

Limit Theory in Monotone Function Estimation

Cécile Durot and Hendrik P. Lopuhaä

Abstract. We give an overview of the different concepts and methods that are commonly used when studying the asymptotic properties of isotonic estimators. After introducing the inverse process, we illustrate its use in establishing weak convergence of the estimators at a fixed point and also weak convergence of global distances, such as the \mathbb{L}_p -distance and supremum distance. Furthermore, we discuss the developments on smooth isotonic estimation.

Key words and phrases: Cox model, current status model, isotonic estimation, limit theory, \mathbb{L}_p -distance, maximum likelihood estimators, monotone density, monotone failure rate, monotone regression, supremum distance.

1. INTRODUCTION

Estimating a mortality intensity is of particular interest in actuarial science and in medicine. While the mortality intensity is an increasing function early in life, it decreases with older people. In reliability studies, interest is on estimating the failure rate of systems. Typically, the failure rate first decreases during a debugging period and then increases after an ageing period. In other cases, we are interested in estimating the distribution function (which is monotone) of a life-time of a human being after a medicinal treatment, based on censored observations: we only observe if the patient is still alive at the (random) observation time. In doses-response experiments, which has applications, for example, in biology or epidemiology, it is often observed that the response tends to increase (or decrease) as the dose increases. In such situations, we are faced with the problem of estimating a monotone function. The typical useful statistical models for the applications described above are regression, current status and right-censored models. In terms of interpretability and accuracy, if the underlying function is monotone, it is useful to construct an estimator that satisfies the monotonicity constraint. This can be achieved by estimating the

function in an appropriate parametric model, but parametric models lack flexibility, since the functions are then restricted to a particular functional form. Hence, it is of interest to build monotone nonparametric estimators.

Early research on nonparametric inference under shape constraints dates back to the 1950s and first focused on estimation of functions under the constraint of monotonicity or unimodality, with the aim of defining nonparametric estimators satisfying the shape constraint. In the sequel, we will focus on nonincreasing functions; the case of nondecreasing functions is most of the time similar (but involves greatest convex minorants instead of least concave majorants below), and the case of unimodal functions with known mode easily reduces to the monotone case. We will focus on estimators obtained as slopes of the least concave majorant (LCM) of an appropriate stochastic process. A number of statistical settings where such estimators arise naturally are given in Section 2.1. We will focus on settings where the underlying function is continuous and observations are independent. We will describe the asymptotic properties of the estimators in such a setting and give, as far as possible, a sketch of proof which could be far from being rigorous, with the hope to provide some intuition on the considered results. We will also give references for the complete proofs. We would like to point out that the rigorous proofs are typically involved since the estimators are nonlinear and nonstandard (with a non-Gaussian limit distribution).

Cécile Durot is Professor, Université Paris Nanterre, 200, avenue de la République, 92000 Nanterre, France (e-mail: cecile.durot@gmail.com). Hendrik P. Lopuhaä is Associate Professor, Delft University of Technology, van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands (e-mail: h.p.lopuhaa@tudelft.nl).

2. MAIN CONCEPTS AND DEFINITIONS

2.1 Estimation Under the Monotonicity Constraint

The popular example of estimation under the constraint of monotonicity is estimating a nonincreasing density function on $[0, \infty)$, which was considered in [37]. See also [51]. Based on a sample X_1, \dots, X_n of i.i.d. observations from the unknown density, the maximum likelihood estimator (MLE), that maximizes

$$f \mapsto \sum_{i=1}^n \log f(X_i)$$

over the set of all nonincreasing densities $f : [0, \infty) \rightarrow [0, \infty)$, is the left-hand slope of the LCM of the empirical distribution function, where the LCM of a function $h : [0, \infty) \rightarrow \mathbb{R}$ is the lowest concave function on $[0, \infty)$ that lies above h . Hence, the MLE is a step function that can have steps only at the order statistics; it is now called *the Grenander estimator*. The MLE of a nonincreasing failure rate based on an i.i.d. sample, is also connected to the LCM of a stochastic process, and can alternatively be described with min-max formulas; see [37, 77].

The regression setting can be described in the Gaussian one-dimensional case as follows; see [14]. Assume that for all $k = 1, \dots, m$, with some $m \in \mathbb{N}$, a sample of n_k random variables with common mean $\mu_0(t_k)$ is observed, where $t_k \in \mathbb{R}$ is fixed and $\mu_0 : \mathbb{R} \rightarrow \mathbb{R}$ is nonincreasing. If all observations are independent Gaussian with common variance, then the MLE of μ_0 exists and can be characterized in terms of min-max formulas. Denoting by \bar{y}_k the average of observations with mean $\mu_0(t_k)$, it coincides with the least squares estimator (LSE) that minimizes

$$\mu \mapsto \sum_{k=1}^m n_k (\bar{y}_k - \mu(t_k))^2$$

over all nonincreasing functions $\mu : \mathbb{R} \rightarrow \mathbb{R}$. More pleasant characterizations arise from the theory of isotonic regression [11], Theorem 1.1: with $t_{(1)} < \dots < t_{(m)}$ the ordered design points, the values of the LSE at $t_{(1)}, \dots, t_{(m)}$ are the left-hand slopes of the LCM of the cumulative sumdiagram (CSD) consisting of the points

$$(1) \quad \begin{aligned} P_0 &= (0, 0) \quad \text{and} \\ P_k &= (G_n(t_{(k)}), \Lambda_n(t_{(k)})), \quad k = 1, \dots, m, \end{aligned}$$

where $\{G_n(t), t \in \mathbb{R}\}$ and $\{\Lambda_n(t), t \in \mathbb{R}\}$ are random processes in the horizontal and vertical direction, respectively, defined by

$$G_n(t) = \int \mathbb{1}_{\{x \leq t\}} d\mathbb{P}_n(x, y) \quad \text{and}$$

$$\Lambda_n(t) = \int \mathbb{1}_{\{x \leq t\}} y d\mathbb{P}_n(x, y)$$

with \mathbb{P}_n the empirical measure (hence G_n is the empirical distribution function of the design points). Any monotone interpolation of these slopes gives a LSE. Analogous to the Grenander estimator, which is a left-continuous step function, the left-continuous step interpolation is usually considered. Note that $\hat{\lambda}_n$ can also be seen as the left-hand slope of the LCM of the CSD $t \mapsto (G_n(t), \Lambda_n(t))$ constructed with piecewise-constant processes G_n and Λ_n that interpolate the points in (1).

In the 1990s there was a sharp rise in interest in estimation under monotonicity constraints. The current status problem studied in [52] aims at estimating the distribution function F_0 of a random variable $X \geq 0$ (a “life time”) based on observations of i.i.d. copies of the pair (Δ, T) , where $T \geq 0$ is a continuous variable interpreted as the observation time independent of X and $\Delta = \mathbb{1}_{\{X \leq T\}}$. The MLE of F_0 maximizes the log-likelihood over the set of all distribution functions F , which are nondecreasing by definition. Again, the MLE can be characterized thanks to min-max formulas; it can be obtained as slopes of the greatest convex minorant (GCM, the highest convex graph that lies below the points in the diagram) of the CSD in (1), where now, $t_{(k)}$ is the k th order statistic of T_1, \dots, T_n ,

$$\begin{aligned} G_n(t_{(k)}) &= k/n \quad \text{and} \\ \Lambda_n(t_{(k)}) &= n^{-1} \sum_{j=1}^n \Delta_j \mathbb{1}_{\{T_j \leq t_{(k)}\}}; \end{aligned}$$

see [52], Proposition 1.2. The second case of interval censoring (see [52]) is related to double censoring: we still aim to estimate the distribution function of $X \geq 0$, but now one only observe i.i.d. copies of $(U, T, \mathbb{1}_{\{X \leq T\}}, \mathbb{1}_{\{X \in (T, U]\}})$, where $0 \leq U \leq T$ are independent observation times independent of X . The characterization of the MLE is more involved in that case and can be given in terms of the slope of the convex minorant of a self-induced cumulative sumdiagram.

The setting of interval censoring generalizes to so-called *panel count data* where several events are recorded across the time for each subject in the study. Then, the parameter of interest is the nondecreasing mean function of the corresponding counting process. Two nonparametric estimators are proposed in [93]: one of them is a pseudo-likelihood estimator that ignores the dependency of the events within a subject and is defined through slopes of the GCM of an appropriate CSD; the other one, which shows a better

performance, takes the dependency into account and can be computed using an iterative convex minorant algorithm.

The problem of estimating the density or the failure rate of a random variable $X \geq 0$ under monotonicity constraints, has been considered in the context of right-censored observations. This means that observations are i.i.d. copies of the pair $(\Delta, \min(X, T))$, where $T \geq 0$ is a censoring time independent of X and $\Delta = \mathbb{1}_{\{X \leq T\}}$. In this setting, [57] provides a characterization of the MLE for the density of X as the LCM of a self-induced stochastic process. Similar to the Grenander estimator, defined as the slope of the LCM of the empirical estimator of the distribution function F_0 corresponding to a nonincreasing density f_0 , it seems natural in the case of right-censoring to start with the Kaplan–Meier estimator for F_0 and define a *Grenander-type estimator* as the slope of the LCM of the Kaplan–Meier estimator. The MLE differs from the Grenander-type estimator, but no significant differences can be seen on simulations; see [57]. In the same setting, [56] study nonparametric estimators of the density f_0 and of the failure rate $f_0/(1 - F_0)$, both under the constraint that it is nonincreasing on $[0, \infty)$, as well as that it is nondecreasing. They consider both the MLE of the monotone function of interest and the Grenander-type estimator defined as the slope of the LCM (or GCM in the nondecreasing case) of either the Kaplan–Meier estimator of the distribution function, or the Nelson–Aalen estimator of the cumulative failure rate. The two approaches are shown to be asymptotically equivalent in [56] (at a fixed point, at $n^{1/3}$ scaling), which supports the empirical findings in [57].

A great advantage of the above estimators is that, in addition to that they satisfy the required monotonicity constraint, they do not require the choice of a tuning parameter. In the sequel, we will focus on the MLE and Grenander-type estimators $\hat{\lambda}_n$ of a monotone function λ_0 on $I = [0, 1]$ or $I = [0, \infty)$, of which the values are given by the slopes of the LCM (or GCM) of either a cumulative sumdiagram built from the observations, or a naive estimator for the primitive Λ_0 corresponding to λ_0 , respectively. For ease of exposition, we give a more formal definition for the nonincreasing case on $[0, 1]$, although monotone functions on $[0, \infty)$, say, are also of considerable interest.

DEFINITION 2.1. Let $\lambda_0 : [0, 1] \rightarrow \mathbb{R}$ be the non-increasing function of interest and let Λ_n be either a piecewise constant or a piecewise linear process on

$[0, 1]$ such that $\Lambda_n(0) = 0$, that estimates the cumulative function $\Lambda_0(t) = \int_0^t \lambda_0(x) dx$ based on n observations. The estimator $\hat{\lambda}_n$ defined as the left-hand slope of the LCM of Λ_n , with $\hat{\lambda}_n(0) = \lim_{t \downarrow 0} \hat{\lambda}_n(t)$, is called a *Grenander-type estimator*.

As λ_0 is the slope of the concave function Λ_0 , it is natural to define Grenander-type estimators as slopes of a concave estimator (the LCM of Λ_n) of Λ_0 , if such an estimator is available. If observations are available only at design points, one has to consider instead isotonic estimators as defined below.

DEFINITION 2.2. Let $\lambda_0 : [0, 1] \rightarrow \mathbb{R}$ be the non-increasing function of interest and let G_n be a step distribution function on $[0, 1]$. Let Λ_n be a piecewise constant process on $[0, 1]$ with the same jump points $t_{(1)} < \dots < t_{(m)}$ as G_n , such that $\Lambda_n(0) = 0$. The estimator $\hat{\lambda}_n$ defined as the left-hand slope of the LCM of $t \mapsto (G_n(t), \Lambda_n(t))$, with $\hat{\lambda}_n(0) = \lim_{t \downarrow 0} \hat{\lambda}_n(t)$, is called an *isotonic estimator*.

REMARK 2.1. In the definition, G_n is typically the empirical distribution function of design points and Λ_n is an estimator of the cumulative function $\Lambda_0(t) = \int_0^t \lambda_0(x) dG(x)$ based on n observations, where G is the limit of G_n . The connection between Grenander-type and isotonic estimators is as follows. Let Γ_n be the piecewise linear process on $[0, 1]$ satisfying $\Gamma_n(0) = 0$ and $\Gamma_n(x_k) = \Lambda_n(t_{(k)})$ for all $k = 1, \dots, m$, where $x_k = G_n(t_{(k)})$. Then, the isotonic estimator in Definition 2.2 is the left-continuous step function such that $\hat{\lambda}_n(t_{(k)}) = \hat{\gamma}_n(x_k)$ for all k , where $\hat{\gamma}_n$ is the Grenander-type estimator of $\gamma_0 := \lambda_0 \circ G^{-1}$ defined as left-hand slope of the LCM of Γ_n . Also, $\hat{\gamma}_n = \hat{\lambda}_n \circ G_n^{-1}$. Here, G_n^{-1} and G^{-1} are the quantile functions corresponding to G_n and G .

The Grenander estimator for a monotone density is included in Grenander-type estimators with $\Lambda_n = F_n$ the empirical distribution function. Grenander-type estimators are considered in [56] for a nonincreasing density or failure rate under right censoring, with Λ_n equal to the Kaplan–Meier or Nelson–Aalen estimator, respectively. The LSE in the regression setting, the MLE for interval censoring case 1, the pseudo-MLE in [93], all boil down to isotonic estimators.

Recently, similar methods have been proposed in models that incorporate covariables: estimation of a monotone baseline hazard in the Cox model [75], a distribution function in current status regression [80, 43], a monotone ridge function in the single index model [6]. Despite similarities with isotonic estimators, they involve additional technicalities and will not be further considered in the present paper.

