Generalized continuous isotonic regression

Piet Groeneboom, Geurt Jongbloed *

Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

ARTICLE INFO

ABSTRACT

The standard isotonic regression of a vector in $\mathbb{R}^n$ is the solution to a least squares projection problem on the cone $\mathcal{C}$ of vectors with 'increasing components' in $\mathbb{R}^n$. Generalized isotonic regression problems are isotonic optimization problems that seem to be quite different from isotonic regression problems, but in fact have the same solution. In problems of maximum smoothed likelihood estimation, often continuous versions of generalized isotonic regression problems emerge. In this paper we show that the solutions to such problems are often given by the relatively simple continuous isotonic regression of a function. We illustrate the result using the problem of maximum smoothed likelihood estimation of an increasing hazard function.

1. Introduction

Consider a vector $y \in \mathbb{R}^n$ and define the closed convex cone $\mathcal{C} \subset \mathbb{R}^n$ by

$$\mathcal{C} = \{x = (x_1, \ldots, x_n) : x_1 \le x_2 \le \cdots \le x_n\}.$$

Furthermore let $w = (w_1, w_2, \ldots, w_n)$ be a weight vector in $(0, \infty)^n$. The *isotonic regression* of y, with weights w, is defined as the minimizer over $\mathcal{C}$ of the following quadratic function:

$$\phi_q(x) = \frac{1}{2} \sum_{i=1}^{n} (x_i - y_i)^2 w_i. \tag{1}$$

Early references on this problem include Ayer et al. (1955) and Grenander (1956). Comprehensive books on isotonic regression are Barlow et al. (1972) and Robertson et al. (1988). The most basic result from these references is that the isotonic regression of y with weights w exists, is unique and can be constructed graphically, in two dimensions. Indeed, defining the set of points P_i, $i = 0, 1, \ldots, n$, by $P_0 = (0, 0)$ and

$$P_i = \left(\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} w_j y_j \right), \quad i = 1, 2, \ldots, n,$$

the kth component of the isotonic regression of y is given by the left derivative of the greatest convex minorant of the set of points $\{P_i : 0 \le i \le n\}$, evaluated at P_k. See e.g. Robertson et al. (1988), Theorem 1.2.1.

Generalized isotonic regression problems are also concerned with minimizing an objective function over the cone $\mathcal{C}$. The objective function is different from (1), but of a specific structure. Due to that structure, the solution to this seemingly more complex optimization problem turns out to be exactly the isotonic regression of y on $\mathcal{C}$ with some choice of weight vector

---

* Corresponding author.
  E-mail addresses: P.Groeneboom@tudelft.nl (P. Groeneboom), G.Jongbloed@tudelft.nl (G. Jongbloed).

w. Examples of these problems are the Poisson extremum problem and the Gamma extremum problem that occur in order constrained maximum likelihood estimation. See section 1.5 in Robertson et al. (1988).

A straightforward continuous version of isotonic regression is concerned with L^2 projections of functions on a compact interval $[a, b]$. Let $w > 0$ be an integrable weight function on $[a, b]$ and $g \in L^2(w)$. Moreover, denote by $\mathcal{K}$ the subset of $L^2(w)$ consisting of the nondecreasing functions on $[a, b]$. The continuous isotonic regression of g on $\mathcal{K}$ is defined as the minimizer of the function

$$\psi_q(f) = \frac{1}{2} \int (f(x) - g(x))^2 w(x) \, \mathrm{d}x. \tag{2}$$

In Mammen (1991), it is mentioned that in case $w \equiv 1$, the solution to this problem is given by the right continuous slope of the convex minorant of the parameterized graph in $\mathbb{R}^2$ given by

$$\left\{ \left(\int_a^t w(x) \, \mathrm{d}x, \int_a^t g(x) w(x) \, \mathrm{d}x \right) : t \in [a, b] \right\} \tag{3}$$

In Anevski and Soulier (2009), Lemma 2, this statement is proved for continuous functions g and weight function $w \equiv 1$. In fact, the proof of that result reveals that it also holds if $w \not\equiv 1$ is a continuous strictly positive weight function on $[a, b]$. This characterization, together with continuity of g, immediately shows that the right continuous version of the solution is unique on $[a, b]$. In Tantiyaswasdikul and Woodroofe (1994), a penalized version of the continuous isotonic regression problem is considered, leading to isotonic smoothing splines.

