Some Developments in the Theory of Shape Constrained Inference

Groeneboom, Piet; Jongbloed, Geurt

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Some Developments in the Theory of Shape Constrained Inference

Piet Groeneboom and Geurt Jongbloed

Abstract.   Shape constraints enter in many statistical models. Sometimes these constraints emerge naturally from the origin of the data. In other situations, they are used to replace parametric models by more versatile models retaining qualitative shape properties of the parametric model. In this paper, we sketch a part of the history of shape constrained statistical inference in a nutshell, using landmark results obtained in this area. For this, we mainly use the prototypical problems of estimating a decreasing probability density on $[0, \infty)$ and the estimation of a distribution function based on current status data as illustrations.

Key words and phrases: Isotonic regression, Grenander estimator, inverse problem, monotonicity, interval censoring, current status regression, single index model, bootstrap, Chernoff's distribution, Airy functions.

1. INTRODUCTION

In the 1950s, research groups around the world started working on order restricted statistical inference. Interestingly, people started to work around the same time on parametric as well as nonparametric problems of this type. The standard parametric model is that of *isotonic regression*, where in its simplest form one observes n independent normally distributed random variables with nondecreasing expectations and known variances. Early references on isotonic regression include Ayer et al. (1955) and van Eeden (1956). In 1956, Ulf Grenander in Grenander (1956) studied the problem of estimating a decreasing probability density function, without imposing parametric assumptions.

In Section 2, it will be seen that the isotonic regression problem can be solved elegantly. Its solution can be constructed graphically, as derivative of the greatest convex minorant of a data-dependent diagram of points. Interestingly, it will be seen that the maximum likelihood estimator for a decreasing density as

Piet Groeneboom is Emeritus Professor of Statistics, Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, Netherlands (e-mail: P.Groeneboom@tudelft.nl). Geurt Jongbloed is Professor of Statistics, Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, Netherlands (e-mail: G.Jongbloed@tudelft.nl).

well as for the distribution function in the current status model, can be computed via similar constructions. While shape constrained statistical inference started off with a number of these "monotone problems" (where the constraint on the function to be estimated was monotonicity), many other shape constrained problems have emerged since. Some of these will also be mentioned in Section 2.

Section 3 discusses the maximum likelihood estimator of a decreasing density function in more detail, reviewing many results that have been obtained during the past decades. Asymptotic properties of the estimator, local as well as global, will be stated with special attention for the role that is played by the so-called switching relation.

In Section 4, related issues for interval censoring models are reviewed, with special attention for the current status model. Ayer et al. (1955) derives the nonparametric MLE of the distribution function for the current status model and proves that it is consistent. Peto (1973) considers the MLE for interval censoring, case 2. There it is suggested that pointwise standard errors for the survival curve can be estimated from the inverse of the Fisher information, which, however, is not correct (we owe this observation to Peter Sasieni). In fact, the beginning of the theory of interval censoring was hampered by many mistakes, due to false analogies with the theory of right censoring and an expected

pointwise normal limit behavior of the MLE. The local asymptotic distribution of the MLE in the current status model is, just as in the case of decreasing densities, governed by the so-called Chernoff distribution. Sometimes, for example, when considering bootstrap procedures, it is important to also study estimators that apart from monotonicity properties, also possess some smoothness. The asymptotic pointwise distribution of the smoothed maximum likelihood estimator is normal and its rate of convergence is faster than that of the MLE. Methods to derive these asymptotic results are local versions of the so-called smooth functional approach that can be used to derive asymptotic distributions of functionals of the MLE that depend on the underlying distribution in a smooth way. For the case 2 interval censoring model, such results are available, but still many questions remain unsolved within this model. Another extension of the current status model to be considered in Section 4 is a regression model where the response variables are of current status type.

For the limiting Chernoff distribution, first encountered in Chernoff (1964), that is omnipresent in monotone problems, analytical results as well as tabulated quantiles are available; see Groeneboom and Wellner (2001). Also in the development of the analytical representation of the distribution, the so-called switch relation is an important tool. In Section 5, the analytical representation of the local limit distribution is discussed and some properties are stated.

In principle, asymptotic distributions can be used as starting point for the construction of confidence intervals for monotone functions of interest. Alternatively, one can use a bootstrap procedure to obtain such intervals. For decreasing densities and the distribution function within the current status model, there are various approaches that seem natural. A lot of recent research has recently been done on these issues. This will be reviewed in Section 6. Open problems will be discussed in Section 7.

2. ISOTONIC REGRESSION AND EXTENSIONS

Order constraints on functions in statistical models often originate from the nature of the problem. One can, for example, measure the height of a sample of children of various ages to estimate the average population heights depending on age. These average population heights can naturally be assumed monotone in age. In its simplest form, isotonic regression is about estimating a nondecreasing regression function r on $\mathbb{R}$ based on data from the model

$$(2.1) \qquad Y_i = r(x_i) + \epsilon_i,$$

where $x_1 < x_2 < \cdots < x_n$ and the ϵ_i's are i.i.d. with expectation zero and finite variance. The values of r at the points x_i can be estimated via least squares:

$$(2.2) \quad \begin{aligned} &(\hat{r}(x_1), \ldots, \hat{r}(x_n)) \\ &\quad = \operatorname*{argmin}_{r(x_1) \leq r(x_2) \leq \cdots \leq r(x_n)} \frac{1}{2} \sum_{i=1}^n (Y_i - r(x_i))^2. \end{aligned}$$

This minimizer is well defined and can be constructed graphically. The construction follows from the Fenchel optimality conditions related to optimization problem (2.2).

LEMMA 2.1.   *Define the points $P_i = (i, \sum_{j=1}^i Y_j)$, $i = 1, \ldots, n$ and $P_0 = (0, 0)$. Let $\hat{R}$ be the greatest convex function on $[0, n]$ lying completely below the set of points $\{P_i : 0 \leq i \leq n\}$. Then $\hat{r}(x_i)$ as defined in (2.2) is the left derivative of $\hat{R}$ evaluated at i.*

Sometimes, the characterization is given in a rescaled diagram where the coordinates are divided by n to fit the diagram on the interval $[0, 1]$, but on the slope of the convex minorant this does not have any effect. Order restricted estimators for more general types of partial ordering also exist; see Chapter 2 in Barlow et al. (1972) or Robertson, Wright and Dykstra (1988). A widely used and efficient algorithm to compute isotonic regressions for a range of orderings is the Pool Adjacent Violators Algorithm (PAVA).

Apart from the situations where monotonicity is a natural modeling assumption, many other models lead to shape constraints on sampling densities. Problems with indirect observations (e.g., in stereology) or censoring (e.g., in survival analysis) lead to inverse problems where the sampling density depends in some way on the distribution of interest. Imposing no assumptions on the distribution of interest still leads to a restricted class of possible sampling distributions. It is the monotonicity of the distribution function of interest that then imposes a shape constraint on the sampling distribution.

One example is related to the aforementioned problem studied in Grenander (1956). Instead of observing a sample from a distribution function of interest G on $(0, \infty)$ with finite mean, one observes uniform random fractions of a sample from the length-biased distribution corresponding to G. More precisely, in the background there is a sample $Z_1, \ldots, Z_n$ from the length-biased distribution function G^w given by

$$G^w(z) = \frac{\int_0^z y \, dG(y)}{\int_0^\infty y \, dG(y)} = \frac{1}{\mu_G} \int_0^z y \, dG(y), \quad z \geq 0$$

and independent of this there is a standard uniformly distributed sample $U_1, \ldots, U_n$ and one observes $X_i = U_i Z_i$. As can be seen from Section 2.2, page 23 in Groeneboom and Jongbloed (2014), the sampling density of the X_i's is then given by
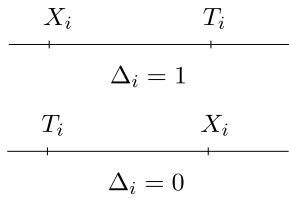
$$f(x) = \frac{1 - G(x)}{\mu_G}.$$

Monotonicity of G immediately implies f to be bounded and monotone. Also note that $f(0) = 1/\mu_G$, leading to the following relation for the survival function:

$$(2.3) \qquad 1 - G(x) = \frac{f(x)}{f(0)}, \quad x \geq 0;$$

see also Exercise 2.4 in Chapter 2 of Groeneboom and Jongbloed (2014).

Using this relation, an estimate of the decreasing sampling density leads to an estimate of the underlying distribution function G of interest. This model is encountered in many situations; see, for example, Keiding et al. (2012), Watson (1971), Vardi (1982) and Vardi (1989).

Another model where monotonicity is a key ingredient comes from survival analysis. Suppose $X_1, X_2, \ldots, X_n$ is a sample from a 1-dimensional distribution with distribution function F_0 on $(0, \infty)$. Instead of observing the X_i's, one only observes for each i whether or not $X_i \leq T_i$ for some sample $T_1, \ldots, T_n$ from a distribution with distribution function G and independent of the X_j's,

$$\begin{array}{c} X_i \qquad\qquad T_i \\ \hline \\ \Delta_i = 1 \end{array}$$

$$\begin{array}{c} T_i \qquad\qquad X_i \\ \hline \\ \Delta_i = 0 \end{array}$$

So, instead of observing the X_i's, one observes pairs

$$(T_i, \Delta_i) = (T_i, 1_{\{X_i \leq T_i\}}), \quad 1 \leq i \leq n.$$

Regarding X_i as event time of subject i and T_i as inspection time for this subject, (T_i, Δ_i) represents the *current status* of subject i at time T_i in the sense that if $\Delta_i = 0$ the event has not yet occurred to subject i at time T_i whereas if $\Delta_i = 1$, the event has already occurred at T_i. This model is also called the *interval censoring, case 1* model. Interval censoring because the information in the data indicates to which interval X_i belongs ($[0, T_i]$ if $\Delta_i = 1$ and $[T_i, \infty)$ if $\Delta_i = 0$) and case 1 because there is only one inspection time per subject.

The data take their value in the set $\{(t, \delta) : t \geq 0, \delta \in \{0, 1\}\}$ and the sampling density is given by

$$(2.4) \qquad h(t, \delta) = F(t)^\delta (1 - F(t))^{1-\delta}$$

with respect to the dominating measure $\mu \times dG$, where μ denotes counting measure on $\{0, 1\}$ and G the distribution function of the inspection time. This means that the sampling density can be represented by means of an underlying distribution function F, which by definition is monotone. This imposes some structure, say shape, on the density h. The current status model will be considered in more detail in Section 4.

Yet another motivation for using shape constrained models is that many of the well known parametric models satisfy shape constraints. Exponential densities are decreasing, bounded Gamma densities and (also multivariate) normal densities are unimodal. Replacing a parametric assumption by a shape constraint on the densities therefore yields more versatile models being able to let the data speak more for themselves. An interesting example of this approach is the model of log concave densities. Log concavity of a function, more so than monotonicity, can readily be defined in arbitrary dimensions. The problem of estimating a density only based on the assumption that it is log concave was introduced in Walther (2001) and further studied in Balabdaoui, Rufibach and Wellner (2009). The multivariate problem is studied in Cule, Samworth and Stewart (2010) and Cule and Samworth (2010).

