests that $Q_x \sim DP(\alpha H_x)$ with H_x being the marginal of H at $x \in \mathcal{X}$. Then based on Theorem 1 and the proposition, we have

$$P_x = \sum_{h=1}^{\infty} p_h \delta_{\Theta_h(x)}, \quad x \in \mathcal{X}, \quad (2.5)$$

which suggests that the induced prior π_P for P has marginal at x being $DP(\alpha H_x)$.

The prior for P in (2.5) is a type of restricted dependent Dirichlet process (rDDP), which modifies the form proposed in Dunson and Peddada (2007) to allow continuous x . Since the Barycenter map is continuous, the rDDP prior in form (2.5) has full support on $C_{\mathcal{X}}$ in the sense of weak convergence defined in the Appendix A if H has full support on $S_{\mathcal{X}}$.

For any two points x and x' in \mathcal{X} with $x < x'$, denote $r_a = H(\Theta_1(x) \in B_a)$ and $q_a = H(\Theta_1(x') \in B_a)$, where $B_a = (a, +\infty)$ for any $a \in \mathfrak{R}$. Since $H(\Theta_1(x) \leq \Theta_1(x')) = 1$, we have $r_a \leq q_a$. It can be shown that under model (2.5)

$$\text{corr}(P_x(B_a), P_{x'}(B_a) | V) = \sqrt{\frac{r_a(1 - q_a)}{q_a(1 - r_a)}},$$

for any $a \in \mathfrak{R}$. This expression suggests several interesting things. First, since the right part does not contain $V = (V_h, h = 1, \dots, \infty)$, the correlation between $P_x(B_a)$ and $P_{x'}(B_a)$ does not depend on the random weights $(p_h, h = 1, \dots, \infty)$. Second, there is a nonnegative correlation between $P_x(B_a)$ and $P_{x'}(B_a)$ for any a and $x, x' \in \mathcal{X}$ since the right part in the above formula always falls inside $(0, 1)$. Third, when x and x' are very close together, the above correlation tends to be close to 1 since r_a is close to q_a . This result is intuitive and appealing in practical situations, which require that the effect of a chemical or drug does not change much if the dose is only slightly changed.

It is worthy to note that a similar integral representation in (2.3) holds on the set of uncountable collections of measures $\mathcal{P}_{\mathcal{X}}$, i.e,

$$P = T(Q) = \int_{\mathcal{R}_{\mathcal{X}}} (\delta_{s(x)}, x \in \mathcal{X}) dQ(s), \quad (2.6)$$

where Q is a probability measure on $(\mathcal{R}_{\mathcal{X}}, \sigma_{\mathcal{R}_{\mathcal{X}}})$, and $\sigma_{\mathcal{R}_{\mathcal{X}}}$ is a σ algebra on $\mathcal{R}_{\mathcal{X}}$. Assigning a $DP(\alpha H)$ prior for Q with H being a probability measure on $(\mathcal{R}_{\mathcal{X}}, \sigma_{\mathcal{R}_{\mathcal{X}}})$, we obtain a DDP prior in the same form as in (2.5).

3. Bayesian inferences and hypothesis testing

3.1 Stochastically ordered conditional densities

In the previous section, we proposed a class of rDDP priors for uncountable collections of stochastically ordered probability measures. Our theoretical results show that the prior has weak support on the space of collections of stochastically ordered random measures. However, the rDDP prior assigns probability one to almost surely discrete distributions, so is not appropriate for modeling of stochastically ordered densities. Here we propose a class of rDDP mixture (rDDPM) models, which are appropriate for continuous response variables.

Using $(P_x, x \in \mathcal{X})$ as a collection of random mixing measures, we obtain the following mixture model for the conditional densities:

$$f(y|x) = \int N(y; \mu, \sigma^2) dP_x(\mu), \quad x \in \mathcal{X}, \quad (3.1)$$

motivated by the fact that any density can be modeled by a mixture of normals with an infinite number of components. There is a long history of using DP mixtures of normals in density estimation (Lo, 1984; Escobar and West, 1995).

Applying the result $P = T(Q)$ in (2.6) to model (3.1) and assigning a $DP(\alpha H)$ prior for

Q , we obtain

$$f(y|x) = \int_{\mathcal{R}_x} N(y; s(x), \sigma^2) dQ(s), \quad \forall x \in \mathcal{X}, \quad Q \sim DP(\alpha H), \quad (3.2)$$

where H is a probability measure on $(\mathcal{R}_x, \sigma_{\mathcal{R}_x})$.

Taking advantage of the stick-breaking representation of the DP as in (2.4), we obtain the following conditional densities

$$f(y|x) = \sum_{h=1}^{\infty} p_h N(y; \Theta_h(x), \sigma^2), \quad \forall x \in \mathcal{X}, \quad (3.3)$$

and the corresponding cumulative distribution functions (CDFs)

$$F_x(y) = \sum_{h=1}^{\infty} p_h \Phi(y; \Theta_h(x), \sigma^2), \quad \forall x \in \mathcal{X}, \quad (3.4)$$

where $\Phi(y; a, b)$ is the CDF of the normal distribution with mean a and variance b .

Let $f_{\mathcal{X}} = (f(\cdot|x), x \in \mathcal{X})$ be the collection of conditional densities indexed by $x \in \mathcal{X}$ in Model (3.2). The DP prior for Q in (3.2) induces a prior on $f_{\mathcal{X}}$. Denote by $D_{\mathcal{X}}$ the space of collections of densities indexed by $x \in \mathcal{X}$ and subject to the nondecreasing stochastic ordering constraint. Regarding such a stochastic ordering constraint, we have the following theorem,

Theorem 2. Under model (3.2), the following statements are equivalent:

- (a) $f_{\mathcal{X}}$ are non-decreasingly stochastically ordered.
- (b) H is a probability measure on $(S_{\mathcal{X}}, \sigma_{S_{\mathcal{X}}})$.

In the sequel, we assume that H is a probability measure on $S_{\mathcal{X}}$. Then based on Theorem 1 and 2, the induced prior for $f_{\mathcal{X}}$ has full support on $D_{\mathcal{X}}$. This is a generalization of Lo (1984), which established a related result for a single density.

3.2 Local and global hypothesis testing

Another primary interest of this paper is hypothesis testing. Since the distributions are stochastically ordered, we want to formalize a framework for testing of equalities in conditional distributions at different predictor values against stochastically ordered alternatives. Such hypothesis tests are not commonly studied. Recent work by Dunson and Peddada (2007) constructed a hypothesis test for finite distributions subject to the stochastic ordering constraint.

In order to test the equality between the distributions F_x and $F_{x'}$ with $x < x'$, we define $d_{TV}(x, x')$ with the following expression to measure the closeness between F_x and $F_{x'}$:

$$d_{TV}(x, x') = \sup_{y \in \mathfrak{R}} |F_x(y) - F_{x'}(y)|$$

and consider the following local hypotheses

$$H_0(x, x') : d_{TV}(x, x') < \epsilon \quad \text{vs.} \quad H_1(x, x') : d_{TV}(x, x') \geq \epsilon,$$

where ϵ is a small positive constant. Global hypothesis tests of near equality in the conditional distributions across a region $A = [a, b] \subset \mathcal{X}$ can be formulated as $H_0(A) = \bigcap_{x, x' \in A} H_0(x, x') = H_0(a, b)$.

The prior probability allocated to $H_0(x, x')$ depends on the prior distribution of $d_{TV}(x, x')$ and on ϵ . We will treat ϵ as a small, fixed constant that is chosen in advance so that values of the distance metric less than ϵ reflect negligible changes in the conditional distributions. We recommend $\epsilon = 0.05$ as a default value. Then, focusing on $d_{TV}(x, x')$, we have

$$d_{TV}(x, x') = \sup_{y \in \mathfrak{R}} \sum_{h=1}^{\infty} p_h \left(\Phi(y; \Theta_h(x), \sigma^2) - \Phi(y; \Theta_h(x'), \sigma^2) \right),$$

where a joint distribution on $\{\Theta(x), \Theta(x')\}$ is induced by H . Note that $\Pr\{H_0(x, x')\} \approx 0$ when x and x' are not close together unless H is chosen to assign non-negligible probability to functions that are flat or close to flat across wide regions of the predictor space. Hence, in applications in which it is uncertain *a priori* whether or not there is any relation between x and y , it is important to choose H carefully. Choice of H is discussed further in the sequel.

3.3 Choice of base measure

The base measure H plays an important role in determining the size of the support of the induced prior for the conditional densities. Model (3.2) induces a prior for the conditional densities with full support on $D_{\mathcal{X}}$ when the base measure H has a full support on the space of nondecreasing functions $S_{\mathcal{X}}$. Since piecewise constant functions are dense in $S_{\mathcal{X}}$, a probability measure H having support on the set of nondecreasing piecewise constant functions also induces a prior for the conditional densities that is dense in $D_{\mathcal{X}}$.

However, piecewise constant functions have discontinuities that induce discontinuities in the collections of conditional densities, which may not be consistent with prior knowledge. To improve efficiency and obtain estimates more consistent with prior knowledge, it is appealing

to consider priors with support on the space of conditional densities subject to the continuity condition: $f(\cdot|x) \rightarrow f(\cdot|x')$ as $x \rightarrow x'$, $\forall x, x' \in \mathcal{X}$.

Lemma 5. Under (3.2) $f_{\mathcal{X}}$ is a stochastically non-decreasing, continuous collection of densities with probability one if and only if H has support on the space of continuous nondecreasing $\mathcal{X} \rightarrow \mathfrak{R}$ functions.

In this situation, we need to choose an H that has support on a dense subspace of the space of continuous nondecreasing functions on \mathcal{X} . For example, we can choose the space of continuous nondecreasing piecewise linear functions or the space of nondecreasing splines.

4. Bayesian computation

Suppose that there is only one predictor x and no other covariates are available at this stage. The observed data are $\{(y_i, x_i), i = 1, \dots, n\}$, where n is the number of subjects in the study. Based on model (3.2), we consider the following model:

$$y_i \sim N(\mu_i(x_i), \sigma^2),$$

$$\mu_i(x) | P_x \stackrel{i.i.d.}{\sim} P_x, \quad P_x = \sum_{h=1}^{\infty} p_h \delta_{\Theta_h(x)}, \quad (4.1)$$

for any $x \in \mathcal{X}$. The p_h and Θ_h terms are defined the same as in model (3.2). In model (4.1), $(\mu_i(x), x \in \mathcal{X})$ can be interpreted as a latent function with $\mu_i(x_i)$ the value at x_i .