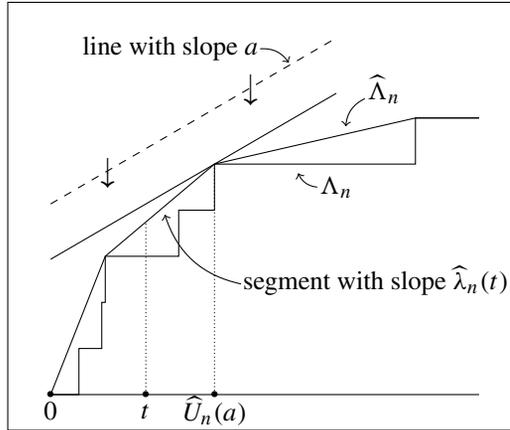


FIG. 1. Switching relation for $\widehat{U}_n(a)$ and $\widehat{\lambda}_n(t)$.

2.2 The Inverse Process

The inverse process, first introduced in [38] for the Grenander estimator, is a very useful tool in deriving their limit behavior. We first give the definition for Grenander-type estimators and then extend to isotonic estimators. The Grenander-type estimator $\widehat{\lambda}_n$ of Definition 2.1 is nonincreasing and left continuous, so for each $a \in \mathbb{R}$ its pseudo-inverse $\widehat{U}_n(a)$ is the largest $t \in [0, 1]$ that satisfies $\widehat{\lambda}_n(t) \geq a$, with the convention that the supremum of an empty set is zero. The resulting *switching relation*, is that

$$(2) \quad \widehat{\lambda}_n(t) \geq a \iff \widehat{U}_n(a) \geq t$$

for all $a \in \mathbb{R}$ and $t > 0$. For any $a \in \mathbb{R}$ and $t > 0$, if one drops a line with slope a from above onto the process Λ_n from Definition 2.1, then $\widehat{\lambda}_n(t) \geq a$ if and only if the line first hits Λ_n at a point $u \geq t$, that is, if and only if $\Lambda_n(z) - az$ is maximal at $z = u$; see Figure 1. Hence,

$$(3) \quad \widehat{U}_n(a) = \operatorname{argmax}_{z \in [0,1]} \{ \Lambda_n(z) - az \},$$

where the argmax is the largest location where the maximum is achieved. In the sequel, we assume that the maximum is achieved for all a , which is the case for instance if Λ_n is nonincreasing and right-continuous, or if Λ_n is piecewise linear. The inverse process \widehat{U}_n is much more tractable than $\widehat{\lambda}_n$. The advantage is that limit behavior of the process $z \mapsto \Lambda_n(z) - az$ is more or less standard and the argmax is a continuous mapping, which enables one to establish the limit behavior of $\widehat{U}_n(a)$. By means of (2) this will then yield the limit behavior of $\widehat{\lambda}_n(t)$.

Now, let $\widehat{\lambda}_n$ be an isotonic estimator as in Definition 2.2. Then,

$$\widehat{\lambda}_n(t) \geq a$$

$$\begin{aligned} \Leftrightarrow & (\exists s \geq t) (\forall w \geq t) : \frac{\Lambda_n(w) - \Lambda_n(s)}{G_n(w) - G_n(s)} \geq a \\ \Leftrightarrow & (\exists s \geq t) (\forall w \geq t) : \Lambda_n(w) - aG_n(w) \\ & \geq \Lambda_n(s) - aG_n(s) \\ \Leftrightarrow & \operatorname{argmax}_{z \in [0,1]} \{ \Lambda_n(z) - aG_n(z) \} \geq t \end{aligned}$$

for all $a \geq 0$ and $t > 0$. Hence, in this case the pseudo inverse for $\widehat{\lambda}_n$ is given by

$$(4) \quad \widehat{U}_n(a) = \operatorname{argmax}_{z \in [0,1]} \{ \Lambda_n(z) - aG_n(z) \}$$

and as before, it satisfies the switching relation in (2). Note that the pseudo inverse (3) is a special case of (4), with $G_n(z) = z$. Note also that in the notation of Remark 2.1, we have $\widehat{U}_n = G_n^{-1} \circ \widehat{\gamma}_n^{-1}$ where for all $a \in \mathbb{R}$,

$$(5) \quad \widehat{\gamma}_n^{-1}(a) = \operatorname{argmax}_{z \in [0,1]} \{ \Gamma_n(z) - az \}.$$

This illustrates the connection between inverses of isotonic estimators and inverses of Grenander-type estimators.

2.3 The Chernoff Distribution and the Argmax Process of a Brownian Motion with Parabolic Drift

This section is concerned with the so-called *Chernoff distribution* and the argmax process of Brownian motion with parabolic drift, as they both emerge from asymptotic theory (as the sample size goes to infinity) of isotonic estimators.

For $a \in \mathbb{R}$, let $V(a)$ be the location of the maximum of $\{W(t) - (t - a)^2, t \in \mathbb{R}\}$, where W is standard two-sided Brownian motion on \mathbb{R} starting at zero.

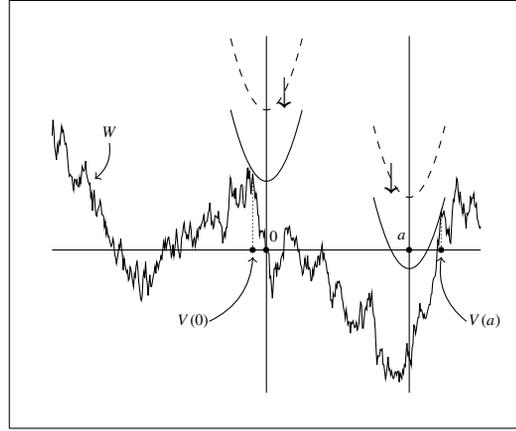


FIG. 2. The construction of $V(a)$.

See Figure 2. It follows from [62], Lemma 2.6, that $V(a)$ is uniquely defined with probability one. See also [40, 41] for an extensive study of V . Since, for any $a \in \mathbb{R}$, the process $\{W(t + a) - W(a), t \in \mathbb{R}\}$ is distributed like W , and the argmax is invariant by addition of a constant, it follows that $\{\zeta(a), a \in \mathbb{R}\}$ is a stationary process, where

$$(6) \quad \zeta(a) := V(a) - a = \operatorname{argmax}_{t \in \mathbb{R}} \{W(t + a) - t^2\}.$$

The distribution of $\zeta(0) = V(0)$, known as the *Chernoff distribution*, first appeared in [20]. It has a density that is connected to a solution of the heat equation. It is expressed in [41] in terms of Airy functions, which proves useful to get accurate numerical approximations. See also [53].

The density f_C of $V(0)$ has a symmetrical bell-shape curve with thin tails,

$$(7) \quad f_C(x) \sim 2\lambda|x| \exp(-2|x|^3/3 - \kappa|x|) \quad \text{as } |x| \rightarrow \infty,$$

for constants $\lambda, \kappa > 0$ (see [41], Corollary 3.4). This yields (see [55], Lemma 2.1)

$$1 - F_C(x) \sim \frac{\lambda}{x} \exp(-2x^3/3 - \kappa x) \quad \text{as } x \rightarrow \infty$$

for the corresponding distribution function F_C . Furthermore, according to [23], Theorem 4,

$$\mathbb{P}(|V(0)| \geq x) \leq 2 \exp(-x^3/2)$$

for all $x > 0$. Note that $V(a)$ can also be seen as the location of the maximum of $\{W(z) - z^2 + 2az, z \in \mathbb{R}\}$, which corresponds to dropping a line with slope $-2a$ onto the process $W(z) - z^2$, for $z \in \mathbb{R}$. Hence, if we define $D(t)$ as the slope at $z = t$ of the LCM of

$\{W(z) - z^2, z \in \mathbb{R}\}$, then by using a switching relation similar to the one in (2), it follows that $V(a) \leq t$ if and only if $D(t) \leq -2a$. See Figure 3. Therefore, by stationarity of $\{V(a) - a, a \in \mathbb{R}\}$, we find that

$$\begin{aligned} \mathbb{P}(V(a) \leq t) &= \mathbb{P}(V(a) - a \leq t - a) \\ &= \mathbb{P}(V(2a - t) - (2a - t) \leq t - a) \\ &= \mathbb{P}(V(2a - t) \leq a) = \mathbb{P}(D(a) \leq 2t - 4a) \\ &= \mathbb{P}(D(a)/2 + 2a \leq t). \end{aligned}$$

It follows that $V(a)$ has the same distribution as $D(a)/2 + 2a$. In particular $2V(0)$ has the same distribution as $D(0)$, being the slope at $z = 0$ of the LCM of $\{W(z) - z^2, z \in \mathbb{R}\}$. This was first pointed out by Prakasa Rao [83].

3. POINTWISE CONVERGENCE

A common feature of Grenander-type and isotonic estimators from Definitions 2.1 and 2.2 is that if λ_0 has a continuous strictly negative derivative in the neighborhood of some $t_0 \in (0, 1)$, then

$$(8) \quad C_0 n^{1/3} (\widehat{\lambda}_n(t_0) - \lambda_0(t_0)) \rightarrow_d 2V(0),$$

where $V(0)$ has the Chernoff distribution (see Section 2.3) and C_0 is a constant that depends on λ_0 and on the other unknown parameters (the precise form of C_0 depends on the setting considered). We compute below the rate of convergence and give two different sketches of the arguments of (8): one based on the “direct approach” that Prakasa Rao [83, 84] first used in this context, and the other one based on the “inverse approach” suggested later on in [40]. The “direct approach” is

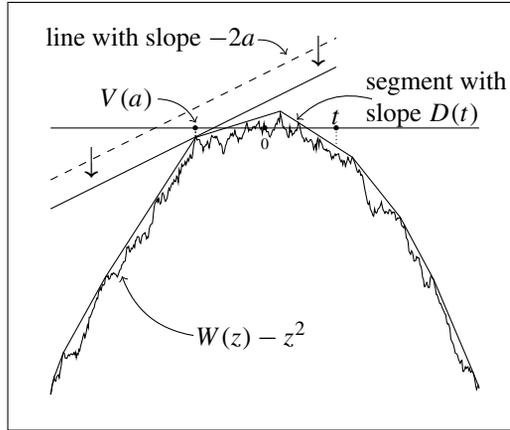


FIG. 3. Switching relation for $V(a)$ and $D(t)$.

based on first considering least concave majorants of the process Λ_n in Definition 2.1 along a sequence of decreasing “truncated” intervals around t_0 and then, together with weak convergence of a suitably scaled version of Λ_n , arguing that asymptotically the truncated intervals may replace \mathbb{R} . The “inverse approach” first makes a detour via the inverse \widehat{U}_n by means of the switch relation (2), establishes the limit behavior of \widehat{U}_n via weak convergence of a suitably scaled version of Λ_n and an argmax continuous mapping theorem, and transfers the limit behavior of \widehat{U}_n to $\widehat{\lambda}_n$ again using (2). Whereas the inverse approach became popular after its introduction in [40], the pioneering direct approach seems more suitable for extensions of pointwise limit behavior to dependent data. On the other hand, the inverse approach seems necessary to deal with global measures of deviation; see Section 4 below.

After having discussed (8), we briefly discuss below the behavior of $\widehat{\lambda}_n$ in a flat region, at a point of discontinuity, and at the boundary of the support of λ_0 .

3.1 Local Rate of Convergence

The underlying mechanism for the $n^{1/3}$ -rate is nicely explained in [62] for argmax type estimators and can be understood from the monotone (inverse) density case, as follows. Suppose one wants to estimate a nonincreasing density f_0 at the point t_0 , such that $a_0 = f_0(t_0) > 0$ and $f'_0(t_0) < 0$. Recall that, with F_n the empirical distribution function of a sample of size n from f_0 , $\widehat{U}_n(a_0)$ maximizes $H_n(t) = F_n(t) - a_0t$, which is close to $H(t) = F_0(t) - a_0t$, where $F_0(t) = \int_0^t f_0(x) dx$. Now, on the one hand, since t_0 maximizes $H(t)$, we have

$$H(t) - H(t_0) \approx -\frac{1}{2}|f'_0(t_0)|(t - t_0)^2,$$

and on the other hand $H_n(t) - H_n(t_0) - (H(t) - H(t_0))$ is approximately $N(0, \sigma_t^2)$ distributed, where

$$\begin{aligned} \sigma_t^2 &= n^{-1}(F_0(t) - F_0(t_0))(1 - (F_0(t) - F_0(t_0))) \\ &\approx n^{-1}f_0(t_0)|t - t_0|. \end{aligned}$$

Hence, $H_n(t) - H_n(t_0)$ is the sum of a negative deterministic trend of the order $(t - t_0)^2$ and a centered random term of the order $O_p(n^{-1/2}|t - t_0|^{1/2})$, as $t \rightarrow t_0$. Because $\widehat{U}_n(a)$ maximizes $H_n(t) - H_n(t_0)$, we must have that at $t = \widehat{U}_n(a)$, the trend is in absolute value of the same order as the random term, whence $\widehat{U}_n(a) - t_0$ is of order $n^{-1/3}$. This gives the intuition for the rate of convergence of \widehat{U}_n , and it turns out that the same rate applies for the direct estimator \widehat{f}_n .

More thorough arguments show that the rate is $n^{1/3}$ also in terms of centered absolute moments: under appropriate assumptions, for $p \geq 1$, there exists $K_p > 0$, such that for all $n, t \in [n^{-1/3}, 1 - n^{-1/3}]$, and $a \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}[|\widehat{\lambda}_n(t) - \lambda_0(t)|^p] &\leq K_p n^{-p/3} \quad \text{and} \\ \mathbb{E}[|\widehat{U}_n(a) - g_0(a)|^p] &\leq K_p n^{-p/3}, \end{aligned} \tag{9}$$

where $\widehat{\lambda}_n$ is a Grenander-type or isotonic estimator, \widehat{U}_n and g_0 are the pseudo-inverses of $\widehat{\lambda}_n$ and λ_0 , respectively. The first result of that type is [25], Theorem 1, for the direct estimator and $p \in [1, 2)$ in a general setting of Grenander-type estimators (the setting covers the MLE of a monotone density, the LSE of a monotone regression function on a uniform fixed design, and also the estimator introduced in [56]). The first inequality in (9) is extended to the current status model in [48] (see (11.32) and (11.33)) for the direct MLE, whereas the second inequality follows from [48], Theorem 11.3, for the inverse process, for all $p \geq 1$. The results in (9)

have been extended for $p = 2$ in [73] to isotonic estimators of a baseline hazard in the Cox model, and for all $p \geq 1$, in [8], Theorems 4.1 and 4.5, to the settings of regression on a random design with subgaussian errors, and estimation of a density or monotone failure rate under right censoring.

