The Annals of Statistics
1998, Vol. 26, No. 6, 2395–2406

ISOTONIC INVERSE ESTIMATORS FOR NONPARAMETRIC DECONVOLUTION

BY BERT VAN ES, GEURT JONGBLOED AND MARTIEN VAN ZUIJLEN

University of Amsterdam, Vrije Universiteit and University of Nijmegen

A new nonparametric estimation procedure is introduced for the distribution function in a class of deconvolution problems, where the convolution density has one discontinuity. The estimator is shown to be consistent and its cube root asymptotic distribution theory is established. Known results on the minimax risk for the estimation problem indicate the estimator to be efficient.

1. Introduction. An often occurring problem in statistics is that we have observations Z_i which are equal to the sum of independent random variables of interest X_i and random variables Y_i, where the distribution of Y_i can be assumed to be known. For instance, consider a value X_i which is measured with measurement error Y_i. Or, consider X_i to be the time of infection of a disease and Y_i the incubation time. The second example is relevant to so-called back calculation problems in AIDS research. The known distribution of Y_i in these two examples will be quite different. An error measurement is usually modelled by a symmetric distribution on the whole real line while the distribution of a time period will be a skewed distribution on the half line of positive reals.

More formally, we have the following model. Let $X_1, X_2, \dots, X_n$ denote a sample from an unknown distribution with distribution function F and, independent of that sample, $Y_1, Y_2, \dots, Y_n$ a sample from a known distribution with density k on $\mathbb{R}$. Consider the problem of estimating F based on the sample $Z_1, Z_2, \dots, Z_n$, where $Z_i = X_i + Y_i$. The density g of Z_1 is the convolution of k and F in the following sense:

$$g(z) = \int_{\mathbb{R}} k(z-x)\, dF(x) =: k * dF(z).$$

For this reason, this estimation problem is known as a *deconvolution problem*.

For the special case where the *kernel* k is a decreasing density on $[0, \infty)$ and $F(0) = 0$, the *nonparametric maximum likelihood estimator* (NPMLE) for F is studied in Groeneboom and Wellner (1992). There this estimator is shown to be consistent and a conjecture is given concerning its asymptotic distribution. Except in a few special cases such as *uniform deconvolution* [see van Es (1991a, b) and van Es and van Zuijlen (1996)], and *exponential deconvolution* [see Jongbloed (1995, 1998)], there is no explicit expression for the NPMLE available and computing the NPMLE requires an iterative proce-

Received August 1995; revised March 1998.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62E20.
Key words and phrases. Convex minorant, cube root asymptotics, isotonic estimation, empirical process.

dure. These maximum likelihood estimators share the same type of cube root asymptotics as, for instance, the Grenander maximum likelihood estimator of a decreasing density and the NPMLE in certain interval censoring problems; see Groeneboom (1996).

We propose an alternative to the NPMLE. Section 2 introduces a nonparametric estimator $\tilde{F}_n^M$ for F for the more general class of deconvolution problems with the known density k concentrated on $[0,\infty)$. The kernel k is *not* assumed to be decreasing. This *isotonic inverse estimator* can, in contrast to the NPMLE, be calculated as the derivative of the convex minorant of a single function depending on the data via a certain function p which is related to k by an integral equation. In Section 3, we state a sufficient condition on k that implies the properties of this function p needed to establish the asymptotic results in Section 4. We prove that the estimator $\tilde{F}_n^M(x_0)$ of $F(x_0)$ is consistent. Moreover, for a class of kernels vanishing on $(-\infty,0)$, having a discontinuity at zero and being smooth on $(0,\infty)$, we derive its asymptotic distribution of $\tilde{F}_n^M(x_0)$ under the assumption that F is differentiable near x_0 with derivative f:

$$n^{1/3}\left(\frac{k(0)^2}{4f(x_0)g(x_0)}\right)^{1/3}\left(\tilde{F}_n^M(x_0)-F(x_0)\right)\to_{\mathscr{D}} Z.$$

Here Z is the last time that the process $t\mapsto W(t)-t^2$ reaches its maximum and W is a standard two-sided Wiener process originating from zero. This asymptotic distribution coincides with the asymptotic distribution conjectured in Groeneboom and Wellner (1992) for the NPMLE in case of decreasing kernels on $[0,\infty)$. This suggests that the estimator might have good properties from the point of view of efficiency. As will be seen in Section 4, this is different in the uniform deconvolution case, where k has two discontinuities. Efficiency is discussed briefly in Section 5.

The convolution structure of the density of the observations allows inversion by Fourier transform techniques. Kernel estimators based on this approach have been introduced and studied by several authors. Some recent references are Fan (1991) and Hall and Diggle (1993). Kernel estimators based on direct inversion formulas for gamma and Laplace deconvolution problems can be found in van Es and Kok (1997). Compared to the maximum likelihood and isotonic inverse estimators, these approaches have both advantages and disadvantages. An advantage is that if the unknown F is smooth, the rate of convergence of the kernel estimators is faster. On the other hand, the resulting estimators of F are not monotone.