4.1. Monotone splines

In model (4.1), the Θ_h 's are iid nondecreasing random functions generated according to H . Hence, to complete a specification of the model it is necessary to choose H . In this section, we propose an approach based on monotone splines. In particular, we assume that the functions have the following basis expansion:

$$\Theta(x) = \sum_{l=0}^k \beta_l b_l(x), \quad \forall x \in \mathcal{X}, \quad (4.2)$$

where $b_0(x) = 1$, $\{b_l\}_{l=1}^k$ is the I-spline basis proposed by Ramsay (1988) for smooth monotone function estimation, and $\{\beta_l\}_{l=0}^k$ are basis coefficients, with $\beta_l \geq 0$, for $l = 1, \dots, k$. As described by Ramsay (1988), each I-spline basis function starts at 0 in an initial flat region, is increasing in a mid region, and then plateaus at 1 at higher values. The placement of the knots determines the locations of these regions, while the degree of the spline controls smoothness.

The specification in (4.2) is an appealing choice, because I-splines can be used to approximate any smooth, non-decreasing function, while also allowing flat regions as corresponding to intervals for which consecutive basis coefficients β_l 's are 0. It is important to allow such flat regions to avoid forcing strict stochastic increases in the conditional densities, which can lead to extremely small probabilities assigned to local null hypotheses and to bias when changes with x are modest. In addition, we can allow for uncertainty in knot selection by starting with many potential knots and then letting knots drop out adaptively by letting their coefficients have zero values (Smith and Kohn, 1996). Using expansion (4.2), the base probability measure H is induced through a prior for the basis coefficients $\{\beta_l\}_{l=0}^k$. One can easily modify the specification to model non-increasing functions by replacing $b_l(x)$ with $1 - b_l(x)$, for $l = 1, \dots, k$.

4.2. Truncation

In performing computation, it is not possible to estimate the infinitely many parameters in (4.1). Hence, we follow Ishwaran and James (2001) and others in replacing the infinite stick-breaking specification with a finite approximation. In particular, in place of P_x , use

$$P_x^N = \sum_{h=1}^N p_h \delta_{\Theta_h(x)}, \quad \Theta_h \stackrel{i.i.d.}{\sim} H, \quad \forall x \in \mathcal{X}, \quad (4.3)$$

where the probabilities (p_1, \dots, p_{N-1}) are defined as in (2.4), and $p_N = \prod_{l=1}^{N-1} (1 - V_l)$ so that all the probability weights add up to 1, which results in $V_N = 1$.

As shown in Ishwaran and James (2001), this approximation is highly accurate for moderate N , particularly if n is not large and $\alpha \leq 1$. As small values of α favor a small number of clusters and a modest number of normal components can approximate any density accurately, it seems adequate to use $N = 20$ in most applications.

4.3. Model and prior specification

Based on (4.2) and (4.3), model (4.1) becomes

$$y_i \sim N(\mu_i(x_i), \sigma^2), \quad \mu_i(x) | P_x^N \stackrel{i.i.d.}{\sim} P_x^N, \\ P_x^N = \sum_{h=1}^N p_h \delta_{\Theta_h(x)}, \quad \Theta_h(x) = \sum_{l=0}^k \beta_{hl} b_l(x), \quad (4.4)$$

for any $x \in \mathcal{X}$, where $\beta_h = (\beta_{h0}, \dots, \beta_{hk})'$ is a vector of coefficients of the basis functions modeling Θ_h , $h = 1, \dots, N$. Denote $\beta = (\beta'_1, \dots, \beta'_N)'$. Model (4.4) is very flexible in capturing conditional densities with varying shapes for different predictor values.

By allowing coefficients β_{hl} 's to take value 0 with positive probability, we can obtain flat regions and allow uncertainty in basis function selection to obtain model-average inferences. Neelon and Dunson (2004) proposed a related idea in the setting of linear spline models for isotonic mean regression. We assign the following prior to $\boldsymbol{\beta}_h$ in model (4.4),

$$\pi(\boldsymbol{\beta}_h; m_0, \nu_0, \lambda, \pi_0) = N(\beta_{h0}; m_0, \nu_0^{-1}) \prod_{l=1}^k \{\pi_0 \delta_0(\beta_{hl}) + (1 - \pi_0) \text{Exp}(\lambda)\},$$

where $m_0 \sim N(\omega_0, \tau_0)$, $\nu_0 \sim \mathcal{G}(a_{\nu_0}, b_{\nu_0})$, $\pi_0 \sim \text{Beta}(a_\pi, b_\pi)$, $\lambda \sim \mathcal{G}(a_\lambda, b_\lambda)$, and $\text{Exp}(\lambda)$ denotes the Exponential distribution with mean λ^{-1} . Here the parameter π_0 controls the prior probability that a basis function is excluded from the model. We assign conjugate priors to the error precision parameter σ^{-2} with $\pi(\sigma^{-2}) = \mathcal{G}(\sigma^{-2}; a_\sigma, b_\sigma)$ and to the DP precision parameter α with $\pi(\alpha) = \mathcal{G}(\alpha; a_\alpha, b_\alpha)$. We let $K_i = h$ if $\mu_i = \Theta_h$ denote that subject i is allocated to functional cluster h , with $\mathbf{K} = (K_1, \dots, K_n)'$.

Before describing the steps in posterior computation, we generalize the model to allow covariates. Let $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})'$ be the observed covariates besides predictor x_i for subject i and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)'$ be the corresponding regression coefficients. Model (4.4) can be generalized with the mean response for subject i , $\mu_i(x_i)$, replaced with $\mu_i(x_i) + \mathbf{z}_i' \boldsymbol{\gamma}$ while everything else remains unchanged. Although other choices can be considered, we choose the unit information prior for $\boldsymbol{\gamma}$ conditional on σ^2 ,

$$\pi(\boldsymbol{\gamma} | \sigma^2) = N(0, n\sigma^2(\mathbf{z}'\mathbf{z})^{-1}), \quad (4.5)$$

where $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)'$ is an $n \times q$ design matrix.

4.4. Posterior computation

With the prior specification in Section 4.3, the joint posterior density of $(\sigma^{-2}, \pi_0, \lambda, m_0, \nu_0)$ is proportional to

$$L(\mathbf{y} | \mathbf{x}, \mathbf{z}, \sigma^2, \mathbf{K}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \pi(\boldsymbol{\beta} | m_0, \nu_0, \lambda, \pi_0) \pi(\boldsymbol{\gamma} | \sigma^2), \pi(\sigma^{-2}) \pi(\pi_0) \pi(\lambda) \pi(m_0) \pi(\nu_0),$$

where

$$L(\mathbf{y} | \mathbf{x}, \mathbf{z}, \sigma^2, \mathbf{K}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \propto (\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left\{ y_i - \sum_{l=0}^k \beta_{K_i l} b_l(x_i) - \mathbf{z}_i' \boldsymbol{\gamma} \right\}^2 \right].$$

Due to the selected priors, all the parameters have closed forms in their full conditionals, which can be sampled from easily. The Ishwaran and James (2001) blocked Gibbs sampler can be implemented through repeating the following steps:

(a) Sample $\boldsymbol{\beta}_h = (\beta_{h0}, \beta_{h1}, \dots, \beta_{hk})'$, $h = 1, \dots, N$ from the conditional posterior given the other parameters and cluster allocation \mathbf{K} . The details are provided in Appendix B.

(b) Sample $K_i, i = 1, \dots, n$, from the multinomial conditional posterior with

$$P(K_i = j | \cdot) = \frac{p_j N(y_i; \sum_{l=0}^k \beta_{jl} b_l(x_i) + \mathbf{z}'_i \boldsymbol{\gamma}, \sigma^2)}{\sum_{h=1}^N p_h N(y_i; \sum_{l=0}^k \beta_{hl} b_l(x_i) + \mathbf{z}'_i \boldsymbol{\gamma}, \sigma^2)}, \quad j = 1, \dots, N, \quad i = 1, \dots, n.$$

(c) Sample the stick-breaking random variables $\mathbf{V} = (V_1, \dots, V_N)$ from their full conditionals. Set $V_N = 1$ and $V_h \sim \text{Beta}(V_h; 1 + m_h, \alpha + \sum_{j=h+1}^N m_j)$, for $h = 1, \dots, N - 1$, where $m_h = \sum_{i=1}^n 1_{(K_i=h)}$, $h = 1, \dots, N$.

(d) Sample $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)'$ as a block from the following full conditional,

$$[\boldsymbol{\gamma} | \cdot] \propto N\left(\boldsymbol{\gamma}; \left(\frac{n+1}{n} \mathbf{z}' \mathbf{z}\right)^{-1} \mathbf{z}'(\mathbf{y} - \boldsymbol{\mu}), \left(\frac{n+1}{n} \mathbf{z}' \mathbf{z}\right)^{-1} \sigma^2\right),$$

where $\mathbf{y} - \boldsymbol{\mu} = (y_1 - \sum_{l=0}^k \beta_{K_{1l}} b_l(x_1), \dots, y_n - \sum_{l=0}^k \beta_{K_{nl}} b_l(x_n))'$.

(e) Sample $\pi_0, \alpha, \sigma^{-2}, \lambda, \nu_0$, and m_0 from the following full conditional posteriors:

$$[\pi_0 | \cdot] \propto \text{Beta}\left(\pi_0; a_{\pi_0} + \sum_{h=1}^N \sum_{l=1}^k 1_{(\beta_{hl}=0)}, b_{\pi_0} + \sum_{h=1}^N \sum_{l=1}^k 1_{(\beta_{hl}>0)}\right),$$

$$[\alpha | \cdot] \propto \mathcal{G}\left(\alpha; a_{\alpha} + N - 1, b_{\alpha} - \sum_{h=1}^{N-1} \log(1 - V_h)\right),$$

$$[\sigma^{-2} | \cdot] \propto \mathcal{G}\left(\sigma^{-2}; a_{\sigma} + \frac{n}{2}, b_{\sigma} + \frac{1}{2} \sum_{i=1}^n [y_i - \sum_{l=0}^k \beta_{K_{il}} b_l(x_i) - \mathbf{z}'_i \boldsymbol{\gamma}]^2 + \frac{1}{2n} (\mathbf{z} \boldsymbol{\gamma})' (\mathbf{z}' \boldsymbol{\gamma})\right).$$

$$[\lambda | \cdot] \propto \mathcal{G}\left(\lambda; a_{\lambda} + \sum_{h=1}^N \sum_{l=1}^k 1_{(\beta_{hl}>0)}, b_{\lambda} + \sum_{h=1}^N \sum_{l=1}^k \beta_{hl}\right),$$

$$[\nu_0 | \cdot] \propto \mathcal{G}\left(\nu_0; a_{\nu_0} + \frac{N}{2}, b_{\nu_0} + \frac{1}{2} \sum_{h=1}^N (\beta_{h0} - m_0)^2\right),$$

and

$$[m_0 | \cdot] \propto N\left(m_0; (\tau_0^{-1} + N\nu_0)^{-1} (\tau_0^{-1} \omega_0 + \nu_0 \sum_{h=1}^N \beta_{h0}), (\tau_0^{-1} + N\nu_0)^{-1}\right).$$

This algorithm is simple to implement and is quite efficient in cases we have considered, with rapid convergence and good mixing.

