

Isotonic smoothing splines under sequential designs

Chim Tantiyaswasdikul

Department of Information Planning and Operations, Bank of Thailand, Bangkok, Thailand.

Michael B. Woodroffe*

University of Michigan, Ann Arbor, MI, USA

Received 17 July 1992; revised manuscript received 17 July 1992

Abstract

Isotonic smoothing splines are introduced as a natural extension of ordinary isotonic estimates in the estimation of a regression function $\mu(x) = E\{Y|X=x\}$. A constructive characterization for the isotonic smoothing splines is given. Conditions are given for consistency under sequential designs, where the observation points are random and the experimenter after observing X_k may choose to observe or to skip observing Y_k .

Key words and phrases: Nonparametric; isotonic regression; sequential design; consistency.

1. Introduction

Since they are necessarily step functions, isotonic regression estimates are not entirely satisfactory when the experimenter is confident that the underlying regression function is smooth. Neither can they provide good interpolations between observation points. A brief literature survey of approaches to resolve these difficulties may be found in Chapter 9 of Robertson et al. (1988). These include the ideas of using moving averages by Friedman and Tibshirani (1984) and of using kernel estimators by Mukerjee (1988). But the method that seems most natural is the smoothing splines. Isotonic smoothing splines can be viewed as a generalization of the isotonic estimates. The isotonic estimates, step functions, may be regarded as isotonic smoothing splines of the first order. But the problem of isotonic smoothing splines or restricted splines in general is nonlinear and is difficult to solve. Utreras (1986) gives the existence,

Correspondence to: M.B. Woodroffe, Dept. of Statistics, University of Michigan, Ann Arbor, MI 48109-1027, USA.

*Research supported by the National Science Foundation.

characterization and convergence rates for isotonic smoothing splines. But a satisfactory algorithm to construct them is yet to be found. In this study, we will investigate the simplest case: isotonic smoothing splines which are piecewise linear, seeking for improvement over the method of ordinary isotonic estimates.

The estimators are presented in Section 2. The algorithm leaves one (real) variable to be found numerically, but is otherwise explicit. The consistency of the estimators is studied in Section 3 for the case in which data are gathered according to a sequential design. The class of designs considered is motivated by the work of Sarkar(1991) and Woodroffe (1979) on the one-armed bandit problem with covariates. For a detailed study of these models, using ordinary isotonic estimators, see Tantiyaswasdikul (1992).

2. Derivation

A Sobolev space. Let $C=C[a, b]$ denote the space of continuous functions on $[a, b]$ endowed with the sup norm $\|\cdot\|_\infty$; let $L_2=L_2[a, b]$ denote the space of (Lebesgue) square integrable functions, endowed with its usual norm $\|\cdot\|_2$ and inner product $(\cdot, \cdot)_2$; let $H=H[a, b]$ denote the space of all absolutely continuous functions $f:[a, b] \mapsto \mathbb{R}$ for which $f' \in L_2$, endowed with the norm

$$\|f\|_H^2 = [f(a)]^2 + \|f'\|_2^2$$

and inner product

$$(f, g)_H = f(a)g(a) + (f', g')_2.$$

Let $H_+ = \{f \in H : f' \geq 0 \text{ a.e.}\}$. Then H is a Banach space and H_+ is a weakly closed convex cone in H .

The criterion function. Let $\omega \neq 0$ be a finite measure on the Borel sets of $[a, b]$. For a given $\alpha > 0$ and $g \in L_2(\omega)$, let

$$\psi(f) = \int (f-g)^2 d\omega + \alpha \int_a^b (f')^2 dx,$$

for all $f \in H$. The problem is to find a function $f \in H_+$ which minimizes ψ in H_+ ; that is,

$$\psi(f) = \min_{h \in H_+} \psi(h).$$

Such an f is called an isotonic smoothing spline.

The existence of an $f \in H_+$ which minimizes ψ may be easily verified by checking that ψ is weakly lower semi continuous and that $\{h : \psi(h) \leq \psi(0)\}$ is (norm) bounded and, therefore, weakly compact. Existence then follows since lower semi-continuous functions attain their minima on compact sets (cf. Wegman, 1984).