2. An isotonic inverse estimator. In this section we introduce a new nonparametric procedure to estimate F. We restrict attention to the case where F has support contained in $[0,\infty)$.

Suppose that, given a kernel k, we have a function p living on $[0,\infty)$, solving the integral equation

$$p*k(x):=\int_0^x p(x-y)k(y)\,dy=\mathbf{1}*\mathbf{1}(x)=x\mathbf{1}(x), \tag{1}$$

where the function $\mathbf{1}$ is defined by

$$\mathbf{1}(x) = 1_{[0,\infty)}(x).$$

Then we can write, for each $x \geq 0$ and a Z having density function $g = k * dF$,

$$\begin{aligned} Ep(x - Z) &= p * g(x) = p * k * dF(x) \\ &= \mathbf{1} * \mathbf{1} * dF(x) = \int_0^x F(s)\,ds =: H(x). \end{aligned} \tag{2}$$

Let G_n denote the empirical distribution function corresponding to a sample $Z_1, Z_2, \ldots, Z_n$ from the density g. The empirical counterpart of the left-hand side of (2) is given by a sample mean:

$$H_n(x) = \int_{[0,x)} p(x - z)\,dG_n(z) = \frac{1}{n}\sum_{i=1}^{n} p(x - Z_i). \tag{3}$$

This function H_n is an estimator for a primitive of F. Taking the derivative of some smoothed version of H_n (H_n itself will in general not be differentiable) would therefore yield an estimator for F. We call such an estimator an *inverse estimator*, since it is based on the inverse relation

$$F(x) = \frac{d}{dx} p * g(x) \quad \text{a.e.}$$

which follows from (2). However, using general smoothing techniques for estimating g, for example, kernel estimation, the information that H is convex, which follows from the monotonicity of F, is not used. Consequently, inverse estimators for the distribution function will in general not be monotone.

For $M \in (0, \infty]$, denote by $\tilde{H}_n^M$ the largest convex function dominated by H_n on $[0, M)$ (the *convex minorant* of H_n on $[0, M)$). At a fixed point $x \in [0, M)$, we define the estimator $\tilde{F}_n^M$ of F as the right derivative of $\tilde{H}_n^M$ evaluated at x,

$$\tilde{F}_n^M(x) = \lim_{h \downarrow 0} \frac{\tilde{H}_n^M(x + h) - \tilde{H}_n^M(x)}{h}.$$

This estimator $\tilde{F}_n^M$ is by construction monotone (isotonic with respect to natural ordering on $\mathbb{R}$), and therefore called an *isotonic inverse estimator*.

Figure 1 shows a picture of the isotonic inverse estimator based on a realization of a sample of size 100 from the convolution of the kernel

$$k(x) = \tfrac{5}{2}(1 - x)^{3/2} 1_{[0,1]}(x)$$

and the uniform distribution function. To obtain this picture we approximated p numerically and computed the convex minorant of the associated function H_n on a fine grid. See Jongbloed (1995) for more examples and details on computational aspects.

One possible choice for M in the definition of $\tilde{F}_n^M$ is $M = \infty$. As we will see in Section 3, we need finiteness of M in order to prove our asymptotic distribution result for a large class of densities k. If we take $M = \infty$, we need monotonicity of k on $[0, \infty)$ in order to make the asymptotics rigorous. For

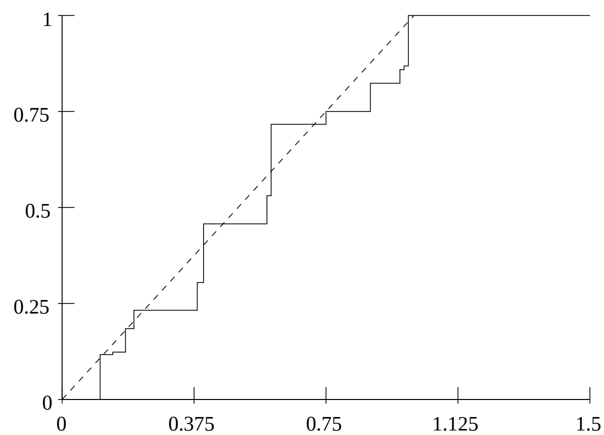

FIG. 1. *Isotonic inverse estimator of the distribution function based on a sample of size* 100; *the dashed curve is the true* (*uniform*) *distribution function.*

practical purposes, there is no difference between $\tilde{F}_n^M$ for finite (but large) and infinite M. See Remark 1.