4.5 Estimation

Under model (4.4), the conditional densities and CDFs are of the form

$$f(y|x) = \sum_{h=1}^N p_h N \left(y; \sum_{l=0}^k \beta_{hl} b_l(x), \sigma^2 \right), \quad \forall x \in \mathcal{X},$$

and

$$F_x(y) = \sum_{h=1}^N p_h \Phi \left(y; \sum_{l=0}^k \beta_{hl} b_l(x), \sigma^2 \right), \quad \forall x \in \mathcal{X},$$

respectively. We consider the following estimate of the conditional density $f(\cdot|x)$,

$$\hat{f}(y|x) = \frac{1}{T} \sum_{r=1}^T \sum_{h=1}^N p_h^{(r)} N \left(y; \sum_{l=0}^k \beta_{hl}^{(r)} b_l(x), \sigma_{(r)}^2 \right),$$

at predictor value x , where $p_h^{(r)}$, $\beta_{hl}^{(r)}$ and $\sigma_{(r)}^2$, $r = 1, \dots, T$, are the MCMC draws after burn-in. The mean regression function $E(y|x)$ is estimated by

$$\hat{M}(x) = \frac{1}{T} \sum_{r=1}^T \sum_{h=1}^N \sum_{l=0}^k p_h^{(r)} \beta_{hl}^{(r)} b_l(x).$$

Denote by $\hat{F}_{x,(r)}(y) = \sum_{h=1}^N p_h^{(r)} \Phi \left(y; \sum_{l=0}^k \beta_{hl}^{(r)} b_l(x), \sigma_{(r)}^2 \right)$ the estimate of conditional CDF at x in the r th iteration of MCMC after burn-in. Then the conditional CDF and the p -th percentile curve $\alpha_p(x)$ can be estimated by

$$\hat{F}_x(y) = \frac{1}{T} \sum_{r=1}^T \hat{F}_{x,(r)}(y) \quad \text{and} \quad \hat{\alpha}_p(x) = \frac{1}{T} \sum_{r=1}^T \hat{F}_{x,(r)}^{-1}(p),$$

respectively.

To estimate posterior probabilities of $H_0(x, x')$, we use $d_{TV}^N(x, x')$ as an approximation to $d_{TV}(x, x')$ based on the truncation approach. $d_{TV}^N(x, x')$ can be calculated based on a grid of y values at each MCMC iteration.

5. A simulation study

A simulation study was conducted to evaluate the performance of the proposed methodology. In the simulation, we focused on one predictor x following $\mathcal{U}(0, 1)$, the uniform distribution on $(0, 1)$. Conditioning on x , the response variable y was generated from the following mixture distribution,

$$f(y|x) = 0.6 N \left(y; 3(x - 0.3)^2 1_{(x>0.3)}, 0.09 \right) + 0.4 N \left(y; 5(x - 0.8)^{3/2} 1_{(x>0.8)}, 0.04 \right).$$

We generated 200 such data sets, each with sample size 500. For each data set, we ran the Gibbs sampler as described in Section 4.4 without covariates. The hyper-parameters considered in Section 4.3 were specified as follows: $a_\alpha = 1$ and $b_\alpha = k$ favoring small values of α , $a_\sigma = 0.1$ and $b_\sigma = 0.1$ giving a vague prior for σ^{-2} , $a_\lambda = b_\lambda = a_{\nu_0} = b_{\nu_0} = 0.5$ resulting in a $\mathcal{G}(\frac{1}{2}, \frac{1}{2})$ prior for both λ and ν_0 , $a_\pi = 1$ and $b_\pi = 1$ leading to an uniform prior for π_0 , and finally $\omega_0 = 1$ and $\tau_0 = 10$ specifying the prior mean and variance of m_0 .

Each chain was started with initial values sampled from the prior, and convergence was fast with excellent mixing. For each chain we summarized results of interest from the 2000 iterations after a 500 iteration burn-in and estimated the mean response curves, p -th percentile curves for different p 's, and conditional densities.

In specifying the I-spline basis, we chose degree 4 and interior knots were taken to be equally spaced in the predictor space $(0, 1)$. We tried different numbers of knots in the simulation (19, 49, 99), but did not observe any noticeable changes in results. The following results were obtained where 49 interior knots were used.

Figure 1 shows the estimated mean regression function. The top plot shows the true mean function and the estimated mean functions of all data sets, while the bottom plot shows the true mean function, the average of 200 mean function estimators, and the average of 95% credible intervals for the estimated mean functions from the 200 data sets. Figure 1 suggests that the estimated mean functions fit the true mean function very well.

Figure 2 shows the results of conditional density estimation at different predictor levels including 10th, 25th, 50th, 75th and 90th percentiles in the rows from the top to the bottom respectively. Each plot in the left column shows 200 estimated densities and the true density while the plot in the right column and the same row shows the true conditional density, the average of the estimated conditional densities and the average of 95% credible intervals of the estimated density functions from the 200 data sets. Clearly, the procedure does an excellent job in estimating the conditional densities, even though the sample size of $n = 500$ is not large and the densities change in shape considerably across the range of x .

Figure 3 shows the estimated and true p -th percentile curves of the conditional response distributions. In each column, p is taken to be 10, 25, 50, 75, 90, and 95 respectively from the top to bottom plots. Each plot in the left column shows the 200 estimated curves (dotted blue) and the true curve (solid red) while each plot in the right column shows the average of the 200 estimated curves along with the true curve.

In estimating the conditional densities of y given x , we have assumed that the conditional

distributions are nondecreasing in x . In addition to estimation, interest often focuses on testing of differences between F_x and F_0 , with F_0 denoting the distribution of y given a minimum of the predictor, such as $x = 0$. In order to solve this problem, we calculate the posterior probability $\Pr(H_1(0, x)|Data) = \Pr(d_{TV}^{(N)}(0, x) \geq \epsilon|Data)$ for each x and $\epsilon = 0.05$ for each of the 200 data sets. The left plot in Figure 4 shows the 200 curves of $\Pr(H_1(0, x)|Data)$ in dotted green lines, the average of the 200 curves in a solid red line and the pointwise empirical 95% confidence interval in solid blue lines. The right plot shows the estimated curves of rejection rate based on the 200 data sets for different values of cut point $r = 0.70, 0.80, 0.90,$ and 0.95 . A data set rejects $H_0 : d_{TV}^{(N)}(0, x) \leq \epsilon$ if $\Pr(H_1(0, x)|Data) > r$, and the estimated rejection rate is calculated as the proportion of the 200 data sets that reject H_0 . For example, if we use 0.90 as a cut point to reject H_0 and conclude H_1 , only 55% of the 200 data sets generated reject H_0 for $x = 0.5$, while 96% of the data sets reject H_0 for $x = 0.6$. The estimated rejection probability is close to 1 when x is larger than 0.6. These results agree with the true model, in which F_x has the same form as F_0 for $x < 0.3$, changes slightly as x increases from 0.3, and becomes dramatically differently for large x .

6. Epidemiologic application

In this section, we apply our method to an epidemiology data set studied by Longnecker et al. (2001). DDT is widely used due to its effectiveness against mosquitoes transmitting human diseases and as an agricultural insecticide. However, DDT also has clear adverse health effects, so is banned from agriculture use worldwide. Here, we focus on the relationship between gestational age at delivery (GAD) and the DDT metabolite DDE with 2313 children whose GAD's are less than 45 weeks.

The relationship between preterm delivery and DDE was studied by Longnecker et al. (2001) by dichotomizing GAD using 37 weeks as a cutoff for preterm delivery. They fitted a logistic regression model with categorized DDE and found a highly significant dose response trend. Dunson and Park (2007) studied the same data using kernel stick-breaking processes and found that there was an increasing risk of premature delivery (corresponding to decreasing GAD) at higher DDE levels. However, the Dunson and Park (2007) approach did not incorporate the stochastic ordering constraint, which is biologically motivated and can improve efficiency. We assume that the conditional densities of GAD given DDE levels are non-increasingly stochastically ordered. Let P_x and F_x denote the mixing measure and CDF of GAD, respectively, given DDE at $x \in \mathcal{X} = (0, 180)$. The above assumption means

$P_{x'} \preceq P_x$ for any $x < x'$ or equivalently $F_x(a) \leq F_{x'}(a)$ for any $a \in \mathfrak{R}$.

Three covariates are considered in the model including a binary indicator of black race and levels of serum total cholesterol and triglycerides. The latter two covariates are standardized before model fitting. Let y be the response variable denoting the GAD, x the predictor representing DDE level, and \mathbf{z} the covariate vector. We analyze the data using the following generalized version of model (4.4) based on the algorithm described in Section 4.4,

$$y_i | x_i, \mathbf{z}_i, \sigma^2 \sim N(\mu_i(x_i) + \mathbf{z}_i' \boldsymbol{\gamma}, \sigma^2), \quad \mu_i(x_i) \sim P_{x_i}$$

$$P_x = \sum_{h=1}^N p_h \delta_{\Theta_h(x)}, \quad \Theta_h(x) = \sum_{l=0}^k \beta_{hl} b_l(x), \quad \forall x \in \mathcal{X}.$$

I splines with 35 interior knots and order 4 are taken to serve as basis functions as in Section 5 resulting in $k = 39$, the number of basis functions. Conditionally on σ^2 , the joint prior of $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)'$ is taken in form of (4.5) with $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)'$, the $n \times 3$ design matrix, where $\mathbf{z}_i = (z_{i1}, z_{i2}, z_{i3})'$ for $i = 1, \dots, n$.

The hyper-parameters are specified as in Section 5 with the exception of ω_0 which we set equal to 250 as a rough guess as mean of β_{h0} . We base our results on 15,000 iterations of the MCMC algorithm of Section 4.4 collected after discarding a 5,000 iteration burn-in.

Table 1 shows posterior summaries of the covariate effects. Both race and serum triglyceride level have a significantly negative association with gestational age at delivery, suggesting that black women and women with high levels of serum triglycerides are more likely to deliver preterm than other women. In contrast, cholesterol does not have a significant effect on gestational age at delivery.

Figure 5 plots the data and the estimated mean function of GAD with respect to DDE values as well as the corresponding 95% credible interval curves. The posterior mean of GAD decreases as DDE increases. The 95% credible interval becomes wider at higher DDE level due to sparse data available there. For the same data, Dunson and Park (2007) observed the same trend using a kernel stick-breaking process but the estimated mean function decreased more than that obtained here and the credible intervals were wider. These estimates may differ due to the stochastic ordering constraint and/or to the covariate adjustment as Dunson and Park (2007) did not consider covariates in their method.

Figure 6 shows the estimated conditional densities of GAD for some fixed DDE levels. As in Dunson and Park (2007), these density plots show a robust right tail but an increasing left tail as DDE increases, which represents increasing risks of premature delivery at higher DDE

values. Also, the credible intervals are wider at very higher DDE values due to sparse data available there. Compared to Dunson and Park (2007), we obtain substantially narrower credible intervals. These narrower intervals likely reflect the improved efficiency due to the stochastic ordering constraint. We also note that our approach is considerably simpler to implement and more computationally efficient.