In this paper we show that for a class of specific continuous projection problems, which we call generalized continuous isotonic regression problems, the solution is given by the continuous isotonic regression in a similar way as in the discrete situation. This implies that so-called maximum smoothed likelihood estimators in various models can be characterized and constructed via a simple convex minorant interpretation.

In Section 2, we state the generalized continuous isotonic regression problem, give two examples from the literature concerning its use and prove that its solution is precisely that of a related continuous isotonic regression problem. In Section 3, we introduce a maximum smoothed likelihood estimator for an increasing hazard rate, characterize it in terms of a convex minorant of a smooth graph in $\mathbb{R}^2$ and use that characterization to compute it based on a dataset of ball bearings from Lieblein and Zelen (1956) and also used in Gijbels and Heckman (2004).

2. Main result and some examples

In this section we describe the generalized continuous isotonic regression problem. We will follow closely the notation and ideas given for the discrete situation in Robertson et al. (1988), section 1.5.

Let g be a fixed real-valued function on $[a, b]$ and $I \subset \mathbb{R}$ be an interval such that $g(x) \in I$ for all $x \in [a, b]$. Moreover, let Φ be a convex function on I with (right) derivative ϕ and define for $u, v \in I$

$$\Delta_\Phi(u, v) = \Phi(u) - \Phi(v) - (u - v)\phi(v).$$

By the subgradient inequality for convex functions, $\Delta_\Phi(u, v) \geq 0$ for all $u, v \in I$. If Φ is strictly convex, this inequality is strict for $u \neq v$. Now, taking a strictly positive continuous weight function w on $[a, b]$, we can define the objective function

$$\psi(f) = \int \Delta_\Phi(g(x), f(x)) \, w(x) \, \mathrm{d}x \tag{4}$$

and minimize this function over the class $\mathcal{K}$. The main result in this paper is that the minimizer of (4) over $\mathcal{K}$ is exactly the minimizer of (2) over $\mathcal{K}$.

Theorem 1. *Suppose that g is a continuous function on $[a, b]$ and w is a strictly positive, continuous function on $[a, b]$. Then the minimizer g^* of ψ_q defined in (2) over $\mathcal{K}$ is continuous and unique. Its value at $x \in (a, b)$ is given by the slope of the greatest convex minorant of the graph defined by (3), evaluated at $t = x$. Moreover, for all $f \in \mathcal{K}$,*

$$\psi(f) \geq \psi(g^*),$$

where ψ is defined in (4). If Φ in the definition of (4) is strictly convex, g^* is the unique continuous minimizer of ψ over $\mathcal{K}$.

Before proving this theorem, we give two examples of situations where this theorem can be applied. A third example will follow in Section 3. All examples are concerned with maximum smoothed likelihood estimation in the spirit of Eggermont and LaRiccia (2001). Denoting the empirical distribution function of a sample by $\mathbb{F}_n$ and the sampling density (w.r.t. some dominating measure) parameterized by a parameter f by g_f, maximum likelihood estimation is usually concerned with maximizing a function of the form

$$f \mapsto \int \log g_f(x) \, \mathrm{d}\mathbb{F}_n(x)$$

over f in the set of parameters. This function is the log likelihood function. The smoothed log likelihood function is a similar function, only the empirical distribution function $\mathbb{F}_n$ is replaced by some smooth alternative estimator. This could be a kernel estimator, spline estimator based on the data or a linear interpolation of $\mathbb{F}_n$.