Characterization. The algorithm is based on the following three elementary lemmas.

Lemma 1. *If $f, h \in H$, then*

$$\psi(f + \varepsilon h) - \psi(f) = 2\varepsilon\psi_1(f, h) + O(\varepsilon^2), \tag{1}$$

as $\varepsilon \rightarrow 0$, where

$$\psi_1(f, h) = \int_a^b (f - g)h \, d\omega + \alpha \int_a^b f' h' \, dx. \tag{2}$$

Proof. For $f, h \in H_+$,

$$\psi(f + \varepsilon h) - \psi(f) = 2\varepsilon\psi_1(f, h) + \varepsilon^2 \left\{ \int_a^b h^2 \, d\omega + \alpha \int_a^b (h')^2 \, dx \right\}. \quad \square$$

For each $f \in H_+$, let

$$H_f = \{h \in H : f + h \in H_+\} = \{h \in H : h' \geq -f'\}.$$

Lemma 2. *$f \in H_+$ minimizes ψ on H_+ if, and only if*

$$\psi_1(f, h) \geq 0 \quad \forall h \in H_f, \tag{3}$$

Proof. If f minimizes ψ on H_+ then (3) follows since $\psi(f + \varepsilon h) - \psi(f) \geq 0$ for all $h \in H_f$ and $0 < \varepsilon < 1$.

Conversely, suppose $\psi_1(f, h) \geq 0 \quad \forall h \in H_f$. If $f_1 \in H_+$, then $h = f_1 - f \in H_f$ and

$$\psi(f_1) - \psi(f) = \Psi(1) - \Psi(0),$$

where $\Psi(\varepsilon) = \psi(f + \varepsilon h) \quad \forall 0 \leq \varepsilon \leq 1$. Here Ψ is a convex function for which

$$\left. \frac{d}{d\varepsilon} \Psi(\varepsilon) \right|_{\varepsilon=0+} = 2\psi_1(f, h) \geq 0,$$

by Lemma 1 and the assumption. So $\Psi'(\varepsilon+) \geq \Psi'(0+) \geq 0, \quad \forall 0 \leq \varepsilon \leq 1$ and therefore $\psi(f_1) \geq \psi(f)$. That is, f minimizes ψ on H_+ . \square

Letting $h = \pm 1$ in (2) and appealing to Lemma 2 shows that if f minimizes ψ on H_+ , then

$$\int_a^b (f - g) \, d\omega = 0. \tag{4}$$

For $a \leq x \leq b$ let

$$F(x) = \int_a^x f(y) \omega(dy) \quad \text{and} \quad G(x) = \int_a^x g(y) \omega(dy)$$

Lemma 3. If $f \in H$ satisfies (4), then for all $h \in H$,

$$\psi_1(f, h) = \int_a^b [\alpha f' - (F - G)] h' dx$$

Proof. If f satisfies (4), then for all $h \in H$,

$$\begin{aligned} \psi_1(f, h) &= \int_a^b [f(x) - g(x)] [h(x) - h(b)] \omega(dx) + \alpha \int_a^b f' h' dx \\ &= - \int_a^b [f(x) - g(x)] \int_x^b h'(y) dy \omega(dx) + \alpha \int_a^b f' h' dx \\ &= - \int_a^b \left\{ \int_a^y [f(x) - g(x)] \omega(dx) \right\} h'(y) dy + \alpha \int_a^b f' h' dx \\ &= \int_a^b [\alpha f' - (F - G)] h' dx. \quad \square \end{aligned}$$

Theorem 1. Necessary and sufficient conditions for $f \in H_+$ to minimize ψ on H_+ are that (4) holds and

$$\alpha f' = (F - G)_+ \quad \text{a.e.}, \quad (5)$$

where $(F - G)_+ = \max\{0, F - G\}$.