As proposed in Sections 3 and 4, we can test the equality of two conditional densities at any two different DDE levels against the stochastic ordering alternatives. Denote F_0 as the CDF of GAD at $x = 2.5$, the observed minimal value of DDE. Let F_x denote the CDF of GAD at DDE level x . We conduct the following hypothesis tests: $H_0(2.5, x) : d_{TV}^N(2.5, x) < 0.05$ vs. $H_1(2.5, x) : d_{TV}^N(2.5, x) > 0.05$ with $x = (12.56, 17.13, 24.68, 36.53, 53.75, 105.61, 178.1)$, corresponding to the 10th, 25th, 50th, 75th, 90th, 99th, and 100th percentiles of the observed DDE distribution. The estimated posterior probabilities of $H_0(2.5, x)$ for these tests are 0.9226, 0.9083, 0.8656, 0.7497, 0.1951, 0.0701, and 0.009, respectively. These results suggest that the weight of evidence of a DDE effect at values less than 53.75 is low, but at 105.6 the evidence increases substantially.

To better see the difference between distributions and how they change. We provide the following heat plot. The heat plot shown in Figure 7 displays estimates of $[F_x(F_0^{-1}(p)) - p]$ for $p \in (0, 1)$ and $x \in \mathcal{X} = [2.5, 180)$. In Figure 7, red color indicates large difference between F_x and F_0 , while blue color indicates little or no difference.

From the heat plot, blue color covers the region where DDE values are small suggesting that F_x is very close to F_0 for x with low DDE levels, and red color covers more and more area as DDE values get larger suggesting a larger difference between F_x and F_0 for x at high DDE levels. Interestingly, the differences between F_x and F_0 for moderate to large DDE dose are much more apparent at low percentiles corresponding to premature and early full term deliveries.

7. Remarks

This article has proposed a Bayesian nonparametric method for inference on stochastically ordered conditional distributions. To our knowledge, we are the first to address this problem for continuous predictors. A key contribution is an extension of Choquet's theorem to prove an integral representation for an uncountable collection of probability measures. Hoff (2003a) considered finite collections, and our work extends his theorem. The integral representation allows the construction of priors for collections of stochastically ordered probability measures

as mixtures over the extreme points in a convex set. By choosing a carefully constructed Dirichlet process prior over the mixture distribution, we obtain a prior with full support, while developing a method that can be implemented with Gibbs samplers for DP mixture models.

In addition to developing an approach for estimating conditional distributions subject to constraints, we have focused on developing methods for hypothesis testing, an under-studied area of Bayesian nonparametrics. Our proposed approach is quite useful in applications, allowing estimation of posterior probabilities of a change in distribution at any given predictor value relative to a reference level. The methods for estimation and hypothesis testing had excellent performance in a simulation study. Hypothesis testing is intimately related to variable selection and in future work it will be interesting to extend these methods to allow multiple candidate predictors. It will also be interesting to study theoretical properties, such as posterior consistency and rates of convergence.

Acknowledgement

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. The authors thank Gregg Dinse and Shyamal Peddada for their critical comments.

Appendix A: Proofs

We first define some notation. Consider the following metric in $\mathcal{R}_{\mathcal{X}}$, the $\mathcal{X} \rightarrow \mathfrak{R}$ function space on \mathcal{X} , $d(s_1, s_2) = \sup_{x \in \mathcal{X}} |s_1(x) - s_2(x)|$ for any $s_1 \in \mathcal{R}_{\mathcal{X}}$ and $s_2 \in \mathcal{R}_{\mathcal{X}}$. Let f be a real valued function on \mathcal{U} , and define $\int f dP = (\int f dP_x, x \in \mathcal{X}) \in \mathcal{R}_{\mathcal{X}}$ for any $P = (P_x, x \in \mathcal{X})$.

For any $P \in \mathcal{P}_{\mathcal{X}}$ and any probability measure H on $\mathcal{B}(\mathcal{X})$, the Borel sets of \mathcal{X} , define

$$G_H(P)(\cdot) = \int_{\mathcal{X}} P_x(\cdot) dH(x).$$

$G_H(P)$ is a probability measure on $\mathcal{B}(\mathcal{U})$, i.e., $G_H(P) \in \mathcal{P}$. Define weak convergence on $\mathcal{P}_{\mathcal{X}}$ as follows: a sequence $\{P^{(n)}\}_{n=1}^{\infty}$ with $P^{(n)} \in \mathcal{P}_{\mathcal{X}}$ for all n converges weakly to $P \in \mathcal{P}_{\mathcal{X}}$, denoted by $P_n \Rightarrow P$, if and only if

$$G_H(P^{(n)}) \Rightarrow G_H(P)$$

for any probability measure H on $\mathcal{B}(\mathcal{X})$.

The above weak convergence implies $P_x^{(n)} \Rightarrow P_x$ for all $x \in \mathcal{X}$, which can be easily seen by taking $H = \delta_x$, the point mass measure at x . Essentially, the weak convergence $P_n \Rightarrow P$ on $\mathcal{P}(\mathcal{X})$ is equivalent to $P_x^{(n)} \Rightarrow P_x$ uniformly for all $x \in \mathcal{X}$. Another equivalent definition of weak convergence is $P_n \Rightarrow P$ if and only if $d\left(\int f dP, \int f dP^{(n)}\right) \rightarrow 0$ for any continuous function f on \mathcal{U} .

Define a mapping $h : S_{\mathcal{X}} \rightarrow exC_{\mathcal{X}}$ with $h(s) = (\delta_{s(x)}, x \in \mathcal{X}) \in exC_{\mathcal{X}}$ for any $s \in S_{\mathcal{X}}$. Let $\sigma_{exC_{\mathcal{X}}}$ be a σ -algebra on $exC_{\mathcal{X}}$ and $\sigma_{S_{\mathcal{X}}}$ be the smallest σ -algebra containing $h^{-1}(\sigma_{exC_{\mathcal{X}}})$ and $\sigma(\{s \in S_{\mathcal{X}} : s(x) < r\}, \forall x \in \mathcal{X}, \forall r \in \mathfrak{R})$, where $h^{-1}(\sigma_{exC_{\mathcal{X}}}) = \sigma(\{h^{-1}(B) : B \in \sigma_{exC_{\mathcal{X}}}\})$. The condition $h^{-1}(\sigma_{exC_{\mathcal{X}}}) \subset \sigma_{S_{\mathcal{X}}}$ is necessary for integral transformation from $exC_{\mathcal{X}}$ to $S_{\mathcal{X}}$, and the condition $\sigma(\{s \in S_{\mathcal{X}} : s(x) < r\}, \forall x \in \mathcal{X}, \forall r \in \mathfrak{R}) \subset \sigma_{S_{\mathcal{X}}}$ ensures the measurability of sets in the form $\{s \in S_{\mathcal{X}} : s(x) < r\}$.

Proof of Lemma 1: The proof of $C_{\mathcal{X}}$ being a convex set is trivial. Now we show that $C_{\mathcal{X}}$ is weakly closed. Let $P = (P_x, x \in \mathcal{X}) \in \bar{C}_{\mathcal{X}}$ and $\{P^{(n)}\}_{n=1}^{\infty}$ be a sequence weakly converging to P , where $P^{(n)} = (P_x^{(n)}, x \in \mathcal{X}) \in C_{\mathcal{X}}$ for each n . We need to show $P \in C_{\mathcal{X}}$. Since the definition of weak convergence on $\mathcal{P}_{\mathcal{X}}$ requires $P \in \mathcal{P}_{\mathcal{X}}$, P_x is automatically a probability measure on $\mathcal{B}(\mathcal{X})$ for any $x \in \mathcal{X}$. We only need to show that $P_x \preceq P_{x'}$ for any two points $x, x' \in \mathcal{X}$ with $x \leq x'$.

For fixed x and x' with $x \leq x'$, the fact that the weak convergence on $\mathcal{P}_{\mathcal{X}}$ implies pointwise weak convergence directly gives $P_x^{(n)} \Rightarrow P_x$ and $P_{x'}^{(n)} \Rightarrow P_{x'}$. Also we have $P_x^{(n)} \preceq P_{x'}^{(n)}$ since $P^{(n)} \in C_{\mathcal{X}}$. Then $P_x \preceq P_{x'}$ follows by using the same argument in the proof for proposition 6 in Hoff (2003a).

Proof of Lemma 2: First, for any nondecreasing function s , $(\delta_{s(x)}, x \in \mathcal{X})$ is an extreme point of $C_{\mathcal{X}}$ since a point mass measure can not be written as a convex combination of any two other points. Now consider a collection $P = (P_x, x \in \mathcal{X}) \in C_{\mathcal{X}}$. Let $F_x(\cdot)$ be the CDF corresponding to measure P_x for any $x \in \mathcal{X}$. Following the idea of Hoff (2003a), define

$$F_x^+(a) = F_x(a) + \min(F_x(a), 1 - F_x(a)) / 2$$

and

$$F_x^-(a) = F_x(a) - \min(F_x(a), 1 - F_x(a)) / 2$$

for any $a \in \mathfrak{R}$. It is easy to check that F_x^+ and F_x^- satisfy the requirements of CDFs for any $x \in \mathcal{X}$. Letting P_x^+ and P_x^- be the probability measures corresponding to F_x^+ and F_x^- , it can be shown that $P^+ = (P_x^+, x \in \mathcal{X}) \in C_{\mathcal{X}}$ and $P^- = (P_x^-, x \in \mathcal{X}) \in C_{\mathcal{X}}$. From this result we obtain that $P \in exC_E$ is equivalent to $P^+ = P^-$, i.e., $F_x(a)$ is 0 or 1 for any $a \in \mathfrak{R}$ and

any $x \in \mathcal{X}$, which suggests that P_x is a point mass measure for any $x \in \mathcal{X}$. Based on this result, Lemma 2 follows when the stochastic ordering constraint is considered.

Proof of Lemma 3: Define $Q_x(\cdot) = \int_{S_{\mathcal{X}}} \delta_{s(x)}(\cdot) dQ(s)$. Then Q_x is a well defined probability measure on $\mathcal{B}(\mathcal{U})$, i.e., $Q_x \in \mathcal{P}$. We only need to show that $(Q_x, x \in \mathcal{X})$ are subject to nondecreasing stochastic ordering constraint. For any two points $x, x' \in \mathcal{X}$ with $x \leq x'$, we have $s(x) \leq s(x')$ for $s \in S_{\mathcal{X}}$. We then have $\{s \in S_{\mathcal{X}} : s(x) > a\} \subseteq \{s \in S_{\mathcal{X}} : s(x') > a\}$ for any $a \in \mathfrak{R}$. By the definition of Q_x and $Q_{x'}$, we have $Q_x(a, \infty) = \int_{S_{\mathcal{X}}} \delta_{s(x)}(a, \infty) dQ(s) = Q(\{s \in S_{\mathcal{X}} : s(x) > a\}) \leq Q(\{s \in S_{\mathcal{X}} : s(x') > a\}) = \int_{S_{\mathcal{X}}} \delta_{s(x')}(a, \infty) dQ(s) = Q_{x'}(a, \infty)$. Hence, we have $Q_x \preceq Q_{x'}$.