3. THE GRENANDER ESTIMATOR

3.1 Characterization and Basic Properties

The log likelihood of a density f on $[0, \infty)$ of a distribution which is absolutely continuous w.r.t. Lebesgue measure is given by

$$(3.1) \qquad \begin{aligned} \ell(f) &= \frac{1}{n} \sum_{i=1}^{n} \log f(X_i) \\ &= \int \log f(x) \, d\mathbb{F}_n(x), \end{aligned}$$

where $\mathbb{F}_n$ is the empirical distribution function of the sample $X_1, \ldots, X_n$. The following theorem, due to Grenander, characterizes the maximum likelihood estimator as maximizer of this function over all decreasing densities f on $[0, \infty)$. The resulting estimator is called the *Grenander estimator*. It is the prototype of a nonparametric isotonic estimator.
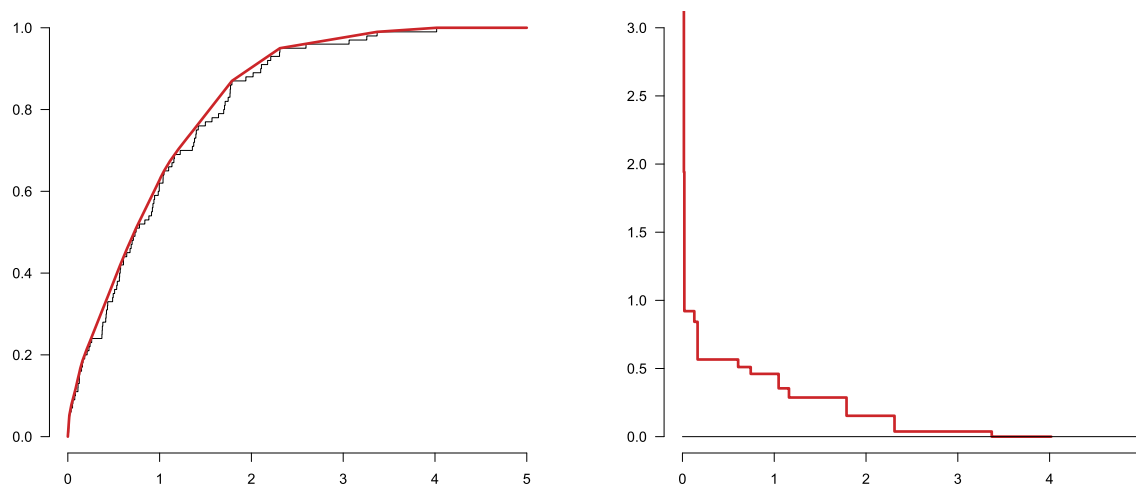
FIG. 1. *Left*: *the empirical distribution function* $\mathbb{F}_n$ *and its least concave majorant* $\hat{F}_n$ *(red) and right*: *its left derivative* $\hat{f}_n$ *(the Grenander estimator)*, *for a sample of size* $n = 100$ *generated from a standard exponential distribution.*

THEOREM 3.1 (Grenander, 1956). *The maximum likelihood estimator* $\hat{f}_n$, *maximizing* (3.1) *over all decreasing densities* f, *is the left derivative of the least concave majorant* $\hat{F}_n$ *of the empirical distribution function* $\mathbb{F}_n$.

Figure 1 shows the functions $\hat{F}_n$ and $\hat{f}_n$ based on a sample of size 100.

A classical lemma from Marshall (1969) shows that using the information that the underlying distribution function F_0 is concave leads to a better estimator of F_0 than the empirical distribution function that does not take the monotonicity of the density into account.

LEMMA 3.1 (Marshall's lemma). *Using the notation of Theorem* 3.1 *and denoting by* F_0 *the underlying concave distribution function of* $X_1, \ldots, X_n$,

$$\|\hat{F}_n - F_0\|_\infty \leq \|\mathbb{F}_n - F_0\|_\infty.$$

Another classical result states that, though in supremum distance the least concave majorant of the empirical distribution function estimates the underlying concave distribution more closely than the empirical distribution function itself, the difference between the two estimators is negligible with respect to the estimation error of rate $n^{-1/2}$. Indeed, in Kiefer and Wolfowitz (1976) the following theorem is proved.

THEOREM 3.2 (Kiefer–Wolfowitz). *Suppose the underlying distribution function* F_0 *attains the value one and is strictly concave on the set* $[0, F_0^{-1}(1)]$. *Then*

$$\frac{n^{2/3}}{\log n} \|\hat{F}_n - \mathbb{F}_n\|_\infty \to 0 \quad \text{almost surely.}$$

Hence, the superiority of $\hat{F}_n$ as estimator for F_0 stated by Marshall's lemma is asymptotically not impressive. Using the shape constraint of concavity therefore seems not to make a huge difference asymptotically. However, for estimating the true underlying density f_0 the situation is of course different. The derivative of $\hat{F}_n$ can be defined and turns out to be the well-behaved (Grenander-) estimator. Differentiating $\mathbb{F}_n$ does not lead to a sensible estimator. Of course, taking a histogram estimator (derivative of linear interpolation of $\mathbb{F}_n$ on a well-chosen grid $0 = t_1 < t_2 < \cdots < t_m$) does yield a consistent estimator for f_0.

The Grenander estimator also has an interpretation as a least squares estimator. Let $\hat{f}_n$ be the Grenander estimator. Then $\hat{f}_n$ *minimizes*

$$(3.2) \qquad \int_0^\infty f(x)^2 \, dx - 2 \int_{[0,\infty)} f(x) \, d\mathbb{F}_n(x)$$

over all decreasing densities f on $[0, \infty)$. From a (certain) minimax point of view, the least squares estimate provided by the Grenander estimator cannot be improved. For details, see Cator (2011). In this sense, the Grenander estimator is pointwise fully adaptive to the smoothness of the underlying distribution.

3.2 Local Limit Distribution

The Grenander estimator has a nonstandard limit behavior. It is given by the following theorem.

THEOREM 3.3 (Prakasa Rao, 1969). *Let* $\hat{f}_n$ *be the Grenander estimate of the density* f_0 *under the monotonicity restriction. Then, if* f_0 *has a strictly negative*
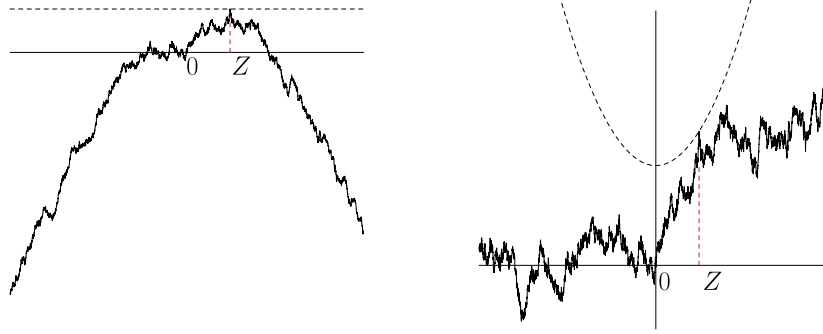
FIG. 2. *Left: Z and $t \mapsto W(t) - t^2$, right: Z and $t \mapsto W(t)$.*

derivative f_0' at the interior point x,

$$n^{1/3}\{\hat{f}_n(x) - f_0(x)\}/|4 f_0(x) f_0'(x)|^{1/3}$$

$$\xrightarrow{\mathcal{D}} Z, \quad n \to \infty,$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, and $Z = \text{argmax}_t\{W(t) - t^2\}$, *that is, Z is the (almost surely unique) location of the maximum of two-sided Brownian motion minus the parabola $y(t) = t^2$.*

A picture of the perhaps somewhat mysterious random variable $Z = \text{argmax}_t\{W(t) - t^2\}$ is shown in Figure 2, for a path of the simulated two-sided Brownian motion W. In Section 5, we will go into more details of the distribution of Z, known as the Chernoff distribution.

3.3 The Switch Relation

A different proof of Prakasa Rao's result was given in Groeneboom (1985). The key observation in this paper is the following *switch relation*:

$$\hat{f}_n(x) \geq a$$

$$\iff \quad x \leq U_n(a) = \text{argmax}\{t \geq 0 : \mathbb{F}_n(t) - at\},$$

$$a \in (0, \hat{f}_n(0)), x > 0,$$

where we take in argmax the last time that the maximum is attained if there are several locations for the maximum, and consider the *process* $\{U_n(a),$ $a \in (0, \hat{f}_n(0))\}$. The left-continuous *slopes* of lines become the *time variable* of the process U_n. See Figure 3.

So we have, if $a = f_0(x)$ and $U_n(a) = \text{argmax}\{t \geq 0 : \mathbb{F}_n(t) - at\}$,

$$\mathbb{P}\{n^{1/3}\{\hat{f}_n(x) - f_0(x)\} \geq t\}$$

$$= \mathbb{P}\{\hat{f}_n(x) \geq a + n^{-1/3}t\}$$

$$\overset{\text{switch relation}}{=} \mathbb{P}\{U_n(a + n^{-1/3}t) \geq x\}.$$

This combined with the "Hungarian embedding" leads to the proof of Prakasa Rao's result in Groeneboom (1985).

3.4 An Application of the Grenander Estimator in Fertility Studies

In Slama et al. (2012) an interesting data set of current durations of pregnancy in France is studied. The aim is to estimate the distribution of the time it takes for a woman to become pregnant after having having started unprotected sexual intercourse. For 867 women, the current duration of unprotected intercourse was recorded and this is the basis of part of the research, reported in Slama et al. (2012).

Given that the woman in the study is currently trying to become pregnant, the actual recorded data (current duration) can be viewed as uniform random fraction of the true, total duration. So we have another instance of model (2.3) for the survival function. The left panel of Figure 4 shows a part of the empirical distribution function of 618 recorded current durations, kindly provided to us by Niels Keiding, where the data are truncated at 36 months and are of a similar nature as the data in Slama et al. (2012). Based on the least concave majorant, the right panel of Figure 4 is computed, showing the resulting MLE of the decreasing density of the observations together with its smoothed version,
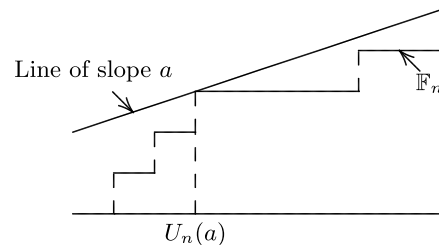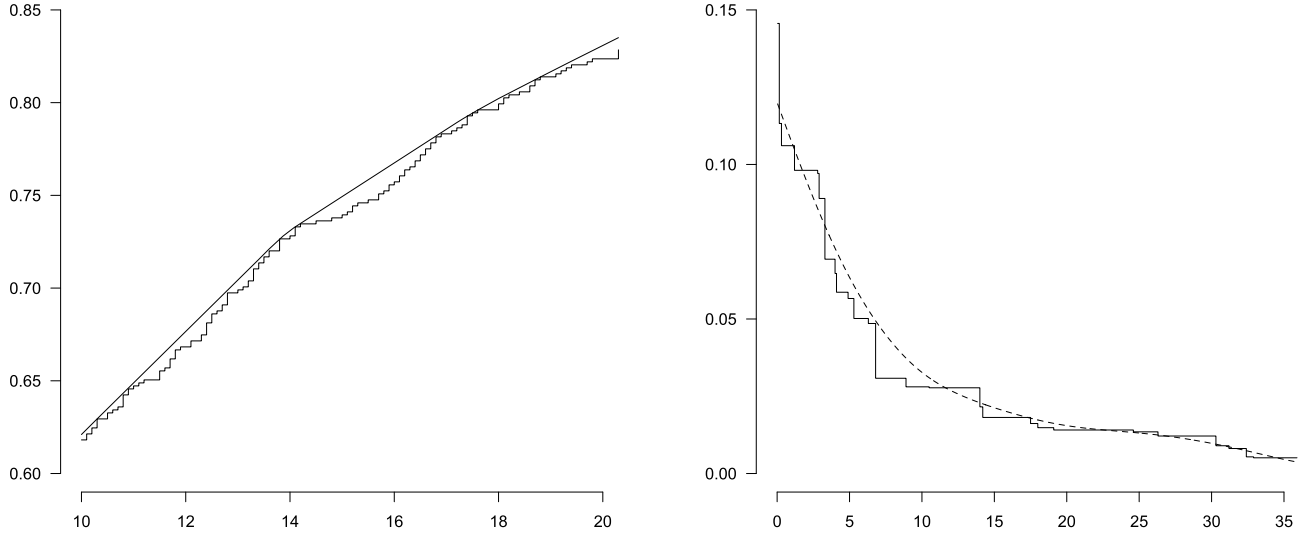

FIG. 3. *The switch relation.*

FIG. 4. *The left panel shows the empirical distribution function and its least concave majorant for the values between* 10 *and* 20 *months of the* 618 *current durations* ≤ 36 *months* (*we give this on a smaller observation interval to highlight the difference between the EDF and the least concave majorant*). *The resulting Grenander estimate* (*the MLE*) *of the observation density on the interval* [0, 36] *is shown in the right panel*, *together with its smoothed version* (*dashed*, *the SMLE*).

the smoothed maximum likelihood estimator (SMLE), defined by

$$
\tilde{f}_{nh}(t) = - \int \mathbb{K}\left(\frac{t - x}{h}\right) d\hat{f}_n(x),
$$
$$
(3.3) \qquad \mathbb{K}(x) = \int_x^\infty K(u)\,du,
$$

where $\hat{f}_n$ is the MLE and K is a symmetric kernel, for which we took the triweight kernel

$$
K(u) = \frac{35}{32}(1 - u^2)^3 1_{[-1,1]}(u), \quad u \in \mathbb{R}.
$$

Note that $-d\hat{f}_n(x)$ corresponds to a nonnegative atomic measure, since the jumps $d\hat{f}_n(x)$ are downward jumps.