It follows from the Fubini theorem that for any integrable random variable Z ,

$$\mathbb{E}|Z| = \int_0^\infty \mathbb{P}(|Z| \geq t) dt,$$

so inequalities as in (9) typically follow from suitable bounds on the tail probabilities $\mathbb{P}(n^{1/3}|\widehat{U}_n(a) - g_0(a)| \geq x)$ for the inverse process. Bounds for $\widehat{\lambda}_n$ are then obtained by combining these bounds with (2). For example, [25], Lemma 2, provides a polynomial bound of the order x^{-3} , whereas in other settings there exist constants K_1, K_2 , such that

$$(10) \quad \mathbb{P}(n^{1/3}|\widehat{U}_n(a) - g_0(a)| \geq x) \leq K_1 \exp(-K_2 x^3)$$

for all $x > 0$; see, for example, [44], Theorem 2.1 or [8], Lemma 8.2, for particular Grenander-type estimators, and [48], Theorem 11.3 or [8], Lemma 8.1, and for other isotonic estimators. The proof of (10) typically relies on martingale theory and Doob's inequality. For a Grenander-type inverse \widehat{U}_n , such an inequality can be obtained based on Remark 2.1 by first proving a similar inequality for the corresponding Grenander-type inverse $\widehat{\gamma}_n^{-1}$ such that $\widehat{U}_n = G_n^{-1} \circ \widehat{\gamma}_n^{-1}$, and combining with a sharp control of the difference between G_n^{-1} and G^{-1} ; see [26, 8].

3.2 Local Asymptotic Distribution: The Direct Approach

Below, we give the main steps of the proof of (8) taken from the pioneering paper [83] in the density case. We consider the Grenander estimator \widehat{f}_n based on an i.i.d. sample of size n from a nonincreasing density f_0 on $[0, \infty)$: \widehat{f}_n is the left-slope of the LCM of F_n , the empirical distribution function. Let $t_0 > 0$ be some fixed point, such that f_0 has a continuous strictly negative derivative in the neighborhood of t_0 [and thus $f_0(t_0) > 0$]. The first step is to localize. By means of [83], Lemma 4.1, $\widehat{f}_n(t_0)$ is equal to the left-slope at t_0 , say $f_{n,c}^*(t_0)$, of the LCM of the restriction of F_n to the interval with center t_0 and width $2cn^{-1/3}$, with arbitrarily large probability by the choice of c . The intuition is that because F_n is close to the distribution function F_0 , which is strictly concave at t_0 , the LCM of F_n at t_0 is not influenced by the values away from

t_0 . The width $2cn^{-1/3}$ is chosen to match the rate of convergence.

Next, $f_{n,c}^*(t_0) - f_0(t_0)$ is the slope at $\delta = 0$ of the LCM of the process

$$\{F_n(t_0 + \delta) - F_n(t_0) - \delta f_0(t_0), \\ \delta \in [-2cn^{-1/3}, 2cn^{-1/3}]\}.$$

With arbitrary $D, r_n > 0$, it follows that $(Dr_n)^{-1} \cdot (f_{n,c}^*(t_0) - f_0(t_0))$ is the slope at $\delta = 0$ of the LCM of the process X_n on $[-d_n, d_n]$, where $d_n = 2cn^{-1/3}/r_n$ and

$$X_n(\delta) = r_n^{-2} D^{-1} (F_n(t_0 + \delta r_n) - F_n(t_0)) \\ - (Dr_n)^{-1} \delta f_0(t_0).$$

With $D = -f'(t_0)/2$ and $r_n^3 = f(t_0)D^{-2}/n$, the process X_n converges on $[-d_c, d_c]$ to X , where $d_c = c(2|f_0'(t_0)|^2/f_0(t_0))^{1/3}$ and $X(\delta) = W(\delta) - \delta^2$. Letting $c \rightarrow \infty$ (so that $d_c \rightarrow \infty$), it can then be derived that $(Dr_n)^{-1}(\widehat{f}_n(t_0) - f_0(t_0))$ has the same asymptotic distribution (as $n \rightarrow \infty$ and $c \rightarrow \infty$) as $(Dr_n)^{-1}(f_{n,c}^*(t_0) - f_0(t_0))$, which is the slope at $\delta = 0$ of the LCM of X on \mathbb{R} . The latter is precisely the distribution of $2V(0)$; see Section 2.3, yielding (8) with $C_0 = (|f_0'(t_0)|f_0(t_0)/2)^{-1/3}$.

The original reasoning in [83] uses that, if a sequence of concave functions x_n on an interval $[-q, q]$ uniformly converges to a function x , which is differentiable at zero (this is the interpretation that we give to the words ‘‘has a unique slope’’ in [83], Section 6), then the left-hand slope at zero of x_n converges to $x'(0)$. Uniqueness of the limiting slope is not given in the paper, whereas the result does not hold without this condition. A counter-example is given by $x_n(t) = |t + 1/n|$, for $n \geq 1$ and $x(t) = |t|$. Hence, there is a small gap in the original proof.

In [84, 95], (8) is proved with a different value of C_0 for the MLE of a monotone failure rate, and the LSE of a monotone regression function. The condition $f_0(t_0) > 0$ in the density case is replaced in the regression case by the condition that the variance of the observations at the design point t_0 is strictly positive. Hence in both models, the condition relies on some variability of the observations. The results in [95] have been extended in [69].

A direct approach similar to the one in [83] is used in [3], who consider Grenander-type estimators obtained as the greatest convex minorants (and derivatives thereof) of partial sum and empirical processes for independent, weakly dependent and long range dependent data. For other extensions of the direct approach to dependent data, see [4, 21, 5].

3.3 Local Asymptotic Distribution: The Inverse Approach

The earliest reference that uses the inverse approach seems to be [40], where the convergence in (8), was reproved using the *inverse process*; see Section 2.2. Thereafter, similar approaches were used in [52, 56, 93, 75] for establishing pointwise convergence of various Grenander-type estimators. We describe the approach again for the Grenander estimator \hat{f}_n for a nonincreasing density f_0 on $[0, \infty)$. It consists in first computing the limit distribution of the inverse process at an appropriate point, and then going back to \hat{f}_n thanks to (2). To be more precise, for all $C_0 > 0$, $x \in \mathbb{R}$ and $t_0 > 0$ we have

$$(11) \quad \begin{aligned} &\mathbb{P}(C_0 n^{1/3}(\hat{f}_n(t_0) - f_0(t_0)) < x) \\ &= \mathbb{P}(\hat{U}_n(a_0 + \delta_n) < t_0), \end{aligned}$$

where $a_0 = f_0(t_0)$ and $\delta_n = x n^{-1/3} / C_0$. Now, for arbitrary $A > 0$, $n^{1/3} A^{-1}(\hat{U}_n(a_0 + \delta_n) - t_0)$ is the location of the maximum of

$$\begin{aligned} &\{F_n(t_0 + A n^{-1/3} t) - (a_0 + \delta_n)(t_0 + A n^{-1/3} t), \\ &t \geq -t_0 n^{1/3} A^{-1}\} \end{aligned}$$

and because the location of the maximum is invariant by addition of constants or multiplication by a positive scalar, it is also the location of the maximum of $\{Z_n(t), t \geq -t_0 n^{1/3} A^{-1}\}$, where

$$\begin{aligned} Z_n(t) &= n^{2/3} (a_0 A)^{-1/2} (F_n(t_0 + A n^{-1/3} t) \\ &\quad - F_n(t_0) - A n^{-1/3} (a_0 + \delta_n) t). \end{aligned}$$

But it follows from the embedding in [63] that uniformly in $t \in \mathbb{R}$,

$$(12) \quad \begin{aligned} F_n(t) &= F_0(t) + n^{-1/2} B_n(F_0(t)) \\ &\quad + O_p(n^{-1} \log n), \end{aligned}$$

where F_0 is the distribution function corresponding to f_0 and B_n is a standard Brownian bridge constructed on the same probability space as F_n . Combining this with a Taylor expansion, one obtains that with $A^{3/2} = 2a_0^{1/2} |f_0'(t_0)|^{-1}$ and $C_0 = (|f_0'(t_0)| f_0(t_0) / 2)^{-1/3}$, Z_n converges in distribution to Z on \mathbb{R} , where

$$Z(t) = W(t) - (t^2 + xt).$$

Hence, $n^{1/3} A^{-1}(\hat{U}_n(a_0 + \delta_n) - t_0)$ converges in distribution to the location of the maximum of $\{Z(t), t \in \mathbb{R}\}$, which is also the location of the maximum $V(-x/2)$

of $\{W(t) - (t + x/2)^2, t \in \mathbb{R}\}$. From (11), we conclude that as $n \rightarrow \infty$,

$$\begin{aligned} &\mathbb{P}(C_0 n^{1/3}(\hat{f}_n(t_0) - f_0(t_0)) < x) \\ &\rightarrow \mathbb{P}(V(-x/2) < 0) \\ &= \mathbb{P}(V(0) < x/2), \end{aligned}$$

by stationarity of $\zeta(a) = V(a) - a$ taken from (6). The result (8) follows because $V(0)$ has a density; see, for example, [39] or the more accessible version in [41].

The convergence of the process Z_n can be established in different ways. One way is invoke the embedding (12), but a similar convergence can be obtained even in models where no embedding is available. In fact, as long as we consider Grenander-type estimators, this approach extends to various settings where a similar embedding holds with $B_n(F_0(t))$ possibly replaced by another Gaussian process (which is typically either a Brownian Bridge as above or a Brownian motion with a given variance function). Another useful source to obtain convergence of the process Z_n are the results in [62], which apply to processes $t \mapsto n^{2/3} \mathbb{P}_n g_0(\cdot; t n^{-1/3})$, where \mathbb{P}_n is the empirical measure corresponding to the observations. This becomes particularly useful in statistical models where no embedding is available, such as estimation of the baseline distribution in the Cox model under monotonicity constraints; see [75].

Now, for isotonic estimators, with inverse of the form $\hat{U}_n = G_n^{-1} \circ \hat{\gamma}_n^{-1}$ (see Section 2.2) where G_n^{-1} typically converges at the rate $n^{-1/2}$ to a quantile function G^{-1} , and $\hat{\gamma}_n$ is a Grenander-type estimator, one can compute the limit distribution of $\hat{\gamma}_n^{-1}$ as above. Then use the delta method to derive the limit distribution of $G^{-1} \circ \hat{\gamma}_n^{-1}$, which is similar to that of $\hat{U}_n = G_n^{-1} \circ \hat{\gamma}_n^{-1}$. The limit distribution of $\hat{\lambda}_n$ in (6) follows from that of \hat{U}_n using the switch relation, as above.

3.4 Behavior on Flat Regions, Near Zero or at a Discontinuity Point

Above, we have shown that at a fixed point t_0 in the interior of the support of the function of interest f_0 , such that $f_0'(t_0) \neq 0$, Grenander-type or isotonic estimators typically converge at the rate $n^{1/3}$ to the Chernoff distribution. This no longer holds on flat regions, at discontinuity points, or at the boundary of the support.

If f_0 is the uniform density on $[0, 1]$, then the Grenander estimator converges at the rate $n^{1/2}$ to the slope of the LCM of the Brownian bridge at t_0 ; see [40], Remark 2.2. More generally, at a fixed point t_0

where the underlying density is constant and nonzero in some open neighborhood of t_0 , it converges at the rate $n^{1/2}$ to the convolution of a closed-form density and a normal density; see [38, 15]. It is shown in [69] that when the derivative vanishes, the rate of convergence and the limiting distribution depend on the order of the first nonzero derivative.

At the boundary of the support of the function of interest, or at discontinuity points, isotonic estimators are typically inconsistent; see [94, 2, 7]. However, the current status model is an exception. Here, the function of interest and its isotonic estimator are distribution functions, so at the boundaries the latter converges to the true values 0 and 1. Several modifications of isotonic estimators have been considered with the aim of consistently estimating the function of interest at the boundaries; see [88] for a penalized MLE of a monotone density, and [66] for a truncated estimator. Grenander-type estimators of a bounded monotone function λ_0 on $[0, 1]$ are typically stochastically bounded at the boundaries 0 and 1 (see [25], Lemma 1), but the rate $n^{1/3}$ is achieved only on the increasing interval $[n^{-1/3}, 1 - n^{-1/3}]$; for $t_n \in (0, n^{-1/3}]$, the rate of convergence at t_n is typically $(nt_n)^{-1/2}$ (and a similar behavior holds on $[1 - n^{-1/3}, 1)$), if the derivative λ'_0 is bounded away from infinity and zero; see [25], Theorem 1. The precise limiting behavior of the Grenander estimator at a point $t_n = cn^{-\alpha}$ for some $\alpha \in (1/3, 1)$ (so that again, $t_n \in (0, n^{-1/3}]$) is given in [66].

4. GLOBAL CONVERGENCE

We consider the limit behavior of global measures of deviation of Grenander-type and isotonic estimators from Definitions 2.1 and 2.2: the L_p -loss (for some $p \geq 1$) and the supremum loss. We discuss rates of convergence and limit distribution of the losses. We point out that in contrast to the methods for obtaining the pointwise limit distribution (see Section 3), all methods known in the literature to prove global convergence make use of the approach based on the inverse process and heavily depend on the availability of an embedding result for the process Λ_n into a Gaussian process, such as in (12).