Proof. That (4) is necessary has already been observed. If f minimizes ψ on H_+ , then by Lemmas 2 and 3

$$0 \leq \psi_1(f, h) = \int_a^b [\alpha f' - (F - G)] h' dx, \quad \forall h \in H_+,$$

and, since h' may be any nonnegative square integrable function, it follows that $\alpha f' - (F - G) \geq 0$ a.e. (by letting h' be the indicator of the set where $\alpha f' - (F - G) < 0$). Since $f' \geq 0$ a.e., it follows that $\alpha f' \geq (F - G)_+$ a.e. So, it suffices to show that $\alpha f' \leq (F - G)_+$ a.e. For this, let

$$B = \{x \in [a, b]: \alpha f'(x) > [F(x) - G(x)]_+\}.$$

Then $f' > 0$ and $\alpha f' - (F - G) > 0$ a.e. on B . As above, there is an $h \in H_f$ for which $h' = -\frac{1}{2} f' I_B$. Then

$$\begin{aligned} 0 \leq \psi_1(f, h) &\leq \int_a^b [\alpha f' - (F - G)] h' dx \\ &= -\frac{1}{2} \int_B [\alpha f' - (F - G)] f' dx, \end{aligned}$$

which is negative, unless B is of measure 0.

Conversely, suppose (4) and (5) are satisfied. Then

$$\psi_1(f, h) = \int_a^b [\alpha f' - (F - G)] h' dx \geq 0 \quad \forall h \in H_f,$$

using Lemma 3 for the equality and (5) for the inequality. It follows from Lemma 2 that f Minimizes ψ on H_+ . \square

A special case. Let $a = x_0 < x_1 < \dots < x_n < x_{n+1} = b$; and suppose that ω is the counting measure on $\{x_1, \dots, x_n\}$. Then, letting

$$f_k = f(x_k), \quad g_k = g(x_k),$$

and

$$F_k = f_1 + \dots + f_k, \quad G_k = g_1 + \dots + g_k$$

for all $k=0, \dots, n$, where an empty sum is to be interpreted as zero (so that $F_0 = G_0 = 0$), one finds that $F(x) - G(x) = F_k - G_k$ for all $x_k \leq x < x_{k+1}$ and for all $k=0, \dots, n$. Thus, for the minimizing f

$$\alpha f'(x) = (F_k - G_k)_+ \quad \forall x_k \leq x < x_{k+1} \quad \forall k=0, \dots, n.$$

So, f' is constant on each of the intervals $[x_k, x_{k+1})$, $k=0, \dots, n$. Let

$$f'_k = f'(x_k +) \quad \forall k=0, \dots, n.$$

Then

$$f_k = f_{k-1} + f'_{k-1} \cdot (x_k - x_{k-1}) \quad \forall k=1, \dots, n,$$

and

$$\alpha f'_k = (F_k - G_k)_+ \quad \forall k=0, \dots, n.$$

In particular, $f'_0 = 0$, so that $f_1 = f_0$. This determines f_1, \dots, f_n in terms of f_0 ; and f_0 may be determined from

$$F_n = f_1 + \dots + f_n = \int_a^b f d\omega = \int_a^b g d\omega = g_1 + \dots + g_n = G_n. \tag{6}$$

Lemma 4.

$$\min_{1 \leq k \leq n} \frac{G_k}{k} \leq f_0 \leq \frac{G_n}{n}.$$

Proof. If $f_0 < \min_{1 \leq k \leq n} G_k/k$, then $F_1 = f_1 = f_0 < G_1$, so that $f'_1 = (F_1 - G_1)_+/\alpha = 0$ and, therefore, $F_2 = f_1 + f_2 = 2f_0 < G_2$. Continuing, one finds that $F_n = nf_0 < G_n$ contradicting (6). Therefore

$$f_0 \geq \min_{1 \leq k \leq n} \frac{G_k}{k}.$$

Conversely, if $f_0 > G_n/n$, then $F_n = f_1 + \dots + f_n \geq nf_0 > G_n$, again contradicting (6). \square

Let \hat{f}_0 be a guess for the value of f_0 and let

$$\hat{f}_k = \hat{f}_{k-1} + \frac{(\hat{F}_{k-1} - G_{k-1})_+}{\alpha} (x_k - x_{k-1}),$$

for $k = 1, \dots, n$, where $\hat{F}_0 = 0$ and $\hat{F}_k = \hat{f}_1 + \dots + \hat{f}_k$, $k = 1, \dots, n$.