The bandwidth h was chosen by least squares cross validation. Near the boundary point 0 we use a local linear boundary correction, to avoid the inconsistency at the boundary

$$
(3.4) \qquad
\begin{aligned}
\tilde{f}_{nh}(t) = & - \int \mathbb{K}\left(\frac{h - x}{h}\right) d\hat{f}_n(x) \\
& + (t - h) \int h^{-1} K\left(\frac{h - x}{h}\right) d\hat{f}_n(x),
\end{aligned}
$$

and near the endpoint 36 we use a similar correction.

The 95% confidence intervals for the survival function at the 99 equidistant points $0.36, 0.72, \ldots, 35.64$, are constructed from 1000 bootstrap samples $T_1^*, \ldots,$ T_n^*, also of size n, drawn from the original sample, and in these samples we computed

$$
(3.5) \qquad
\begin{aligned}
& \tilde{f}_{n,h_{\text{bootstrap}}}^*(t) / \tilde{f}_{n,h_{\text{bootstrap}}}^*(0) \\
& \quad - \tilde{f}_{n,h_{\text{original}}}(t) / \tilde{f}_{n,h_{\text{original}}}(0),
\end{aligned}
$$

where $\tilde{f}_{nh}$ and $\tilde{f}_{nh}^*$ are the SMLEs in the original sample and the bootstrap sample, respectively. That this procedure will work is nontrivial, but can be argued along the lines of Groeneboom and Hendrickx (2017b) (see also Section 6). The bandwidths were chosen by least squares cross validation, both in the original and in the bootstrap samples. The 95% asymptotic confidence intervals are given by

$$
\left[\tilde{f}_{nh}(t) / \tilde{f}_{nh}(0) - U_{0.975}^*, \tilde{f}_{nh}(t) / \tilde{f}_{nh}(0) - U_{0.025}^*\right],
$$

where $U_{0.025}^*$ and $U_{0.975}^*$ are the 2.5% and 97.5% percentiles of the bootstrap values (3.5), and h is the bandwidth in the original sample. The result is shown on the left of Figure 5 and should be compared with the confidence intervals in part A of Figure 2, page 1495 of Slama et al. (2012), based on a parametric (generalized gamma) model. Note that our treatment is for two reasons different from that in Groeneboom and Jongbloed (2014) and Groeneboom and Jongbloed (2015):

1. Least squares cross validation is used to make the procedure fully automatic.
2. The local linear boundary correction is used (3.4) to deal with the inconsistency at zero.
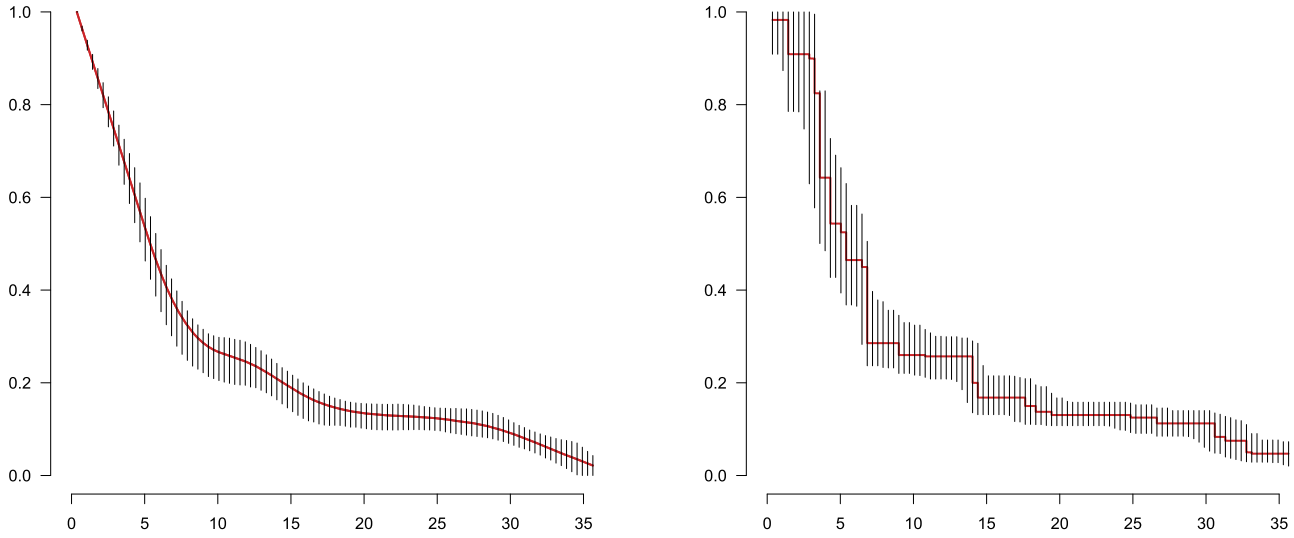
FIG. 5. *Left: 95% confidence intervals, based on the SMLE, for the data in Slama et al.* (2012), *at the points* 0.36, 0.72, . . . , 35.64. *The bandwidth was determined by cross validation, both in the original sample and in the bootstrap samples. Right: Banerjee–Wellner type 95% confidence intervals, based on LR test for the* (*restricted*) *MLE's, as proposed in Groeneboom and Jongbloed* (2015).

In Groeneboom and Jongbloed (2015) also a procedure is discussed to compute confidence intervals, as suggested by Banerjee and Wellner (Banerjee and Wellner, 2001). It is proved in Groeneboom and Jongbloed (2015) that this method can also be applied to the Grenander estimator and this method is also used to analyze the present data there. One has to avoid dividing by an inconsistent estimate at zero, and can condition the Grenander estimate to have a value at zero which is consistent.

We took the value produced by the boundary-corrected SMLE for determining the density estimate at zero. To this end, one has to solve a nonlinear equation, for which Brent's algorithm was used. For the likelihood ratio test, on which the Banerjee–Wellner method is based, one also has to fix the value of the Grenander estimate at the point where one constructs a confidence interval, and for this purpose one has to solve two nonlinear equations (see Lemma 3.4 on page 2045 of Groeneboom and Jongbloed, 2015). To this end, we presently use the Hooke–Jeeves method; the corresponding code is given in Groeneboom (2015). The result is shown on the right of Figure 5.

3.5 Limits of Global Functionals

The Grenander estimator has also been studied from a more global perspective. How well does the estimator approximate the underlying density in global measures of distance? The first result on this was the following limit relation, proved in Groeneboom and Pyke (1983) and Groeneboom (1983).

Let f_0 be the uniform density on [0, 1] and let $\hat{f}_n$ be the Grenander estimator, based on a sample of size n from f_0. Then

$$(3.6) \qquad \frac{1}{\sqrt{3 \log n}} \{n \| \hat{f}_n - f_0 \|_2^2 - \log n \} \xrightarrow{\mathcal{D}} N(0, 1).$$

Noting that

$$\frac{1}{\sqrt{3}} \{ \sqrt{n} \| \hat{f}_n - f_0 \|_2 - \sqrt{\log n} \}$$

$$= \sqrt{\frac{\log n}{3}} \left\{ \sqrt{\frac{n}{\log n}} \| \hat{f}_n - f_0 \|_2 - 1 \right\}$$

we get by the delta method, using the square root transformation:

$$(3.7) \qquad \frac{2}{\sqrt{3}} \{ n^{1/2} \| \hat{f}_n - f_0 \|_2 - \sqrt{\log n} \} \xrightarrow{\mathcal{D}} N(0, 1).$$

For a *strictly monotone* density f_0, the following result was stated in Groeneboom (1985) for the L_1-distance, with a sketch of proof:

$$(3.8) \quad \begin{aligned} & n^{1/6} \{ n^{1/3} \| \hat{f}_n - f_0 \|_1 - \mu \} \\ & = n^{1/2} \| \hat{f}_n - f_0 \|_1 - n^{1/6} \mu \xrightarrow{\mathcal{D}} N(0, \sigma^2), \end{aligned}$$

for constants μ and σ, depending on the inverse process

$$(3.9) \quad V(c) = \sup\{t : W(t) - (t - c)^2 \text{ is maximal}\}.$$

In fact, $\mu = 2E|V(0)| \int_0^1 |\frac{1}{2} f'(t) f(t)|^{1/3} dt$, and

$$(3.10) \qquad \sigma^2 = 8 \int_0^\infty \text{covar}(|V(0)|, |V(c) - c|) \, dc.$$

A rigorous proof is given in Groeneboom, Hooghiemstra and Lopuhaä (1999) and this result is also discussed in Groeneboom and Jongbloed (2014); see Theorem 13.1 on page 379. The result has been generalized to more general L_p-distances by Durot, Kulikov, Lopuhaä and the paper of Cécile Durot and Rik Lopuhaä in this issue gives detailed information on these developments. Note that $\sqrt{n}\|\hat{f}_n - f_0\|_p$ after centering by a constant which increases with n, is $O_p(1)$, both in (3.7) and (3.8). Also note that, by (3.10), the variance σ^2 in (3.8) is independent of f_0.
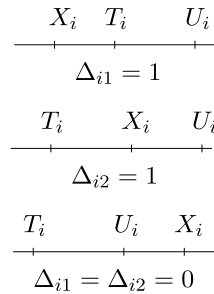
Relations (3.7) and (3.8) were the first results on the limit distribution of L_p-norms of the distance of the Grenander estimator to the true density.

4. INTERVAL CENSORING

4.1 The Model

In interval censoring models, there is a sample $X_1, X_2, \ldots, X_n$ from a distribution with distribution function F_0 in the background. Instead of observing this sample, one observes a set of n intervals $I_1, \ldots, I_n$ with the information on the X_i's that $X_i \in I_i$. In Section 2, current status model was introduced. There the intervals are all of the form $[0, T_i]$ or (T_i, ∞).

In the interval censoring case 2 model, one only observes for each i whether $X_i \leq T_i$ or $X_i \in (T_i, U_i]$ or $X_i > U_i$, for some random pair (T_i, U_i), where $T_i < U_i$, where the (T_i, U_i)'s are independent of the X_i's, yielding intervals of the form $[0, T_i]$, $(T_i, U_i]$ or (U_i, ∞).

$$\begin{array}{ccc} X_i & T_i & U_i \\ \hline \end{array}$$
$$\Delta_{i1} = 1$$

$$\begin{array}{ccc} T_i & X_i & U_i \\ \hline \end{array}$$
$$\Delta_{i2} = 1$$

$$\begin{array}{ccc} T_i & U_i & X_i \\ \hline \end{array}$$
$$\Delta_{i1} = \Delta_{i2} = 0$$

The data can be represented as

$$(T_i, U_i, \Delta_{i1}, \Delta_{i2}) = (T_i, U_i, 1_{\{X_i \leq T_i\}}, 1_{\{X_i \in (T_i, U_i]\}}).$$

The analysis of the case 2 model is much more complicated than the analysis of the current status model. Already in computing the MLE, the difference is apparent. While the MLE in the current status model can be constructed explicitly, as will be seen below, for the interval censoring case 2 problem iterative procedures are needed to compute it. In this paper, we concentrate on the current status model.

We want to estimate the unknown distribution function F_0 of X_i, based on the data $(T_i, \Delta_i) = (T_i, 1_{\{X_i \leq T_i\}})$. If the X_i are independent of the T_i, using density (2.4), the log likelihood function in F (conditional on the T_i's) is given by

$$\ell(F) = \sum_{i=1}^{n}\{\Delta_i \log F(T_i) + (1 - \Delta_i) \log(1 - F(T_i))\}.$$

The (nonparametric) maximum likelihood estimator (MLE) $\hat{F}_n$ of F_0 maximizes $\ell(F)$ over the class of *all* distribution functions F.

Again, just as in the case of the Grenander estimator, the MLE $\hat{F}_n$ has an interpretation as a least squares estimator. Writing $y_i = F(T_{(i)})$, $1 \leq i \leq n$, the MLE $\hat{F}_n$ *minimizes* the sum

$$\sum_{i=1}^{n}\{\Delta_{(i)} - y_i\}^2$$

over $(y_1, \ldots, y_n)$ such that $0 \leq y_1 \leq \cdots \leq y_n \leq 1$ where $\Delta_{(i)}$ denotes the indicator variable corresponding to the order statistic $T_{(i)}$. Using Lemma 2.1, the solution $(y_1, \ldots, y_n)$ is given by the isotonic regression of $(\Delta_{(1)}, \ldots, \Delta_{(n)})$.