Example 1. Maximum Smoothed Likelihood estimation of a decreasing density. Let g be some smooth density estimate on $[0, M]$ (e.g. a kernel estimate based on a sample). The MSLE of the underlying density based on this g is then defined as the maximizer of

$$f \mapsto \int \log f(x) g(x)\, \mathrm{d}x - \int f(x)\, \mathrm{d}x$$

over the class $\mathcal{K}_d$, the class of nonincreasing nonnegative functions on $[0, M]$. Note that the second term in this expression is a relaxation term, guaranteeing the solution to have integral equal to 1. The isotonized kernel estimator $\tilde{f}_n$ discussed in Van der Vaart and Van der Laan (2003) is in fact an MSLE, where the initial density estimate is the standard kernel estimator as studied in Silverman (1986). To see this, note that $\Phi(u) = u \log u$ yields

$$\begin{aligned}\Delta_\Phi\left(g(x), f(x)\right) &= g(x) \log g(x) - f(x) \log f(x) - (g(x) - f(x))(1 + \log f(x)) \\ &= c_g(x) - g(x) \log f(x) + f(x)\end{aligned}$$

and observe that the MSLE based on g coincides with the continuous isotonic regression of g, with weight function $w \equiv 1$.

Example 2. Maximum Smoothed Likelihood estimation in the current status model. This model is studied in Groeneboom et al. (in press). The maximizer of their smoothed log likelihood (3.1),

$$F \mapsto \int \left(\frac{g_{n,1}(x)}{g_n(x)} \log F(x) + \left(1 - \frac{g_{n,1}(x)}{g_n(x)}\right) \log(1 - F(x)) \right) g_n(x)\, \mathrm{d}x$$

is characterized in their Theorem 3.1 via a direct argument. Viewing their optimization problem as generalized continuous isotonic regression problem with $\Phi(u) = u \log u + (1 - u) \log(1 - u)$, the characterization follows immediately. Indeed, this choice leads to

$$\Delta_\Phi\left(\frac{g_{n,1}(x)}{g_n(x)}, F(x)\right) = c(x) - \frac{g_{n,1}(x)}{g_n(x)} \log F(x) - \left(1 - \frac{g_{n,1}(x)}{g_n(x)}\right) \log(1 - F(x))$$

where c does not depend on F. Using their notation, the solution of this optimization problem is the continuous isotonic regression of $g_{n,1}/g_n$ with weight function g_n, i.e. the minimizer of

$$F \mapsto \frac{1}{2} \int \left(\frac{g_{n,1}(x)}{g_n(x)} - F(x) \right)^2 g_n(x)\, \mathrm{d}x$$

over all functions F in $\mathcal{K}$.

We now prove Theorem 1 along the lines of the proof given in Robertson et al. (1988) in the discrete situation. In order to do so, we use the following two lemmas. The proofs of these results can be found in the Appendix. First we prove that the continuous isotonic regression g^* of a continuous function g is continuous.

Lemma 1. *Let g be a continuous function on $[a, b]$, w a continuous strictly positive function on $[a, b]$. Then the continuous isotonic regression g^* of g with weight function w exists, is unique and continuous on $[a, b]$. Its value at x is given by the slope of the greatest convex minorant of the graph defined by (3), evaluated at $t = x$.*

The next lemma is a generalization of Theorem 1.3.5 in Robertson et al. (1988).

Lemma 2. *Let g be a continuous function on $[a, b]$ and w a continuous strictly positive function on $[a, b]$. Let g^* be the continuous isotonic regression of g with weight function w. Define, for each $c \in \mathbb{R}$, the level set $L_c = \{x \in [a, b] : g^*(x) = c\}$. If L_c consists of exactly one point τ, necessarily $g^*(\tau) = g(\tau)$. If L_c has positive Lebesgue measure, then $L_c = [\sigma, \tau]$, where $a \le \sigma < \tau \le b$. Moreover,*

$$\frac{\int_\sigma^\tau g(x) w(x)\, \mathrm{d}x}{\int_\sigma^\tau w(x)\, \mathrm{d}x} = c. \tag{5}$$

Proof of Theorem 1. Let g^* be the continuous isotonic regression of g with weight function w. Its continuity, uniqueness and representation as derivative of the convex minorant of the graph defined in (3) follow from Lemma 1. Before showing that this g^* also minimizes ψ over $\mathcal{K}$, we first establish a useful equality.