Lemma 5. *If $\hat{f}_0 < f_0$ then $\hat{F}_n < G_n$, and if $\hat{f}_0 > f_0$ then $\hat{F}_n > G_n$.*

Proof. Similar to that of Lemma 4. \square

Lemma 4 provides lower and upper bounds for the value of f_0 and Lemma 5 provides an indication of whether an estimate of f_0 is too low or too high. The two lemmas justify a simple bisection method for determining the value of f_0 .

Example. Consider the model

$$Y_k = \mu(X_k) + \varepsilon_k,$$

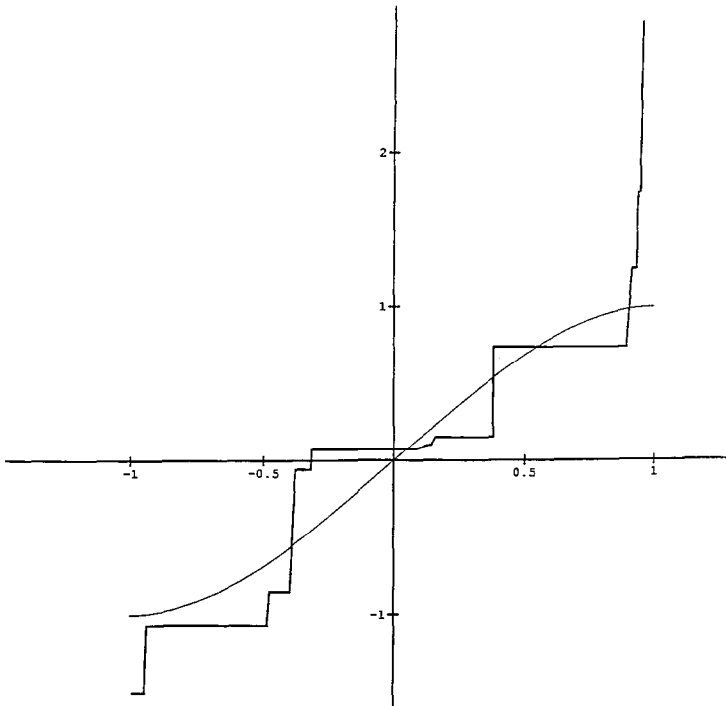


Fig. 1. The isotonic estimator.

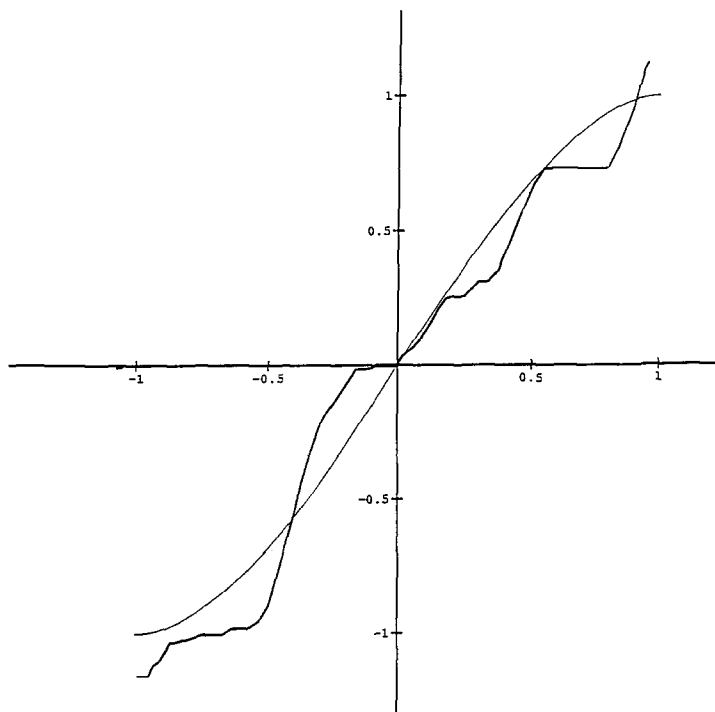


Fig. 2. An isotonic smoothing spline.