4.2 Limit Distribution of (Smoothed) Maximum Likelihood Estimator in the Current Status Model

The following result gives the limit distribution of the MLE $\hat{F}_n$ in the current status model.

THEOREM 4.1 (Groeneboom, 1987). *Let F_0 and G be differentiable at t with strictly positive derivatives $f_0(t)$ and $g(t)$. Let $\hat{F}_n$ be the MLE of F_0. Then, as $n \to \infty$,*

$$\frac{n^{1/3}\{\hat{F}_n(t) - F_0(t)\}}{\{4F_0(t)(1 - F_0(t))f_0(t)/g(t)\}^{1/3}} \xrightarrow{\mathcal{D}} Z,$$

where $Z = \text{argmax}_t\{W(t) - t^2\}$.

Note the similarity with Theorem 3.3. Proofs, using the switch relation (see Section 3.3 that are given in van der Vaart and Wellner, 1996, Section 3.2.15, and Groeneboom and Jongbloed, 2014, Section 3.8 and Figure 6).

As is clear from the construction of the MLE, and from Figure 7, the MLE is a discrete distribution function. If one assumes the underlying distribution function to be smooth, one can use the MLE to define a
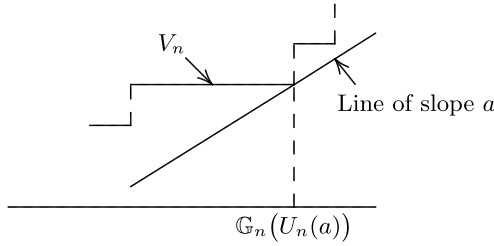
FIG. 6. *The switch relation*; $\mathbb{G}_n$ *is the empirical distribution function of the* T_i, $V_n(t) = n^{-1}\Delta_i 1_{\{T_i \leq t\}}$ *and* $U_n(a) = \text{argmin}\{t : V_n(t) - a\mathbb{G}_n(t)\}$.

smooth estimator of F_0. This smoothed maximum likelihood estimator (SMLE) $\hat{F}_{n,h}^{(\text{SMLE})}$ is (modulo a boundary correction), defined by

$$\hat{F}_{n,h}^{(\text{SMLE})}(t) = \int \mathbb{K}\left(\frac{t-x}{h}\right) d\hat{F}_n(x),$$

(4.1)

$$\mathbb{K}(y) = \int_{-\infty}^{y} K(u)\,du,$$

where K is (for example) the triweight kernel

$$K(u) = \frac{35}{32}\{1 - u^2\}^3 1_{[-1,1]}(u).$$

One generally takes $h \asymp n^{-1/5}$ (the usual bandwidth in density estimation). The SMLE simply smoothes the MLE, preserving the monotonicity. As will be seen in Section 6, the SMLE also plays an important role in the bootstrap approach to finding confidence sets for $F_0(t)$.

For the SMLE, we have the following limit result.
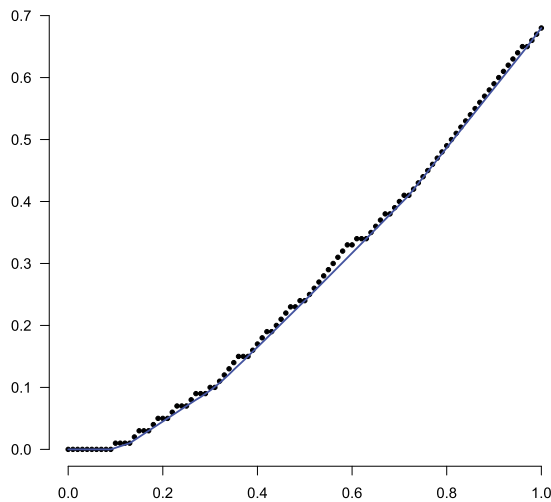
THEOREM 4.2 (Groeneboom, Jongbloed and Witte, 2010). *Let* F_0 *be differentiable at* t *with second derivative* $f_0'(t) \neq 0$ *and let* $g(t) > 0$ *and have a bounded derivative at* t. *If* $h_n = cn^{-1/5}$, *for some* $c > 0$, *then*

$$n^{2/5}\big(\hat{F}_{n,h_n}^{(\text{SMLE})}(t) - F_0(t)\big) \xrightarrow{\mathcal{D}} N(\mu, \sigma^2),$$

where

$$\mu = \frac{1}{2}c^2 f_0'(t) \int u^2 K(u)\,du,$$

$$\sigma^2 = \frac{F_0(t)\{1 - F_0(t)\}}{cg(t)} \int K(u)^2\,du.$$

4.3 Smooth Functionals and Current Status Regression

The SMLE of Theorem 4.2 is an example of a local smooth functional, which converges at a faster rate than the MLE. We now turn to smooth functionals of the MLE which attain the $\sqrt{n}$-rate. The following result is the first of this kind for the current status model, proved during a summer course given in Stanford in 1990 (Groeneboom, 1991) and also given in Groeneboom and Wellner (1992) (the second part of this book is in fact virtually the same as the lecture notes of the Stanford course, but contains more misprints).

THEOREM 4.3 (Groeneboom, 1991). *Let* F_0 *be differentiable on* $[0, B]$ *and let the density* g *of the inspection times stay away from zero on* $[0, B]$. *Then*

$$\sqrt{n}\left\{\int x\,d\hat{F}_n(x) - \int x\,dF_0(x)\right\} \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$
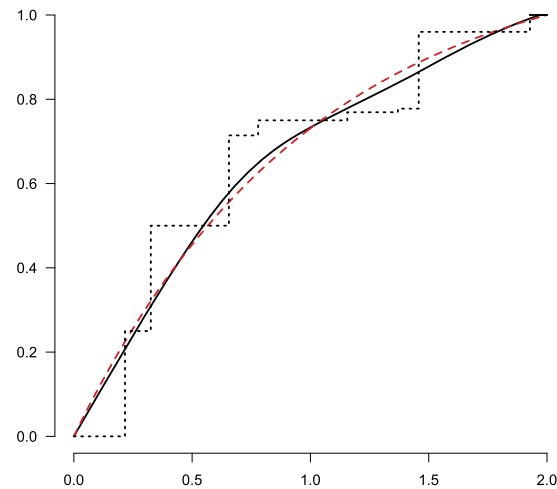

FIG. 7. *Left*: *the cusum diagram with points* $(\frac{i}{n}, \frac{1}{n}\sum_{j=1}^{i}\Delta_{(j)})$ *and its greatest convex minorant. Right*: *the MLE* (*dotted*), *the SMLE* (*solid*) *and the real distribution function* (*dashed*).

where

$$\sigma^2 = \int \frac{F_0(x)\{1 - F_0(x)\}}{g(x)} \, dx.$$

The mapping $F \mapsto K(F) = \int x \, dF(x)$ is a *smooth functional* for the current status model. On the other hand, $F \mapsto F(t)$ is, for example, *not* a smooth functional in the current status model and cannot be expected to have estimators that converge at $\sqrt{n}$ rate (see van der Vaart, 1991).

Proving results of the type of Theorem 4.3 for the interval censoring, case 2, model is however much more difficult. The reason for the increased difficulty is the fact that the behavior has to be deduced from the qualitative properties of solutions of integral equations for which no explicit solutions are known.

The general pattern for the proof of the $\sqrt{n}$ rate and efficiency of a smooth functional of the MLE is as follows, which we demonstrate here for the current status model:

- Show that the nonlinear aspect of the functional is negligible. This means

$$\sqrt{n}\{K(\hat{F}_n) - K(F_0)\}$$
$$= \sqrt{n} \int \kappa_{F_0} \, d(\hat{F}_n - F_0) + o_p(1),$$

for some "score" κ_{F_0}.
- Transformation to the observation space measure

$$\int \kappa_{F_0} \, d(\hat{F}_n - F_0) = -\int \theta_{\hat{F}_n}(t, \delta) \, dQ_0(t, \delta),$$

where Q_0 is the probability measure of (T_i, Δ_i),

$$\theta_{\hat{F}_n}(t, \delta) = k_{F_0}(t) \frac{\delta - \hat{F}_n(t)}{g(t)},$$

and $k_{F_0}(t) = \kappa'_{F_0}(t)$. As an example, for the mean function, $\kappa_{F_0}(t) = t$ and $k_{F_0} \equiv 1$. More generally (e.g., in the interval censoring, case 2, case) $\theta_{\hat{F}_n}(t, \delta)$ is the solution of a particular integral equation (see Groeneboom, 2013).
- Use that $\hat{F}_n$ is the MLE. Replace $\theta_{\hat{F}_n}$ by

$$\bar{\theta}_{\hat{F}_n}(t, \delta) = \bar{k}_{F_0}(t)(\delta - \hat{F}_n(t))/\bar{g}(t),$$

where $\bar{k}_{F_0}$ and $\bar{g}$ are constant on the same intervals as $\hat{F}_n$. Then

$$(4.2) \qquad \int \bar{\theta}_{\hat{F}_n}(t, \delta) \, dQ_n = 0,$$

and

$$\int \kappa_{F_0} \, d(\hat{F}_n - F_0)$$
$$\overset{\text{preceding step}}{=} -\int \theta_{\hat{F}_n} \, dQ_0$$
$$\overset{(4.2)}{=} \int \bar{\theta}_{\hat{F}_n} \, d(Q_n - Q_0) - \int \{\theta_{\hat{F}_n} - \bar{\theta}_{\hat{F}_n}\} \, dQ_0.$$

- Asymptotic variance equals information lower bound. Show

$$\int \bar{\theta}_{\hat{F}_n} \, d(Q_n - Q_0)$$
$$= \int \theta_{F_0} \, d(Q_n - Q_0) + o_p(n^{-1/2}),$$

where

$$\theta_{F_0}(t, \delta) = k_{F_0}(t) \frac{\delta - F_0(t)}{g(t)},$$

and show

$$\int \{\theta_{\hat{F}_n} - \bar{\theta}_{\hat{F}_n}\} \, dQ_0 = o_p(n^{-1/2}).$$

Instead of the completely nonparametric current status model, one can also consider the current status regression model, where the observations are of the form $(\mathbf{X}_i, T_i, \Delta_i)$, and where

$$\Delta_i = 1_{\{Y_i \le T_i\}}, \quad Y_i = \mathbf{X}_i'\boldsymbol{\beta}_0 + \varepsilon_i,$$

where $\mathbf{X}_i$ is a k-dimensional covariate and ε_i an observation error with expectation α_0 (the intercept), independent of the T_i and $\mathbf{X}_i$. In this case, we cannot observe Y_i and only have the indirect information via $(\mathbf{X}_i, T_i, \Delta_i)$. Then the (relevant part) of the log likelihood becomes

$$\sum_{i=1}^n \{\Delta_i \log F(T_i - \mathbf{X}_i'\boldsymbol{\beta}) + (1 - \Delta_i)$$
$$\cdot \log\{1 - F(T_i - \mathbf{X}_i'\boldsymbol{\beta})\}\},$$

where F_0 is the distribution function of ε_i. Here and in the following, we will denote the parameters by boldface letters is they are higher dimensional, but not if the are just 1-dimensional.

The following 1-dimensional example is taken from Groeneboom and Hendrickx (2018b): X_i and T_i are uniform on $[0, 2]$, $\beta_0 = 1/2$, $\alpha_0 = \mathbb{E}\varepsilon_i = 1/2$, and ε_i has a density which is a rescaled version of the density $6x(1 - x)$ on $[0, 1]$. We rescale it to a density on $[3/8, 5/8]$.

One can compute the profile log likelihood. Pick a β: maximize for this β the log likelihood

$$(4.3) \quad \ell_\beta(F) = \sum_{i=1}^{n} \{\Delta_i \log F(T_i - X_i\beta) + (1 - \Delta_i) \\ \cdot \log\{1 - F(T_i - X_i\beta)\}\}$$

over F, this gives $\hat{F}_{n,\beta}$ via one step (convex minorant) algorithm. Now find an argmax for β over all $\hat{F}_{n,\beta}$ so obtained. Note that $\ell_\beta(\hat{F}_{n,\beta})$ is piecewise constant and can only change if there is a change of ordering of the observations $T_i - \beta X_i$ as a consequence of a change of β. A picture of the profile log likelihood is shown on the left of Figure 8.