Proof of Theorem 1: Following the idea in Hoff (2003a), define $s_x(\omega) = F_x^{-1}(\omega)$ for any $x \in \mathcal{X}$ and $\omega \in [0, 1]$, where F_x is the CDF w.r.t. P_x and $F_x^{-1}(\omega) = \inf\{u : F_x(u) \geq \omega\}$. For any $x, x' \in \mathcal{X}$ with $x \leq x'$, $P_x \preceq P_{x'}$ gives us $\{u : F_{x'}(u) \geq \omega\} \subset \{u : F_x(u) \geq \omega\}$ and thus $s_x(\omega) \leq s_{x'}(\omega) \forall \omega$. Let Q be the canonical measure of $s = (s(x), x \in \mathcal{X})$ on $(S_{\mathcal{X}}, \sigma_{S_{\mathcal{X}}})$ induced by a uniform distribution on ω . Then Q represents P with $Q_x = P_x$ for any $x \in \mathcal{X}$.

Proof of Lemma 4: The continuity of T means that if $Q^{(n)} \Rightarrow Q$ with $P^{(n)} = T(Q^{(n)})$ and $P = T(Q)$, then $P^{(n)} \Rightarrow P$. To this end we need to show $G_H(P^{(n)}) \Rightarrow G_H(P)$ for any probability measure H on $\mathcal{B}(\mathcal{X})$, which is equivalent to

$$\int_{\mathcal{U}} f dG_H(P^{(n)}) \Rightarrow \int_{\mathcal{U}} f dG_H(P)$$

for any bounded uniformly continuous function f on \mathcal{U} and any probability measure H on $\mathcal{B}(\mathcal{X})$ by Parthasarathy (1967) since \mathcal{U} is a separable metric space. Let $U(\mathcal{U})$ denote the set of bounded uniformly continuous function f on \mathcal{U} in the following.

First, for any $f \in U(\mathcal{U})$ and any probability measure H on $\mathcal{B}(\mathcal{X})$, we have the following facts: $G_H(P)(\cdot) = \int_{\mathcal{X}} Q_x(\cdot) dH(x) = \int_{\mathcal{X}} \int_{S_{\mathcal{X}}} \delta_{s(x)}(\cdot) dQ(s) dH(x)$ and

$$\begin{aligned} \int_{\mathcal{U}} f dG_H(P) &= \int_{\mathcal{U}} f(u) \int_{\mathcal{X}} \int_{S_{\mathcal{X}}} dQ(s) dH(x) \delta_{s(x)}(du) \\ &= \int_{S_{\mathcal{X}}} \int_{\mathcal{X}} \int_{\mathcal{U}} f(u) \delta_{s(x)}(du) dH(x) dQ(s) \\ &= \int_{S_{\mathcal{X}}} \int_{\mathcal{X}} f(s(x)) dH(x) dQ(s). \end{aligned}$$

Define $g_{(H,f)}(s) = \int_{\mathcal{X}} f(s(x)) dH(x)$, a function on $S_{\mathcal{X}}$ given $H \in \mathcal{B}(\mathcal{X})$ and $f \in U(\mathcal{U})$. Then we have

$$\int_{\mathcal{U}} f dG_H(P) = \int_{S_{\mathcal{X}}} g_{(H,f)}(s) dQ(s).$$

Following the same argument, we have

$$\int_{\mathcal{U}} f dG_H(P^{(n)}) = \int_{S_{\mathcal{X}}} g_{(H,f)}(s) dQ^{(n)}(s).$$

Since $Q^{(n)} \Rightarrow Q$, we only need to show that $g_{(H,f)}$ is a bounded continuous function of s over $S_{\mathcal{X}}$ for any $f \in U(\mathcal{U})$ and any measure H on $\mathcal{B}(\mathcal{X})$. This is trivial based on the metric d on $\mathcal{R}_{\mathcal{X}}$.

Proof of Theorem 2: Under model (3.2), the CDF's are in the form of (3.4) with each Θ_h being a random function without any monotone constraint. Then for any two predictor values x and x' with $x \leq x'$,

$$\begin{aligned} F_x(y) - F_{x'}(y) &= \sum_{h=1}^{\infty} p_h \left\{ \Phi\left(y; \Theta_h(x), \sigma^2\right) - \Phi\left(y; \Theta_h(x'), \sigma^2\right) \right\} \\ &= \sum_{h=1}^{\infty} p_h \left\{ \Phi_0\left(\frac{y - \Theta_h(x)}{\sigma}\right) - \Phi_0\left(\frac{y - \Theta_h(x')}{\sigma}\right) \right\}, \end{aligned}$$

where $\Phi_0(\cdot)$ denotes the CDF of standard normal distribution.

(b) \Rightarrow (a) is obvious due to the monotonicity of Θ_h sampled from H having support on $S_{\mathcal{X}}$, the space of nondecreasing functions.

Suppose (a) holds, i.e., $F_x(y) - F_{x'}(y) \geq 0$ for any $x \leq x'$ and $y \in \mathfrak{R}$ based on the definition. Then

$$\sum_{h=1}^{\infty} p_h \left\{ \Phi_0\left(\frac{y - \Theta_h(x)}{\sigma}\right) - \Phi_0\left(\frac{y - \Theta_h(x')}{\sigma}\right) \right\} \geq 0$$

holds for any positive weights $\mathbf{p} = (p_h, h = 1, \dots, \infty)'$ subject to $\sum_{h=1}^{\infty} p_h = 1$, which requires $\Phi_0\left(\frac{y - \Theta_h(x)}{\sigma}\right) - \Phi_0\left(\frac{y - \Theta_h(x')}{\sigma}\right) \geq 0$ for all $h \geq 1$. To see this more clearly, consider $\mathbf{p} \in \{\mathbf{p} : \sum_{h=1}^{\infty} p_h = 1, \text{ and } p_l > 1 - \epsilon\}$ for some small ϵ and a fixed l . Due to the fact that Φ_0 as a CDF is bounded, the left part of the above inequality has an upper bound

$$p_l \left\{ \Phi_0\left(\frac{y - \Theta_l(x)}{\sigma}\right) - \Phi_0\left(\frac{y - \Theta_l(x')}{\sigma}\right) \right\} + \epsilon.$$

By taking ϵ arbitrarily small, we obtain $p_l \left\{ \Phi_0\left(\frac{y - \Theta_l(x)}{\sigma}\right) - \Phi_0\left(\frac{y - \Theta_l(x')}{\sigma}\right) \right\} \geq 0$ and further obtain $\Phi_0\left(\frac{y - \Theta_l(x)}{\sigma}\right) - \Phi_0\left(\frac{y - \Theta_l(x')}{\sigma}\right) \geq 0$ since p_l is positive. Since the same argument holds for any positive integer l , we have $\Phi_0\left(\frac{y - \Theta_l(x)}{\sigma}\right) - \Phi_0\left(\frac{y - \Theta_l(x')}{\sigma}\right) \geq 0$ for all $l \geq 1$. We then have $H(\Theta_h \text{ is nondecreasing}) = 1$.

Proof of Lemma 5: Under model (3.2), $f(\cdot|x)$ is in the form of (3.3) for any x . For any x and x' , $f(\cdot|x) - f(\cdot|x') = \sum_{h=1}^{\infty} p_h \{N(\cdot; \Theta_h(x), \sigma^2) - N(\cdot; \Theta_h(x'), \sigma^2)\}$. Based on the fact that the normal density function is bounded and following a similar argument as in the proof of Theorem 2, we can show that $f(\cdot|x) \rightarrow f(\cdot|x')$ as $x \rightarrow x'$ is equivalent to $\Theta_h(x) - \Theta_h(x') \rightarrow 0$ as $x \rightarrow x'$ for any h , which means Θ_h is continuous at x' . Since Θ_h is a random function from H , the continuity of $f_{\mathcal{X}}$ w.r.t. x is equivalent to $H(s \text{ is continuous}) = 1$ on $(S_{\mathcal{X}}, \sigma_{S_{\mathcal{X}}})$.

Appendix B: Gibbs sampler details of step (a) in Section 4.4:

Step (a) Sample $\beta_h = (\beta_{h0}, \beta_{h1}, \dots, \beta_{hk})$, $h = 1, \dots, N$.

Let \mathbf{K}^* be the set of unique values in $\mathbf{K} = (K_1, \dots, K_n)'$ in the current iteration of Gibbs sampler.

(a1) If $h \notin K^*$ or $\sum_{i:K_i=h} b_l^2(x_i) = 0$, sample β_{h0} from $N(m_0, \nu_0)$, and sample β_{hl} from the mixed prior $\pi_0 \delta_0(\beta_{hl}) + (1 - \pi_0) \text{Exp}(\beta_{hl}; \lambda)$ for $l = 1, \dots, k$.

(a2) If $h \in K^*$ and $\sum_{i:K_i=h} b_l^2(x_i) > 0$, sample β_{h0} from $N(E_{h0}, W_{h0})$, where

$$W_{h0} = (\sigma^{-2} \sum_{i:K_i=h} b_0^2(x_i) + \nu_0)^{-1},$$

and

$$E_{h0} = W_{h0} \left[\sigma^{-2} \sum_{i:K_i=h} b_0(x_i) \{y_i - \sum_{l \neq 0} \beta_{hl} b_l(x_i) - \mathbf{z}'_i \boldsymbol{\gamma}\} + \nu_0 m_0 \right].$$

For $l = 1, \dots, k$,

$$\beta_{hl} | \cdot \propto \pi_{hl} \delta_0(\beta_{hl}) + (1 - \pi_{hl}) N_+(E_{hl}, W_{hl}),$$

where

$$\pi_{hl} = \left(1 + \frac{(1 - \pi_0) \lambda (1 - \Phi(0; E_{hl}, W_{hl}))}{\pi_0 N(0; E_{hl}, W_{hl})} \right)^{-1},$$

$$W_{hl} = \left(\sigma^{-2} \sum_{i:K_i=h} b_l^2(x_i) \right)^{-1},$$

and

$$E_{hl} = W_{hl} \left[\sigma^{-2} \sum_{i:K_i=h} b_l(x_i) \{y_i - \sum_{l' \neq l} \beta_{hl'} b_{l'}(x_i) - \mathbf{z}'_i \boldsymbol{\gamma}\} - \lambda \right].$$

In the above, $\Phi(0; a, b)$ and $N(0; a, b)$ denote the cumulative distribution function and density function of $N(a, b)$ evaluated at 0, respectively, and $N_+(a, b)$ denotes the truncated normal distribution of $N(a, b)$ above 0.

