Abstract. Shape constrained regression analysis has applications in dose-response modeling, environmental risk assessment, disease screening and many other areas. Incorporating the shape constraints can improve estimation efficiency and avoid implausible results. We propose two novel methods focusing on Bayesian monotone curve and surface estimation using Gaussian process projections. The first projects samples from an unconstrained prior, while the second projects samples from the Gaussian process posterior. Theory is developed on continuity of the projection, posterior consistency and rates of contraction. The second approach is shown to have an empirical Bayes justification and to lead to simple computation with good performance in finite samples. Our projection approach can be applied in other constrained function estimation problems including in multivariate settings.

Keywords: Asymptotics; Bayesian nonparametrics; Isotonic regression; Projective Gaussian process; Shape constraint.

1. Introduction

In a rich variety of applications, prior knowledge is available on the shape of a surface, with examples including monotonicity, unimodality and convexity. Incorporating such shape constraints can often substantially improve estimation efficiency and stability, while producing results consistent with prior knowledge. We propose two novel approaches based on Gaussian process projections. Gaussian processes are routinely applied but have the disadvantage of not allowing constraints. Although we focus on monotone curves and surfaces, the approach can be applied directly in much broader settings including additive models, multivariate regression with monotonicity constraints only in certain directions, and other types of shape constraints.

There is a rich frequentist literature on monotone curve and isotonic regression estimation, with a common approach minimizing a least squares loss subject to a restriction (Barlow et al. 1972, Robertson et al. 1988). For more recent references, refer to Bhattacharya & Kong (2007) and Bhattacharya & Lin (2010, 2011). Alternatively, restricted kernel (Müller & Schmitt 1988), Dette et al. (2005) and Mammen (1991)) and spline (Ramsay 1988 and Kong & Eubank 2006) methods have been proposed.

From a Bayesian perspective, one specifies a prior on the regression function and inference is based on the posterior distribution. Gelfand & Kuo (1991) use an ordered Dirichlet prior on a strictly monotone dose-response function. Neelon & Dunson (2004) use an additive model with a prior imposed on the slope of the piecewise linear functions. Shively et al. (2009) and Shively et al. (2011) use restricted splines. Bornkamp & Ickstadt (2009) adopt mixture modeling.

Although there is a rich existing literature on Bayes monotone curve estimation, our work has two key motivations: (1) There is a lack of theory supporting these methods beyond...
consistency; (2) Current approaches involve basis expansions and challenges arise in multivariate cases. Gaussian processes have a rich theoretical foundation, can easily incorporate prior information, and can be implemented routinely in multivariate settings. We define a class of projective Gaussian processes which inherit these advantages.

2. GAUSSIAN PROCESS PROJECTIONS

Let \( w \sim \text{GP}(\mu, R) \) denote the sample path of a ‘mother’ Gaussian process indexed on \( \mathcal{X} \subset \mathbb{R}^p \), with \( \mu : \mathcal{X} \to \mathbb{R} \) the mean function and \( R : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+ \) the covariance kernel. Let \( \mathcal{M} \) be a subset of the space of continuous functions mapping from \( \mathcal{X} \) to \( \mathbb{R} \) having some constraint. We define the projective Gaussian process \( P_w \) on the constrained space \( \mathcal{M} \) as

\[
(2.1)\quad P_w = \arg\min_{F \in \mathcal{M}} \int_{\mathcal{X}} \{w(t) - F(t)\}^2 dt.
\]

Let \( \mathcal{M}[0,1]^p \) denote the space of monotone functions on \([0,1]^p\). Focusing initially on the \( p = 1 \) case, \(2.1\) has the solution

\[
(2.2)\quad P_w(x) = \inf_{v \geq x} \sup_{u \leq x} \frac{1}{v-u} \int_u^v w(t) dt, \text{ for } x \in [0,1].
\]

The existence and uniqueness of the projection follow from Theorem 1 in Rychlik (2001).

**Remark 2.1.** The projection in \(2.2\) can be well approximated using the pooled adjacent violators algorithm (Barlow et al. (1972)).

Some properties of the projection function include:

1. \( P_w(x) = w(x) \) if \( w \) is a monotone function. Therefore, \( P_w \) is surjective.

2. \( P_w(x) = c \) if \( w \) is a decreasing function where \( c = \int_0^1 w(s) ds \) which is the slope of the line joining \((0,0)\) and \((1, \int_0^1 w(s) ds)\).

3. \( P_w(x) \) is a continuous function given \( w \) is continuous (Groeneboom & Jongbloed (2010)).

Hence, in projecting the Gaussian process from \( C[0,1] \) to \( \mathcal{M}[0,1] \) one induces a valid measure on the set of continuous monotone functions \( \mathcal{M}[0,1] \).

The following lemma on the continuity of the projection as an operator is key to showing the projective Gaussian process inherits concentration and approximation properties of the mother Gaussian process.

**Lemma 2.1.** Let \( w_1, w_2 \) be continuous functions on \([0,1]\). Then the following holds:

\[
(2.3)\quad \sup_{x \in [0,1]} |P_{w_1}(x) - P_{w_2}(x)| \leq \sup_{x \in [0,1]} |w_1(x) - w_2(x)|.
\]

Monotone curve estimation under the projective Gaussian process is easily extended to monotone surface estimation. As an illustration, suppose that \( F \in \mathcal{M}[0,1]^2 \) is a monotone continuous function on \([0,1]^2\) with respect to partial orderings, so that given \( s_1 \leq s_2 \) and \( t_1 \leq t_2 \), \( F(s_1, t_1) \leq F(s_2, t_2) \). Since \( \mathcal{M}[0,1]^2 \) is a closed convex cone, equation (1) can be solved to obtain \( P_w \) from the sample path \( w \sim \text{GP}(\mu, R) \) of a two-dimensional Gaussian process. Apply Algorithm 1 to obtain the solution.

Theorem 1 characterizes the solution to Algorithm 1.
Algorithm 1

Given any fixed \( t \), \( w(s, t) \) is a function of \( s \) and apply (2.2) to obtain \( \hat{w}^{(1)}(s, t) \) by projecting \( w \) along the \( s \) direction. Letting \( S^{(1)} = \hat{w}^{(1)} - w \), project \( w + S^{(1)} \) onto \( M[0,1] \) with respect to the \( t \) direction to obtain \( \hat{w}(s, t) \). Let \( T^{(1)} = \hat{w}^{(1)} - (w + S^{(1)}) \). Letting \( i = 2, \ldots, k \), in the \( i \)th step we obtain \( \hat{w}^{(i)} \) by projecting \( w + T^{(i-1)} \) along the \( s \) direction for any \( t \) and \( \hat{w}^{(i)} \) as the projection of \( w + S^{(i)} \) along the \( t \) direction for any \( s \). The algorithm terminates when \( \hat{w}^{(i)} \) or \( \hat{w}^{(i)} \) is monotone with respect to both \( s \) and \( t \) for some step \( i \).