What are the properties of the MLE of β_0 so obtained? This is still unknown, although estimators of this type have been studied for at least 25 years. It is, for example, unknown whether the MLE of β_0 is $\sqrt{n}$-consistent. Li and Zhang (1998) conjecture that the MLE will give a $\sqrt{n}$-consistent but inefficient estimate. Murphy, van der Vaart and Wellner (1999) prove that, in a 1-dimensional model in which one only has observations from a part of F_0 where F_0 stays away from 0 and 1, the MLE gives a $n^{1/3}$-consistent estimate of β_0.

Returning to the general case where the regression parameter is a vector, it is still not completely clear whether it really helps in the estimation of $\boldsymbol{\beta}_0$ to use an isotonic estimator of F_0. Groeneboom and Hendrickx (2018b) show that it is possible to use the isotonic estimate for constructing $\sqrt{n}$-rate estimates of the regression parameter which are arbitrarily (depending on a

truncation parameter) close to an efficient estimate. But they also show this for a nonisotonic Nadaraya–Watson plug-in estimate of F_0, which is similar to the estimates used in the econometric literature for the binary choice model (e.g., Klein and Spady, 1993, Cosslett, 2007), and is defined as the Nadaraya–Watson estimator

$$(4.4) \quad F_{nh,\boldsymbol{\beta}}(t - \boldsymbol{\beta}'\mathbf{x}) \\ = \frac{\sum_{i=1}^{n} \Delta_i K_h(t - \boldsymbol{\beta}'\mathbf{x} - T_i + \boldsymbol{\beta}'\mathbf{X}_i)}{\sum_{i=1}^{n} K_h(t - \boldsymbol{\beta}'\mathbf{x} - T_i + \boldsymbol{\beta}'\mathbf{X}_i)},$$

where $K_h = h^{-1}K(\cdot/h)$ is a scaled version of a probability density function K. Note the smoothness of the log likelihood function in Figure 9 (also for a simulated sample of size $n = 1000$) in comparison with Figure 8.

Figure 9 shows the derivative w.r.t. β of the log likelihood, if F is of the form (4.4), and for these estimators the maximization of the likelihood gives the same result as the root of the score equation. This is completely different for the estimators which use the MLE $\hat{F}_{n,\boldsymbol{\beta}}$, where we cannot differentiate in this way. Nevertheless, Groeneboom and Hendrickx (2018b) define in this case also a "score function" (see the right picture in Figure 8) and show that the root of this score function produces $\sqrt{n}$ consistent estimates of the regression parameter. But the relation between log likelihood and the score as its derivative is no longer there. In this case, it is also not clear that the maximizer of the log likelihood is $\sqrt{n}$-consistent.

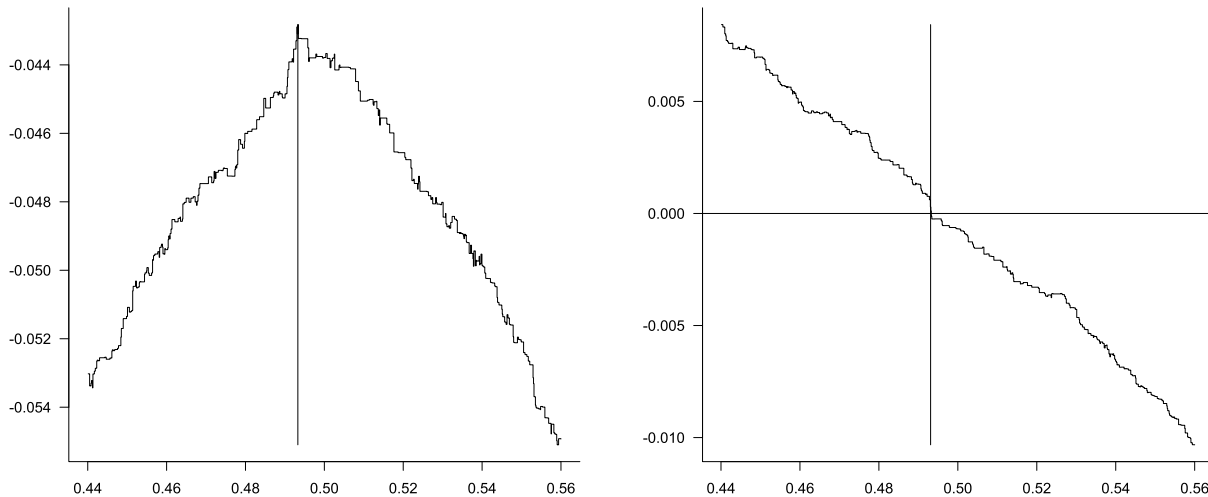The conditions used in the econometric literature are much stronger than the conditions in Groeneboom

FIG. 8. *Left: the profile log likelihood function $\beta \mapsto \ell_\beta(\hat{F}_{n,\beta})$, where ℓ_β is as in (4.3) and $\hat{F}_{n,\beta}$ is the MLE for fixed β. Right: the "score function" $\beta \mapsto n^{-1}\sum_{i=1}^{n} X_i\{\hat{F}_{n,\beta}(T_i - X_i\beta) - \Delta_i\}$. The vertical line denotes the location of the maximum of $\ell_\beta(\hat{F}_{n,\beta})$, respectively, the crossing of zero of the score function. The sample size $n = 1000$.*
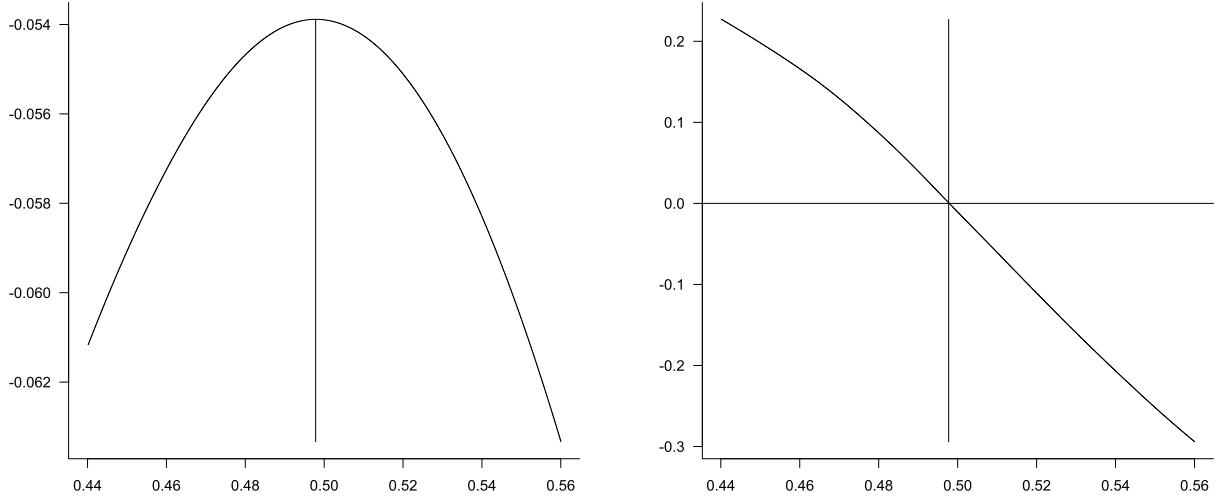
FIG. 9.   *Left: the function* $\beta \mapsto \ell_\beta(F_{nh,\beta})$, *where* ℓ_β *is as in* (4.3) *and* $F_{nh,\beta}$ *is the plug-in estimator for fixed* β. *Right: the "score function"* $\beta \mapsto \frac{\partial}{\partial \beta} \ell_\beta(F_{nh,\beta})$. *The vertical line denotes the location of the maximum of* $\ell_\beta(F_{n,\beta})$, *respectively, the location of the zero of the score function. The sample size* $n = 1000$.

and Hendrickx (2018b) and also use heavier smoothing (e.g., Klein and Spady, 1993 use higher-order kernels and bandwidths h of order $n^{-1/6} \lesssim h \lesssim n^{-1/8}$). The corresponding result using the Nadaraya–Watson estimator in Groeneboom and Hendrickx (2018b) only uses the existence of second derivatives and a bandwidth of order $n^{-1/5}$ (the usual bandwidth order in density estimation).

A typical result, based on the score function of the type shown in Figure 8, runs as follows. Let, for $\varepsilon \in (0, 1/2)$, the function $\psi_n^{(\varepsilon)}$ be defined by defined by

$$
\psi_n^{(\varepsilon)}(\beta)
$$

$$
(4.5) \quad \stackrel{\text{def}}{=} \int_{\hat{F}_{n,\beta}(t-\beta'\mathbf{x}) \in [\varepsilon, 1-\varepsilon]} \mathbf{x}
$$

$$
\cdot \left\{ \delta - \hat{F}_{n,\beta}(t - \beta'\mathbf{x}) \right\} d\mathbb{P}_n(t, \mathbf{x}, \delta),
$$

where $\hat{F}_{n,\beta}$ is the MLE based on the order statistics of the values $T_i - \beta'\mathbf{X}_i$, $i = 1, \ldots, n$ for chosen β. Then we have the following.

THEOREM 4.4.   *Under the regularity assumptions in Groeneboom and Hendrickx* (2018b), *we have*:

(i) (Existence of a root) *For all large* n, *a crossing of zero* $\hat{\beta}_n$ *of* $\psi_n^{(\varepsilon)}$ *exists with probability tending to one.*

(ii) (Consistency)

$$
\hat{\beta}_n \xrightarrow{p} \beta_0, \quad n \to \infty.
$$

(iii) (Asymptotic normality) $\sqrt{n}\{\hat{\beta}_n - \beta_0\}$ *is asymptotically normal with mean zero and variance*

$A^{-1}BA^{-1}$, *where*

$$
A = \mathbb{E}_\varepsilon \left[f_0(T - \beta_0'\mathbf{X}) \operatorname{Cov}(\mathbf{X}|T - \beta_0'\mathbf{X}) \right],
$$

and

$$
B = \mathbb{E}_\varepsilon \left[F_0(T - \beta_0'\mathbf{X})\{1 - F_0(T - \beta_0'\mathbf{X})\} \right.
$$

$$
\left. \cdot \operatorname{Cov}(\mathbf{X}|T - \beta_0'\mathbf{X}) \right],
$$

defining $\mathbb{E}_\varepsilon(w(T, \mathbf{X}, \Delta)) = \mathbb{E}1_{\{F_0(t-\beta_0'\mathbf{x}) \in [\varepsilon, 1-\varepsilon]\}} \cdot w(T, \mathbf{X}, \Delta)$ *for functions* w, *and assuming that* A *is nonsingular.*

Probably the ε-truncation in Theorem 4.4 is not really necessary (i.e., we can take $\varepsilon = 0$) and only concerns a technical point. Simulations in Groeneboom and Hendrickx (2018b) show that the simple estimators which use the isotonic MLE $\hat{F}_{n,\beta}$ might generally have a better behavior for ordinary sample sizes (e.g., in the range $n = 100$ to $n = 1000$) than the estimators, based on the nonisotonic Nadaraya–Watson estimators of F_0.

Theorem 4.4 is a first result showing that one can construct $\sqrt{n}$-consistent estimators on the basis of the isotonic MLE's, using the "score" (4.5). Note that a similar result is still unknown to exist for the corresponding profile likelihood estimate. One will need a better insight in how the process $\beta \mapsto \ell_\beta(\hat{F}_{n,\beta})$ [see (4.3) and the left picture in Figure 8] changes as a function of the order changes of $T_1 - \beta'\mathbf{X}_1, \ldots, T_n - \beta'\mathbf{X}_n$ with varying β.

5. ANALYTIC CHARACTERIZATION OF THE DISTRIBUTION OF Z

In Theorem 3.3, the random variable Z was defined as the location the maximum of two-sided Brownian motion minus a parabola. The first step in the analytic characterization of this distribution was done by Chernoff in a paper on an estimate of the mode of a distribution.

We sketch the line of argument in that paper, leading to the characterization.