4.1 Rate of Convergence of the L_p -Loss

The global rate of convergence can typically be obtained using entropy arguments if the estimator is the maximizer of an empirical criterion, which is the case for instance for the Grenander estimator \hat{f}_n of a nonincreasing bounded density function f_0 on

$[0, 1]$. The main tool for that is Theorem 2.7.5 in [91], which proves that the class of all nonincreasing functions $f : [0, 1] \rightarrow [0, 1]$ has ε -bracketing entropy for the $L_2(\mu)$ -norm of maximal order $1/\varepsilon$, where μ denotes the Lebesgue measure. Based on this entropy result, [91], Example 3.4.6, proves that the Hellinger distance $h(\hat{f}_n, f_0)$ between \hat{f}_n and f_0 , that is, the $L_2(\mu)$ -distance on the square roots of the functions, is of the order $O_p(n^{-1/3})$. Since f_0 is bounded by assumption and $\hat{f}_n(0)$ is stochastically bounded (see Section 3.4), we also obtain the rate of convergence in the L_2 -sense: by monotonicity,

$$\begin{aligned} \|\hat{f}_n - f_0\|_2 &\leq (\sqrt{\hat{f}_n(0)} + \sqrt{f_0(0)})h(\hat{f}_n, f_0) \\ &= O_p(n^{-1/3}). \end{aligned}$$

The above entropy bound can be used in various settings, where the function of interest is not necessarily a density function. For instance, it was used in [6] to compute the $n^{1/3}$ -rate of convergence of the LSE in the monotone single index model. Alternatively, one can integrate the inequalities in (9) over t in the interval $[n^{-1/3}, 1 - n^{-1/3}]$, together with bounds of the order $[n(t \wedge (1 - t))]^{-p/2}$ that hold outside this interval; see [25], Theorem 1, to get

$$\begin{aligned} &\mathbb{E}\left[\int_0^1 |\hat{\lambda}_n(t) - \lambda_0(t)|^p dt\right] \\ &= \int_0^1 \mathbb{E}[|\hat{\lambda}_n(t) - \lambda_0(t)|^p] dt \\ &\leq Kn^{-p/3}, \end{aligned}$$

so the L_p -loss of $\hat{\lambda}_n$ is $O_p(n^{-1/3})$. A similar result holds for the inverse \hat{U}_n .

4.2 Rate of Convergence of the Supremum Loss

Grenander-type estimators $\hat{\lambda}_n$ are typically inconsistent at the boundaries of their interval of support $[0, 1]$; see Section 3.4. Therefore, the supremum of $|\hat{\lambda}_n - \lambda_0|$ over $[0, 1]$ is typically dominated by the behavior at the boundaries and the range of the supremum has to be restricted to a subinterval in order to obtain pertinent results. In [60], Lemma 2.1, the rate of the supremum loss is established in a semi-parametric model for censored observations where the MLE is a Grenander-type estimator. In [29], Theorem 2.1, the rate of convergence for the supremum of $|\hat{\lambda}_n - \lambda_0|$ is shown to be of the order $(n/\log n)^{-1/3}$, on subintervals that can grow towards $[0, 1]$, as long as one stays away sufficiently far from the boundaries. The result is proved for Grenander-type estimators in a general setup that

includes Grenander’s estimator of a monotone density, the LSE for a monotone regression function on a uniform fix design, as well as the estimator for a monotone failure rate under random censoring, where λ_0 is assumed to have a first derivative that is bounded both away from zero and from infinity. In the course of the proof of their Lemma 5.9, [52], page 120, obtained the same rate for the supremum loss of the MLE for the distribution function of interest in the current status model. In this particular model, the supremum can be taken over the whole interval of interest since the MLE for the distribution function is consistent at the boundaries, due to known values 0 and 1.

Below is a brief sketch of proof for the rate, extracted from [29]. It follows from the triangle inequality that for two consecutive jump points τ_{i-1}, τ_i of $\widehat{\lambda}_n$,

$$(13) \quad \begin{aligned} & \sup_{u \in (\tau_{i-1}, \tau_i]} |\widehat{\lambda}_n(u) - \lambda_0(u)| \\ & \leq |\widehat{\lambda}_n(\tau_i) - \lambda_0(\tau_i)| + K|\tau_{i-1} - \tau_i|, \end{aligned}$$

where $K = \|\lambda'_0\|_\infty$, using that $\widehat{\lambda}_n(u) = \widehat{\lambda}_n(\tau_i)$ for all $u \in (\tau_{i-1}, \tau_i]$. Now, using monotonicity and dividing $[\lambda_0(1), \lambda_0(0)]$ into a union of disjoint intervals of length of order $n^{-1/3}$, one can derive from (10) that

$$(14) \quad \begin{aligned} & \sup_{a \in [\lambda_0(1), \lambda_0(0)]} |\widehat{U}_n(a) - g_0(a)| \\ & = O_p(n/\log n)^{-1/3}. \end{aligned}$$

The height of a jump of the piecewise constant process \widehat{U}_n is precisely the width of a step of $\widehat{\lambda}_n$, so for a continuous g_0 , (14) implies that $\max_i |\tau_{i-1} - \tau_i| = O_p(n/\log n)^{-1/3}$. Moreover, with $\gamma_i := \widehat{\lambda}_n(\tau_i)$, we have $\tau_i = \widehat{U}_n(\gamma_i)$ and $\gamma_i = \lambda_0 \circ g_0(\gamma_i)$, provided that γ_i belongs to the range of λ_0 , which can be proved to happen with high probability if τ_i is far enough from the boundaries of $[0, 1]$. So (13) combined with the Taylor expansion yields that uniformly,

$$\begin{aligned} \sup_{u \in (\tau_{i-1}, \tau_i]} |\widehat{\lambda}_n(u) - \lambda_0(u)| & \leq K|g_0(\gamma_i) - \widehat{U}_n(\gamma_i)| \\ & + O_p\left(\frac{\log n}{n}\right)^{1/3}. \end{aligned}$$

Taking the maximum over appropriate indexes i , and using again (14), proves that for α_n, β_n of order $n^{-1/3}(\log n)^{-2/3}$, the supremum of $|\widehat{\lambda}_n - \lambda_0|$ over $(\alpha_n, 1 - \beta_n)$ is of order $O_p(n/\log n)^{-1/3}$; see [29], Theorem 2.1. Note that the switch relation is not used here to make the connection between $\widehat{\lambda}_n$ and the inverse \widehat{U}_n . Instead, the connection between the jump points and steps of the two estimators is used.

4.3 Limit Distribution of the L_p -Loss

One motivation for studying the limit distribution of the L_p -loss comes from goodness-of-fit tests. Assume that we wish to test $H_0 : \lambda = \lambda_0$ for a given monotone λ_0 , against the nonparametric alternative that λ is monotone, but different from λ_0 . When the properly normalized L_p -distance between a monotone estimate $\widehat{\lambda}_n$ and λ_0 can be shown to have a limiting distribution under H_0 , one can build a test of level α that rejects H_0 if this normalization exceeds the $(1 - \alpha)$ -quantile of the limiting distribution; see [33]. In this section we will describe that with proper standardization, the L_p -loss converges in distribution under H_0 to a standard Gaussian law.

Assume that λ_0 has a first derivative that is bounded away from zero and from infinity on $[0, 1]$. We show below that the normalized L_p -loss of isotonic estimators

$$(15) \quad \mathcal{J}_n := n^{p/3} \int_0^1 |\widehat{\lambda}_n(t) - \lambda_0(t)|^p dt$$

typically converges, after appropriate scaling, at $n^{1/6}$ -rate to a centered Gaussian distribution. The earliest reference [40] of this type of result concerns the L_1 -loss of the Grenander estimator of a monotone density on $[0, 1]$. A complete proof of the result announced in [40] is provided in [44]. Other papers in this direction are concerned with monotone regression [23], and the L_p -loss for monotone density [65], and the L_p -loss [25] in a general setup. This setup considers Grenander-type estimators as in Definition 2.1, and assumes that an embedding holds for Λ_n in the sense that there exists either a Brownian motion or a Brownian bridge B_n , such that Λ_n can be approximated by a suitable Gaussian process:

$$(16) \quad \Lambda_n(t) \approx \Lambda_0(t) + n^{-1/2} B_n \circ L(t),$$

where the possibly unknown $L : [0, 1] \rightarrow \mathbb{R}$ is smooth increasing such that $L(0) = 0$. The setup covers several standard settings of isotonic estimation, such as monotone density, monotone regression on a fix uniform design, and monotone failure rate. The link between the Brownian motion and the Brownian bridge cases can be made using the representation

$$(17) \quad B_n(t) = W_n(t) - \xi_n t, \quad t \in [0, 1],$$

where W_n is a standard Brownian motion, $\xi_n \equiv 0$ if B_n is a Brownian motion, and ξ_n is a standard Gaussian variable independent of B_n , if B_n is a Brownian

bridge. In this setup, it is proved that under appropriate assumptions,

$$(18) \quad n^{1/6}(\mathcal{J}_n - m_p) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_p^2),$$

as $n \rightarrow \infty$, where \mathcal{J}_n is defined in (15), $m_p = \mathbb{E}|\zeta(0)|^p \int_0^1 |4\lambda'_0(t)L'(t)|^{p/3} dt$, and

$$(19) \quad \begin{aligned} \sigma_p^2 &= 8k_p \int_0^1 |4\lambda'_0(t)L'(t)|^{2(p-1)/3} L'(t) dt, \\ k_p &= \int_0^\infty \text{cov}(|\zeta(0)|^p, |\zeta(a)|^p) da \end{aligned}$$

with ζ defined in (6). Note that in the monotone density model, $L' = \lambda_0$, so that in the case $p = 1$, the limiting variance is $8k_1$, which does not depend on λ_0 .

We begin with heuristics in the simplest case where L is the identity function and $\widehat{\lambda}_n$ is a Grenander-type estimator. Let $g_0 = \lambda_0^{-1}$. The first step is to use the switch relation to go from $\widehat{\lambda}_n$ to the inverse \widehat{U}_n :

$$(20) \quad \mathcal{J}_n \approx n^{p/3} \int_{\lambda_0(1)}^{\lambda_0(0)} |\widehat{U}_n(a) - g_0(a)|^p |g'_0(a)|^{1-p} da.$$

For all a , $n^{1/3}(\widehat{U}_n(a) - g_0(a))$ is stochastically bounded (see Section 3.3) and equals

$$\begin{aligned} &\operatorname{argmax}_{g_0(a)+n^{-1/3}u \in [0,1]} \{ \Lambda_n(g_0(a) + n^{-1/3}u) \\ &\quad - a(g_0(a) + n^{-1/3}u) \}, \end{aligned}$$

due to (3). The location of the maximum does not change by a vertical shift of the whole function. Likewise, we may also multiply the process by $n^{2/3}$, so $n^{1/3}(\widehat{U}_n(a) - g_0(a))$ is the location of the maximum of

$$\begin{aligned} Z_n(u) &= n^{2/3}(\Lambda_n(g_0(a) + n^{-1/3}u) - \Lambda_n(g_0(a))) \\ &\quad - an^{1/3}u, \end{aligned}$$

for $-n^{1/3}g_0(a) \leq u \leq n^{1/3}(1 - g_0(a))$. To control the range of u , we do a localization step and restrict ourselves to $|u| \leq \log n$. This is possible provided that $g_0(a)$ is far enough from the boundaries, because the location of the maximum of the process $Z_n(u)$, for $|u| \leq \log n$, can only be different from $n^{1/3}(\widehat{U}_n(a) - g_0(a))$ if the latter is greater than $\log n$ in absolute value. Due to the bounds on the tail probabilities for $n^{1/3}(\widehat{U}_n(a) - g_0(a))$, such as the one in (10), this only happens with very small probability.

If (16) holds with L the identity function, then by means of Taylor's expansion and representation (17), using that $a = \Lambda'_0(g_0(a))$ we get

$$\begin{aligned} Z_n(u) &\approx n^{1/6} \{ W_n(g_0(a) + n^{-1/3}u) - W_n(g_0(a)) \} \\ &\quad + \frac{1}{2} \lambda'_0(g_0(a)) u^2 \end{aligned}$$

for $|u| \leq \log n$, and therefore, $n^{1/3}|\lambda'_0(g_0(a))/2|^{2/3} \cdot (\widehat{U}_n(a) - g_0(a))$ can be approximated by

$$(21) \quad \begin{aligned} &\zeta_n(g_0(a)) \\ &:= \operatorname{argmax}_{|u| \leq |\lambda'_0(g_0(a))/2|^{2/3} \log n} \{ W_{g_0(a)}(u) - u^2 \}, \end{aligned}$$

where for every $t \in [0, 1]$, W_t is the Brownian motion defined by

$$\begin{aligned} W_t(u) &= n^{1/6} |\lambda'_0(t)/2|^{1/3} \{ W_n(t + n^{-1/3} |\lambda'_0(t)/2|^{-2/3} u) \\ &\quad - W_n(t) \}. \end{aligned}$$

Hence,

$$(22) \quad \begin{aligned} \mathcal{J}_n &\approx 2^{2p/3} \int_{\lambda_0(1)}^{\lambda_0(0)} |\zeta_n(g_0(a))|^p |g'_0(a)|^{1-p/3} da \\ &= \int_0^1 h(t) |\zeta_n(t)|^p dt, \end{aligned}$$

where $h = (4|\lambda'_0|)^{p/3}$. Now, note that $\zeta_n(t)$ depends only on the increments of W_n on an interval with center t and width of order $n^{-1/3} \log n$. The increments of the Brownian motion W_n are independent, so for all $t > s + n^{-1/3}(\log n)^\beta$, with some $\beta > 1$, the processes $\{\zeta_n(u), u \leq s\}$ and $\{\zeta_n(u), u \geq t\}$ are independent. Hence,

$$\begin{aligned} &\operatorname{var} \left(\int_0^1 h(x) |\zeta_n(x)|^p dx \right) \\ &= 2 \int_0^1 \int_s^{s+n^{-1/3}(\log n)^\beta} h(t)h(s) \operatorname{cov}(|\zeta_n(t)|^p, \\ &\quad |\zeta_n(s)|^p) dt ds, \end{aligned}$$

using Fubini's theorem. Moreover, for every $s \in [0, 1]$ and $a \in \mathbb{R}$, let $\zeta_s(a)$ be the location of the maximum of $\{W_s(u+a) - u^2, u \in \mathbb{R}\}$. Then, the processes ζ_s all have the same distribution as ζ in (6) and

$$(23) \quad \zeta_n(s + n^{-1/3}t) \approx \zeta_s(|\lambda'_0(s)/2|^{2/3}t).$$

Therefore, by change of variable, the above variance is approximately equal to

$$\begin{aligned} &2n^{-1/3} \int_0^1 \int_0^{(\log n)^\beta} h^2(s) \operatorname{cov}(|\zeta_s(0)|^p, \\ &\quad |\zeta_s(|\lambda'_0(s)/2|^{2/3}t)|^p) dt ds, \end{aligned}$$

provided that h is smooth. When multiplied by $n^{1/3}$, this converges to σ_p^2 in (19), taking L equal to the identity. This explains the rate of convergence in (18).