The number of 'level intervals' of g^* of positive length is at most countable. Continuity of g^* implies that these intervals are closed. Denote the intervals by $[\sigma_n, \tau_n]$, $n = 1, 2, \ldots$. From Lemma 2, it immediately follows that for any real-valued function Ψ,

$$\int_a^b \left(g(x) - g^*(x)\right) \Psi(g^*(x)) w(x)\, \mathrm{d}x = \int_{\{x \,:\, g^*(x) \ne g(x)\}} \left(g(x) - g^*(x)\right) \Psi(g^*(x)) w(x)\, \mathrm{d}x$$

$$= \sum_n \int_{\sigma_n}^{\tau_n} \left(g(x) - g^*(x) \right) \Psi \left(g^*(x) \right) w(x) \, \mathrm{d}x$$

$$= \sum_n \Psi \left(g^*(\tau_n) \right) \int_{\sigma_n}^{\tau_n} \left(g(x) - g^*(x) \right) w(x) \, \mathrm{d}x = 0. \qquad (6)$$

This means that for each $f \in \mathcal{K}$

$$\psi(f) - \psi(g^*) = \int \left(\Delta_\Phi \left(g(x), f(x) \right) - \Delta_\Phi \left(g(x), g^*(x) \right) \right) w(x) \, \mathrm{d}x$$

$$= \int \left(\Phi(g^*(x)) - \Phi(f(x)) - (g(x) - f(x)) \phi(f(x)) + (g(x) - g^*(x)) \phi(g^*(x)) \right) w(x) \, \mathrm{d}x$$

$$\overset{(i)}{=} \int \left(\Delta_\Phi(g^*(x), f(x)) - (g(x) - g^*(x)) \phi(f(x)) \right) w(x) \, \mathrm{d}x \overset{(ii)}{\geq} 0.$$

In (i) we use (6), in (ii) that $\Delta_\Phi(\cdot, \cdot) \geq 0$, the fact that $x \to \phi(f(x))$ is monotone and the fact that for each nondecreasing function m on $[a, b]$,

$$\int (g(x) - g^*(x)) m(x) w(x) \, \mathrm{d}x = \lim_{\epsilon \downarrow 0} \epsilon^{-1} \left(\psi_q(g^*) - \psi_q(g^* + \epsilon m) \right) \leq 0.$$

In case Φ is strictly convex, inequality (ii) is strict if $f \neq g^*$ on a set of positive Lebesgue measure. □

3. Application: Smooth increasing hazard

Consider an i.i.d. sample $X_1, \ldots, X_n$ from an unknown density g on $[0, \infty)$ with distribution function G and increasing hazard function f. The functions g, G and f satisfy the following relations:

$$G(x) = \int_0^x g(y) \, \mathrm{d}y = 1 - \exp \left\{ - \int_0^x f(y) \, \mathrm{d}y \right\}, \qquad f(x) = -\frac{\mathrm{d}}{\mathrm{d}x} \log(1 - G(x)) = \frac{g(x)}{1 - G(x)}.$$

One approach to estimate f would be to use the maximum likelihood estimator. This estimator is well defined and described in Robertson et al. (1988), Section 7.4. It is a step function, piecewise constant between successive observed data points. In some situations it is reasonable to assume that the underlying hazard rate is smooth and monotone. See, e.g., Gijbels et al. (2001) for examples and a "biased bootstrap" procedure to estimate a hazard rate under such an assumption. Alternatively, one could study a maximum smoothed likelihood estimator in the spirit of Eggermont and LaRiccia (2001) where a smoothed log likelihood function is maximized. Like in the examples described in Section 2, the difference between this function and the log likelihood function is that the empirical distribution function of the data that occurs in the log likelihood, is replaced by a smooth estimator of the distribution function. More precisely, let $x_1 < x_2 < \cdots < x_n$ be the ordered realized sample. Let G_n be a smooth estimator of the distribution function G, with continuous density g_n. The smoothed log likelihood is then defined by

$$\ell(f) = \int g_n(x) \log g_f(x) \, \mathrm{d}x = \int g_n(x) \log f(x) \, \mathrm{d}x - \int g_n(x) \int_{y=0}^x f(y) \, \mathrm{d}y \, \mathrm{d}x$$

$$= \int g_n(x) \log f(x) \, \mathrm{d}x - \int f(x)(1 - G_n(x)) \, \mathrm{d}x.$$