(i) Define $u(s, x)$ by

$$u(s, x) = \mathbb{P}\{W(t) > t^2 \text{ for some } t > s \,|\, W(s) = x\}.$$

Then for $x < s^2$ fixed,

$$u(s, x)$$
$$= \mathbb{P}\{W(t) > t^2 \text{ for some } t > s + \epsilon \,|\, W(s) = x\}$$
$$\quad + \mathbb{P}\{W(t) > t^2 \text{ for some } t \in (s, s + \epsilon],$$
$$\quad W(t) \leq t^2 \text{ for all } t > s + \epsilon \,|\, W(s) = x\}$$
$$= \mathbb{E}\{u(s + \varepsilon, W(s + \varepsilon)) \,|\, W(s) = x\} + o(\epsilon)$$
$$= u(s, x) + \frac{\partial}{\partial s} u(s, x)\epsilon + \frac{1}{2}\frac{\partial^2}{\partial x^2} u(s, x)\epsilon$$
$$\quad + o(\epsilon), \quad \epsilon \downarrow 0.$$

Hence

$$\frac{\partial}{\partial s} u(s, x) = -\frac{1}{2}\frac{\partial^2}{\partial x^2} u(s, x)$$

and

$$u(s, x) = 1, \quad \forall x \geq s^2, \forall s \in \mathbb{R},$$
$$\forall s \in \mathbb{R} : u(s, x) \to 0, x \to -\infty.$$

(ii) Define

$$M_h = \max_{t \in [s-h, s]} W(t).$$

Then, given $W(s)$, $W(s - h)$ and M_h, the maximum value of $W(t) - t^2$ over the range $[s - h, s]$ is $M_h - s^2 + O(h)$, and, using the spatial homogeneity of Brownian motion, we obtain [for notational clarity writing W' for Brownian motion, starting at $s^2 + W(s) - M_h$, in the second line]:

$$\mathbb{P}\Big\{\max_{t \geq s}\{W(t) - t^2\} > M_h - s^2 \,|$$
$$\quad W(s), W(s - h), M_h\Big\}$$
$$= \mathbb{P}\Big\{\max_{t \geq s}\{W'(t) - t^2\} > 0 \,|$$

$$\quad W'(s) = s^2 + W(s) - M_h\Big\}$$
$$= u(s, s^2 + W(s) - M_h)$$
$$= u(s, s^2)(= 1) + \{W(s) - M_h\}\partial_2 u(s, s^2)$$
$$\quad + O_p(h).$$

Similarly,

$$\mathbb{P}\Big\{\max_{t \leq s-h}\{W(t) - t^2\} \geq M_h - (s - h)^2 \,|$$
$$\quad W(s), W(s - h), M_h\Big\}$$
$$= u(-s, s^2)(= 1)$$
$$\quad + \{W(s - h) - M_h\}\partial_2 u(-s, s^2) + O_p(h).$$

(iii) From this, it follows that

$$\mathbb{P}\{Z \in [s - h, s]\}$$
$$\sim \mathbb{E}\Big\{\mathbb{P}\Big\{\max_{t \notin [s-h, s]}\{W(t) - t^2\} < M_h - s^2 \,|$$
$$\quad W(s), W(s - h), M_h\Big\}\Big\}$$
$$\sim \mathbb{E}\{M_h - W(s)\}$$
$$\quad \cdot \{M_h - W(s - h)\}\partial_2 u(s, s^2)\partial_2 u(-s, s^2)$$
$$\sim h\mathbb{E}\Big(\max_{x \in [0, 1]} B(x)\Big)^2 \partial_2 u(s, s^2)\partial_2 u(-s, s^2)$$

($B = $ Brownian Bridge)

$$\sim \frac{1}{2}h\partial_2 u(s, s^2)\partial_2 u(-s, s^2), \quad h \downarrow 0.$$

This leads to the following.

THEOREM 5.1 (Chernoff, 1964). *The density f_Z of $Z = \mathrm{argmax}\{W(t) - t^2\}$ is given by*

$$f_Z(s) = \frac{1}{2}\partial_2 u(-s, s^2)\partial_2 u(s, s^2),$$

where $u(s, x)$ solves the heat equation

$$\frac{\partial}{\partial s} u(s, x) = -\frac{1}{2}\frac{\partial^2}{\partial x^2} u(s, x)$$

subject to

$$u(s, x) = 1, \quad x \geq s^2, \quad u(s, x) \to 0, \quad x \to -\infty.$$

The original computations of this density were based on numerically solving Chernoff's heat equation. This was done by Chernoff himself (personal communication), with the help of people from numerical mathematics, and by Groeneboom in 1982 at the Mathematical Centre, Amsterdam (now CWI), also with the help

of the Numerical Mathematics Department. The latter mathematicians (in particular, Ben Sommeijer) noticed the instability of the solutions in the region where the time argument s is negative, if the then rather fashionable "multigrid method" for the solution of partial differential equation was used. This phenomenon was explained by Groeneboom (1984):

$$\partial_2 u(-s, s^2) \sim c_1 \exp\left\{-\frac{2}{3}s^3 - cs\right\}, \quad s \to \infty,$$

where $c \approx 2.9458$ and $c_1 \approx 2.2638$. This fast decay entails that a numerical solution of this partial differential equation on a grid will not give a really accurate solution. Theorem 5.2 provides a representation of the density of $Z = \operatorname{argmax} W(t) - t^2$ in terms of the Airy function.

THEOREM 5.2 (Groeneboom, 1984, Daniels and Skyrme, 1985). *The probability density f of the location of the maximum of the process $t \mapsto W(t) - t^2$, $t \in \mathbb{R}$, is given by*

$$f(s) = \frac{1}{2}g(s)g(-s),$$

where

$$g(s) = \frac{1}{2^{2/3}\pi} \int_{-\infty}^{\infty} \frac{e^{-ius}}{\operatorname{Ai}(i2^{-1/3}u)} \, du,$$

where Ai *is the Airy function* Ai.

Using this theorem, the density can be computed efficiently by two lines in Mathematica; see Figure 10.

Theorem 5.2 is also given in Janson (2013) and Groeneboom, Lalley and Temme (2015). The proof in

the latter paper seems at present the easiest way to obtain the result and is based on the maximum principle for parabolic partial differential equations. The distribution of the maximum itself, $\max_t W(t) - t^2$, was studied in Janson, Louchard and Martin-Löf (2010), Groeneboom (2010) and Groeneboom and Temme (2011).

Another interesting result is the following:

$$\operatorname{var}(Z) = \frac{1}{3}\mathbb{E}\max_t\{W(t) - t^2\}.$$

This is proved in Groeneboom (2011) and Janson (2013), and (not using the relation with Airy functions) in Pimentel (2014).

The Airy functions enter via the Cameron–Martin–Girsanov formula and the Feynman–Kac or Itô formula. The proof of Chernoff's result (Chernoff, 1964) and an exposition of how the Airy functions enter and of the recent proof of Theorem 5.2 in Groeneboom, Lalley and Temme (2015) is given in Groeneboom (2018).

6. CAN WE USE THE BOOTSTRAP?

For the Grenander estimator, the following negative result holds.

THEOREM 6.1 (Kosorok, 2008). *The nonparametric bootstrap, where one resamples with replacement from the (T_i, Δ_i) and recomputes the Grenander estimator for the bootstrap sample, is* inconsistent, *that is,*

$$n^{1/3}\{\hat{f}_n^*(t) - \hat{f}_n(t)\} \xrightarrow{\mathcal{D}} |4f'(t)f(t)|^{1/3}Z$$

```
In[1]:= f[x_] := (1 / (2 * Pi)) * 2 ^ (1 / 3) *
          Re[NIntegrate[Exp[-I * u * x] / AiryAi[I * 2 ^ (-1 / 3) * u], {u, -10, 10}]]

In[2]:= g[x_] := (1 / 2) * f[x] * f[-x]

In[3]:= Plot[g[x], {x, -2, 2}]
```
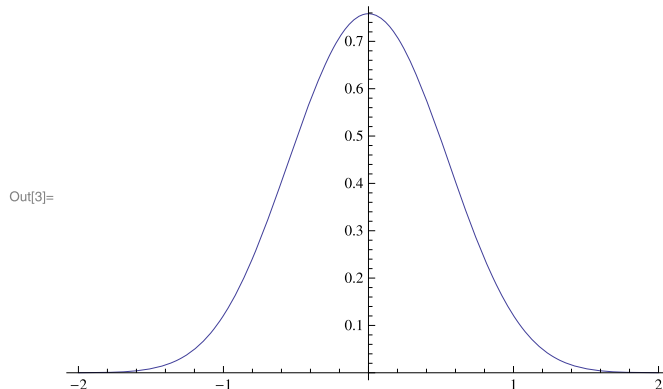
FIG. 10.   *Density of $Z = \operatorname{argmax}\{W(t) - t^2\}$.*

does not hold (*in probability*), *conditionally on the data, where*

$$Z = \operatorname{argmax}\{W(x) - x^2 : x \in \mathbb{R}\}.$$

The result has a rather interesting proof by contradiction.

(i) Suppose $n^{1/3}\{\hat{f}_n^*(t) - \hat{f}_n(t)\} \xrightarrow{\mathcal{D}} |4f'(t)f(t)|^{1/3}Z$ (in probability), conditionally on the data. Then

$$n^{1/3}\{\hat{f}_n^*(t) - f(t)\}$$
$$= n^{1/3}\{\hat{f}_n^*(t) - \hat{f}_n(t)\} + n^{1/3}\{\hat{f}_n(t) - f(t)\}$$
$$\xrightarrow{\mathcal{D}} |4f'(t)f(t)|^{1/3}\{Z_1 + Z_2\},$$

where the Z_i are independent copies of $\operatorname{argmax}\{W(x) - x^2\}$.

(ii) On the other hand,

$$n^{1/3}\{\hat{f}_n^*(t) - f(t)\}$$
$$\xrightarrow{\mathcal{D}} |4f'(t)f(t)|^{1/3}$$
$$\cdot \operatorname*{argmax}_x\{W_1(x) + W_2(x) - x^2\}.$$

The right-hand side [with *smaller variance* than the limit in (i)] comes from

$$\mathbb{F}_n^*(t + n^{-1/3}x) - \mathbb{F}_n^*(t)$$
$$- (\mathbb{F}_n(t + n^{-1/3}x) - \mathbb{F}_n(t)) + \mathbb{F}_n(t + n^{-1/3}x)$$
$$- \mathbb{F}_n(t) - f(t)n^{-1/3}x,$$

where the first line corresponds with $x \mapsto W_1(x)$ and the second line with $x \mapsto W_2(x) - x^2$.

Sen, Banerjee and Woodroofe (2010) show that generating bootstrap samples for the Grenander estimator itself also does not work. In this case, we have

$$n^{1/3}\{\hat{f}_n^*(t) - f(t)\}$$
$$(6.1) \qquad \to |4f'(t)f(t)|^{1/3}$$
$$\cdot \operatorname*{argmax}_x\{W_1(x) + V_2(x)\},$$

where W_1 is two-sided Brownian motion and V_2 is the least concave majorant of the drifting two-sided Brownian motion $W_2(x) - x^2$, independent of W_1. The right-hand side of (6.1) comes from

$$\mathbb{F}_n^*(t + n^{-1/3}x) - \mathbb{F}_n^*(t) - (\hat{F}_n(t + n^{-1/3}x) - \hat{F}_n(t))$$
$$+ \hat{F}_n(t + n^{-1/3}x) - \hat{F}_n(t) - f(t)n^{-1/3}x,$$

and $\hat{F}_n$ is the *least concave majorant* of the empirical distribution function $\mathbb{F}_n$.

It is suggested in Groeneboom and Jongbloed (2015) however, that one can use the nonparametric bootstrap with a smoothed Grenander estimator. Results of this type for the nonparametric bootstrap are proved in Groeneboom and Hendrickx (2017b) for the current status model.

In the current status model, one can perform a bootstrap experiment in which one only resamples the indicators Δ_i and keeps the observation times T_i fixed. This idea is used by Sen and Xu (2015) who resample the indicators Δ_i as Bernoulli random variables with success probability $\tilde{F}_{nh}(T_i)$, where $\tilde{F}_{nh}$ is for example the SMLE, based on the original sample. This gives a bootstrap sample

$$(T_1, \Delta_1^*), \ldots, (T_n, \Delta_n^*),$$

for which one can compute the MLE $\hat{F}_n^*$. They then use the following result.