Convergence to a Gaussian distribution can then be proved using the method of big-blocks-small-blocks that we describe now. Partition $[0, 1]$ into intervals of alternating “big” and “small” size, say B_i and S_i , so that thanks to (22),

$$\mathcal{J}_n \approx \sum_i \int_{B_i} h(t)|\zeta_n(t)|^p dt + \sum_i \int_{S_i} h(t)|\zeta_n(t)|^p dt.$$

The size of the small blocks should be chosen large enough such that the integrals over the big blocks B_i become independent, but small enough in comparison with the size of the big blocks so that the summation over the small blocks S_i is negligible. One possibility (see [25]) is to take big blocks of length $n^{-1/3}(\log n)^5$ and small blocks of length $n^{-1/3}(\log n)^2$. When the contribution of the small blocks is negligible, then $n^{1/6}(\mathcal{J}_n - \mathbb{E}[\mathcal{J}_n])$ is asymptotically equivalent to the contribution of the big blocks, which is a sum of independent centered variables. Under appropriate assumptions, this converges to a centered Gaussian law with variance σ_p^2 , by the Lindeberg–Feller central limit theorem. Now, due to (23), $\zeta_n(t) \approx \zeta_t(0)$, which is distributed like ζ in (6), so (at least heuristically) $\mathbb{E}|\zeta_n(t)|^p \approx \mathbb{E}|\zeta(0)|^p$ and from (22), $\mathbb{E}[\mathcal{J}_n] \approx \mathbb{E}|\zeta(0)|^p \int_0^1 h(t) dt = m_p$, whence (18).

We point out some of the difficulties to make the above heuristics rigorous. If $\hat{\lambda}_n$ is stochastically bounded, but not consistent at the boundaries, then the contribution of $n^{p/3} \int_0^{n^{-1}} |\hat{\lambda}_n(t) - \lambda_0(t)|^p dt$ is of order $n^{p/3-1}$, which is negligible with respect to $n^{-1/6}$ if $p < 5/2$. The same comment holds for the right boundary. In such a case, one has to consider $p \in [1, 5/2)$. Another difficulty is that we need a much more refined result than the approximation by (21) to ensure (22): we need to connect the rate of convergence of processes to the rate of convergence of their location of maximum. This can be done thanks to [23], Proposition 1.

We now consider the more difficult case of isotonic estimators. As a typical example, we consider the LSE of a decreasing regression function λ_0 on $[0, 1]$ described in Section 2.1, where (18) holds with a different form for m_p and σ_p^2 ; see [26], Theorem 3. To deal with the inverse Grenander-type estimator $\hat{\gamma}_n^{-1}$ from (5), which is connected to \hat{U}_n through $\hat{U}_n = G_n^{-1} \circ \hat{\gamma}_n^{-1}$ (see Section 2.2), we would like to approximate Γ_n by a Gaussian process, similar to (16). However, we are not aware of such an embedding for Γ_n . Instead, we have

$$(24) \quad \Gamma_n(t) \approx \int_0^t \lambda_0 \circ G_n^{-1}(u) du + \frac{1}{\sqrt{n}} W_n(L_n(t))$$

conditionally on the design points, where $L_n(t) = \int_0^t \sigma^2 \circ G_n^{-1}(u) du$, $\sigma^2(X)$ is the conditional variance of an observation given the design point X , and G_n^{-1} is the empirical quantile function of the design points. The function L_n is close to the smooth function $t \mapsto \int_0^t \sigma^2 \circ G^{-1}(u) du$, where G^{-1} is the common quantile function of the design points, but even so, it is not smooth. Likewise, the centering $t \mapsto \int_0^t \lambda_0 \circ G_n^{-1}(u) du$ of the Gaussian process in (24) is close to the smooth function $t \mapsto \int_0^t \lambda_0 \circ G^{-1}(u) du$ but is not smooth either, so Taylor’s expansions cannot be performed as before. Nevertheless, at the price of additional technicalities, the big-blocks-small-blocks method described above shows that the L_p -loss of $\hat{\gamma}_n^{-1}$ in (5) is asymptotically Gaussian. Going from this L_p -loss to that of $\hat{U}_n = G_n^{-1} \circ \hat{\gamma}_n^{-1}$ then requires a sharp control of the difference $G_n^{-1} - G^{-1}$.

Recently, a CLT for the Hellinger-loss of a Grenander-type estimator $\hat{\lambda}_n$ was established in [70] in the same general setup as in [25] using that the squared Hellinger distance between $\hat{\lambda}_n$ and the function of interest λ_0 is approximately equal to a weighted L_2 -loss:

$$(25) \quad \int_0^1 \left(\frac{\hat{\lambda}_n(t) - \lambda_0(t)}{\sqrt{\hat{\lambda}_n(t) + \lambda_0(t)}} \right)^2 dt \approx \int_0^1 (\hat{\lambda}_n(t) - \lambda_0(t))^2 (4\lambda_0(t))^{-1} dt.$$

It was proven along the lines of the arguments given above, that

$$n^{1/6} \{n^{1/3} h(\hat{\lambda}_n, \lambda_0) - \tilde{m}_2\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\sigma}^2),$$

as $n \rightarrow \infty$, where h denotes the Hellinger distance,

$$\tilde{m}_2^2 = \mathbb{E}[|\zeta(0)|^2] \int_0^1 \frac{|\lambda_0'(t)L'(t)|^{2/3}}{4^{1/3}\lambda_0(t)} dt, \quad \tilde{\sigma}^2 = \frac{2^{1/3}k_2}{4\tilde{m}_2^2} \int_0^1 \frac{|\lambda_0'(t)L'(t)|^{2/3}L'(t)}{\lambda_0(t)^2} dt,$$

where L and k_2 are taken from (16) and (19). Note that in statistical models where $L' = \lambda_0$, such as the monotone density model (see [25], Theorem 6), the limiting variance $\tilde{\sigma}^2 = k_2/(2\mathbb{E}[|\zeta(0)|^2])$ does not depend on λ_0 . Hence, as conjectured in [92], the limiting variance of the Hellinger loss for the Grenander estimator does not depend on the underlying density. This coincides with that of the limiting variance in the central limit theorem for the L_1 -error for the Grenander estimator.

4.4 Limit Distribution of the Supremum Loss

While pointwise confidence intervals for a monotone function are available using the limiting distribution of the isotonic estimator at the fixed point, nonparametric confidence bands have remained a formidable challenge. Although the limiting distribution of the L_p -loss may have applications in goodness-of-fit testing, it cannot be used straightforwardly to construct uniform confidence bands for a monotone function λ_0 . For this purpose, the limit distribution of the supremum loss seems more appropriate. It is the purpose of this section to describe how this limit distribution can be obtained in the same general framework as considered in [25].

Again, assume that λ_0 has a first derivative that is bounded away from zero and from infinity on $[0, 1]$. By [29], Theorem 2.2, a suitably standardized supremum of $|\widehat{\lambda}_n - \lambda_0|$ converges to a Gumbel law, where $\widehat{\lambda}_n$ is a Grenander-type estimator as in Definition 2.1 in a general setup where (16) holds with B_n a Brownian motion or a Brownian bridge and L a strictly increasing function. The supremum is restricted to subintervals of $[0, 1]$, so that the inconsistency at the boundaries does not dominate the supremum; see Section 4.2. For all sequences $(\alpha_n)_n$ and $(\beta_n)_n$ satisfying $\alpha_n, \beta_n \rightarrow 0$, and $1 - v + \beta_n, u + \alpha_n \gg n^{-1/3}(\log n)^{-2/3}$, for some fixed $0 \leq u < v \leq 1$, we have

$$\mathbb{P}\left(\log n \left\{ \left(\frac{n}{\log n}\right)^{1/3} \sup_{t \in (u+\alpha_n, v-\beta_n]} \frac{|\widehat{\lambda}_n(t) - \lambda_0(t)|}{|2\lambda'_0(t)L'(t)|^{1/3}} - \mu_n \right\} \leq x\right) \rightarrow e^{-e^{-x}},$$

for all $x \in \mathbb{R}$, as $n \rightarrow \infty$, under appropriate assumptions, where

$$\begin{aligned} \mu_n &= 1 - \frac{\kappa}{2^{1/3}(\log n)^{2/3}} \\ &+ \frac{1}{\log n} \left[\frac{\log \log n}{3} \right. \\ &\left. + \log\left(2\lambda \int_u^v \frac{|\lambda'_0(t)|^{2/3}}{L'(t)^{1/3}} dt\right) \right], \end{aligned} \tag{26}$$

with λ, κ taken from (7). The largest possible interval, with $u = 0, v = 1$, stays away from the boundaries at distance larger than $n^{-1/3}(\log n)^{-2/3}$.

We outline the proof in the simplest case, where B_n in (16) is a Brownian motion and $L(t) = t$. Again, the approach consists of first obtaining an analogous result for the supremum between the inverses \widehat{U}_n and $g_0 :=$

λ_0^{-1} :

$$S_n := n^{1/3} \sup_{a \in [\lambda_0(v-\beta_n), \lambda_0(u+\alpha_n)]} \left| \frac{\lambda'_0(g_0(a))}{2} \right|^{2/3} \cdot |\widehat{U}_n(a) - g_0(a)|. \tag{27}$$

Here, the normalisation is such that variables in the supremum can be approximated by variables $\zeta_n(g_0(a))$ as in (21). Similar to Section 4.3, we partition $[\lambda_0(v - \beta_n), \lambda_0(u + \alpha_n)]$ into blocks of alternating big and small size, where here the big blocks have length $2n^{-1/3}(\log n)^2$, whereas the small blocks have length of order $n^{-1/3}(\log n)$. It can be shown that the contribution of the small blocks is negligible in the sense that the above supremum over $[\lambda_0(v - \beta_n), \lambda_0(u + \alpha_n)]$ is asymptotically equivalent to the supremum over the union of the big blocks. Moreover, the suprema over the various big blocks are asymptotically independent.

Now, consider a particular big block B_i and denote its center by b . Similar to (21), with $V_n(a) = n^{1/3}(\widehat{U}_n(a) - g_0(a))$, for all $a \in B_i$ we have

$$V_n(a) \approx \operatorname{argmax}_{|u| \leq \log n} \{W_a(u) - D_n(a, u)\}, \tag{28}$$

where $D_n(a, u) \approx |\lambda'_0(g_0(a))|u^2/2$, and W_a is the Brownian motion defined by

$$W_a(u) = n^{1/6} \{W_n(g_0(a) + n^{-1/3}u) - W_n(g_0(a))\}.$$

Because $|a - b| \leq n^{-1/3}(\log n)^2$, for $a \in B_i$, the drift $D_n(a, u)$ in (28) can be approximated by the same drift $u \mapsto -|\lambda'_0(g_0(b))|u^2/2$. The idea now, is to sandwich $V_n(a)$ between two similar quantities, for which we can show that they have the same limit behavior. To this end, we use a trick due to [66], Lemma 2.1, to show that

$$V_n^-(a, b) \lesssim V_n(a) \lesssim V_n^+(a, b) \tag{29}$$

for all $a \in B_i$, where $V_n^-(a, b)$ and $V_n^+(a, b)$ are defined similarly to the right-hand side of (28), but with $D_n(a, u)$ replaced by $(|\lambda'_0(g_0(b))|/2 + 2\varepsilon_n)u^2$ and $(|\lambda'_0(g_0(b))|/2 - 2\varepsilon_n)u^2$, respectively, where $\varepsilon_n = 1/\log n$. The purpose of this is that when we will vary a over B_i and fix b to be the midpoint of this interval, we will obtain variables $V_n^+(a, b)$ each defined with the same drift

$$u \mapsto -\left(\frac{|\lambda'_0(g_0(b))|}{2} - 2\varepsilon_n\right)u^2,$$

and a Brownian motion W_a only depending on a . Similarly for $V_n^-(a, b)$. Now, by definition of W_a and since

the argmax is invariant by addition of constants,

$$V_n^+(a, b) = \operatorname{argmax}_{|u| \leq \log n} \left\{ W_b(u + n^{1/3}(g_0(a) - g_0(b))) - \left(\frac{|\lambda'_0(g_0(b))|}{2} - 2\varepsilon_n \right) u^2 \right\},$$

where only the Brownian motion W_b is involved. Then, similar to (21) and (23), the process $a \mapsto |\lambda'_0(g_0(b))/2|^{2/3} V_n^+(a, b)$ on B_i can be approximated by ζ_i^+ on an interval Δ_{ni} , with length $(\log n)^2 \cdot |\lambda'_0(g_0(b))/2|^{-1/3}$, where ζ_i^+ is distributed as ζ in (6). In [55], Theorem 1.1, an extremal limit theorem has been obtained for suprema of the process ζ over increasing intervals. From this result, one can derive that for all $\delta_n \rightarrow \infty$, $\tau_n \rightarrow 0$, and $u_n \rightarrow \infty$, such that $u_n/\delta_n \rightarrow 0$, $\log(\tau_n)/\delta_n^3 \rightarrow 0$, and $\delta_n f_C(u_n)/\tau_n \rightarrow 1$, where f_C denotes the density of $\zeta(0)$, one has

$$-(2\tau_n)^{-1} \log \mathbb{P} \left(\sup_{c \in [0, \delta_n]} |\zeta_i^+(c)| \leq u_n \right) - 1 \rightarrow 0.$$