Theorem 2.1. Let \( P_w \) be the projection of \( w \) solving (2.1). Then one has

\[
P_w(s, t) = \lim_{k \to \infty} w^{(k)}(s, t) = \lim_{k \to \infty} \tilde{w}^{(k)}(s, t) \quad \text{as} \quad k \to \infty,
\]

and

\[
\sup_{s, t} |P_{w_1}(s, t) - P_{w_2}(s, t)| \leq \sup_{s, t} |w_1(s, t) - w_2(s, t)|.
\]

Theorem 2.1 implies that higher dimensional projections can be obtained by sequentially projecting the adjusted \( w \) along each of its directions. This approach can be trivially extended to \( p > 2 \) dimensional problems with monotonicity constraints in one or more directions.

3. Bayesian inference under projective Gaussian process

3.1. Model and notation. We carry out Bayesian inference under the projective Gaussian process, focusing on estimation of the \( p \)-dimensional monotone function \( F(x) \) assuming the following model

\[
y_i = F(x_i) + \epsilon_i, \quad 1 \leq i \leq n \quad (x_1 \preceq x_2 \preceq \cdots \preceq x_n),
\]

where \( \epsilon_i \sim N(0, \sigma^2) \) and \( F(x) \) is monotone increasing under partial orderings in the sense that \( F(x_1) \leq F(x_2) \) whenever \( x_1 \preceq x_2 \). Without loss of generality, assume \( x \) lies in the compact set \([0, 1]^p\). The design of \( x_1, \ldots, x_n \) can be fixed or random where \( x_i \sim G_0 \) for some distribution \( G_0 \). Although we focus on Gaussian residuals for simplicity, the methods can be automatically applied in general settings.

We first define some notions of neighborhoods. Let \( \eta = (F, \sigma), \eta_0 = (F_0, \sigma_0) \) denote the true value, and \( \Pi \) denote the prior on \( \eta \), which is expressed as \( \Pi_F \Pi_\sigma \), where \( \Pi_F \) and \( \Pi_\sigma \) are independent priors on \( F \) and \( \sigma \) respectively. As shorthand, let \( f_{x \mid 0} \) denote the conditional density \( N(F(x), \sigma^2) \) with \( f_{x \mid 0} \) the true conditional density. For random design, Hellinger distances \( d_H(\eta, \eta_0) \) are defined as

\[
d_H^2(\eta, \eta_0) = \int d_h^2(f_{x \mid F}, f_{x \mid 0})G_0(dx),
\]

with \( d_h^2(f_{x \mid F}, f_{x \mid 0}) = \frac{1}{2} \int \left( \sqrt{f_{x \mid F}} - \sqrt{f_{x \mid 0}} \right)^2 dy \). We let \( U_\epsilon(\eta_0) \) denote an \( \epsilon \) Hellinger neighborhood around \( \eta_0 \) with respect to \( d_H \). The Kullback-Leibler divergence between \( \eta \) and \( \eta_0 \) is

\[
d_{KL}(\eta, \eta_0) = \int \int f_{x \mid 0} \log \frac{f_{x \mid F}}{f_{x \mid 0}} dyG_0(dx).
\]

An \( \epsilon \) Kullback-Leibler neighborhood around \( \eta_0 \) is denoted by \( K_\epsilon(\eta_0) \).
3.2. Projective Gaussian process prior. We first use a projective Gaussian process, $F \sim \text{pGP}_M(\mu, R)$, as a prior on the monotone regression function $F(x)$ for $x \in [0, 1]$. We assume the mother Gaussian process $w \sim \text{GP}(\mu, R)$ is continuous, with $\mathcal{H} = C[0,1]$ the closure of the reproducing kernel Hilbert space corresponding to $R$. Our proofs assume $\mu = 0$. The following Theorems show posterior consistency and convergence rates under our projective Gaussian process prior in the $p = 1$ special case; these Theorems can be generalized to arbitrary $p$.

**Theorem 3.1.** Let $w_0$ be in the pre-image of $F_0$ so that $F_0 = P_{w_0}$. Let $\Pi_F$ be the projective Gaussian process prior on $\mathcal{M}[0,1]$. Assume $\Pi_\sigma$ has a positive continuous density including $\sigma_0$ in its support. Then under a random design, for all $\epsilon > 0$,

$$
\Pi\{U_\epsilon^C(\eta_0)(a_1, y_1), \ldots, (a_n, y_n)\} \to 0 \text{ a.s. } \prod_{i=1}^{n} P_{f_{a_i0}},
$$

where $U_\epsilon^C(\eta_0)$ is the complement of $U_\epsilon(\eta_0)$ in $\mathcal{M}[0,1]$.

A similar consistency theorem holds for the case of fixed designs.

**Theorem 3.2.** We maintain the same conditions as in Theorem 3.1. For the case of a fixed design, the posterior under the projective Gaussian process prior is consistent, that is, for all $\epsilon > 0$

$$
\Pi\{U_\epsilon(\eta_0)^C(a_1, y_1), \ldots, (a_n, y_n)\} \to 0 \text{ a.s. } \prod_{i=1}^{n} P_{f_{a_i0}},
$$

where $U_\epsilon(\eta_0)$ is the average (empirical) Hellinger neighborhood.

For the rates theorem, we assume a fixed design and that $\sigma^2$ follows a log-uniform prior $\Pi_{\sigma^2}$ on a compact interval $[l, u]$ including $\sigma_0^2$ with $l > 0$. Let $\phi_{w_0}(\epsilon n)$ denote the Gaussian process concentration function defined in the Appendix.

**Theorem 3.3.** Let $F_0$ be the true monotone function and $w_0$ be any element in the pre-image of $F_0$. Let $\Pi_F = \text{pGP}_M(0, R)$. If $\phi_{w_0}(\epsilon n) \leq n \epsilon_n, \Pi_{\sigma^2}\{\sigma_0^2(1 - \epsilon_n^2/3, 1 + \epsilon_n^2/3)\} \geq e^{-C_1 \epsilon_n^2}$ and $\frac{1}{\sqrt{n} \epsilon_n} \leq e^{C_0 \epsilon_n}$ for some constants $C_1$ and $C_0$, the posterior distribution of $\eta$ satisfies

$$
\Pi_n\{\eta : d_H(\eta, \eta_0) > M \epsilon_n|(a_1, y_1), \ldots, (a_n, y_n)\} \to 0 \text{ a.s. } \prod_{i=1}^{n} P_{f_{a_i0}}
$$

for $M$ large enough where $d_H(\cdot, \cdot)$ is the empirical Hellinger distance.

Let $W_t = \text{GP}(0, R)$ with squared exponential covariance kernel $R(t_1, t_2) = e^{-(t_1 - t_2)^2}$. Define a scaled Gaussian process $W_A = \{W_{at}\}$. As an example, we consider the rate of contraction for the projection prior using $W_A$ with $A$ having a Gamma prior as in van der Vaart & van Zanten (2007, 2009).

**Corollary 3.4.** Let $\Pi = \Pi_F \Pi_\sigma$ with $\Pi_F$ the projective Gaussian process prior induced from the projection of $W_A$. One has the following results on the convergence rate of the posterior.