References

- Abrevaya, I. (2005), "Isotonic Quantile Regression," *Sankhya*, 67, 187-199.
- Arjas, E. and Gasbarra, D. (1996), "Bayesian Inference of Survival Probabilities, Under Stochastic Ordering Constraints," *Journal of the American Statistical Association*, 91, 1101-1109.
- De Iorio, M., Müller, P., Rosner, G.L. and MacEachern, S.N. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205-215.
- Dunson, D.B. and Park, J-H. (2006), "Kernel Stick-Breaking Processes," *ISDS Discussion Paper*, 22, Duke University, Durham, NC, USA.
- Dunson, D.B. and Peddada, S. (2007), "Bayesian Nonparametric Inferences on Stochastic Ordering," *ISDS Discussion Paper*, 2, Duke University, Durham, NC, USA.
- Dunson, D.B., Pillai, N. and Park, J-H. (2007), "Bayesian Density Regression," *Journal of the Royal Statistical Society B*, 69, 163-183.
- 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.
- Ferguson, T.S. (1974), "Prior Distributions on Spaces of Probability Measures" *The Annals of Statistics*, 2, 615-629.
- Gelfand, A.E. and Kottas, A. (2000), "Nonparametric Bayesian Modeling for Stochastic Order," *The Annals of the Institute of Statistical Mathematics*, 53, 865-876.

- Gelfand, A.E., Kottas, A. and MacEachern, S.N. (2005), "Bayesian Nonparametric Spatial Modeling with Dirichlet Process Mixing," *Journal of the American Statistical Association*, 100, 1021-1035.
- Gelfand, A.E., Smith, A.F.M. and Lee, T.-M. (1992), "Bayesian Analysis of Constrained Parameter and Truncated Data Problems," *Journal of the American Statistical Association*, 87, 523-532.
- Griffin, J.E. and Steel, M.F.J. (2006), "Order-based Dependent Dirichlet Processes," *Journal of the American Statistical Association*, 101, 179-194.
- Hoff, P. (2003a), "Nonparametric estimation of convex models via mixtures," *The Annals of Statistics*, 31, 174-200.
- Hoff, P. (2003b), "Bayesian methods for partial stochastic ordering," *Biometrika*, 90, 303-317.
- Holmes, C.C. and Heard, N.A. (2003), "Generalized Monotonic Regression using Random Change Points," *Statistics in Medicine*, 22, 623-638.
- Ishwaran, H. and James, L.F. (2001), "Gibbs Sampling Methods for Stick-breaking Priors," *Journal of the American Statistical Association*, 101, 179-194.
- Lavine, M. and Mockus, A. (1995), "A Nonparametric Bayes Method for Isotonic Regression," *Journal of Statistical Planning and Inference*, 46, 235-248.
- Lo, A. (1984), "On a Class of Bayesian Nonparametric Estimates: I. Density Estimation," *The Annals of Statistics* vol. 12, pp. 351-357.
- Longnecker, M.P. Klebanoff, M. A., Zhou, H.B. and Brock, J.W. (2001), "Association between Maternal Serum Concentration of the DDT and Preterm and Small-for-gestational-age Babies at Birth," *Lancet*, 358, 110-114.
- MacEachern, S.N. (1999), "Dependent Dirichlet Processes," in *Proceedings of the Bayesian Statistical Science Section*, Journal of the American Statistical Association.
- MacEachern, S.N. (2001), "Decision Theoretic Aspects of Dependent Nonparametric process," In *Bayesian Methods with Applications to Science, Policy and Official Statistics* (ed. E. George), pp. 551-560. Creta: International Society for Bayesian Analysis.

- Neelon B. and Dunson, D.B. (2004), "Bayesian Isotonic Regression and Trend Analysis," *Biometrics*, 60, 398-406.
- Parthasarathy, K.R. "Probability Measures on Metric Spaces," Academic Press Inc., New York, 1967.
- Ramsay, J.O. (1988) "Monotone Regression Splines in Action," *Statistical Science*, 3, 425-441.
- Sethuraman, J. (1994) "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639-650.
- Smith, M. and Kohn, R. (1996) "Nonparametric regression using Bayesian variable selection," *Journal of Econometrics*, 75, 317-343.

Table 1: Estimation of covariate effects

covariate	posterior mean	95% credible interval
race=black	-4.4655	(-5.7927 -3.1222)
cholesterol level	0.1294	(-0.5157 0.7902)
triglyceride level	-1.7030	(-2.3765 -1.0325)

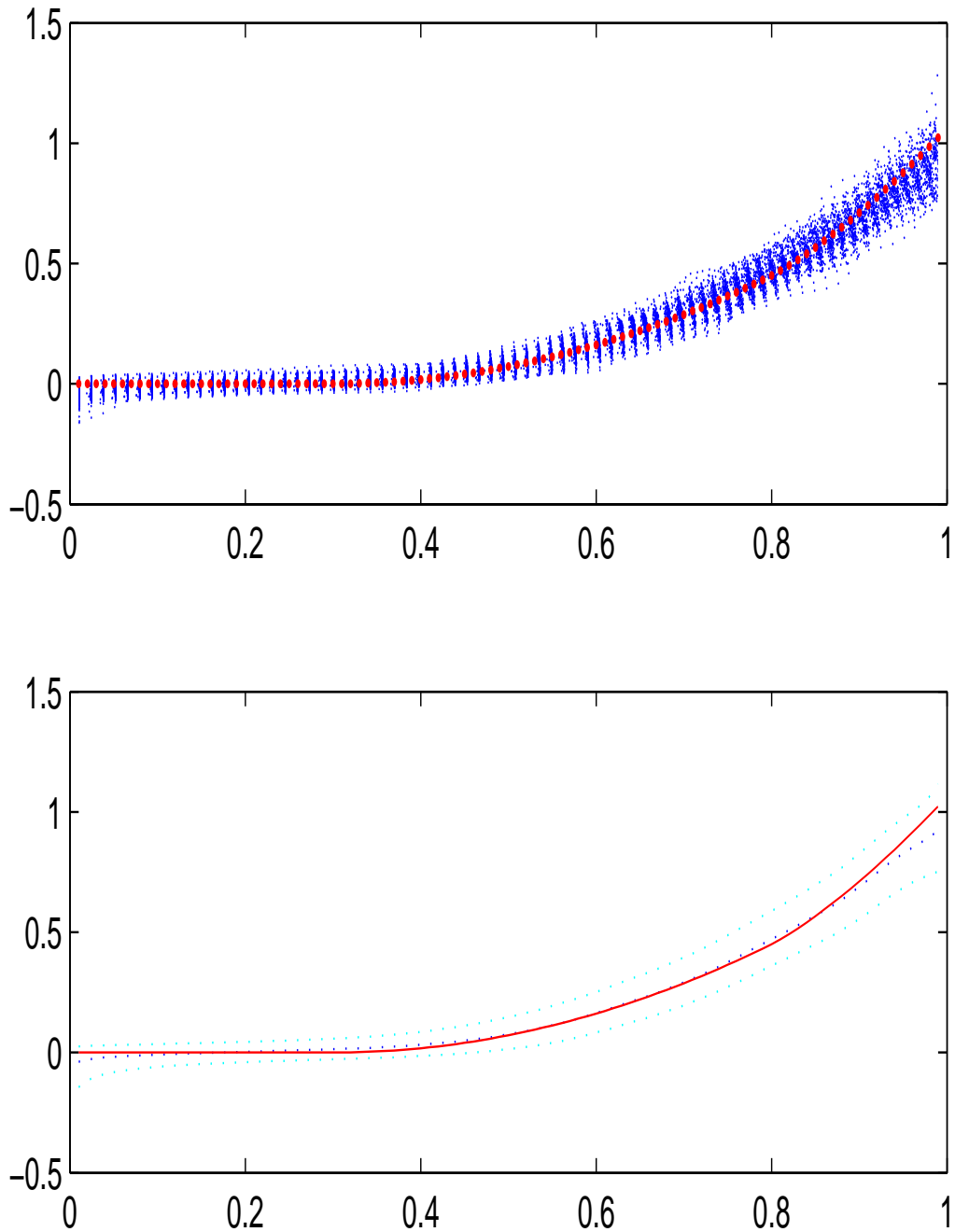


Figure 1: Top: True (red lines) and estimated (blue dotted lines) mean regression functions for each of the 200 simulated data sets. Bottom: True (red line), the average of mean regression functions (blue dotted), and the average of 95% credible intervals of mean functions from the 200 data sets.