where X_k , $k=1, 2, \dots$, are independent and uniformly distributed over $(-1, 1)$, ε_k , $k=1, 2, \dots$, are independent standard normal random variables, and

$$\mu(x) = \frac{3}{2}x - \frac{1}{2}x^3, \quad -1 < x < 1.$$

This function was chosen to be nearly flat near the endpoints of the interval. The unsmoothed isotonic estimator has difficulty with such functions.

The unsmoothed isotonic estimator ($\alpha=0$) and the isotonic smoothing spline with $\alpha=1$ are compared in Figures 1 and 2.

Just a little smoothing ($\alpha=1$) has a dramatic effect on the estimators.

3. Consistency under sequential designs

A covariate model. Let (X_k, Y_k) , $k=1, 2, \dots$, be a sequence of independent bivariate random variables identically distributed as (X, Y) where

1. X has a known distribution F with support $[a, b]$,
2. the conditional distribution of Y given $X=x$ has mean $\mu(x) = E(Y|X=x)$ and variance $\sigma^2(x) = E\{[Y - \mu(x)]^2 | X=x\}$,

3. μ is continuous and nondecreasing on $[a, b]$.

Suppose that after observing X_k , we may choose to observe or not to observe Y_k for all $k = 1, 2, \dots$. This model is closely related to ones considered by Sarkar (1991) and Woodroffe (1979) in their study of one-armed bandit problems with covariates.

By a decision rule, we mean a sequence of binary-valued functions $\delta = \{\delta_k, k = 1, 2, \dots\}$ where $\delta_k = 0$ or 1 corresponds to the choice of not observing or observing Y_k , as a function of information available at stage k , i.e.,

$$\mathcal{F}_{k-1}^\delta = \sigma(X_j, \delta_j, \delta_j Y_j, j = 1, \dots, k-1),$$

and the value of X_k . An interesting class of decision rules are naturally those based on the estimates of μ constructed from the information contained in \mathcal{F}_{k-1}^δ .

For a fixed decision rule $\delta = \{\delta_k, k = 1, 2, \dots\}$, define the isotonic smoothing spline for each $n = 1, 2, \dots$ as the function $m_n \in H_+$ which minimizes

$$\begin{aligned} \psi_n(f) &= \sum_{k=1}^n \delta_k [f(X_k) - Y_k]^2 + \alpha_n \int_a^b [f'(x)]^2 dx \\ &= \sum_{j=1}^{K_n} [f(X_{n,j}) - Y_{n,j}]^2 + \alpha_n \int_a^b [f'(x)]^2 dx, \end{aligned}$$

where $X_{n,1} \leq \dots \leq X_{n,K_n}$ denote the ordered values of X_1, \dots, X_n , $Y_{n,1}, \dots, Y_{n,K_n}$ denote the concomitant order statistics, and $\alpha_n > 0$. For $a \leq x \leq b$, let

$$\begin{aligned} G_n(x) &= \sum_{k=1}^n \delta_k Y_k I\{X_k \leq x\}, \\ M_n(x) &= \sum_{k=1}^n \delta_k \mu(X_k) I\{X_k \leq x\}. \end{aligned}$$

Proposition 1. Suppose that $\int_a^b \sigma^2(x) dF(x) < \infty$. If δ is a decision rule, then

$$\sup_{a \leq t \leq b} \frac{|G_n(t) - M_n(t)|}{n^{(1/2) + \beta}} \xrightarrow{a.s.} 0, \tag{7}$$

as $n \rightarrow \infty$ for all $\beta > 0$.

Proof. See Tantiyaswasdikul (1992). \square

In the following theorem, O_e denotes exact order.