1. If the true monotone function $F_0 \in C_0[0,1] \cap \mathcal{M}[0,1]$ for some $\alpha \geq 0$, then the posterior converges at rate at least $n^{-\alpha/(2\alpha+1)}(\log n)^{(4\alpha+1)/(4\alpha+2)}$. 


(2) If \( F_0 \in C^0[0,1] \cap \mathcal{M}[0,1] \) which is continuous but not differentiable, then the convergence rate is at least \( n^{-\alpha/(2\alpha+1)}(\log n)^{(4\alpha+1)/(4\alpha+2)} \) \( \alpha \geq 1 \), if there exists \( w_0 \in C^0[0,1] \) such that \( P_{w_0} = F_0 \).

Remark 3.1. If the covariates are random from \( G_0 \), the rates hold with norm \( d_2(\eta, \eta_0) = \left[ \int_{0}^{1} \{F(x) - F_0(x)\}^2 G_0(dx) \right]^{1/2} + |\sigma - \sigma_0| \).

3.3. Inference by projecting the Gaussian process posterior. In this section we propose an alternative approach that relies on projecting draws from the posterior under a Gaussian process prior onto the space of monotone functions \( \mathcal{M}[0,1]^p \). This approach is easy to implement and has excellent performance in applications we have considered.

We first impose a Gaussian process on \( F \) and a prior \( \Pi_\sigma \) on \( \sigma \), and then project the posterior of \( F \) onto \( \mathcal{M}[0,1]^p \). This induces a probability measure on \( \mathcal{M}[0,1]^p \) based on which the inference is carried out. We denote by \( \tilde{\Pi}(\cdot|y_1, \ldots, y_n) \) the induced distribution on \( \Gamma = \mathcal{M}[0,1]^p \times (0, \infty) \). We first present the following Theorem which shows the existence of a prior on \( \Gamma \) whose posterior is \( \tilde{\Pi}(\cdot|y_1, \ldots, y_n) \). Hence, our inference scheme fits in the Bayesian paradigm. Assume \( \sigma \) is compactly supported.

Theorem 3.5. Given \( \tilde{\Pi}(\cdot|y_1, \ldots, y_n) \), a probability measure on \( \Gamma \) obtained by projecting the posterior of a Gaussian process onto \( \mathcal{M}[0,1]^p \), there exists a prior \( \tilde{\Pi}(dF, d\sigma) \) on \( \Gamma \) whose posterior is \( \tilde{\Pi}(\cdot|y_1, \ldots, y_n) \).

Let \( F_0 \in \mathcal{M}[0,1]^p \) be the true monotone function. Let \( \eta = (F, \sigma) \) and \( \eta_0 = (F_0, \sigma_0) \). In proving the theory, we consider the random design with the covariates sampled from a distribution \( G_0 \) with distance \( d_2(\eta, \eta_0) \) the same as in Remark 3.1. Theory is shown for the special case \( p = 1 \) which can be generalized to arbitrary \( p \).

Since the covariates are from a distribution \( G_0 \), we consider the projection of the function \( w(t) \) onto the monotone space by minimizing

\[
\int_{0}^{1} \{w(t) - F(t)\}^2 G_0(dt).
\]

The solution to (3.5) is given by

\[
P_w(x) = \inf_{v \geq x \leq x} \sup \frac{1}{G_0(v) - G_0(u)} \int_{u}^{v} w(t)G_0(dt), \text{ for } x \in [0,1],
\]

which is a weighted version of (2.2). In terms of implementing the projection, one can use the pooled adjacent violators algorithm with non-constant weights. The following lemma shows continuity of the projection.

Lemma 3.1. Let \( w_1 \) and \( w_2 \) be two functions on \( [0,1] \). Then one has

\[
||P_{w_1} - P_{w_2}||_{2G_0(dx)} \leq ||w_1 - w_2||_{2G_0(dx)}
\]

where \( ||f - g||_{2G_0(dx)} = \{\int (f - g)^2 G_0(dx)\}^{1/2} \).

Theorem 3.6. Given the scaled Gaussian process \( W^A \) with \( A \) from some Gamma distribution, the convergence rates of \( \tilde{\Pi} \) with respect to \( d_2(\eta, \eta_0) \) are given as follows:

1. If the true monotone function \( F_0 \in C^0[0,1] \cap \mathcal{M}[0,1] \), then the posterior converges at rate at least \( n^{-\alpha/(2\alpha+1)}(\log n)^{(4\alpha+1)/(4\alpha+2)} \).
If $F_0$ is a flat function, so that $F_0 = C$ for some constant $C$, then the rate of convergence is at least $n^{-1/2} (\log n)^2$.

Our theory of projecting the posteriors applies naturally to the higher dimensional cases.

4. Posterior computation

4.1. Monotone estimation of curves with simulated data. We apply the approach proposed in §3.3 by projecting the posterior of a Gaussian process. Let $w \sim \text{GP}(0, R)$ with $R(x_1, x_2) = \beta^{-1} \exp\{-\gamma(x_1 - x_2)^2\}$, where $\beta \sim \text{Ga}(4, 1)$, $\gamma \sim \text{Ga}(4, 1)$ and $\sigma^{-2} \sim \text{Ga}(4, 1)$. In a first stage, we run a Markov chain Monte Carlo algorithm to obtain draws from the joint posterior of covariance parameters $(\beta, \gamma, \sigma)$ and the pre-projection curve evaluated at the data points $w_n^* = \{w(x_1), \ldots, w(x_n)\}$. This can proceed using any of a wide variety of algorithms developed for Gaussian process regression models; we use Vanhatalo et al. (2012, arXiv:1206.5754v1). The number of Markov chain Monte Carlo iterations is taken to be 5,000 with a burn in of 1,000. After convergence, sample paths $w_n^*$ are then projected to the monotone space using the pooled adjacent violators algorithm.

Data of size $n = 100$ are simulated from a normal error model with standard deviation $\sigma = 1$. The true mean functions given below are proposed by Holmes & Heard (2003) and Neelon & Dunson (2004) and are also used in a comparative study in Shively et al. (2009).

(a) $F_1(x) = 3$, $x \in (0, 10]$ (flat function).
(b) $F_2(x) = 0.32\{x + \sin(x)\}$, $x \in (0, 10]$ (sinusoidal function).
(c) $F_3(x) = 3$ if $x \in (0, 8]$ and $F_3(x) = 6$ if $x \in (8, 10]$ (step function).
(d) $F_4(x) = 0.3x$, $x \in (0, 10]$ (linear function).
(e) $F_5(x) = 0.15 \exp(0.6x - 3)$, $x \in (0, 10]$ (exponential function).
(f) $F_6(x) = 3/\{1 + \exp(-2x + 10)\}$, $x \in (0, 10]$ (logistic function).