THEOREM 6.2 (Sen and Xu, 2015). *Suppose F_n is any estimate of F_0 such that, almost surely,*

$$(6.2) \qquad \lim_{n \to \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)| = 0.$$

Moreover, assume that $f_0(t_0) > 0$, $g(t_0) > 0$, where g is the density of the T_i and, almost surely,

$$(6.3) \qquad \lim_{n \to \infty} n^{1/3}|F_n(t_0 + n^{-1/3}t)$$
$$- F_n(t_0) - f_0(t_0)n^{-1/3}t| = 0.$$

Then the smooth bootstrap from F_n is consistent, that is, $n^{1/3}\{\hat{F}_n^*(t_0) - F_n(t_0)\}$ converges, conditionally on the data, in distribution to the same (nonnormal) limit distribution as $n^{1/3}\{\hat{F}_n(t_0) - F_0(t_0)\}$.

Note that the problem of resampling from a nondifferentiable distribution function, such as the MLE, is removed locally in condition (6.3). Next, one constructs the bootstrap confidence intervals:

$$[\hat{F}_n(t_0) - U_{1-\alpha/2}^*, \hat{F}_n(t_0) - U_{\alpha/2}^*],$$

where U_α^* is the αth percentile of the bootstrap values $\hat{F}_n^*(t_0) - \tilde{F}_{nh}(t_0)$, $\tilde{F}_{nh}$ being (e.g.) the SMLE which satisfies condition (6.2) locally, and where $\hat{F}_n$ is the (discrete) nonparametric maximum likelihood estimator. This gives a *consistent bootstrap*, (remarkably) reproducing the nonnormal limiting distribution of the MLE. Figure 11 shows the result for a sample of size $n = 1000$, where F_0 is the standard truncated exponential distribution function on $[0, 2]$ and the T_i are uniform on $[0, 2]$. Notice that the confidence intervals are not monotone in the time variable.
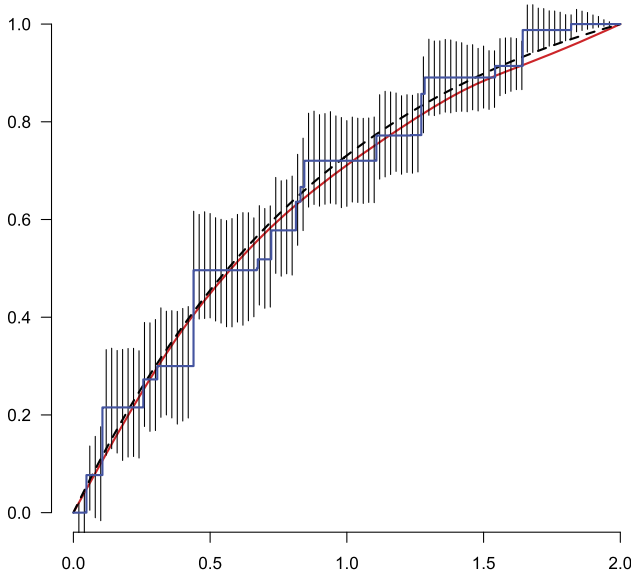
FIG. 11. *Sen–Xu 95% confidence intervals for the (truncated exponential) distribution function F_0, $n = 1000$, observations are (T_i, Δ_i), where T_i is Uniform(0, 2). Construction uses 1000 (smooth) bootstrap samples of indicators Δ_i^*, using the SMLE (red) of the original sample. Dashed curve: truncated exponential distribution function.*

FIG. 12. *95% confidence intervals produced by the R package* curstatCI *for the (truncated exponential) distribution function F_0, $n = 1000$, observations are (T_i, Δ_i), where T_i is Uniform(0, 2). The construction uses 1000 nonparametric bootstrap samples, as described in the text. The locally optimal bandwidth h (in principle different for each evaluation point) is also computed by a bootstrap experiment. Dashed curve: truncated exponential distribution function, solid curve: SMLE. A boundary correction is used in neighborhoods of 0 and 2.*

However, if one assumes the extra smoothness, as in (6.3), it may be more natural to construct confidence intervals around the SMLE instead of the MLE. For the current status model, one can do that in two ways: either by only resampling the Δ_i^* in the way discussed, but now using the bootstrap values

$$(6.4) \quad n^{2/5} \left\{ \tilde{F}_{nh}^*(t_0) - \int \mathbb{K}\big((t_0 - u)/h\big) \, d\tilde{F}_{nh}(u) \right\},$$

where $\mathbb{K}$ and $\tilde{F}_{nh} = \tilde{F}_{nh}^{(\text{SMLE})}$ are defined as in (4.1); here, $\tilde{F}_{nh}^*$ is the SMLE in the bootstrap sample, based on the MLE $\hat{F}_n^*$ in the bootstrap sample, using bandwidth h. Next, one constructs the bootstrap confidence intervals

$$(6.5) \qquad \big[\tilde{F}_{nh}(t_0) - U_{1-\alpha/2}^*, \tilde{F}_{nh}(t_0) - U_{\alpha/2}^* \big],$$

where U_α^* is the αth percentile of the bootstrap values (6.4). This also leads to a consistent bootstrap, as proved in Groeneboom and Hendrickx (2018a). One can also use a Studentized version of (6.4). Note that to attain the $n^{2/5}$ convergence in (6.4) we need to postulate the existence of a second continuous derivative of F_0 at the point of the interval, as in Theorem 4.2 of Groeneboom, Jongbloed and Witte (2010).

Another possibility is to use the nonparametric bootstrap and to resample with replacement the values
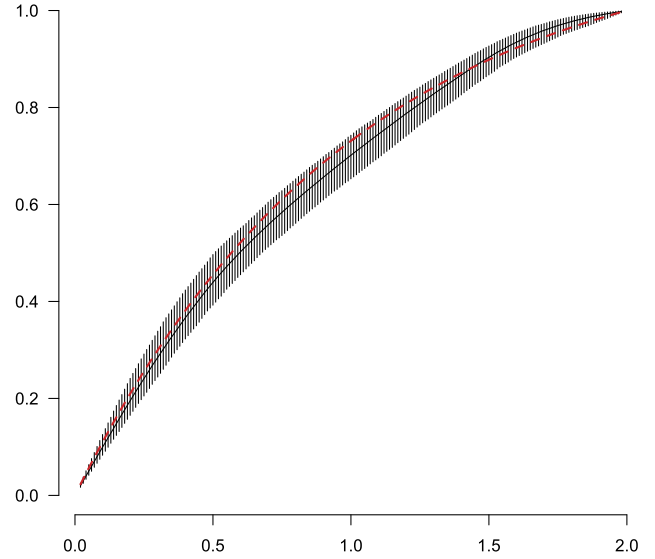
(T_i, Δ_i). In this case, one uses the bootstrap values

$$(6.6) \qquad n^{2/5} \big\{ \tilde{F}_{nh}^*(t_0) - \tilde{F}_{nh}(t_0) \big\},$$

where $\tilde{F}_{nh}$ is the SMLE in the original sample and $\tilde{F}_{nh}^*$ the SMLE on the basis of a bootstrap sample $(T_1^*, \Delta_1^*), \ldots, (T_n^*, \Delta_n^*)$, with bandwidth h of order $n^{-1/5}$, drawn with replacement from the original sample. Next, one computes again (6.5). As in the preceding method, one can use Studentized values instead of (6.6). The latter method is implemented in the R package curstatCI Groeneboom and Hendrickx (2017a). That this also give a consistent bootstrap is proved in Groeneboom and Hendrickx (2017b).

Figure 12 shows the result for a sample of size $n = 1000$, where F_0 is the standard truncated exponential distribution function on $[0, 2]$ and the T_i are uniform on $[0, 2]$.

7. DISCUSSION

In this paper, we review problems and results from shape constrained statistical inference. Using the examples of estimating a decreasing density and interval censoring, estimators were considered, characterized and computed. Also asymptotic distributions were

discussed. In this section, we will put these results in a somewhat broader context and state some interesting (sometimes long time) open questions. Moreover, some problems and approaches not mentioned in this paper will be briefly touched upon.

The computations of the Grenander estimator and the MLE in the current status model are somewhat special. The estimators can be computed via an explicit geometric construction based on necessary and sufficient optimality conditions. In many other problems, like the convex and log concave density estimation problem and interval censoring case 2, such explicit constructions do not exist. Iterative algorithms can be employed to approximate the MLEs. Popular algorithms developed within the field of statistics can be used in many examples. For instance, the Expectation Maximization (EM) algorithm (Dempster, Laird and Rubin, 1977) when there is a missing data interpretation of the model, the Iterative Convex Minorant (ICM) algorithm (Groeneboom and Wellner, 1992 and Jongbloed, 1998) in case one wants to maximize over a cone of monotone functions, and Vertex Direction algorithms (Böhning, 1986) for mixture models can be used in many examples. Sometimes these algorithms can be combined to even yield faster algorithms (Wellner and Zhan, 1997, Jongbloed, 2001). Also more general approaches from optimization theory, such as interior point methods (Wright, 1997) and active set algorithms (Dümbgen and Rufibach, 2011) have been applied.

The idea that estimators using shape constraints are better than those not using the information as seen for the estimation of a concave distribution function in Lemma 3.1 arises in more situations. In particular in the situation of convex density estimation (Dümbgen, Rufibach and Wellner, 2007) and in a discrete convex estimation setting (Balabdaoui and Durot, 2015), at the cost of a (sharp) factor 2 at the right hand side.

The asymptotic distribution theory of the two models focused in this paper has been quite well studied. For the convex and log concave density estimators, much less is known. In Groeneboom, Jongbloed and Wellner (2001a), the asymptotic (pointwise) distribution of convex least squares and maximum likelihood estimators is derived, which is further characterized in Groeneboom, Jongbloed and Wellner (2001b). The limit is characterized as the *second* derivative of the "invelope" of integrated Brownian motion $+t^4$ (the terminology is due to the first author of the present paper; the envelope is in fact a cubic spline lying inside and touching the integrated Brownian motion $+t^4$). So, instead of an "envelope" of Brownian motion with a

parabolic drift, we have the "invelope" of integrated Brownian motion $+t^4$.

The limit distribution of the log concave densities has a similar characterization, see Balabdaoui, Rufibach and Wellner (2009), but the finite sample characterization is somewhat different, since in this case the changes of slope are at the observation points, in contrast with the situation for the convex densities and least squares estimators where the changes of slope are *never* at the observation points (see Groeneboom, Jongbloed and Wellner, 2001a). Recently, global rate of convergence results for log concave density estimation in higher dimensions have also been derived in Kim and Samworth (2016). Much less is known about this limiting distribution, even in dimension 1, than for the Chernoff distribution. No analytic representation of the limit distribution is known, one only has a construction based on cubic splines. It seems that the limit density is not symmetric, in contrast with the Chernoff density. Asymptotic properties and also an R package `logcondiscr` for constructing confidence intervals are discussed in Balabdaoui et al. (2013) and Azadbakhsh, Jankowski and Gao (2014).

Another long standing open problem is that of the asymptotic distribution of the MLE in the interval censoring case 2 model, for the so-called nonseparated case, where the observation intervals $[T_i, U_i]$ can be arbitrarily small. The conjecture for the limit distribution in Groeneboom and Wellner (1992) still stands. For the separated case, the limit distribution was derived in Groeneboom (1996). Simulations for both the separated and nonseparated case are given in Groeneboom and Ketelaars (2011). It is conjectured that the rate of convergence of the MLE is $(n \log n)^{1/3}$ for the nonseparated and proven in Groeneboom (1996) that it is $n^{1/3}$ for the separated case. Birgé (1999) has constructed a histogram-type estimator for which the rate can be proven to be of order $(n \log n)^{1/3}$ in the nonseparated case, a rate that also was expected on the basis of a minimax calculation (see, e.g., Theorem 2.1 in Groeneboom and Ketelaars, 2011).

The current status model has also been studied in the context of the competing risk model; see, for example, Groeneboom, Maathuis and Wellner (2008a) and Groeneboom, Maathuis and Wellner (2008b). This is also a situation where an analytic characterization of the limit distribution is not known.

The methods used for the current status regression model in Groeneboom and Hendrickx (2018b) have also been applied to construct estimators of the regression parameter based on the least squares estimators

in the single index model, converging at $\sqrt{n}$-rate, in Balabdaoui, Groeneboom and Hendrickx (2017). Here again, a "score function" is used instead of an argmin functional. Whether the full profile least squares estimator of the regression parameter is $\sqrt{n}$-consistent is still an open problem. In any case, even if it would be $\sqrt{n}$-consistent, its behavior in simulations is not very good; this bad behavior even inspired Tanaka (2008) to conjecture that the rate is $n^{0.45}$.