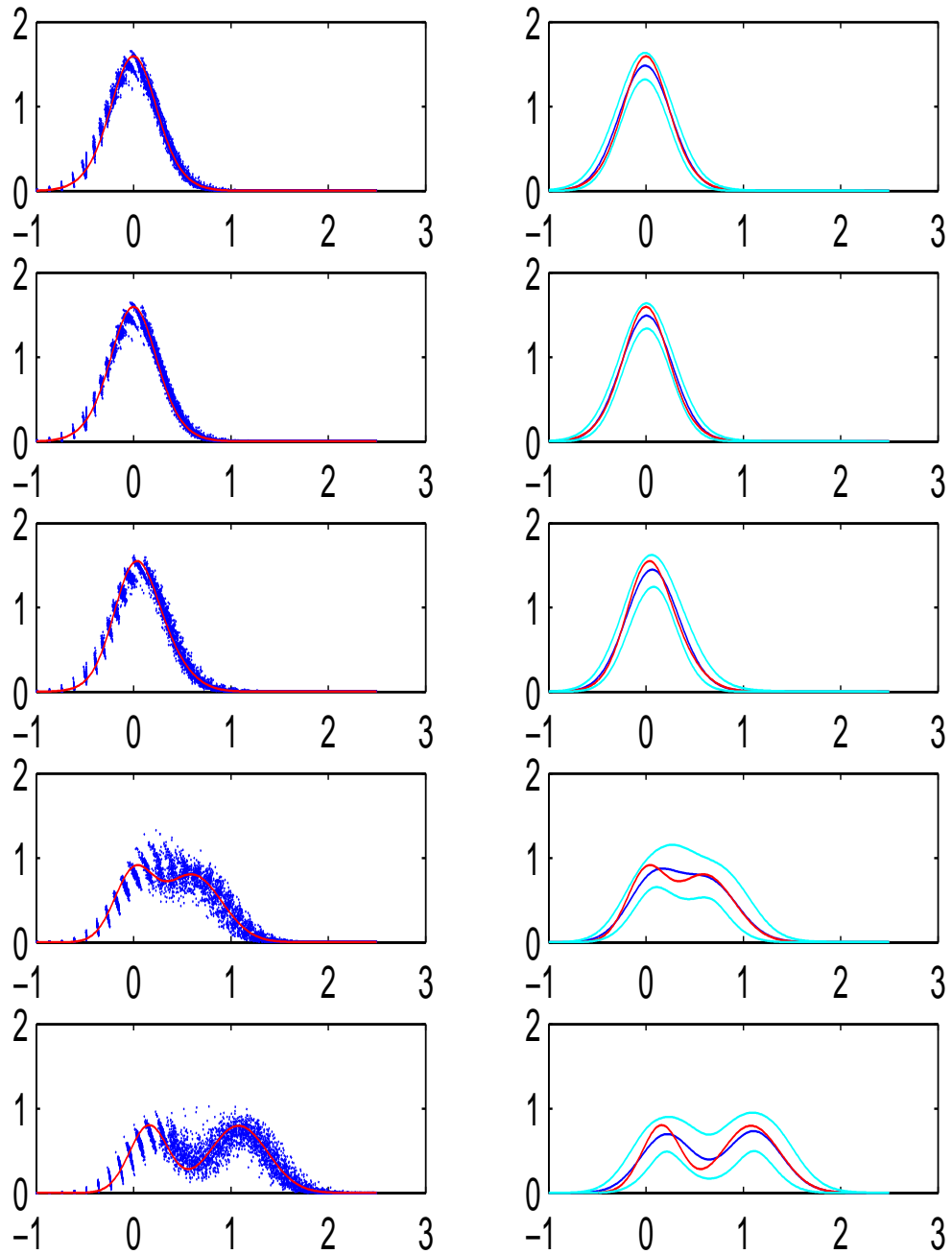


Figure 2: True (red lines) and estimated (blue dotted lines) conditional densities of the 10th, 25th, 50th, 75th, and 90th percentiles of the empirical distribution of x . The left panel shows the estimate for each simulated data set, while the right panel shows the average of these estimates together with the average of the estimated 95% credible intervals.

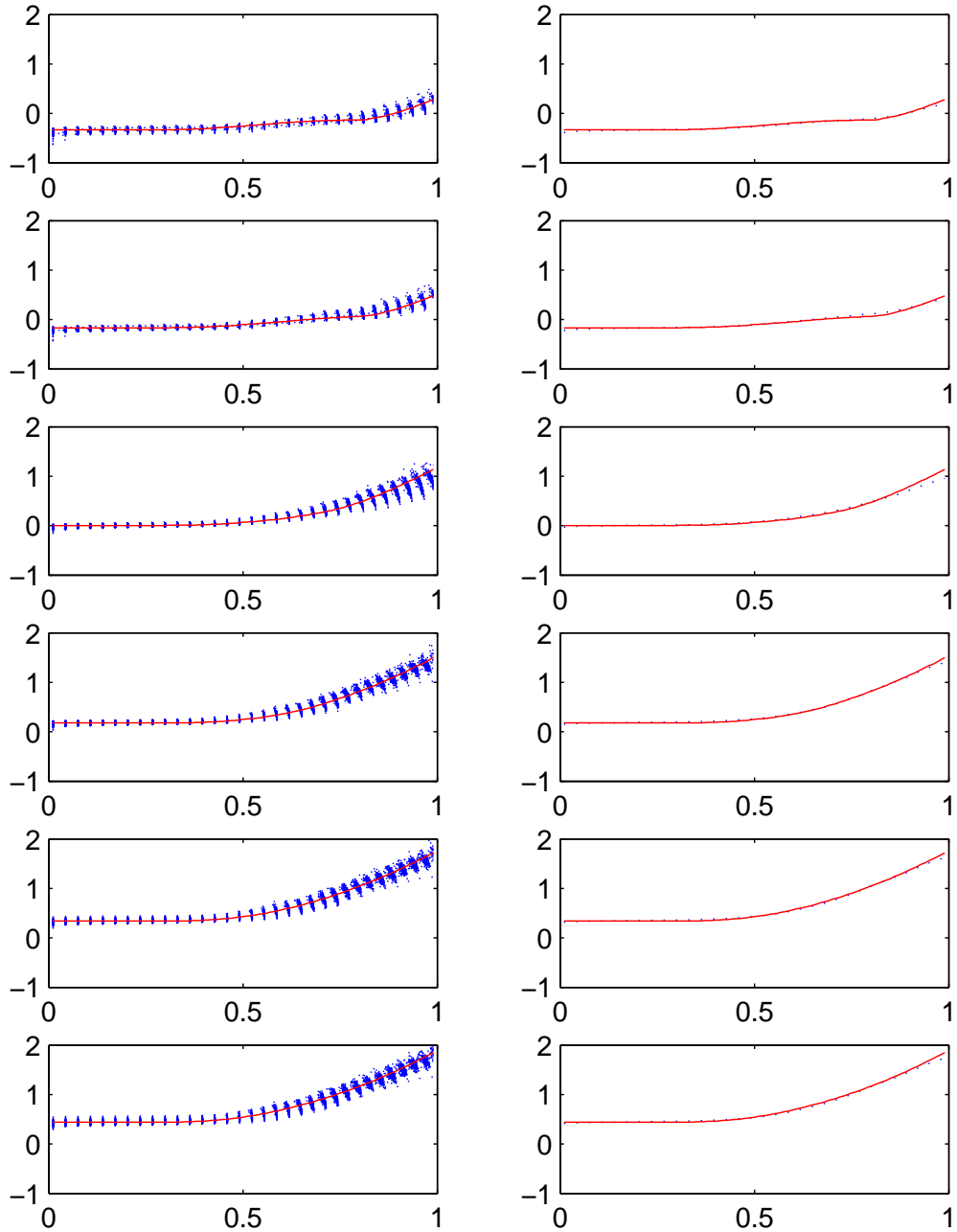


Figure 3: True (red lines) and estimated (blue dotted lines) quantile regression curves for the 10th, 25th, 50th, 75th, 90th, and 95th percentiles. The left panel shows the estimate for each simulated data set, while the right panel shows the average of these estimates.

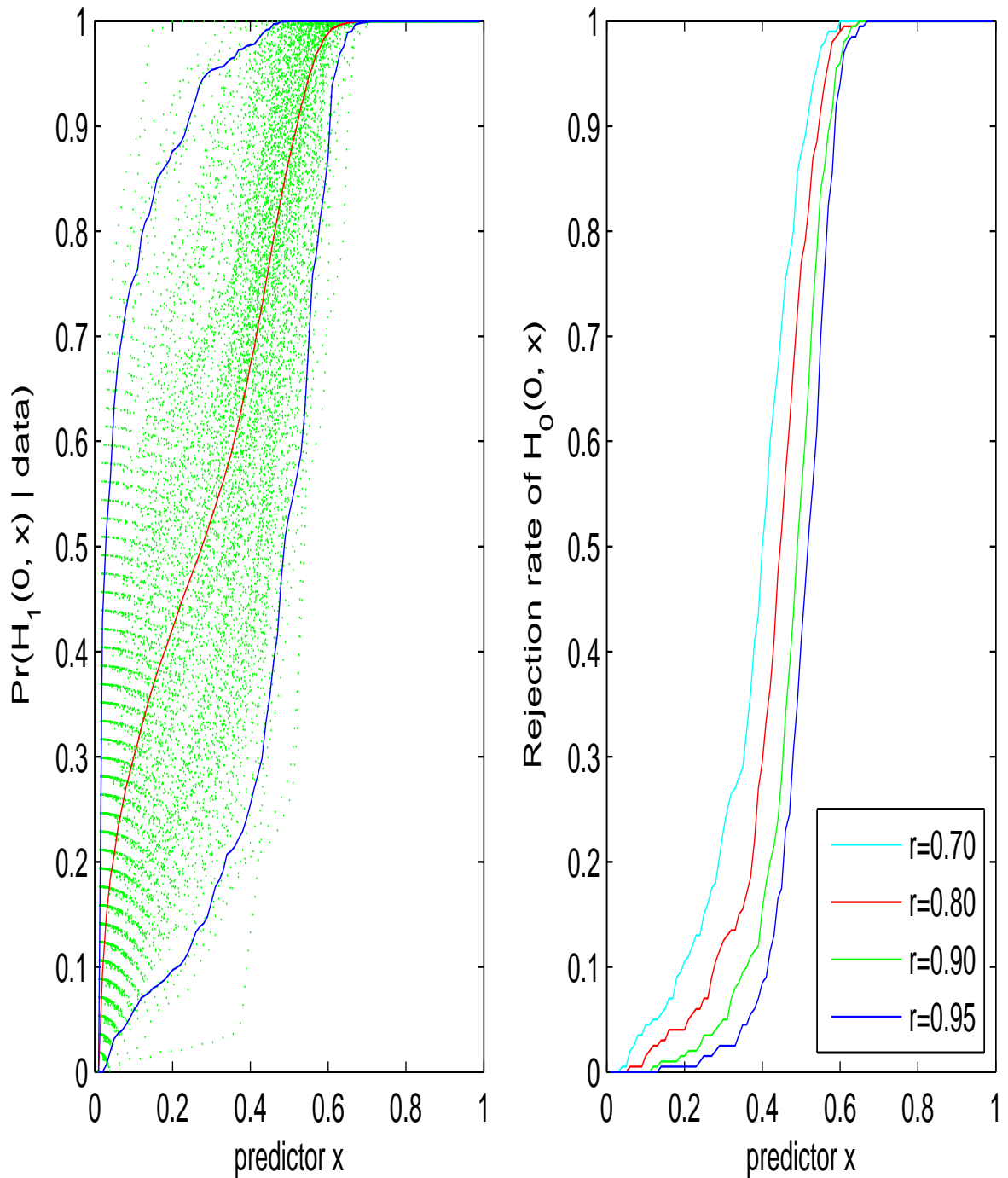


Figure 4: The left plot shows the estimated posterior probability of $H_1(0, x)$ with $\epsilon = 0.05$ as a function of x for each of the 200 simulated data sets in green, with the mean in red and 95% intervals in blue. The right plot shows the proportion of the data sets having the $\Pr(H_1(0, x) | \text{Data})$ above a threshold $r = 0.70, 0.80, 0.90$ or 0.95 .