Theorem 2. Let $\mu \in H_+$; suppose that $\int_a^b \sigma^2(x) dF(x) < \infty$; and let δ be a fixed decision rule. If there are $p > \frac{1}{2}$, and $1 - p < q < p$ for which $K_n > n^p$ for almost every n w.p.1 and $\alpha_n = O_e(n^q)$ as $n \rightarrow \infty$, then

$$\frac{1}{n^p} \sum_{k=1}^n \delta_k [m_n(X_k) - \mu(X_k)]^2 \xrightarrow{a.s.} 0 \quad \text{as } n \rightarrow \infty.$$

Proof. Since m_n minimizes ψ_n over H_+ and $\mu \in H_+$,

$$\begin{aligned} & \sum_{k=1}^n \delta_k [m_n(X_k) - Y_k]^2 + \alpha_n \int_a^b [m'_n(x)]^2 dx \\ & \leq \sum_{k=1}^n \delta_k [\mu(X_k) - Y_k]^2 + \alpha_n \int_a^b [\mu'(x)]^2 dx. \end{aligned}$$

By simple rearrangements

$$\begin{aligned} & \sum_{k=1}^n \delta_k [m_n(X_k) - \mu(X_k)]^2 + \alpha_n \int_a^b [m'_n(x)]^2 dx \\ & \leq 2 \sum_{k=1}^n \delta_k [m_n(X_k) - \mu(X_k)] [Y_k - \mu(X_k)] + \alpha_n \int_a^b [\mu'(x)]^2 dx. \end{aligned} \tag{8}$$

Summation by parts yields

$$\begin{aligned} \sum_{k=1}^n \delta_k [m_n(X_k) - \mu(X_k)] [Y_k - \mu(X_k)] &= \sum_{j=1}^{K_n} [m_n(X_{n,j}) - \mu(X_{n,j})] [Y_{n,j} - \mu(X_{n,j})] \\ &= A_n + B_n, \end{aligned}$$

where

$$\begin{aligned} A_n &= [G_n(b) - M_n(b)] [m_n(X_{n,\kappa_n}) - \mu(X_{n,\kappa_n})], \\ B_n &= \sum_{j=1}^{K_n-1} [G_n(X_{n,j}) - M_n(X_{n,j})] \\ & \quad \times [\mu(X_{n,j+1}) - \mu(X_{n,j}) - m_n(X_{n,j+1}) + m_n(X_{n,j})]. \end{aligned}$$

Let

$$D_n = \sup_{a \leq x \leq b} |G_n(x) - M_n(x)|.$$

By Lemma 4, $m_n(X_{n,1}) \leq \sum_{j=1}^{K_n} Y_{n,j} / K_n$; and since μ is nondecreasing, $\mu(X_{n,\kappa_n}) \geq \sum_{j=1}^{K_n} \mu(X_{n,j}) / K_n$, so that

$$\begin{aligned} m_n(X_{n,1}) - \mu(X_{n,\kappa_n}) &\leq \frac{1}{K_n} \sum_{j=1}^{K_n} [Y_{n,j} - \mu(X_{n,j})] \\ &\leq \frac{1}{K_n} [G_n(b) - M_n(b)] \leq \frac{D_n}{K_n}. \end{aligned}$$

Therefore,

$$m_n(X_{n,\kappa_n}) - \mu(X_{n,\kappa_n}) \leq m_n(X_{n,\kappa_n}) - m_n(X_{n,1}) + \frac{D_n}{K_n},$$

and

$$\begin{aligned} |A_n| &\leq D_n \left[m_n(X_{n, K_n}) - m_n(X_{n, 1}) + \frac{D_n}{K_n} \right] \\ &= D_n \int_a^b m'_n(x) dx + \frac{D_n^2}{K_n}. \end{aligned}$$

Also

$$\begin{aligned} |B_n| &\leq D_n [\mu(X_{n, K_n}) - \mu(X_{n, 1}) + m_n(X_{n, K_n}) - m_n(X_{n, 1})] \\ &\leq D_n \left[\int_a^b \mu'(x) dx + \int_a^b m'_n(x) dx \right]. \end{aligned}$$