Since the processes $a \mapsto V_n^+(a, b)$ over different blocks B_i are asymptotically independent, this can be used to show that for all sequences $(u_n)_n$, such that $u_n \rightarrow \infty$ in such a way that $n^{1/3} f_C(u_n) \rightarrow \tau > 0$,

$$\begin{aligned} & \mathbb{P} \left(\sup_i \sup_{a \in B_i} \left| \frac{\lambda'_0(g_0(b))}{2} \right|^{2/3} |V_n^+(a, b)| \leq u_n \right) \\ & \approx \prod_i \mathbb{P} \left(\sup_{c \in \Delta_{ni}} |\zeta_i^+(c)| \leq u_n \right) \\ & \rightarrow \exp \left\{ -2\tau \int_u^v |\lambda'_0(t)/2|^{2/3} dt \right\}, \end{aligned}$$

as $n \rightarrow \infty$. The same result holds with V_n^- instead of V_n^+ , so thanks to (29), it also holds with V_n , that is, for S_n defined in (27),

$$\mathbb{P}\{S_n \leq u_n\} \rightarrow \exp \left\{ -2\tau \int_u^v |\lambda'_0(t)/2|^{2/3} dt \right\}.$$

Now, fix $x \in \mathbb{R}$ and choose τ , such that $2\tau \int_u^v |\lambda'_0(t)/2|^{2/3} dt = e^{-x}$. Then, using that $n^{1/3} f_C(u_n) \rightarrow \tau$, together with the expansion of $f_C(u)$ in (7), we choose $u_n = x/a_n + b_n + o(\log n)^{-2/3}$, with $a_n = 2^{1/3}(\log n)^{2/3}$ and $b_n = 2^{-1/3}(\log n)^{1/3} \mu_n$, where μ_n is defined in (26). It then follows that, for all $x \in \mathbb{R}$,

$$\begin{aligned} (30) \quad & \mathbb{P} \left\{ \log n \left\{ \left(\frac{2}{\log n} \right)^{1/3} S_n - \mu_n \right\} \leq x \right\} \\ & \rightarrow \exp \{ -e^{-x} \}. \end{aligned}$$

By using the connection between the jump points and steps of $\widehat{\lambda}_n$ and \widehat{U}_n (see Section 4.2), the scaled supremum distance between the inverses \widehat{U}_n and g_0 can be

related to a scaled supremum between the functions $\widehat{\lambda}_n$ and λ_0 themselves,

$$S_n = \sup_{t \in [u + \alpha_n, v - \beta_n]} \frac{n^{1/3} |\widehat{\lambda}_n(t) - \lambda_0(t)|}{|4\lambda'_0(t)|^{1/3}} + o_p(\log n)^{2/3}$$

(see [29], Lemma 5.2). As a consequence, the result in (30) can be transferred to the one for the scaled supremum between $\widehat{\lambda}_n$ and λ_0 .

5. SMOOTH ISOTONIC ESTIMATION

The isotonic estimators of Definitions 2.1 and 2.2 are step functions that exhibit a non normal limit distribution at rate $n^{1/3}$. On the other hand, if one is willing to assume more regularity on the monotone function of interest, smooth estimators are sometimes preferred to piecewise constant ones, because they can be used to achieve a faster rate of convergence to a Gaussian distributional law and to estimate derivatives. Typically, these estimators are constructed by combining an isotonization step with a smoothing step. Below we discuss estimators constructed by first obtaining a smooth estimate and then isotonize, and estimators for which the order of the two steps are interchanged. Alternative methods, such as shape-constrained splines (see [89, 78]) will not be discussed in the paper.

5.1 Smoothing Followed by Isotonization

Smooth isotonic estimators of a monotone function λ_0 that are constructed by smoothing followed by an isotonization step have been considered in [19, 96, 36, 85], for the regression setting, in [35, 90] for estimating a monotone density, and in [71], for estimation of the baseline hazard in the Cox model. Comparisons between isotonized smooth estimators and smoothed isotonic estimators were made in [76] for the regression setting, and in [50] for the current status model. Isotonization of kernel smoothed piecewise constant estimators for both the regression setting and for estimating an increasing density, is considered by [3] in a general framework that includes independent, weakly dependent, and long range dependent observations.

One approach for first smoothing and then isotonize, are maximum smoothed likelihood estimators (MSLE), obtained by smoothing the loglikelihood and then maximizing the smoothed likelihood over all monotone functions; see [35, 50, 71]. For example, in the current status problem (see also Section 2.1), the loglikelihood

$$\ell(F) = \int (\delta \log F(t) + (1 - \delta) \log(1 - F(t))) d\mathbb{P}_n(\delta, t)$$

is smoothed by replacing the empirical measure \mathbb{P}_n by $\widehat{\mathbb{P}}_n(\delta, t) = \delta d\widehat{G}_{n,1}(t) + (1 - \delta)d\widehat{G}_{n,0}(t)$, where $\widehat{G}_{n,0}$ and $\widehat{G}_{n,1}$ are kernel smoothed versions of

$$\mathbb{G}_{n,0}(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{T_j \leq t, \Delta_j = 0\}} \quad \text{and}$$

$$\mathbb{G}_{n,1}(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{T_j \leq t, \Delta_j = 1\}},$$

which yields the smoothed loglikelihood

$$\ell^s(F) = \int \log(1 - F(t)) d\widehat{G}_{n,0}(t) + \int \log F(t) d\widehat{G}_{n,1}(t).$$

Another approach are Grenander-type estimators constructed by taking slopes of the LCM (or GCM) of a smooth estimator Λ_n^s for the primitive Λ_0 corresponding to the nonincreasing (or nondecreasing) function λ_0 of interest.

Both approaches lead to continuous isotonic estimators that typically can be characterized as slopes of the LCM (or GCM), either of a continuous random process Λ_n^s or of a CSD $t \mapsto (\mathbb{X}_n(t), \mathbb{Y}_n(t))$, constructed with continuous random processes \mathbb{X}_n and \mathbb{Y}_n . An example is given by the MSLE $\widehat{F}_n^{\text{MS}}$ for the distribution function F_0 in the current status problem, with $\mathbb{X}_n = \widehat{G}_{n,0} + \widehat{G}_{n,1}$ and $\mathbb{Y}_n = \widehat{G}_{n,1}$ (see [50], Theorem 3.1), and also for the MSLE of a nondecreasing baseline hazard λ_0 in the Cox model (see Lemma 3.1 in [71]). These authors also investigated a Grenander-type estimator, defined as the slope of the GCM of

$$(31) \quad \Lambda_n^s(t) = \int k_b(t - u) \Lambda_n(u) du,$$

where $k_b(s) = b^{-1}k(s/b)$ is a scaled kernel function with bandwidth b (depending on n) and Λ_n is the Breslow estimator for the cumulative baseline hazard Λ_0 . This is similar to the estimator for a nonincreasing density in [90], defined as the slope of the GCM of a kernel smoothed empirical distribution function.

To establish the limit behavior of the isotonized smooth estimators that emerge from both approaches, the key idea is that these estimators are the solution of a continuous isotonic regression problem considered in [45]. For example, the Grenander-type estimator for a nondecreasing baseline hazard λ_0 in the Cox model minimizes

$$\lambda \mapsto \int (\lambda_n^s(t) - \lambda(t))^2 dt,$$

over all nondecreasing functions λ , where $\lambda_n^s(t) = d\Lambda_n^s(t)/dt$; see Lemma 1 in [45]. This suggests λ_n^s as a naive smooth estimator for λ_0 . In the current status problem, Theorem 1 in [45] shows that the MSLE for F_0 is also the minimizer of

$$F \mapsto \int \left(\frac{\widehat{g}_{n,1}(t)}{\widehat{g}_n(t)} - F(t) \right)^2 \widehat{g}_n(t) dt,$$

over all distribution functions, where $\widehat{g}_n(t) = d\widehat{G}_{n,0}(t)/dt + d\widehat{G}_{n,1}(t)/dt$ and $\widehat{g}_{n,1}(t) = d\widehat{G}_{n,1}(t)/dt$ (see also [45], Example 2). This suggests $\widehat{F}_n^{\text{naive}} = \widehat{g}_{n,1}/\widehat{g}_n$ as a naive smooth estimator. A similar situation occurs for the MSLE of a nondecreasing baseline hazard in the Cox model; see (15) in [71].

Typically, the naive smooth estimators are not necessarily monotone, but they are much simpler to analyze and they can be shown to be asymptotically equivalent to the corresponding smooth isotonic estimator. Let us illustrate things for the MSLE $\widehat{F}_n^{\text{MS}}$ in the current status problem. The first step is to show for intervals $[m, M]$ in the interior of the support of F_0 , that (under suitable conditions)

$$(32) \quad \mathbb{P}(\widehat{F}_n^{\text{naive}}(t) = \widehat{F}_n^{\text{MS}}(t) \text{ for all } t \in [m, M]) \rightarrow 1;$$

see [50], Corollary 3.4, and [71] for general isotonized smooth estimators. Hence, the pointwise limit behavior of $\widehat{F}_n^{\text{MS}}$ at a point t_0 in the interior of the support is equivalent to that of the corresponding naive smooth estimator, whose limit behavior is often relatively standard.

5.2 Isotonization Followed by Smoothing: Using Kiefer–Wolfowitz

Another approach to obtain smooth monotone estimators, is to take an isotonic estimator $\widehat{\lambda}_n$ for the problem at hand and then smooth it:

$$(33) \quad \widehat{\lambda}_n^s(t) = \int k_b(t - u) \widehat{\lambda}_n(u) du.$$

See [79] for a kernel smoothed LSE in the regression context, [50] for the smoothed MLE in the current status model, [46, 72] for estimating a monotone failure rate, [28] to bootstrap from a smooth decreasing density estimate, and [73] for the smoothed MLE and smoothed Grenander-type estimator in the Cox model. It can be seen from the argument on page 956 in [28] that when the kernel function is genuine probability density, smoothing after isotonization preserves monotonicity.

Asymptotic normality for the kernel smoothed LS regression estimator was first established by [79], for

the triangular kernel. In [90], it is shown that the isotonized kernel density estimator has the same limit normal distribution at the usual rate $n^{m/(2m+1)}$ as the ordinary kernel density estimator, when the density is m times continuously differentiable. Similar results were obtained by [50, 72] for the smoothed MLE and for a smoothed Grenander-type estimator, respectively.

In the notation of Definition 2.1, with $\widehat{\Lambda}_n$ the LCM of Λ_n , smoothed Grenander-type estimators for a monotone function λ_0 , can be written as

$$(34) \quad \widehat{\lambda}_n^s(t) = \int k_b(t - u) d\widehat{\Lambda}_n(u).$$

To obtain the asymptotic behavior of (34) at a point t_0 , one can decompose

$$(35) \quad \begin{aligned} \widehat{\lambda}_n^s(t_0) &= \int k_b(t_0 - u)\lambda_0(u) du \\ &+ \int k_b(t_0 - u) d(\Lambda_n - \Lambda_0)(u) \\ &+ \int k_b(t_0 - u) d(\widehat{\Lambda}_n - \Lambda_n)(u). \end{aligned}$$

The first term on the right-hand side determines the asymptotic bias, that is, when $b = cn^{1/5}$, for some $c > 0$, then

$$(36) \quad \begin{aligned} n^{2/5} \left\{ \int k_b(t_0 - u)\lambda_0(u) du - \lambda_0(t_0) \right\} \\ \rightarrow \frac{1}{2}c^2\lambda_0''(t_0) \int y^2k(y) dy. \end{aligned}$$

The last term in (35) is bounded from above in absolute value by

$$\begin{aligned} \frac{1}{b} \int |\widehat{\Lambda}_n(t_0 - by) - \Lambda_n(t_0 - by)| |k'(y)| dy \\ \leq \frac{1}{b} \sup_s |\widehat{\Lambda}_n(s) - \Lambda_n(s)| \int |k'(y)| dy. \end{aligned}$$

The supremum on the right-hand side is of the order $n^{-2/3}(\log n)^{2/3}$; see Kiefer–Wolfowitz [61] for the density model, [82] for the regression model, and [30] for the more general setup in [25]. This implies that the second term in (35) determines the limiting distribution.

We illustrate things for estimating a monotone failure rate λ_0 at a point t_0 in the interior of the support of λ_0 , as in [46]. Here, $\Lambda_n = -\log(1 - F_n)$ is the empirical cumulative failure rate and the second term in (35) is equal to

$$\begin{aligned} b^{-1} \int k(y) d(\Lambda_n - \Lambda_0)(t_0 - by) \\ = n^{-2/5}c^{-1/2} \int k'(y) d\widetilde{W}_n(y), \end{aligned}$$

where $\widetilde{W}_n(y) = \sqrt{n/b}(\Lambda_n(t_0 - by) - \Lambda_n(t_0) - \Lambda_0(t_0 - by) + \Lambda_0(t_0))$, which can be shown to converge weakly to the process $y \mapsto \lambda_0(t_0) f_0(t_0)^{-1/2} W(y)$, where W is standard two-sided Brownian motion and f_0 and λ_0 are the underlying density and failure rate, respectively. It follows that $n^{2/5}(\widehat{\lambda}_n^s(t_0) - \lambda_0(t_0))$ converges to a Gaussian law of which expectation and variance are

$$\begin{aligned} \mu &= \frac{1}{2}c^2\lambda_0''(t_0) \int y^2k(y) dy \quad \text{and} \\ \sigma^2 &= \frac{\lambda_0(t_0)^2}{cf_0(t_0)} \int k(y)^2 dy. \end{aligned}$$

A similar argument was used to determine the limit behavior of the smoothed Grenander-type estimator of a monotone failure rate under random censoring in [72], who applied a Kiefer–Wolfowitz type of result from [30].