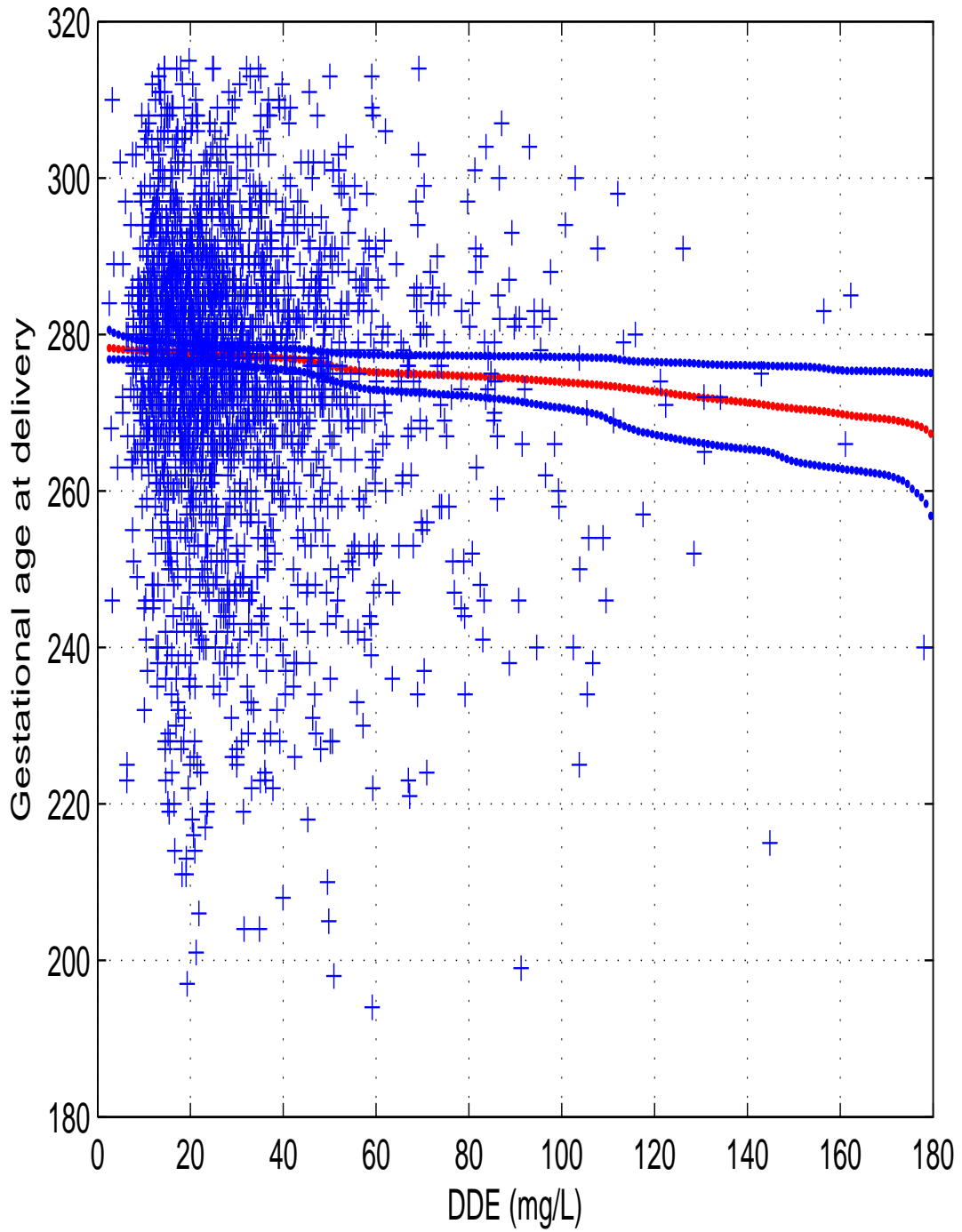


Figure 5: Scatter plot and mean function estimation for epidemiology application. The posterior mean curve is in red and the 95% credible intervals are in blue.

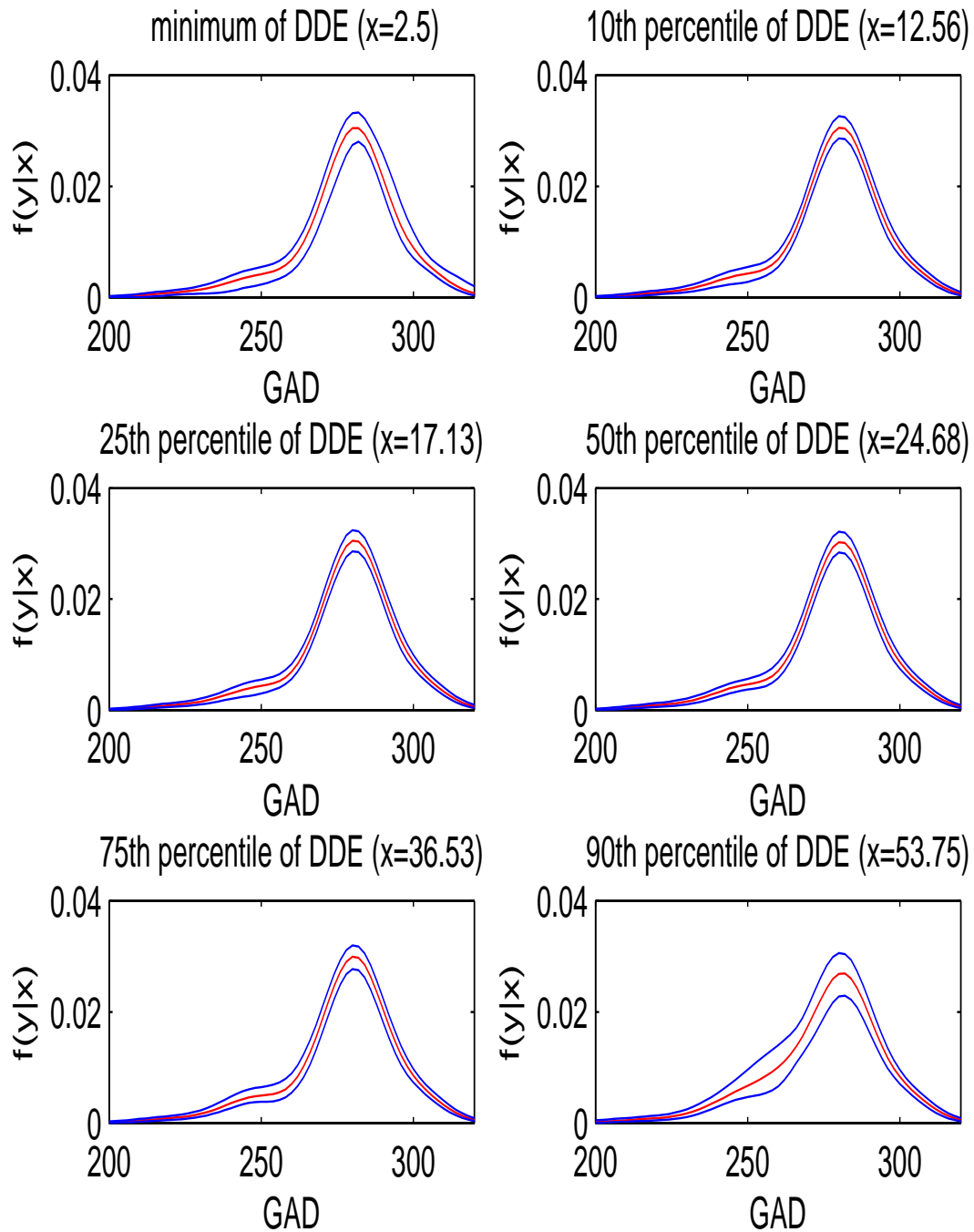


Figure 6: Estimated conditional densities of gestational age at delivery given DDE for different DDE values. The posterior mean density curves are in red and the 95% credible intervals are in blue in each plot.

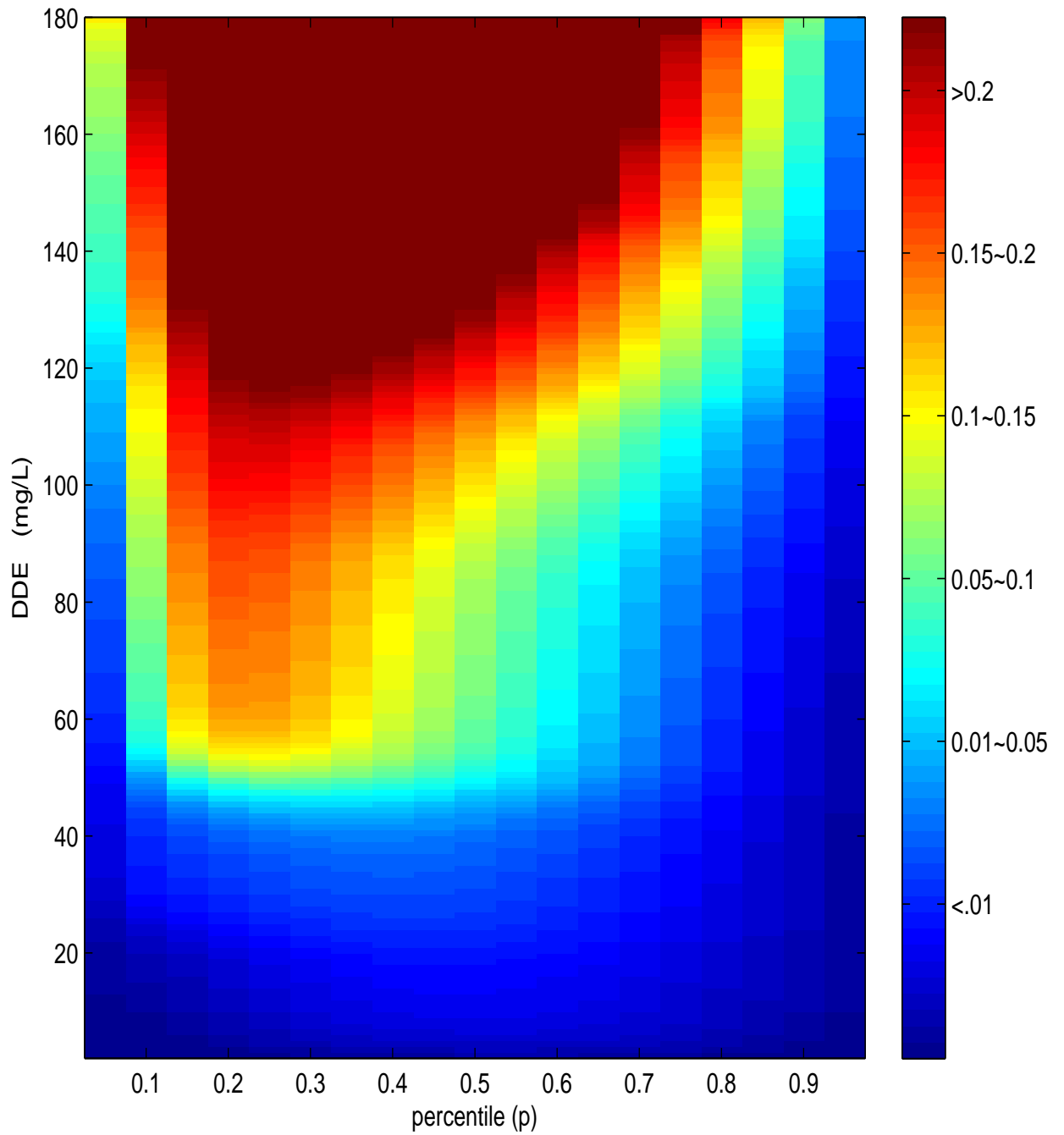


Figure 7: Heat plot for the difference between F_x and F_0 for different DDE doses and percentiles (p) of the response distribution.