Maximizing this function over $\mathcal{K}$ is equivalent to computing the continuous isotonic regression of the function $g_n/(1 - G_n)$ with weight function $w = 1 - G_n$. This follows from Theorem 1, using the same function Φ as in Example 1 in the previous section and taking weight function $w = 1 - G$. Note that $g_n/(1 - G_n)$ is the plug-in estimator of the hazard rate based on g_n. Having computed the MSLE $\hat{f}$ of the hazard rate, the plug-in principle can be used to compute the MSLE of G and g within this class:

$$\hat{G}(x) = 1 - \exp \left\{ - \int_0^x \hat{f}(y) \, \mathrm{d}y \right\}, \qquad \hat{g}(x) = (1 - \hat{G}(x)) \hat{f}(x).$$

In Groeneboom and Jongbloed (in preparation), the MSLE of f is used to construct a test on monotonicity of a hazard function. This approach is an alternative to approaches to this testing problem recently adopted by Gijbels and Heckman (2004) and Hall and Van Keilegom (2005).

To illustrate the MSLE, we use the ball bearing dataset from Lieblein and Zelen (1956), also used in Gijbels and Heckman (2004). This dataset consists of $n = 23$ failure times (although more could be said on this, see Catoni (2002)). The left panel of Fig. 1 shows the MLE of Robertson et al. (1988) on the interval [0, 150], a step function, as well as the MSLE based on a kernel density estimator with bandwidth $h = 100n^{-1/4} \approx 46$ (using a boundary kernel near zero). The middle panel of Fig. 1 shows the empirical distribution function of the data as well as the estimates based on the MLE and the MSLE. The right panel of Fig. 1 shows the density estimates corresponding to the MLE and the MSLE.
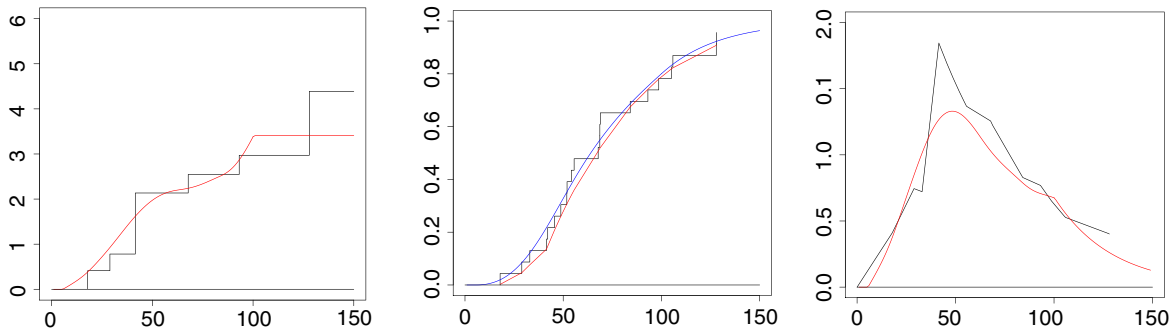
Fig. 1. The MLE's and MSLE's based on the ball bearing data on the interval [0, 150], $h = 46$. From left to right: the hazard functions; the empirical distribution function, the estimate corresponding to the MLE (red) and the MSLE (blue); the density estimates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Acknowledgements

Appendix

In this section we prove Lemmas 1 and 2.