The $x$ values are taken to be equidistant in the interval $(0, 10]$. The root mean squared error of the estimates is calculated in the simulation study for the Gaussian process with and without projection, with the results shown in Table 1. The results presented in the following table are the average root mean squared error of 50 samples of data. We compare our results with the root mean squared error results of the regression spline provided in Shively et al. (2009). Figs 1 and 2 show projection estimates and 99% pointwise credible intervals for some of the regression functions and randomly selected simulated data sets along with the true curves. In each case the estimated curve was close to the truth and 99% intervals mostly enclosed the true curves.

| Method                  | Flat       | Sinusoidal | Step       | Linear     | Exponential | Logistic  |
|------------------------|------------|------------|------------|------------|-------------|-----------|
| Gaussian process       | 0.151      | 0.219      | 0.271      | 0.167      | 0.197       | 0.255     |
| Gaussian process projection | 0.113      | 0.211      | 0.253      | 0.163      | 0.191       | 0.224     |
| Regression spline      | 0.097      | 0.229      | 0.285      | 0.240      | 0.213       | 0.194     |
Figure 1. Dash lines correspond to true curves, circles are data points, solid lines are posterior mean curves under the Gaussian process projection, dashes with dots are 99% pointwise credible intervals.

Figure 2. Dash lines correspond to true curves, circles are data points, solid lines are posterior mean curves under the Gaussian process projection, dashes with dots are 99% pointwise credible intervals.

4.2. Monotone estimation of surfaces. In this section, we consider estimation of monotone surfaces. We choose a Gaussian process prior with covariance kernel $R(x, x') = \beta^{-1} \exp\left(-\sum_{k=1}^{2} \gamma_k (x_k - x'_k)^2\right)$, the posteriors of which are then projected to the monotone space. The hyperpriors are independent with $\beta \sim \text{Ga}(4, 1)$, $\gamma_1 \sim \text{Ga}(4, 1)$, $\gamma_2 \sim \text{Ga}(4, 1)$ and $\sigma^{-2} \sim \text{Ga}(4, 1)$. The Markov chain Monte Carlo algorithm was run for 3,000 iterations, with the initial 500 iterations discarded. The pre-projection curve is first evaluated at the $m_1m_2$ points $w(s_i, t_j)$ with $i = 1, \ldots m_1$ and $j = 1, \ldots m_2$ which are then projected to the
space of monotone surfaces. We briefly describe the projection scheme in the following steps in which \( w \) is only evaluated at the points \((s_i, t_j)\). This projection scheme was first introduced in [Robertson et al. (1988)] for their matrix partial ordering data.

**Step 1** For any \( t_j \) \((j = 1, \ldots, m_2)\), project \( w \) along the \( s \) direction by applying the pooled adjacent violators algorithm to each vector \( \{w(s_1, t_j), w(s_2, t_j), \ldots, w(s_{m_1}, t_j)\} \). Denote the projection of \( w \) by \( \tilde{w}^{(1)} \). Calculate the residual \( S^{(1)} = \tilde{w}^{(1)} - w \).

**Step 2** For any \( s_i \) \((i = 1, \ldots, m_1)\), project \( w + S^{(1)} \) along the \( t \) direction using the pooled adjacent violators algorithm, calculate the residual \( T^{(1)} = \tilde{w}^{(1)} - (w + S^{(1)}) \).

**Step 3** Iterate Step 1 and Step 2 by starting projecting \( w + T^{(1)} \) along the \( s \) direction.

In the \( i \)th iteration \((i = 1, \ldots, k)\), \( \tilde{w}^{(i)} \) is obtained by projecting \( w + T^{(i-1)} \) along the \( s \) direction and \( \tilde{w}^{(i)} \) is obtained by projecting \( w + S^{(i)} \) along the \( t \) direction.

For all our examples, this algorithm, which is a finite approximation to Algorithm 1, converged to a monotone solution in under 20 iterations. By the proof of Theorem [2.1] one can show that the solution obtained using the above projection scheme converges to the solution minimizing \( \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} (w(s_i, t_j) - F(s_i, t_j))^2 \) over the class of \( F \) that are monotone with respect to the partial ordering on \((s_i, t_j)\).

In the first seven examples, data of size \( n = 1,024 \) are simulated from a normal error model with true error \( \sigma = 0.5, 0.1 \) and true mean regression surfaces \( F_1 - F_7 \), the first six of which are also used in [Saarela & Arjas (2011)]. The model fit is checked for each of our models in terms of posterior mean \( \sigma \), the standard deviation of the posterior mean residuals, the correlations between the true and the posterior mean residuals and the correlation between the true and posterior mean predicted responses. We also look at the discrepancy between the true and estimated surface in terms of the mean squared error of our estimates. The results shown in Table [2] indicate good model fit using our projection estimates. For the case when \( \sigma = 0.1 \), the estimates of some models are plotted below in Figures [2] and [3] with the corresponding true surfaces. More plots are available in the supplementary appendix including more models and the higher noise case with \( \sigma = 0.5 \).

We illustrate the application to non-Gaussian data through analyzing pneumoconiosis risk in mine workers ([Ashford & Smith (1964)]). In epidemiology and toxicology studies, it is often of interest to assess joint risk as a function of multiple exposures, with risk increasing as dose of each exposure increases. Under this assumption, the probability of pneumoconiosis is a monotone function.

The data were collected for coal miners who had been employed only as coal getters on the coal face and haulage workers in the underground roadways. The exposures are defined as the length of time spent at these two types of work, \( t = (t_1, t_2)^T \), with records obtained on whether each miner developed pneumoconiosis. We let \( \text{pr}(y = 1|t) = \Phi\{F(t)\} \) which is the probability for a worker to develop pneumoconiosis under level \( t \), with \( F \) a real-valued bivariate monotone function and \( \Phi(\cdot) \) the standard normal cumulative distribution function. We give \( F \) a Gaussian process prior as described in the simulation examples, with the draws from the posterior projected to the constrained space.

We apply our method in estimating the monotone surface of response probability. The dose-response surface is estimated by projecting \( \Phi(w) \) where \( w \) is the posterior sample path of of a Gaussian process and \( \Phi(\cdot) \) is the Probit link function. The likelihood is given by the binomial model instead of the normal model. The estimated dose-response surface and its corresponding 95% pointwise credible intervals are plotted in Figure [5].
Table 2. $\sigma$-true normal error; $F$-true surface; $\bar{\sigma}$-posterior mean $\sigma$; $SD(\bar{\epsilon})$-standard deviation of posterior mean residuals; $\text{cor}(\epsilon, \bar{\epsilon})$-correlation between true and posterior mean residuals; $\text{cor}(y, \bar{y})$-correlation between true and posterior mean predicted responses; $\text{mse}$-mean squared error.