5.3 Isotonization Followed by Smoothing: Using L_2 -Bounds

The previous approach heavily depends on a Kiefer–Wolfowitz type of result, which is not always available. Recently, [42] (see also [48], Chapter 11) developed a different method for finding the limit distribution of smoothed isotonic estimators, which is mainly based on uniform L_2 -bounds for the distance between the nonsmoothed isotonic estimator and the true function, in the case of i.i.d. observations. Starting from (33), we have

$$\begin{aligned} \widehat{\lambda}_n^s(t) &= \int k_b(t - u)\lambda_0(u) du \\ &+ \int k_b(t - u)(\widehat{\lambda}_n(u) - \lambda_0(u)) du. \end{aligned}$$

The first term on the right-hand side determines the asymptotic bias; see (36). To handle the second term on the right-hand side, the first step is to approximate it as follows:

$$(37) \quad \begin{aligned} n^{2/5} \int k_b(t - u)(\widehat{\lambda}_n(u) - \lambda_0(u)) du \\ \approx -n^{2/5} \int \theta_{n,t}(x) d\mathbb{P}(x), \end{aligned}$$

where \mathbb{P} is the common distribution of the i.i.d. observations. The idea is to take for $\theta_{n,t}$ a suitable function, whose piecewise constant version integrates to zero with respect to the empirical measure \mathbb{P}_n , that is, $\int \bar{\theta}_{n,t}(x) d\mathbb{P}_n(x) = 0$, where $\bar{\theta}_{n,t}$ is taken constant between successive points of jump of the isotonic estimator $\widehat{\lambda}_n$ in (33). Hence, the right-hand term in (37) is

equal to

$$(38) \quad n^{2/5} \int \bar{\theta}_{n,t}(x) d(\mathbb{P}_n - \mathbb{P})(x) + n^{2/5} \int (\bar{\theta}_{n,t}(x) - \theta_{n,t}(x)) d\mathbb{P}(x).$$

The key step is to bound the second integral by means of bounds on the L_2 -distance between the isotonic estimator $\hat{\lambda}_n$ and λ_0 . The last step is to replace $\bar{\theta}_{n,t}$ by a deterministic function $\eta_{n,t}$ and determine the limit behavior of

$$(39) \quad n^{2/5} \int \eta_{n,t}(x) d(\mathbb{P}_n - \mathbb{P})(x).$$

To find a suitable function $\theta_{n,t}$ is not trivial. In [42] and [48], Chapter 11, some guidelines are given for the interval censoring problem, deconvolution, and estimation of a monotone failure rate, where the choice of the function $\theta_{n,t}$ is related to solutions of integral equations that emerge from the theory of smooth functionals; see [42], Lemma 2.1 and Lemma 5.1, or [48], Lemma 11.9 and Lemma 11.17.

Sometimes $\theta_{n,t}$ can be deduced from the structure of the statistical model at hand. For example, [73] considers smooth isotonic estimators of a nondecreasing baseline hazard in the Cox model, where no Kiefer–Wolfowitz type of result is available. Let us illustrate things for the simple (special) case of no covariates, which coincides with estimation of a nondecreasing failure rate λ_0 under random censoring. In this model, the cumulative failure rate Λ_0 satisfies

$$(40) \quad d\Lambda_0(u) = \frac{\delta}{1 - H(u)} d\mathbb{P}(\delta, u).$$

Here \mathbb{P} denotes the distribution of the pair $(\Delta, \min(X, T))$ in the notation of Section 2.1 and H is the distribution function of $\min(X, T)$. This means

$$\int k_b(t - u) d\Lambda_0(u) = \int k_b(t - u) \frac{\delta}{1 - H(u)} d\mathbb{P}(\delta, u).$$

Hence, a natural candidate for $\theta_{n,t}$ can be constructed by combining δ and $a_{n,t}(u) := k_b(t - u)/(1 - H(u))$. In fact, the function

$$(41) \quad \theta_{n,t}(\delta, u) = \delta a_{n,t}(u) - \int \mathbb{1}_{\{v \leq u\}} a_{n,t}(v) d\hat{\Lambda}_n(v)$$

yields the desired approximation (37), by application of Fubini’s theorem:

$$\begin{aligned} & \int \theta_{n,t}(\delta, u) d\mathbb{P}(\delta, u) \\ &= \int k_b(t - u) d\Lambda_0(u) \\ & \quad - \int (1 - H(v)) a_{n,t}(v) d\hat{\Lambda}_n(v) \\ &= - \int k_b(t - u) d(\hat{\Lambda}_n - \Lambda_0)(u). \end{aligned}$$

Similar constructions were used in [73] for the smoothed MLE and the smoothed Grenander estimator for a monotone baseline hazard in the Cox model. For the smoothed MLE for a monotone failure rate under random censoring, [48], Section 11.6, use a different $a_{n,t}$, which is the solution of an integral equation.

To illustrate, how to bound the second integral in (38) in the same setting by means of the L_2 -distance between the isotonic estimator $\hat{\lambda}_n$ and the true monotone function, let us continue with the example above about estimation of a nondecreasing failure rate λ_0 under random censoring. Let $0 < \tau_1 < \dots < \tau_m$ denote the points of jump of λ_n . For $u \in (\tau_i, \tau_{i+1}]$, let

$$A_n(u) = \begin{cases} \tau_i & \text{if } \lambda_0(t) > \hat{\lambda}_n(\tau_{i+1}), \\ & \text{for all } t \in (\tau_i, \tau_{i+1}]; \\ s & \text{if } \lambda_0(s) = \hat{\lambda}_n(s), \\ & \text{for some } s \in (\tau_i, \tau_{i+1}]; \\ \tau_{i+1} & \text{if } \lambda_0(t) < \hat{\lambda}_n(\tau_{i+1}), \\ & \text{for all } t \in (\tau_i, \tau_{i+1}]. \end{cases}$$

With $\bar{a}_{n,t}(u) = a_{n,t}(A_n(u))$, define $\bar{\theta}_{n,t}$ in the same manner as $\theta_{n,t}$ in (41), with $a_{n,t}$ replaced by $\bar{a}_{n,t}$. If the kernel k is supported on $[-1, 1]$, then (40) yields that the second integral in (38), with $x = (\delta, u)$, is equal to

$$\begin{aligned} & \int (\bar{a}_{n,t}(u) - a_{n,t}(u))(1 - H(u)) d\Lambda_0(u) \\ & \quad - \int (1 - H(v)) (\bar{a}_{n,t}(v) - a_{n,t}(v)) d\hat{\Lambda}_n(v) \\ &= \int_{t-b}^{t+b} (1 - H(v)) (\bar{a}_{n,t}(v) - a_{n,t}(v)) \\ & \quad \cdot (\lambda_0(v) - \hat{\lambda}_n(v)) dv. \end{aligned}$$

By Cauchy–Schwarz, this is bounded from above in absolute value by

$$\|(\bar{a}_{n,t} - a_{n,t})\mathbb{1}_{[t-b, t+b]}\|_{L_2} \|(\lambda_0 - \hat{\lambda}_n)\mathbb{1}_{[t-b, t+b]}\|_{L_2}.$$

Now, if $u \in (\tau_i, \tau_{i+1}]$ and $A_n(u) > \tau_i$, then $\widehat{\lambda}_n(u) = \widehat{\lambda}_n(A_n(u))$, which implies

$$\begin{aligned} |u - A_n(u)| &\leq \frac{1}{\inf |\lambda'_0|} |\lambda_0(u) - \lambda_0(A_n(u))| \\ &\leq \frac{2}{\inf |\lambda'_0|} |\lambda_0(u) - \widehat{\lambda}_n(u)|. \end{aligned}$$

It can be seen that a similar inequality holds if $A_n(u) = \tau_i$. It follows that

$$\begin{aligned} &\|(\bar{a}_{n,t} - a_{n,t})\mathbb{1}_{[t-b,t+b]}\|_{L_2}^2 \\ &\leq C_1 b^{-4} \int_{t-b}^{t+b} (A_n(u) - u)^2 du \\ &\leq C_2 b^{-4} \int_{t-b}^{t+b} (\widehat{\lambda}_n(u) - \lambda_0(u))^2 du. \end{aligned}$$

It follows from (9) with $p = 2$ that the expectation of the right-hand side is of the order $O(b^{-3}n^{-2/3})$, and that of $\|(\lambda_0 - \widehat{\lambda}_n)\mathbb{1}_{[t-b,t+b]}\|_{L_2}^2$ is of the order $O(bn^{-2/3})$. This means that the second term in (38), where $x = (\delta, u)$, is of order $n^{2/5} O_p(b^{-1}n^{-2/3})$, which tends to zero, if the bandwidth has the usual rate $b \sim n^{-1/5}$.

Finally, we have to choose a suitable deterministic function $\eta_{n,t}$ such that

$$\begin{aligned} &n^{2/5} \int (\bar{\theta}_{n,t}(\delta, u) - \eta_{n,t}(\delta, u)) d(\mathbb{P}_n - \mathbb{P})(\delta, u) \\ (42) \quad &= o_P(1). \end{aligned}$$

We define $\eta_{n,t}$ in the same manner as $\theta_{n,t}$ in (41), with $\widehat{\Lambda}_n$ replaced by Λ_0 . Establishing (42) involves the use of entropy results for classes of bounded functions. For the current status problem, the monotone function of interest is clearly bounded as being a distribution function; see [50], Lemma A.7. This is not the case when estimating a monotone baseline hazard in the Cox model, or a monotone failure rate under random censoring. To prove (42), one has to deal with

$$\begin{aligned} &\frac{1}{b} \left| k\left(\frac{t - A_n(u)}{b}\right) - k\left(\frac{t - u}{b}\right) \right| \\ &\leq C_1 b^{-2} |A_n(u) - u| \\ &\leq C_2 b^{-2} |\lambda_0(u) - \widehat{\lambda}_n(u)|. \end{aligned}$$

To have the left-hand side bounded, requires a bound on $\mathbb{E}[\sup_u (\lambda_0(u) - \widehat{\lambda}_n(u))^2]$, which is more demanding than (9). When a strong embedding such as (16) is available, then one can obtain a bound of the order $O((n/\log n)^{-2/3})$, which still suffices to compensate

b^{-2} . This is the case for estimation of a monotone failure rate under random censoring, but no longer in the more general setup of estimating a monotone baseline hazard in the Cox model. However, [73], Lemma 5.6, manage to obtain a bound of the order $O(n^{-4/9})$, which suffices for their purposes.

We conclude, that for $b = cn^{1/5}$, $n^{2/5}(\widehat{\lambda}_n^s(t_0) - \lambda_0(t_0))$ is asymptotically equivalent to the sum of the asymptotic bias in (36) and the term in (39). Asymptotic normality of the term in (39) can then be established using standard arguments.

6. MISCELLANEA

This section includes some other topics where limit theory of isotonic estimators plays a role but that cannot be detailed here, due to space constraints.

Minimax risk bounds for the least squares estimator of a monotone regression function have been obtained in [97]. Recently, similar results have been obtained by [17], for isotonic and other shape constrained regression problems, in [18] for matrix isotonic estimation, in [13] for aggregation of affine estimators, and in [54] for isotonic regression in general dimensions. These results also provide oracle inequalities and describe the adaptation properties as well as behavior of the least squares estimator in higher dimensions.

Although [25] allows for some form of dependence, in the present paper we have mainly considered setups of independent data. Results about the pointwise limiting distribution for isotonic estimators in the regression setting with short and long range dependent errors can be found in [21, 97], whereas [5] considers the construction of confidence intervals in a similar setting. Estimation of a monotone spectral density is considered in [4], for short memory linear processes and long memory Gaussian processes. Results on the pointwise limiting distribution of isotonic estimators can be found in [3], who considers a general framework which includes weakly dependent and long range dependent data.

With a completely different asymptotic behavior, the discrete case as been considered in [59]: in that case, the rate of convergence is the square-root of the sample size, and the local limit distribution of the estimator is again nonnormal. It is obtained by concatenation of normal vectors (where the underlying p.m.f. is strictly monotone) and isotonic regression of normal vectors (where the underlying p.m.f. is flat). Another interesting extension is the limiting behavior of the Grenander estimator under misspecification [58].

Following the pioneering paper [61] by Kiefer and Wolfowitz, the limit theory for the difference between a naive estimator for a concave cumulative function and its least concave majorant has been considered in [34, 82, 67, 68, 30, 74]. As seen above, this could be used to build smooth monotone estimators. Other Kiefer–Wolfowitz type of results can be found in [87] for densities that are in a subclass of bounded variation functions or a Hölder ball.

Another application is connected to the problem of testing monotonicity of a function of interest [1, 24, 27, 47, 12]. Other tests can be built based on the limiting global behavior of isotonic estimators, to test goodness of fit [32, 31], or equality of several monotone functions [28], whereas some other tests are more connected to the local limit behavior of isotonic estimators, such as likelihood ratio tests [9, 10, 81]. Another application of local limiting distribution of isotonic estimator is construction of confidence intervals for monotone functions [49].

To conclude the section, we would like to point out that the standard bootstrap typically does not work for Grenander-type estimators; for example, see [64, 86]. However, [64, 16, 22, 28] discuss various smoothed bootstrap methods that are consistent.