Proof of Lemma 1. Existence of g^* and its representation as derivative of the convex minorant of the graph defined in (3) follows from Lemma 1 and Lemma 2 in Anevski and Soulier (2009), with obvious modification to incorporate strictly positive continuous $w \not\equiv 1$. Uniqueness of the right continuous minimizer follows from strict convexity of the function ψ_q.

Suppose the right continuous function g^* has a point of discontinuity $x \in (a, b)$, i.e. $g^*(x) - g^*(x^-) = \delta > 0$. Then x belongs to the support of the Stieltjes measure $d(g^*)$ and hence, by Lemma 1 in Anevski and Soulier (2009),

$$\int_a^t \left(g(y) - g^*(y) \right) w(y) \, dy = 0 \quad \text{if } t = x.$$

Since for each $t \in [a, b]$ this integral is greater than or equal to zero and it is (as an integral) differentiable on (a, b), $g^*(x) = g(x)$ necessarily. Now define, for small $\epsilon > 0$,

$$g^*_\epsilon(y) = g^*(y) \left(1_{[a,x-\epsilon)}(y) + 1_{[x,b]}(y) \right) + g(x) 1_{[x-\epsilon,x)}(y).$$

Note that $g^*_\epsilon \in \mathcal{K}$. Using continuity of g and monotonicity of g^*, we can choose ϵ sufficiently small to guarantee that $| g(y) - g(x) | < \delta/4$ and $| g^*(y) - g(y) | > 3\delta/4$ for $y \in [x - \epsilon, x)$. This yields for $\epsilon > 0$ sufficiently small

$$\psi_q(g^*_\epsilon) - \psi_q(g^*) = \frac{1}{2} \int_{x-\epsilon}^x \left((g(x) - g(y))^2 - (g^*(y) - g(y))^2 \right) w(y) \, dy$$

$$\leq \left(\frac{\delta^2}{32} - \frac{9\delta^2}{32} \right) \int_{x-\epsilon}^x w(y) \, dy = -\frac{1}{4}\delta^2 \int_{x-\epsilon}^x w(y) \, dy < 0$$

contradicting $\psi_q(g^*_\epsilon) \geq \psi_q(g^*)$. □

Proof of Lemma 2. If $L_c = \tau$ is a singleton, τ is a point of increase of g^*, and hence belongs to the support of the Stieltjes measure $d(g^*)$. By Lemma 1 in Anevski and Soulier (2009), this implies that $\int_a^\tau g^*(y)w(y) \, dy = \int_a^\tau g(y)w(y) \, dy$ and similar to in the proof of our Lemma 1, it follows that $g^*(\tau) = g(\tau)$.

If L_c has positive Lebesgue measure, it is necessarily a closed interval. This follows from monotonicity of g^* and continuity of both g and g^*. Consider such an interval $[\sigma, \tau]$, with $a < \sigma < \tau < b$ and define c as in (5). Now suppose that the value of g^* on the interval $[\sigma, \tau]$ is strictly smaller than c. For $\epsilon > 0$, define

$$g^*_\epsilon(x) = \begin{cases} g^*(x) & \text{if } x \notin (\sigma, \tau + \epsilon) \\ g^*(\tau + \epsilon) & \text{if } x \in (\sigma, \tau + \epsilon). \end{cases}$$

Then

$$\psi_q(g^*) - \psi_q(g^*_\epsilon) = \frac{1}{2} \left(\int_\sigma^\tau + \int_\tau^{\tau+\epsilon} \right) \left((g(x) - g^*(x))^2 - (g(x) - g^*_\epsilon(x))^2 \right) w(x) \, dx = I^{(1)}_\epsilon + I^{(2)}_\epsilon$$

where (using that $g^*(x) = g^*(\tau)$ for $x \in [\sigma, \tau]$)

$$I_\epsilon^{(1)} = \frac{1}{2}(g^*(\tau + \epsilon) - g^*(\tau)) \int_\sigma^\tau \left(2g(x) - g^*(\tau + \epsilon) - g^*(\tau)\right) w(x)\, dx$$

and

$$I_\epsilon^{(2)} = \frac{1}{2} \int_\tau^{\tau+\epsilon} (g^*(\tau + \epsilon) - g^*(x))(2g(x) - g^*(x) - g^*(\tau + \epsilon)) w(x)\, dx.$$