| $\sigma$ | $F_1$ | $\bar{\sigma}$ | $SD(\bar{\epsilon})$ | $\text{cor}(\epsilon, \bar{\epsilon})$ | $\text{cor}(y, \bar{y})$ | $\text{mse}$ |
|----------|-------|----------------|----------------------|-----------------------------|----------------------|-----------|
| 0.5      | 0.5031 | 0.5029 | 0.9962 | 0.9893 | 0.0014   |           |
| $F_2$    | 0.4993 | 0.4993 | 0.9998 | 0.9990 | 0.0004   |           |
| $F_3$    | 0.4997 | 0.4996 | 0.9946 | 0.9821 | 0.0016   |           |
| $F_4$    | 0.4937 | 0.4932 | 0.9745 | 0.9463 | 0.0035   |           |
| $F_5$    | 0.5113 | 0.5110 | 0.9767 | 0.9328 | 0.0035   |           |
| $F_6$    | 0.5014 | 0.5009 | 0.9879 | 0.9758 | 0.0025   |           |
| $F_7$    | 0.5034 | 0.5028 | 0.9976 | 0.9907 | 0.0011   |           |

| 0.1      | $F_1$ | 0.0997 | 0.0996 | 0.9909 | 0.9986 | 0.0004   |           |
| $F_2$    | 0.1050 | 0.1050 | 0.9987 | 0.9997 | 0.0002   |           |
| $F_3$    | 0.1008 | 0.1007 | 0.9859 | 0.9976 | 0.0005   |           |
| $F_4$    | 0.1205 | 0.1204 | 0.8457 | 0.9821 | 0.0020   |           |
| $F_5$    | 0.1145 | 0.1143 | 0.8554 | 0.9808 | 0.0019   |           |
| $F_6$    | 0.1045 | 0.1044 | 0.9384 | 0.9949 | 0.0010   |           |
| $F_7$    | 0.0997 | 0.0996 | 0.9977 | 0.9996 | 0.0002   |           |

Appendix 1

Lemma 2.1.

Proof. Let $x$ be any real number in $[0, 1]$. We see that

(1) $P_{w_1}(x) - P_{w_2}(x) = \inf_{v \geq x} \sup_{u \leq x} \frac{1}{v-u} \int_u^v w_1(t)dt - \inf_{v \geq x} \sup_{u \leq x} \frac{1}{v-u} \int_u^v w_2(t)dt.$

For each $\epsilon > 0$, there exists an element $v_0 \geq x$ such that

$\sup_{u \leq x} \frac{1}{v_0-u} \int_u^{v_0} w_2(t)dt < \inf_{v \geq x} \sup_{u \leq x} \frac{1}{v-u} \int_u^v w_2(t)dt + \epsilon,$

and hence

$-\inf_{v \geq x} \sup_{u \leq x} \frac{1}{v-u} \int_u^v w_2(t)dt < -\sup_{v \leq x} \frac{1}{v_0-u} \int_u^{v_0} w_2(t)dt + \epsilon.$

It follows from (1) that

(2) $P_{w_1}(x) - P_{w_2}(x) < \inf_{v \geq x} \sup_{u \leq x} \frac{1}{v-u} \int_u^v w_1(t)dt - \sup_{v \leq x} \frac{1}{v_0-u} \int_u^{v_0} w_2(t)dt + \epsilon.$

Note that

$\inf_{v \geq x} \sup_{u \leq x} \frac{1}{v-u} \int_u^v w_1(t)dt \leq \sup_{u \leq x} \frac{1}{v_0-u} \int_u^{v_0} w_1(t)dt$

for all $v \geq x$. In particular, this implies that the inequality above holds for $v = v_0$. Hence we see from (2) that

(3) $P_{w_1}(x) - P_{w_2}(x) < \sup_{v \leq x} \frac{1}{v_0-u} \int_u^{v_0} w_1(t)dt - \sup_{u \leq x} \frac{1}{v_0-u} \int_u^{v_0} w_2(t)dt + \epsilon.$
For each $\delta > 0$, there exists an element $u_0 \leq x$ such that
\[
\sup_{u \leq x} \frac{1}{v_0 - u} \int_u^{v_0} w_1(t)dt < \frac{1}{v_0 - u_0} \int_{u_0}^{v_0} w_1(t)dt + \delta.
\]
Since $u_0 \leq x$, we have
\[
\sup_{u \leq x} \frac{1}{v_0 - u} \int_u^{v_0} w_2(t)dt \geq \frac{1}{v_0 - u_0} \int_{u_0}^{v_0} w_2(t)dt.
\]
Thus it follows from (3) that
\[
P_{w_1}(x) - P_{w_2}(x) < \frac{1}{v_0 - u_0} \int_{u_0}^{v_0} w_1(t)dt - \frac{1}{v_0 - u_0} \int_{u_0}^{v_0} w_2(t)dt + \epsilon + \delta
\]
\[
< \frac{1}{v_0 - u_0} \int_{u_0}^{v_0} |w_1(t) - w_2(t)|dt + \epsilon + \delta
\]
\[
< \frac{1}{v_0 - u_0} \sup_{t \in [0,1]} |w_1(t) - w_2(t)|dt + \epsilon + \delta
\]
\[
< \sup_{t \in [0,1]} |w_1(t) - w_2(t)| + \tau.
\]
where $\tau = \epsilon + \delta$, and $\epsilon$ and $\delta$ are arbitrarily positive numbers. Therefore, for every $\tau > 0$ and $x \in [0, 1]$, we have

$$P_{w_1}(x) - P_{w_2}(x) < \sup_{t \in [0,1]} |w_1(t) - w_2(t)| + \tau.$$ 

Thus we see that

$$|P_{w_1}(x) - P_{w_2}(x)| < \sup_{t \in [0,1]} |w_1(t) - w_2(t)| + \tau$$

for every $\tau > 0$ and $x \in T$, and hence, for every $\tau > 0$, we have that

$$\sup_{x \in [0,1]} |P_{w_1}(x) - P_{w_2}(x)| < \sup_{t \in [0,1]} |w_1(t) - w_2(t)| + \tau.$$ 

Upon letting $\tau \to 0$ in the above inequality, the lemma follows. \hfill \Box

We first prove a lemma which is used in proving Theorem 2.1.

**Lemma 1.** Let $C_s$ be the cone of continuous functions $f(s,t)$ which are monotone with respect to $s$ for any $t$ and $C_t$ be the cone of continuous functions which are monotone with respect to $t$ for any $s$. Define their dual cones $C^*_s$ and $C^*_t$ as

$$C^*_s = \left\{ g(s,t) \in C[0,1]^2 : \int f(s,t)g(s,t)ds \leq 0, \text{ for all } t \text{ and } f \in C_s \right\},$$

$$C^*_t = \left\{ h(s,t) \in C[0,1]^2 : \int g(s,t)h(s,t)ds \leq 0, \text{ for all } s \text{ and } g \in C_t \right\}.$$
and

\[ C^*_t = \left\{ g(s,t) \in C[0,1]^2 : \int f(s,t)g(s,t)dt \leq 0, \text{ for all } s \text{ and } f \in C_t \right\}. \]

Denote \( P(w|C_s) \) as the projection of \( w \) over \( C_s \) by minimizing \( \int (w - f)^2 ds \) for all \( f \in C_s \) and any fixed \( t \). Denote \( P(w|C_t) \) as the projection of \( w \) over \( C_t \) by minimizing \( \int (w - f)^2 dt \) for all \( f \in C_t \) and any fixed \( s \). Then

\[ (4) \quad P(w|C^*_s) = w - P(w|C_s) \quad \text{and} \quad P(w|C^*_t) = w - P(w|C_t). \]

Furthermore, \( P(w|C^*_s) \) turns out to be the solution to the projection by minimizing \( \int (w - f)^2 ds dt \) over all \( f \in C_s \) and \( P(w|C^*_t) \) is the solution to the projection by minimizing \( \int (w - f)^2 ds dt \) over all \( f \in C_t \).