So,

$$\begin{aligned} &\sum_{k=1}^n \delta_k [m_n(X_k) - \mu(X_k)] [Y_k - \mu(X_k)] \\ &\leq 2D_n \left[\int_a^b \mu'(x) dx + \int_a^b m'_n(x) dx \right] + 2 \frac{D_n^2}{K_n} \\ &\leq 2\sqrt{b-a} D_n \left[\sqrt{\int_a^b [\mu'(x)]^2 dx} + \sqrt{\int_a^b [m'_n(x)]^2 dx} \right] + 2 \frac{D_n^2}{K_n}. \end{aligned}$$

Therefore

$$\begin{aligned} &\sum_{k=1}^n \delta_k [m_n(X_k) - \mu(X_k)]^2 + \alpha_n \int_a^b [m'_n]^2 dx \\ &\leq 4\sqrt{b-a} D_n \left[\sqrt{\int_a^b [\mu']^2 dx} + \sqrt{\int_a^b [m'_n]^2 dx} \right] + \alpha_n \int_a^b [\mu']^2 dx + \frac{4D_n^2}{K_n} \\ &= \frac{4\sqrt{b-a} D_n}{\sqrt{\alpha_n}} \left[\sqrt{\alpha_n \int_a^b [\mu']^2 dx} + \sqrt{\alpha_n \int_a^b [m'_n]^2 dx} \right] \\ &\quad + \alpha_n \int_a^b [\mu']^2 dx + \frac{4D_n^2}{K_n}. \end{aligned}$$

Rewrite the above inequality as

$$L_n^2 \leq \frac{4\sqrt{b-a} D_n}{\sqrt{\alpha_n}} (L_n + R_n) + R_n^2,$$

where

$$L_n^2 = \sum_{k=1}^n \delta_k [m_n(X_k) - \mu(X_k)]^2 + \alpha_n \int_a^b [m'_n]^2 dx,$$

and

$$R_n^2 = \alpha_n \int_a^b [\mu']^2 dx + 4 \frac{D_n^2}{K_n}.$$

One finds that

$$L_n \leq \frac{4\sqrt{b-a}D_n}{\sqrt{\alpha_n}} + R_n$$

or since $(x+y)^2 \leq 2x^2 + 2y^2$

$$L_n^2 \leq \frac{32(b-a)D_n^2}{\alpha_n} + 2R_n^2.$$

That is

$$\begin{aligned} \frac{1}{n^p} \sum_{k=1}^n \delta_k [m_n(X_k) - \mu(X_k)]^2 &\leq \frac{32(b-a)D_n^2}{\alpha_n n^p} + \frac{2\alpha_n}{n^p} \int_a^b [\mu']^2 dx + 8 \frac{D_n^2}{n^p K_n} \\ &= \frac{32(b-a)D_n^2}{O(n^{p+q})} + \frac{2}{O(n^{p-q})} \int_a^b [\mu']^2 dx + 8 \frac{D_n^2}{n^p K_n} \\ &\stackrel{\text{a.s.}}{\rightarrow} 0, \end{aligned}$$

by the proposition above. \square

Uniform consistency. We can improve the results of the previous theorem to uniform consistency of the isotonic splines, if the observation points become sufficiently dense in $[a, b]$ as $n \rightarrow \infty$.

Let $\delta = \{\delta_k, k \geq 1\}$ be a fixed decision rule. For each $n = 1, 2, \dots$, define for $a \leq x \leq b$,

$$K_n(x) = K_n^\delta(x) = \sum_{k=1}^n \delta_k I\{X_k \leq x\}$$

and

$$K_n = K_n^\delta = K_n^\delta(b).$$

For notational simplicity, we will assume δ as given and omit the superscript δ . $K_n(\cdot)$ is a step function with jumps of magnitude 1 at $X_{n,j}, j = 1, \dots, K_n$. Of course, if F is continuous, then there is strict inequality with probability one. $K_n(\cdot)$ is also used to denote the counting measure on the set $\{X_{n,j}, j = 1, \dots, K_n\}$.