Define $\delta = c - g^*(\tau) > 0$. Using continuity of g^*, we can make sure by taking $\epsilon > 0$ sufficiently small, that

$$I_\epsilon^{(1)} = \frac{1}{2}(g^*(\tau + \epsilon) - g^*(\tau)) \int_\sigma^\tau \left(2c - g^*(\tau + \epsilon) - g^*(\tau)\right) w(x)\, dx > \frac{1}{4}(g^*(\tau + \epsilon) - g^*(\tau))\delta.$$

On the other hand,

$$\mid I_\epsilon^{(2)} \mid \le (\|g\|_\infty + \|g^*\|_\infty)(g^*(\tau + \epsilon) - g^*(\tau))\|w\|_\infty \epsilon.$$

Combining these inequalities, we obtain that, for $\epsilon > 0$ sufficiently small,

$$\psi_q(g^*) - \psi_q(g_\epsilon^*) > \frac{1}{8} \int_\sigma^\tau w(x)\, dx \cdot (g^*(\tau + \epsilon) - g^*(\tau))\delta,$$

contradicting that $g^* = \mathrm{argmin}_{f \in \mathcal{K}} \psi_q(f)$.

If $g^*(\tau) > c$, a similar argument, using the point $\sigma - \epsilon$ instead of $\tau + \epsilon$, can be used. If $\sigma = a$ and / or $\tau = b$, the argument is a bit simpler. Instead of using $\sigma - \epsilon$ or $\tau + \epsilon$ to define g_ϵ^*, this function can be defined directly on the first (in case of $\sigma = a$) or last (in case $\tau = b$) interval. $\square$

References

Anevski, D., Soulier, P., (2009) Monotone spectral density estimation. ArXiv:0901.3471v1.

Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T., Silverman, E., 1955. An empirical distribution function for sampling with incomplete information. Ann. Math. Stat. 26, 641–647.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D., 1972. Statistical inference under order restrictions. John Wiley & Sons, New York.

Catoni, C., 2002. The correct ball bearings data. Lifetime Data Anal. 8, 395–399.

Eggermont, P.P.B., LaRiccia, V.N., 2001. Maximum penalized likelihood estimation. Springer, New York.

Gijbels, I., Heckman, N., 2004. Nonparametric testing of a monotone hazard function via normalized spacings. J. Nonparametr. Stat. 16, 463–478.

Grenander, U., 1956. On the theory of mortality measurement, part II. Skand. Aktuar. 39, 125–153.

Groeneboom, P., Jongbloed, G., 2009. Testing for an increasing hazard (in preparation).

Groeneboom, P., Jongbloed, G., Witte, B.I., 2009. Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. Ann. Statist. (in press).

Hall, P., Huang, L.-S., Gifford, J.A., Gijbels, I., 2001. Nonparametric estimation of hazard rate under the constraint of monotonicity. J. Comp. Graph. Stat. 10, 592–614.

Hall, P., Keilegom, I.van, 2005. Testing for monotone increasing hazard rate. Ann. Statist. 33, 1109–1137.

Lieblein, J., Zelen, M., 1956. Statistical investigation of the fatugue life of deep-groove ball bearings. J. Res. Natl. Bur. Stand. 57, 273–316.

Mammen, E., 1991. Estimating a smooth monotone regression function. Ann. Stat. 19, 724–740.

Robertson, T., Wright, F.T., Dykstra, R.L., 1988. Order restricted statistical inference. John Wiley & Sons, New York.

Silverman, B.W., 1986. Density estimation for statistics and data analysis. Chapman & Hall, London.

Tantiyaswasdikul, C., Woodroofe, M.B., 1994. Isotonic smoothing splines under sequential designs. J. Statist. Plann. Inference 38, 75–88.

Van der Vaart, A.W., Van der Laan, M.J., 2003. Smooth estimation of a monotone density. Statistics 37, 189–203.