**Proof of Lemma** First note that \( P(w|C_s) \) is obtained by minimizing \( \int (w - f)^2 ds \) for all \( f \in C_s \) and any fixed \( t \). Then according to Theorem 1 of [Rychlik (2001)](http://example.com), one has \( \int \{w - P(w|C_s)\} f ds \leq 0 \) and \( \int P(w|C_s) \{w - P(w|C_s)\} ds = 0 \) by the properties of \( P(w|C_s) \). The first property implies \( w - P(w|C_s) \in C^*_s \). For any \( h \in C^*_s \), one has \( \int \{w - (w - P(w|C_s))\} h ds = \int P(w|C_s) h ds \leq 0 \). One can then deduce that \( P(w|C^*_s) = w - P(w|C_s) \). With a similar argument, one can show that \( P(w|C^*_t) = w - P(w|C_t) \).
Since \( \int \{ w - P(w|C_s) \}^2 ds \leq \int (w - f)^2 ds \) for any fixed \( t \), then one has \( \int \{ w - P(w|C_s) \}^2 ds dt \leq \int (w - f)^2 ds dt \). Therefore, \( P(w|C_s) \) minimizes \( \int (w - f)^2 ds dt \) for all \( f \in C_s \) and \( P(w|C_i^*) \) minimizes \( \int (w - f)^2 ds dt \) for all \( f \in C_t \) by the same argument.

Proof of Theorem 2.7: Define the norm \( ||f|| = \langle f, f \rangle^{1/2} = \left[ \int \{ f^2(s,t) \} ds dt \right]^{1/2} \) with \( \langle , \rangle \) denoting the inner product.

One has \(- S^{(k+1)} = (w + T^{(k)}) - \hat{w}^{(k)} = (w + T^{(k)}) - P(w + T^{(k)}|C_s) = P(w + T^{(k)}|C_s^*) \) where the last equality follows from Lemma A1. Here \( P(w + T^{(k)}|C_s) \) denotes the projection of \( w + T^{(k)} \) onto \( C_s \) and \( P(w + T^{(k)}|C_s^*) \) is the projection onto \( C_s^* \). Therefore \(- S^{(k+1)} \) minimizes \( \| (w + T^{(k)}) - f \| \) for all \( f \in C_s^* \) and \(- T^{(k)} \) minimizes \( \| (w + S^{(k)}) - f \| \) for all \( f \in C_i^* \). Then one concludes that

\[
\| \hat{w}^{(k)} \| = \| w + S^{(k)} - (-T^{(k-1)}) \| \geq \| w + S^{(k)} - (-T^{(k)}) \| \geq \| w + T^{(k)} - (-S^{(k+1)}) \|
\]

for all \( k \). Therefore, one has \( \| \hat{T}^{(k)} \| \geq \| \hat{w}^{(k)} \| \geq \| \hat{w}^{(k+1)} \| \). Now we wish to show that \( \{ S^{(k)} \} \) and \( \{ T^{(k)} \} \) are bounded and that \( \| S^{(k+1)} - S^{(k)} \| \to 0 \) and \( \| T^{(k+1)} - T^{(k)} \| \to 0 \) as \( k \to \infty \).

Assume that \( \{ S^{(k)} \} \) or \( \{ T^{(k)} \} \) is not bounded. Take an arbitrary large number \( M > 0 \). Then there exists an integer \( N \) such that \( |S^{(N)}| \geq M \) or \( |T^{(N)}| \geq M \). One then partitions \([0,1] \times [0,1]\) into \( m_1 m_2 \) squares of equal areas. The vertices of the squares are of the form \((s_i, t_j)\), where \( s_i = \frac{i - 1}{m_1} \) and \( t_j = \frac{j - 1}{m_2} \) with \( i = 1, \ldots, m_1 + 1 \) and \( j = 1, \ldots, m_2 + 1 \). Let \((s_{i_0}, t_{j_0})\) be the point such that \( |S^{(N)}| \geq M \) or \( |T^{(N)}| \geq M \) over the square between \((s_{i_0}, t_{j_0})\) and \((s_{i_0+1}, t_{j_0+1})\) for the first time with respect to the partial ordering. Without loss of generality, assume \( |S^{(N)}| \geq M \). For \( t = t_{j_0} \), let \( f \) be the monotone function such that \( f = -1 \) for \( s \leq s_{i_0}, f = 0 \) for \( s \geq s_{i_0+1} \) and \( f \) is linearly interpolated between \((s_{i_0}, t_{j_0})\) and \((s_{i_0+1}, t_{j_0+1})\). Note that for any \( \epsilon > 0 \), one can also partition \([0,1] \times [0,1]\) finely enough such that \( \sum_{i=1}^{m_1} S(s_i, t_{j_0}) f(s_i, t_{j_0}) - \int S(s, t_{j_0}) f(s, t_{j_0}) ds < \epsilon \) and \( \sum_{j=1}^{m_2} S(s_{i_0}, t_j) f(s_{i_0}, t_j) - \int S(s_{i_0}, t) f(s_{i_0}, t) dt < \epsilon \). By the properties of the dual cones and our construction of the function \( f \), one can partition \([0,1] \times [0,1]\) finely enough such that \( \sum_{i=1}^{m_1} S(s_i, t_{j_0}) \leq 0 \), which implies that \( S^{(N)} \leq -M \) in the square between \((s_{i_0}, t_{j_0})\) and \((s_{i_0+1}, t_{j_0+1})\) up to an arbitrary small number \( \epsilon \). Since the norm of \( \hat{w}^{(N)} = w + S^{(N)} + T^{(N)} \) is bounded, it follows from the continuity of the estimates that \( T^{(N)} \geq M \) up to an arbitrary small number \( \epsilon \) over the square of \((s_{i_0}, t_{j_0})\). On the other hand, we know that \( \sum_{j=1}^{m_2} T(s_{i_0}, t_j) \leq 0 \), which contradicts the fact that \( T^{(N)} \geq M \) and \((s_{i_0}, t_{j_0})\) is the point such that \( |S^{(N)}| \geq M \) or \( |T^{(N)}| \geq M \) over the square between \((s_{i_0}, t_{j_0})\) and \((s_{i_0+1}, t_{j_0+1})\) for the first time.