ACKNOWLEDGEMENTS

The contribution of the first author has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

REFERENCES

- [1] AKAKPO, N., BALABDAOUI, F. and DUROT, C. (2014). Testing monotonicity via local least concave majorants. *Bernoulli* **20** 514–544. [MR3178508](#)
- [2] ANEVSKI, D. and HÖSSJER, O. (2002). Monotone regression and density function estimation at a point of discontinuity. *J. Nonparametr. Stat.* **14** 279–294. [MR1905752](#)
- [3] ANEVSKI, D. and HÖSSJER, O. (2006). A general asymptotic scheme for inference under order restrictions. *Ann. Statist.* **34** 1874–1930. [MR2283721](#)
- [4] ANEVSKI, D. and SOULIER, P. (2011). Monotone spectral density estimation. *Ann. Statist.* **39** 418–438. [MR2797852](#)
- [5] BAGCHI, P., BANERJEE, M. and STOEV, S. A. (2016). Inference for monotone functions under short- and long-range dependence: Confidence intervals and new universal limits. *J. Amer. Statist. Assoc.* **111** 1634–1647. [MR3601723](#)
- [6] BALABDAOUI, F., DUROT, C. and JANKOWSKI, H. (2016). Least squares estimation in the monotone single index model. Preprint. Available at [arXiv:1610.06026](#).
- [7] BALABDAOUI, F., JANKOWSKI, H., PAVLIDES, M., SEREGIN, A. and WELLNER, J. (2011). On the Grenander estimator at zero. *Statist. Sinica* **21** 873–899. [MR2829859](#)
- [8] BANERJEE, M., DUROT, C. and SEN, B. (2016). Divide and conquer in non-standard problems and the super-efficiency phenomenon. Preprint. Available at [arXiv:1605.04446](#).
- [9] BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. [MR1891743](#)
- [10] BANERJEE, M. and WELLNER, J. A. (2005). Confidence intervals for current status data. *Scand. J. Stat.* **32** 405–424. [MR2204627](#)
- [11] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression*. Wiley, New York. [MR0326887](#)
- [12] BEARE, B. K. and SCHMIDT, L. D. W. (2016). An empirical test of pricing kernel monotonicity. *J. Appl. Econometrics* **31** 338–356. [MR3481366](#)
- [13] BELLEC, P. C. (2018). Optimal bounds for aggregation of affine estimators. *Ann. Statist.* **46** 30–59. [MR3766945](#)
- [14] BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* **26** 607–616. [MR0073894](#)
- [15] CAROLAN, C. and DYKSTRA, R. (1999). Asymptotic behavior of the Grenander estimator at density flat regions. *Canad. J. Statist.* **27** 557–566. [MR1745821](#)
- [16] CARROLL, R. J., DELAIGLE, A. and HALL, P. (2011). Testing and estimating shape-constrained nonparametric density and regression in the presence of measurement error. *J. Amer. Statist. Assoc.* **106** 191–202. [MR2816713](#)
- [17] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. [MR3357878](#)
- [18] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. [MR3706788](#)
- [19] CHENG, K. F. and LIN, P. E. (1981). Nonparametric estimation of a regression function. *Z. Wahrsch. Verw. Gebiete* **57** 223–233. [MR0626817](#)
- [20] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. [MR0172382](#)
- [21] DEDECKER, J., MERLEVÈDE, F. and PELIGRAD, M. (2011). Invariance principles for linear processes with application to isotonic regression. *Bernoulli* **17** 88–113. [MR2797983](#)
- [22] DU, P., PARMETER, C. F. and RACINE, J. S. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statist. Sinica* **23** 1347–1371. [MR3114717](#)
- [23] DUROT, C. (2002). Sharp asymptotics for isotonic regression. *Probab. Theory Related Fields* **122** 222–240. [MR1894068](#)
- [24] DUROT, C. (2003). A Kolmogorov-type test for monotonicity of regression. *Statist. Probab. Lett.* **63** 425–433. [MR1996191](#)
- [25] DUROT, C. (2007). On the \mathbb{L}_p -error of monotonicity constrained estimators. *Ann. Statist.* **35** 1080–1104. [MR2341699](#)
- [26] DUROT, C. (2008). Monotone nonparametric regression with random design. *Math. Methods Statist.* **17** 327–341. [MR2483461](#)
- [27] DUROT, C. (2008). Testing convexity or concavity of a cumulated hazard rate. *IEEE Trans. Reliab.* **57** 465–473.
- [28] DUROT, C., GROENEBOOM, P. and LOPUHAÄ, H. P. (2013). Testing equality of functions under monotonicity constraints. *J. Nonparametr. Stat.* **25** 939–970. [MR3174305](#)

- [29] DUROT, C., KULIKOV, V. N. and LOPUHAÄ, H. P. (2012). The limit distribution of the L_∞ -error of Grenander-type estimators. *Ann. Statist.* **40** 1578–1608. [MR3015036](#)
- [30] DUROT, C. and LOPUHAÄ, H. P. (2014). A Kiefer–Wolfowitz type of result in a general setting, with an application to smooth monotone estimation. *Electron. J. Stat.* **8** 2479–2513. [MR3285873](#)
- [31] DUROT, C. and REBOUL, L. (2010). Goodness-of-fit test for monotone functions. *Scand. J. Stat.* **37** 422–441. [MR2724504](#)
- [32] DUROT, C. and TOCQUET, A.-S. (1998). Goodness of fit test for isotonic regression. *C. R. Acad. Sci. Paris Sér. I Math.* **327** 199–204. [MR1645120](#)
- [33] DUROT, C. and TOCQUET, A.-S. (2001). Goodness of fit test for isotonic regression. *ESAIM Probab. Stat.* **5** 119–140. [MR1875667](#)
- [34] DUROT, C. and TOCQUET, A.-S. (2003). On the distance between the empirical process and its concave majorant in a monotone regression framework. *Ann. Inst. Henri Poincaré Probab. Stat.* **39** 217–240. [MR1962134](#)
- [35] EGGERMONT, P. P. B. and LARICCIA, V. N. (2000). Maximum likelihood estimation of smooth monotone and unimodal densities. *Ann. Statist.* **28** 922–947. [MR1792794](#)
- [36] FRIEDMAN, J. and TIBSHIRANI, R. (1984). The monotone smoothing of scatter plots. *Technometrics* **26** 243–350.
- [37] GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. [MR0093415](#)
- [38] GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* **11** 1016–1027. [MR0714964](#)
- [39] GROENEBOOM, P. (1984). Brownian Motion With a Parabolic Drift and Airy Functions. CWI technical report, Dept. Mathematical Statistics-R 8413, CWI, Amsterdam.
- [40] GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*. 539–555. Wadsworth, Belmont, CA. [MR0822052](#)
- [41] GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. [MR0981568](#)
- [42] GROENEBOOM, P. (2013). Nonparametric (smoothed) likelihood and integral equations. *J. Statist. Plann. Inference* **143** 2039–2065. [MR3106623](#)
- [43] GROENEBOOM, P. and HENDRICKX, K. (2018). Current status linear regression. *Ann. Statist.* **46** 1415–1444. [MR3819105](#)
- [44] GROENEBOOM, P., HOOGHIEMSTRA, G. and LOPUHAÄ, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.* **27** 1316–1347. [MR1740109](#)
- [45] GROENEBOOM, P. and JONGBLOED, G. (2010). Generalized continuous isotonic regression. *Statist. Probab. Lett.* **80** 248–253. [MR2575453](#)
- [46] GROENEBOOM, P. and JONGBLOED, G. (2013). Smooth and non-smooth estimates of a monotone hazard. In *From Probability to Statistics and Back: High-Dimensional Models and Processes. Inst. Math. Stat. (IMS) Collect.* **9** 174–196. IMS, Beachwood, OH. [MR3202633](#)
- [47] GROENEBOOM, P. and JONGBLOED, G. (2013). Testing monotonicity of a hazard: Asymptotic distribution theory. *Bernoulli* **19** 1965–1999. [MR3129041](#)
- [48] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics. Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. [MR3445293](#)
- [49] GROENEBOOM, P. and JONGBLOED, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43** 2019–2054. [MR3375875](#)
- [50] GROENEBOOM, P., JONGBLOED, G. and WITTE, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.* **38** 352–387. [MR2589325](#)
- [51] GROENEBOOM, P. and LOPUHAÄ, H. P. (1993). Isotonic estimators of monotone densities and distribution functions: Basic facts. *Stat. Neerl.* **47** 175–183. [MR1243853](#)
- [52] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation. DMV Seminar* **19**. Birkhäuser, Basel. [MR1180321](#)
- [53] GROENEBOOM, P. and WELLNER, J. A. (2001). Computing Chernoff’s distribution. *J. Comput. Graph. Statist.* **10** 388–400. [MR1939706](#)
- [54] HAN, Q., WANG, T., CHATTERJEE, S. and SAMWORTH, R. J. (2017). Isotonic regression in general dimensions. Preprint. Available at [arXiv:1708.09468](#).
- [55] HOOGHIEMSTRA, G. and LOPUHAÄ, H. P. (1998). An extremal limit theorem for the argmax process of Brownian motion minus a parabolic drift. *Extremes* **1** 215–240. [MR1814624](#)
- [56] HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. [MR1334065](#)
- [57] HUANG, Y. and ZHANG, C.-H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22** 1256–1274. [MR1311975](#)
- [58] JANKOWSKI, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.* **42** 625–653. [MR3210981](#)
- [59] JANKOWSKI, H. K. and WELLNER, J. A. (2009). Estimation of a discrete monotone distribution. *Electron. J. Stat.* **3** 1567–1605. [MR2578839](#)
- [60] JONKER, M. A. and VAN DER VAART, A. W. (2001). A semi-parametric model for censored and passively registered data. *Bernoulli* **7** 1–31. [MR1811742](#)
- [61] KIEFER, J. and WOLFOWITZ, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. Verw. Gebiete* **34** 73–85. [MR0397974](#)
- [62] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. [MR1041391](#)
- [63] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV’s and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. [MR0375412](#)
- [64] KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen. Inst. Math. Stat. (IMS) Collect.* **1** 282–292. IMS, Beachwood, OH. [MR2462212](#)
- [65] KULIKOV, V. N. and LOPUHAÄ, H. P. (2005). Asymptotic normality of the L_k -error of the Grenander estimator. *Ann. Statist.* **33** 2228–2255. [MR2211085](#)

- [66] KULIKOV, V. N. and LOPUHAÄ, H. P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist.* **34** 742–768. [MR2283391](#)
- [67] KULIKOV, V. N. and LOPUHAÄ, H. P. (2006). The limit process of the difference between the empirical distribution function and its concave majorant. *Statist. Probab. Lett.* **76** 1781–1786. [MR2274141](#)
- [68] KULIKOV, V. N. and LOPUHAÄ, H. P. (2008). Distribution of global measures of deviation between the empirical distribution function and its concave majorant. *J. Theoret. Probab.* **21** 356–377. [MR2391249](#)
- [69] LEURGANS, S. (1982). Asymptotic distributions of slope-of-greatest-convex-minorant estimators. *Ann. Statist.* **10** 287–296. [MR0642740](#)
- [70] LOPUHAÄ, H. P. and MUSTA, E. (2016). A central limit theorem for the Hellinger loss of Grenander type estimators. Available at [arXiv:1612.06647](#).
- [71] LOPUHAÄ, H. P. and MUSTA, E. (2017). Isotonized smooth estimators of a monotone baseline hazard in the Cox model. *J. Statist. Plann. Inference* **191** 43–67. [MR3679108](#)
- [72] LOPUHAÄ, H. P. and MUSTA, E. (2017). Smooth estimation of a monotone hazard and a monotone density under random censoring. *Stat. Neerl.* **71** 58–82. [MR3605364](#)
- [73] LOPUHAÄ, H. P. and MUSTA, E. (2018). Smoothed isotonized estimators of a monotone baseline hazard in the Cox model. *Scand. J. Stat.* **45** 753–791.
- [74] LOPUHAÄ, H. P. and MUSTA, E. (2018). The distance between a naive cumulative estimator and its least concave majorant. *Statist. Probab. Lett.* **139** 119–128. [MR3802192](#)
- [75] LOPUHAÄ, H. P. and NANE, G. F. (2013). Shape constrained non-parametric estimators of the baseline distribution in Cox proportional hazards model. *Scand. J. Stat.* **40** 619–646. [MR3091700](#)
- [76] MAMMEN, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.* **19** 724–740. [MR1105841](#)
- [77] MARSHALL, A. W. and PROSCHAN, F. (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Ann. Math. Stat.* **36** 69–77. [MR0170436](#)
- [78] MEYER, M. C. (2008). Inference using shape-restricted regression splines. *Ann. Appl. Stat.* **2** 1013–1033. [MR2516802](#)
- [79] MUKERJEE, H. (1988). Monotone nonparametric regression. *Ann. Statist.* **16** 741–750. [MR0947574](#)
- [80] MURPHY, S. A., VAN DER VAART, A. W. and WELLNER, J. A. (1999). Current status regression. *Math. Methods Statist.* **8** 407–425. [MR1735473](#)
- [81] NANE, G. F. (2015). A likelihood ratio test for monotone baseline hazard functions in the Cox model. *Statist. Sinica* **25** 1163–1184. [MR3410303](#)
- [82] PAL, J. K. and WOODROOFE, M. (2006). On the distance between cumulative sum diagram and its greatest convex minorant for unequally spaced design points. *Scand. J. Stat.* **33** 279–291. [MR2279643](#)
- [83] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. [MR0267677](#)
- [84] PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Ann. Math. Stat.* **41** 507–519. [MR0260133](#)
- [85] RAMSAY, J. O. (1998). Estimating smooth monotone functions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 365–375. [MR1616049](#)
- [86] SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. [MR2676880](#)
- [87] SÖHL, J. (2015). Uniform central limit theorems for the Grenander estimator. *Electron. J. Stat.* **9** 1404–1423. [MR3360732](#)
- [88] SUN, J. and WOODROOFE, M. (1996). Adaptive smoothing for a penalized NPMLE of a non-increasing density. *J. Statist. Plann. Inference* **52** 143–159. [MR1392133](#)
- [89] TANTIYASWADIKUL, C. and WOODROOFE, M. B. (1994). Isotonic smoothing splines under sequential designs. *J. Statist. Plann. Inference* **38** 75–87. [MR1256849](#)
- [90] VAN DER VAART, A. W. and VAN DER LAAN, M. J. (2003). Smooth estimation of a monotone density. *Statistics* **37** 189–203. [MR1986176](#)
- [91] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. [MR1385671](#)
- [92] WELLNER, J. A. (2015). Musings about shape constrained estimation and inference: Some problems. Presentation at workshop *Shape Constrained Inference: Open Problems and New Directions*, Lorentz Centre, Leiden, The Netherlands.
- [93] WELLNER, J. A. and ZHANG, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28** 779–814. [MR1792787](#)
- [94] WOODROOFE, M. and SUN, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing. *Statist. Sinica* **3** 501–515. [MR1243398](#)
- [95] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. [MR0606630](#)
- [96] WRIGHT, F. T. (1982). Monotone regression estimates for grouped observations. *Ann. Statist.* **10** 278–286. [MR0642739](#)
- [97] ZHAO, O. and WOODROOFE, M. (2012). Estimating a monotone trend. *Statist. Sinica* **22** 359–378. [MR2933180](#)