The situation is better for the profile least squares estimator in the single index model with a *convex* link function instead of a *monotone* link function. It is proved in Kuchibhotla, Patra and Sen (2017) that this estimator is $\sqrt{n}$-consistent and efficient if the LS estimators are constrained by a uniform Lipschitz condition. It is conjectured, on the basis of simulations, that this even will be true without the uniform Lipschitz condition. Of course, in this case we have differentiability of the criterion function, which is missing in the case of the monotone link function.

As a side effect of the proof of relation (3.6), it was proved in Groeneboom (1983) that Brownian motion can be decomposed into its concave majorant, generated by an inhomogeneous Poisson process, and Brownian excursions on intervals between jumps. This has generated a lot of interesting probabilistic research, beginning with Pitman (1983) (interpretation in terms of Bessel processes and path decomposition results of David Williams). We further mention: Çinlar (1992) (connection with queueing systems), Carolan and Dykstra (2001) [marginal joint densities of Brownian motion and (slope) of concave majorant], Balabdaoui and Pitman (2011) (maximal difference between Brownian bridge and its concave majorant), Pitman and Ross (2012) (greatest convex minorant of Brownian motion, meander, and bridge) and Pitman and Uribe Bravo (2012) (the convex minorant of a Lévy process).

ACKNOWLEDGMENTS

REFERENCES

AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Stat.* **26** 641–647. MR0073895

AZADBAKHSH, M., JANKOWSKI, H. and GAO, X. (2014). Computing confidence intervals for log-concave densities. *Comput. Statist. Data Anal.* **75** 248–264. MR3178372

BALABDAOUI, F. and DUROT, C. (2015). Marshall lemma in discrete convex estimation. *Statist. Probab. Lett.* **99** 143–148. MR3321508

BALABDAOUI, F., GROENEBOOM, P. and HENDRICKX, K. (2017). Score estimation in the monotone single index model. Submitted.

BALABDAOUI, F. and PITMAN, J. (2011). The distribution of the maximal difference between a Brownian bridge and its concave majorant. *Bernoulli* **17** 466–483. MR2798000

BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075

BALABDAOUI, F., JANKOWSKI, H., RUFIBACH, K. and PAVLIDES, M. (2013). Asymptotics of the discrete log-concave maximum likelihood estimator and related applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 769–790. MR3091658

BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. MR1891743

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 Series in Probability and Mathematical Statistics*. Wiley, London. MR0326887

BIRGÉ, L. (1999). Interval censoring: A nonasymptotic point of view. *Math. Methods Statist.* **8** 285–298. MR1735467

BÖHNING, D. (1986). A vertex-exchange-method in *D*-optimal design theory. *Metrika* **33** 337–347. MR0868043

CAROLAN, C. and DYKSTRA, R. (2001). Marginal densities of the least concave majorant of Brownian motion. *Ann. Statist.* **29** 1732–1750. MR1891744

CATOR, E. (2011). Adaptivity and optimality of the monotone least-squares estimator. *Bernoulli* **17** 714–735. MR2787612

CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. MR0172382

ÇINLAR, E. (1992). Sunset over Brownistan. *Stochastic Process. Appl.* **40** 45–53. MR1145458

COSSLETT, S. R. (2007). Efficient estimation of semiparametric models by smoothed maximum likelihood. *Internat. Econom. Rev.* **48** 1245–1272. MR2375624

CULE, M. and SAMWORTH, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.* **4** 254–270. MR2645484

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 545–607. MR2758237

DANIELS, H. E. and SKYRME, T. H. R. (1985). The maximum of a random walk whose mean path has a maximum. *Adv. in Appl. Probab.* **17** 85–99. MR0778595

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

DÜMBGEN, L. and RUFIBACH, K. (2011). logcondens: Computations related to univariate log-concave density estimation. *J. Stat. Softw.* **39** 1–28.

DÜMBGEN, L., RUFIBACH, K. and WELLNER, J. A. (2007). Marshall's lemma for convex density estimation. In *Asymptotics: Particles, Processes and Inverse Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **55** 101–107. IMS, Beachwood, OH. MR2459933

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153 (1957). MR0093415

GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* **11** 1016–1027. MR0714964

GROENEBOOM, P. (1984). Brownian motion with a parabolic drift and Airy functions. CWI Technical Report, Dept. Mathematical Statistics-R 8413, CWI. Available at http://oai.cwi.nl/oai/asset/6435/6435A.pdf.

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). *Wadsworth Statist./Probab. Ser.* 539–555. Wadsworth, Belmont, CA. MR0822052

GROENEBOOM, P. (1987). Asymptotics for incomplete censored observations. Report 87-18, Mathematical Institute, Univ. Amsterdam.

GROENEBOOM, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. Technical Report 378, Dept. Statistics, Stanford Univ. Available at https://statistics.stanford.edu/research/nonparametric-maximum-likelihood-estimators-interval-censoring-and-deconvolution.

GROENEBOOM, P. (1996). Lectures on inverse problems. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1994). *Lecture Notes in Math.* **1648** 67–164. Springer, Berlin. MR1600884

GROENEBOOM, P. (2010). The maximum of Brownian motion minus a parabola. *Electron. J. Probab.* **15** 1930–1937. MR2738343

GROENEBOOM, P. (2011). Vertices of the least concave majorant of Brownian motion with parabolic drift. *Electron. J. Probab.* **16** 2234–2258. MR2861676

GROENEBOOM, P. (2013). Nonparametric (smoothed) likelihood and integral equations. *J. Statist. Plann. Inference* **143** 2039–2065. MR3106623

GROENEBOOM, P. (2015). Rcpp scripts and C++ code. Available at https://github.com/pietg/book.

GROENEBOOM, P. (2018). Chernoff's distribution and differential equations of parabolic and Airy type. Talk at meeting Shape-Constrained Methods: Inference, Applications, and Practice. Banff, 1-28 to 2-2, 2018.

GROENEBOOM, P. and HENDRICKX, K. (2017a) curstatCI. R package, 2017a. Available at https://cran.r-project.org/web/packages/curstatCI/index.html. Version 0.1.1.

GROENEBOOM, P. and HENDRICKX, K. (2017b). The nonparametric bootstrap for the current status model. *Electron. J. Stat.* **11** 3446–3484. MR3709860

GROENEBOOM, P. and HENDRICKX, K. (2018a). Confidence intervals for the current status model. *Scand. J. Stat.* **45** 135–163. MR3764289

GROENEBOOM, P. and HENDRICKX, K. (2018b). Current status linear regression. *Ann. Statist.* **46** 1415–1444. MR3819105

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

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

GROENEBOOM, P. and JONGBLOED, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43** 2019–2054. MR3375875

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001a). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. MR1891742

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001b). A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.* **29** 1620–1652. MR1891741

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

GROENEBOOM, P. and KETELAARS, T. (2011). Estimators for the interval censoring problem. *Electron. J. Stat.* **5** 1797–1845. MR2870151

GROENEBOOM, P., LALLEY, S. and TEMME, N. (2015). Chernoff's distribution and differential equations of parabolic and Airy type. *J. Math. Anal. Appl.* **423** 1804–1824. MR3278229

GROENEBOOM, P., MAATHUIS, M. H. and WELLNER, J. A. (2008a). Current status data with competing risks: Consistency and rates of convergence of the MLE. *Ann. Statist.* **36** 1031–1063. MR2418648

GROENEBOOM, P., MAATHUIS, M. H. and WELLNER, J. A. (2008b). Current status data with competing risks: Limiting distribution of the MLE. *Ann. Statist.* **36** 1064–1089. MR2418649

GROENEBOOM, P. and PYKE, R. (1983). Asymptotic normality of statistics based on the convex minorants of empirical distribution functions. *Ann. Probab.* **11** 328–345. MR0690131

GROENEBOOM, P. and TEMME, N. M. (2011). The tail of the maximum of Brownian motion minus a parabola. *Electron. Commun. Probab.* **16** 458–466. MR2831084

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation. DMV Seminar* **19**. Birkhäuser, Basel. MR1180321

GROENEBOOM, P. and WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388–400. MR1939706

JANSON, S. (2013). Moments of the location of the maximum of Brownian motion with parabolic drift. *Electron. Commun. Probab.* **18** no. 15, 8. MR3037213

JANSON, S., LOUCHARD, G. and MARTIN-LÖF, A. (2010). The maximum of Brownian motion with parabolic drift. *Electron. J. Probab.* **15** 1893–1929. MR2738342

JONGBLOED, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.* **7** 310–321. MR1646718

JONGBLOED, G. (2001). Sieved maximum likelihood estimation in Wicksell's problem and related deconvolution problems. *Scand. J. Stat.* **28** 161–183. MR1844355

KEIDING, N., HANSEN, O. K. H., SØRENSEN, D. N. and SLAMA, R. (2012). The current duration approach to estimating time to pregnancy. *Scand. J. Stat.* **39** 185–204. MR2927018

KIEFER, J. and WOLFOWITZ, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. Verw. Gebiete* **34** 73–85. MR0397974

KIM, A. K. H. and SAMWORTH, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** 2756–2779. MR3576560

KLEIN, R. W. and SPADY, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61** 387–421. MR1209737

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

KUCHIBHOTLA, A. K., PATRA, R. K. and SEN, B. (2017). Score estimation in the convex single index models. Submitted. Available at https://arxiv.org/abs/1708.00145.

LI, G. and ZHANG, C.-H. (1998). Linear regression with interval censored data. *Ann. Statist.* **26** 1306–1327. MR1647661

MARSHALL, A. W. (1969). Discussion on Barlow and van Zwet's paper. In *Nonparametric Techniques in Statistical Inference. Proceedings of the First International Symposium on Nonparametric Techniques Held at Indiana Univ.* 174–176.

MURPHY, S. A., VAN DER VAART, A. W. and WELLNER, J. A. (1999). Current status regression. *Math. Methods Statist.* **8** 407–425. MR1735473

PETO, R. (1973). Experimental survival curves for interval-censored data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **22** 86–91.

PIMENTEL, L. P. R. (2014). On the location of the maximum of a continuous stochastic process. *J. Appl. Probab.* **51** 152–161. MR3189448

PITMAN, J. W. (1983). Remarks on the convex minorant of Brownian motion. In *Seminar on Stochastic Processes*, 1982 (*Evanston, Ill.*, 1982). *Progr. Probab. Statist.* **5** 219–227. Birkhäuser, Boston, MA. MR0733673

PITMAN, J. and ROSS, N. (2012). The greatest convex minorant of Brownian motion, meander, and bridge. *Probab. Theory Related Fields* **153** 771–807. MR2948693

PITMAN, J. and URIBE BRAVO, G. (2012). The convex minorant of a Lévy process. *Ann. Probab.* **40** 1636–1674. MR2978134

PRAKASA RAO, B. L. S. (1969). Estkmation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference. Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR0961262

SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880

SEN, B. and XU, G. (2015). Model based bootstrap methods for interval censored data. *Comput. Statist. Data Anal.* **81** 121–129. MR3257405

SLAMA, R., HANSEN, O. K. H., DUCOT, B., BOHET, A., SORENSEN, D., ALLEMAND, L., EIJKEMANS, M. J., ROSETTA, L., THALABARD, J. C., KEIDING, N. et al. (2012). Estimation of the frequency of involuntary infertility on a nation-wide basis. *Hum. Reprod.* **27** 1489–1498.

TANAKA, H. (2008). Semiparametric least squares estimation of monotone single index models and its application to the iterative least squares estimation of binary choice models.

VAN DER VAART, A. (1991). On differentiable functionals. *Ann. Statist.* **19** 178–204. MR1091845

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York. MR1385671

VAN EEDEN, C. (1956). Maximum likelihood estimation of ordered probabilities. *Nederl. Akad. Wetensch. Proc. Ser. A* **18** 444–455. MR0083859

VARDI, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10** 616–620. MR0653536

VARDI, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76** 751–761. MR1041420

WALTHER, G. (2001). Multiscale maximum likelihood analysis of a semiparametric model, with applications. *Ann. Statist.* **29** 1297–1319. MR1873332

WATSON, G. S. (1971). Estimating functionals of particle size distributions. *Biometrika* **58** 483–490. MR0312639

WELLNER, J. A. and ZHAN, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J. Amer. Statist. Assoc.* **92** 945–959. MR1482125

WRIGHT, S. J. (1997). *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, PA. MR1422257