Since \( \{ S^{(k)} \} \) and \( \{ T^{(k)} \} \) are bounded, then there exists convergent subsequences indexed by \( n_i \) such that \( \{ S^{(n_i)} \} \to S \) and \( \{ T^{(n_i)} \} \to T \). Then one has for \( n_i \to \infty \),

\[
(5) \quad \hat{w}^{(n_i)} = w + S^{(n_i)} + T^{(n_i)} \to w + S + T; \quad \hat{w}^{(n_i+1)} = w + S^{(n_i+1)} + T^{(n_i)} \to w + S + T.
\]

Denote the limit as \( w_L = w + S + T \). One claims that \( w_L \) is the projection of \( w \) which is the solution to (2.1) under the partial ordering constraint. First note that \( w_L \in C_s \) since \( \hat{w}^{(k)} \in C_s \) and \( w_L \in C_t \) since \( \hat{w}^{(k)} \in C_t \). This implies that \( w_L \in C_s \cap C_t \) which says that
Proof of Theorem 5. \( w_L \) is monotone with respect to the partial ordering on \((s, t)\). Now,
\[
\langle w - w_L, w_L \rangle = \langle w - w_L + S, w_L \rangle - \langle S, w_L \rangle = \lim_{n_i \to \infty} \langle w + S(n_i) - \hat{w}(n_i), w_L(n_i) \rangle + \lim_{n_i \to \infty} \langle w + T(n_i) - \hat{w}(n_i+1), \hat{w}(n_i+1) \rangle = 0 + 0 = 0.
\]
Let \( h \) be any element in \( C_s \cap C_t \), one looks at
\[
\langle w - w_L, h \rangle = \langle w - w_L + S, h \rangle - \langle S, h \rangle = \lim_{n_i \to \infty} \langle w + S(n_i) - \hat{w}(n_i), h \rangle + \lim_{n_i \to \infty} \langle w + T(n_i) - \hat{w}(n_i+1), h \rangle \leq 0 + 0 = 0.
\]
Then by Theorem 1 in [Rychlik (2001)], \( w_L \) is indeed the projection of \( w \). Now we will show that \( \|S^{(i+1)} - S^{(i)}\|^2 \to 0 \) and \( \|T^{(i+1)} - T^{(i)}\|^2 \to 0 \) with which we can conclude that \( S^{(k)} \to S \), \( T^{(k)} \to T \) and both \( \hat{w}^{(k)} \) and \( \hat{w}^{(k)} \) converge to \( w_L \). First by the projection property, one can show that
\[
\|T^{(i)} - T^{(i-1)}\|^2 = \|w + T^{(i)} - (w + T^{(i-1)})\|^2 \geq \|S^{(i+1)} - S^{(i)}\|^2 = \|w + S^{(i+1)} - (w + S^{(i)})\|^2 \geq \|T^{(i+1)} - T^{(i)}\|^2.
\]
Therefore, \( S^{(i+1)} - S^{(i)} \) and \( T^{(i+1)} - T^{(i)} \) converge to the same limit. Assume on the contrary that \( \|S^{(i+1)} - S^{(i)}\|^2 \) does not converge to zero. Then over some Lebesgue measure non-zero set \( U \), there exists \( \epsilon > 0 \) such that for \((s, t) \in U\)
\[
|S^{(i+1)} - S^{(i)}| > \epsilon \quad \text{for infinitely many } i.
\]
Now look at
\[
\|S^{(i+1)} - S^{(i)}\|^2 - \|T^{(i+1)} - T^{(i)}\|^2 = \|w + T^{(i)} + S^{(i)} - (w + T^{(i+1)} + S^{(i+1)})\|^2 + 2\langle w + T^{(i)} + S^{(i)} - (w + T^{(i+1)} + S^{(i+1)}), S^{(i+1)} - S^{(i)} \rangle.
\]
Note that \( \|S^{(i+1)} - S^{(i)}\|^2 - \|T^{(i+1)} - T^{(i)}\|^2 \to 0 \) and the last term on the right hand side of the above equation is non-negative. Therefore, one can conclude that
\[
(T^{(i+1)} - T^{(i)}) - (S^{(i+1)} - S^{(i)}) \to 0.
\]
By a similar argument, one has
\[
(T^{(i+2)} - T^{(i+1)}) - (S^{(i+2)} - S^{(i)}) \to 0.
\]
Subtracting (7) from (8), one has
\[
(T^{(i+2)} - T^{(i+1)}) - (T^{(i+1)} - T^{(i)}) \to 0.
\]
This implies that there exist \( i, j \) large enough with \( |i - j| \) finite such that \( T^{(i+1)} - T^{(i)} \) can be made arbitrarily close to \( T^{(j+1)} - T^{(j)} \). However this contradicts (4) and the fact that \( \{T^{(k)}\} \) is bounded such that there exists constant \( C \) such that \( |T^{(i)} - T^{(j)}| < C \) for all \( i, j \). By the same argument, one can show that \( \|T^{(i+1)} - T^{(i)}\|^2 \to 0 \). Therefore, \( S^{(k)} \to S \), \( T^{(k)} \to T \) which implies \( \hat{w}^{(k)} \to w_L \) and \( \hat{w}^{(k)} \to w_L \).

The inequality in the Theorem can be shown combining Lemma 2.3 and the properties of the projection.

\[ \blacksquare \]

Proof of Theorem 5. We wish to find a probability \( \Pi \), say, on the space \( \Gamma \) (thought of as a ‘prior’ for \((F, \sigma)\), but which may depend on the data \( y_i \)) such that the projection of the posterior \( \Pi(-|y_1, \ldots, y_n) \) of a Gaussian process is the posterior on \( \Gamma \) with prior \( \Pi \). Let
\( \Pi(dF, d\sigma | y_1, \ldots, y_n) \) denote this probability. Since the (conditional) density of the observations \( y_1, \ldots, y_n \), given \( (F, \sigma) \), is the joint Normal density as before, say \( f(y_1, \ldots, y_n | F, \sigma) \), one needs to have \( \Pi(dF, d\sigma | y_1, \ldots, y_n) \) satisfy

\[
\Pi(dF, d\sigma | y_1, \ldots, y_n) \int_{\Gamma} f(y_1, \ldots, y_n | F, \sigma) \Pi(dF, d\sigma) = f(y_1, \ldots, y_n | F, \sigma) \Pi(dF, d\sigma).
\]

Let

\[
g(F, \sigma | y_1, \ldots, y_n) = \frac{1}{f(y_1, \ldots, y_n | F, \sigma)} \left\{ \int_{\Gamma} f(y_1, \ldots, y_n | F, \sigma)^{-1} \Pi(dF, d\sigma | y_1, \ldots, y_n) \right\}^{-1},
\]

which is well-defined since \( \int_{\Gamma} f(y_1, \ldots, y_n | F, \sigma)^{-1} \Pi(dF, d\sigma | y_1, \ldots, y_n) < \infty \). First note that \( g \) is a density on \( \Gamma \) with respect to the measure \( \Pi(dF, d\sigma | y_1, \ldots, y_n) \) since one can easily check that \( \int_{\Gamma} g(F, \sigma | y_1, \ldots, y_n) \Pi(dF, d\sigma | y_1, \ldots, y_n) = 1 \). Define

\[
\Pi(dF, d\sigma) = g(F, \sigma | y_1, \ldots, y_n) \Pi(dF, d\sigma | y_1, \ldots, y_n).
\]

We will show that \( \Pi(dF, d\sigma) \) satisfies equation (9) above which is equivalent to showing

\[
\int_{\Gamma} f(y_1, \ldots, y_n | F, \sigma) \Pi(dF, d\sigma) = f(y_1, \ldots, y_n | F, \sigma) g(F, \sigma | y_1, \ldots, y_n).
\]