Theorem 3. Let $\mu \in H_+$ and $\int_a^b \sigma^2(x) dF(x) < \infty$; and let δ be a decision rule. Suppose that there are $p > \frac{1}{2}$ and $1 - p < q < p$ for which $\alpha_n = O_c(n^q)$ and

$$\liminf_{n \rightarrow \infty} \frac{K_n(J)}{n^p} > 0 \quad \text{a.s.},$$

for any nondegenerate compact subinterval J of $[a, b]$. Then for any $c, d: a < c < d < b$,

$$\sup_{c \leq x \leq d} |m_n(x) - \mu(x)| \stackrel{\text{a.s.}}{\rightarrow} 0,$$

as $n \rightarrow \infty$.

Proof. From Theorem 2, we know that

$$\frac{1}{n^p} \int_a^b [m_n(x) - \mu(x)]^2 dK_n(x) \xrightarrow{\text{a.s.}} 0, \tag{9}$$

as $n \rightarrow \infty$.

For a fixed $x_0: a < x_0 < b$, suppose (9) holds but $m_n(x_0)$ fails to converge to $\mu(x_0)$. Then there are $\varepsilon > 0$ and a subsequence $N_0 \subset \{1, 2, \dots\}$ such that for all $n \in N_0$

$$m_n(x_0) - \mu(x_0) > 2\varepsilon.$$

By continuity of μ , there exists $\eta > 0$ such that for all $x: x_0 - \eta \leq x \leq x_0 + \eta$

$$|\mu(x) - \mu(x_0)| \leq \varepsilon.$$

Consequently, for all $x: x_0 \leq x \leq x_0 + \eta$ and for all $n \in N_0$

$$\begin{aligned} m_n(x) - \mu(x) &\geq m_n(x_0) - \mu(x_0) - \varepsilon \\ &> \varepsilon. \end{aligned}$$

But then

$$\frac{1}{n^p} \int_{x_0}^{x_0 + \eta} [m_n(x) - \mu(x)]^2 dK_n(x) > \varepsilon^2 \frac{K_n([x_0, x_0 + \eta])}{n^p},$$

for all $n \in N_0$. Therefore

$$\frac{1}{n^p} \int_{x_0}^{x_0 + \eta} [m_n(x) - \mu(x)]^2 dK_n(x)$$

is bounded away from 0 as $n \rightarrow \infty$ through N_0 contradicting (9). Therefore for all $x: a < x < b$

$$\limsup_{n \rightarrow \infty} m_n(x) - \mu(x) \leq 0 \quad \text{a.s.}$$

Similarly, one can verify that for all $x: a < x < b$, $\liminf_{n \rightarrow \infty} [m_n(x) - \mu(x)] \geq 0$ a.s. That is for all $x: a < x < b$

$$\lim_{n \rightarrow \infty} |m_n(x) - \mu(x)| = 0 \quad \text{a.s.}$$

Uniform convergence in any compact subinterval of $[a, b]$ is provided by the monotonicity and continuity of μ (see Breiman, 1968, p. 160). \square

4. References

Breiman, L. (1968). *Probability*, Addison-Wesley, New York.
 Friedman, J. and R. Tibshirani (1984). The monotone smoothing of scatterplots. *Technometrics* **26**, 243–250
 Mukerjee, H. (1988). Monotone nonparametric regression, *Ann. Statist.* **16**, 741–750.

- Robertson, T.F., F. Wright and R. Dykstra (1988). *Order-Restricted Inference*. Wiley, New York.
- Sarkar, J. (1991). One-armed bandit problems with covariates. *Ann. Statist.* **19**, 1978–2003.
- Tantiyaswasdikul, C. (1992). Isotonic regression under sequential designs. Submitted to *Sequential Analysis*.
- Utreras, F.I. (1986). Smoothing noisy data using monotonicity constraints. *Numerische Mathematik* **47**, 611–625.
- Wegman, E.J. (1984). Optimal nonparametric function estimation. *J. Statist. Plan. Inference* **8**, 375–387.
- Woodroffe, M. (1979). A one-armed bandit problem with a concomitant variable. *J. Amer. Statist. Assoc.* **74**, 799–806.