One has

\[
\int_{\Gamma} f(y_1, \ldots, y_n | F, \sigma) \Pi(dF, d\sigma)
\]

\[
= \int_{\Gamma} f(y_1, \ldots, y_n | F, \sigma) g(F, \sigma | y_1, \ldots, y_n) \Pi(dF, d\sigma | y_1, \ldots, y_n)
\]

\[
= \int_{\Gamma} \left\{ \int_{\Gamma} f(y_1, \ldots, y_n | F, \sigma)^{-1} \Pi(dF, d\sigma | y_1, \ldots, y_n) \right\}^{-1} \Pi(dF, d\sigma | y_1, \ldots, y_n)
\]

\[
= \left\{ \int_{\Gamma} f(y_1, \ldots, y_n | F, \sigma)^{-1} \Pi(dF, d\sigma | y_1, \ldots, y_n) \right\}^{-1} = f(y_1, \ldots, y_n | F, \sigma) g(F, \sigma | y_1, \ldots, y_n).
\]

Then our contention follows. \( \square \)

Acknowledgement

Lizhen Lin thanks Professor Rabi Bhattacharya for useful discussions. This work was supported by grant R01ES017240 from the National Institute of Environmental Health Sciences (NIEHS) of the National Institute of Health.

References

[Anevksi & Soulier (2011)] Anevksi, D. & Soulier, P. (2011). Monotone spectral density estimation. Ann. Statist. 39(1), 418-438.

[Ashford & Smith (1964)] Ashford, J.R. & Smith, C.S. (1964). General models for quantal response to the joint action of a mixture of drugs. Biometrika 51, 3 and 4, 413-428.

[Barlow et al. (1972)] Barlow, R. E., Bartholomew, D., Bremner, J. M. & Brunk, H. D. (1972). Statistical Inference under Order Restrictions; the Theory and Application of Isotonic Regression. Wiley, New York.

[Bhattacharya & Kong (2007)] Bhattacharya, R. & Kong, M. (2007). Consistency and asymptotic normality of the estimated effective dose in bioassay. J. Statist. Plann. Inference 137, 643-658.

[Bhattacharya & Lin (2010)] Bhattacharya, R.N., & Lin, L. (2010). An adaptive nonparametric method in benchmark analysis for bioassay and environmental studies. Statist. Probab. Lett. 80, 1947–1953.
[Bhattacharya & Lin (2011)] Bhattacharya, R.N & Lin, L. (2011). Nonparametric benchmark analysis in risk assessment: a comparative study by simulation and data analysis. Sankhyā, Ser. B 73, 144–163.

[Bornkamp & Ickstadt (2009)] Bornkamp, B. & Ickstadt, K. (2009). Bayesian nonparametric estimation of continuous monotone functions with applications to dose-response analysis. Biometrics 65(1), 198-206.

[Dette et al. (2005)] Dette, H., Neumeyer, N. & Pliz, K.F. (2005). A note on nonparametric estimation of the effective dose in quantal bioassay. J. Amer. Statist. Assoc. 100, 503-510.

[Eggermont & Lariccia (2000)] Eggermont, P.P.B. & Lariccia, V. N. (2000). Maximum likelihood estimation of smooth monotone and unimodal densities. Ann. Statist. 28(3), 922-947.

[Eghbal & Kuo (1991)] Gelfand, A.E. & L. Kuo (1991). Nonparametric Bayesian bioassay including ordered polytomous response. Biometrika 78, 657–666.

[Ghosal & van der Vaart (2007)] Ghosal, S. & Van der Vaart, A.A. (2007). Convergence rates of posterior distributions for noniid observations. Ann. Statist. 35(1) 192-223.

[Ghosal & Roy (2006)] Ghosal, S. & Roy, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. Ann. Statist. 34, 2413–2429.

[Ghosh & Ramamoorthi (2002)] Ghosh, J.K. & Ramamoorthi, R.V. (2002). Bayesian Nonparametrics. Springer Series in Statistics. Springer, New York.

[Groeneboom & Jongbloed (2010)] Groeneboom, P. & Jongbloed, G. (2010). Generalized continuous isotonic regression. Statist. Probab. Lett. 80, 248-253.

[Holmes & Heard (2003)] Holmes, C.C. & Heard, N.A. (2003). Generalised monotonic regression using random change points. Stat. Med. 22, 623–638.

[Kong & Eubank (2006)] Kong, M. & Eubank, R.L. (2006) Monotone smoothing with application to dose-response curve. Comm. Statist. Simulation Comput. 35(4), 991-1004.

[Mammen (1991)] Mammen, E (1991). Estimating a smooth monotone regression function. Ann. Statist. 19(2), 724–740.

[Müller & Schmitt (1988)] Müller, H.G. & Schmitt, T. (1988). Kernel and probit estimation in quantal bioassay. J. Amer. Statist. Assoc. 83(403), 750-759.

[Neelon & Dunson (2004)] Neelon, B. & Dunson, D. B. (2004). Bayesian isotonic regression and trend analysis. Biometrics 60, 177–191.

[Ramsay (1988)] Ramsay, J. O. (1988) Monotone regression splines in action. Statist. Sci. 3(4), 425-441.

[Rychlik (2001)] Rychlik, T. (2001). Projecting Statistical Functionals. Lecture Notes in Statistics. Springer.

[Robertson et al. (1988)] Robertson, T., Wright, F. T. & Dykstra, R. L. (1988). Order Restricted Statistical Inference. Wiley, New York.

[Saarela & Arjas (2011)] Saarela, O. & Arjas, E. (2011). A method for Bayesian monotonic multiple regression. Scand. J. Stat. 38, 499-513.

[Shively et al. (2009)] Shively, T.S., Sager, T. W., & Walker, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. J. Roy. Statist. Soc. Ser. B 71(1), 159-175

[Shively et al. (2011)] Shively, T.S., Walker, S.G., & Damien, P. (2011.) Nonparametric function estimation subject to monotonicity, convexity and other shape constraints. J. Econometrics 161 (2), 166–181.

[van der Vaart & van Zanten (2007)] Van der Vaart, A. W. & Van Zanten, J. H. (2007). Bayesian inference with rescaled Gaussian process priors. Electron. J. Stat. 1, 433–448.

[van der Vaart & van Zanten (2008)] Van der Vaart, A. W. & Van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. Ann. Statist. 36(3), 1435–1463.

[van der Vaart & van Zanten (2009)] Van der Vaart, A. W. & Van Zanten, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth. Ann. Statist. 37(5B), 2655–2675.

[Walker & Hjort (2001)] Walker, S.G. & Hjort, N.L. (2001). On Bayesian consistency. J. Roy. Statist. Soc. Ser. B 63, 811–821.

E-mail address: lizhen@stat.duke.edu

E-mail address: dunson@stat.duke.edu

DEPARTMENT OF STATISTICAL SCIENCE, DUKES UNIVERSITY, DURHAM, NC 27708-0251, USA.