3. The integral equation. Integral equation (1) is a Volterra equation of the first kind, of convolution type. The function p is sometimes called the *resolvent of the first kind* of k [see Gripenberg, Londen and Staffans (1990), page 158]. To prove consistency of $\tilde{F}_n^M$ in Section 4, we have to impose a condition on p.

CONDITION 1. On bounded intervals, the function p has only finitely many discontinuities. All these discontinuities are finite in size.

For the cube root asymptotics of $\tilde{F}_n^M(x)$ for $x < M$, as stated in Theorem 2, we need a slightly stronger condition on p.

CONDITION 2. The function p is Hölder continuous of order $\alpha > 1/2$ on $(0, \infty)$ and $0 < p(0) < \infty$.

In Section 5 we will see that, under the weaker assumption that p satisfies Condition 1, has more than one discontinuity, is Hölder continuous of order $\alpha > 1/2$ between the successive points of jump and has $0 < p(0) < \infty$, the estimator is still $n^{1/3}$-consistent, but it is not efficient anymore.

Since it is more natural to impose conditions on the kernel k rather than on the function p, we state Lemma 1. It gives a sufficient condition for Condition 2 and thus also for Condition 1 to hold.

LEMMA 1. *Let* $0 < M < \infty$. *Suppose the density* k *can be written as*

$$k(x) = k(0)\left(1 + \int_0^x l(u)\,du\right), \qquad x \in [0, M]$$

for some bounded Borel measurable function l: $[0, M] \to \mathbb{R}$. *Then the unique continuous* (*on* $(0, M)$) *solution* p *of* (1) *allows the representation*

$$p(x) = \frac{1}{k(0)}\left(1 + \int_0^x q(u)\,du\right), \qquad x \in [0, M],$$

where q *is a bounded Borel measurable function on* $[0, M]$.

PROOF. Consider the type-II Volterra convolution equation

$$q(x) + \int_0^x l(x-u)q(u)\,du = -l(x),$$

or, equivalently

$$q + l * q = -l.$$

By Theorem 3.5 in Gripenberg, Londen and Staffans (1990), it follows that the solution q of this equation is unique. It is bounded and Borel measurable on $[0, M]$ whenever l is. Now define $p = k(0)^{-1}(\mathbf{1} + \mathbf{1} * q)$ and observe

$$k * p = k(0)(\mathbf{1} + \mathbf{1} * l) * \frac{\mathbf{1} + \mathbf{1} * q}{k(0)} = \mathbf{1} * \mathbf{1} + \mathbf{1} * \mathbf{1} * (q + l + q * l) = \mathbf{1} * \mathbf{1}. \qquad \square$$

4. Asymptotic results. The first theorem below establishes the almost sure consistency of the estimator under the weak Condition 1. The next theorems give the asymptotic distribution, first for the case that Condition 2 is satisfied and k is allowed only a jump at zero, and second for uniform deconvolution where k has jumps at zero and 1.

THEOREM 1. *Let* p *satisfy Condition* 1. *Then, for all* $0 < M < \infty$ *and* $x_0 \in [0, M)$,

$$F(x_0^-) \le \liminf_{n\to\infty} \tilde{F}_n^M(x_0) \le \limsup_{n\to\infty} \tilde{F}_n^M(x_0) \le F(x_0) \quad a.s. \tag{4}$$

If F *is continuous on* $[0, M)$, *then* $\sup_{0\le x\le M} |\tilde{F}_n^M(x) - F(x)| \to 0$ *almost surely*.

PROOF. Fix $0 < M < \infty$. By Condition 1, p is uniformly continuous on each of the finitely many open intervals between the successive finite jumps of p in $[0, M)$. Therefore, as $n \to \infty$,

$$\sup_{x\in[0,M)} |H_n(x) - H(x)| = \sup_{x\in[0,M)} \left| \int p(x-z)\,d(G_n - G)(z) \right| \to 0 \quad \text{a.s.} \tag{5}$$

where G is the distribution function corresponding to g. Since the operation of taking the right derivative of the convex minorant of a function on $[0, M]$ at a fixed point $x_0 \in (0, M)$ is continuous with respect to the supremum norm [see, e.g., the lemma preceding Theorem 7.2.2. in Robertson, Wright and Dykstra (1988)], the theorem follows. □

REMARK 1. If $M = \infty$, then (4) cannot be derived from (5). A localization argument to ensure that the convex minorant of H_n on $\mathbb{R}$ evaluated at x_0 is determined by H_n on a bounded interval, together with (5) for each finite $M > 0$, would imply consistency of the estimator with $M = \infty$. For this localization, an additional property of p is needed: $\lim_{x\to\infty} x^{-1} p(x) = 1$. Taking the Laplace transform of (1), it follows that $\hat{p}(s) = (s^2 \hat{k}(s))^{-1}$. If k has a finite second moment, permitting a local expansion of $\hat{k}(s) = 1 - m_1 s + m_2 s^2/2$ near zero, $\hat{p}(s) = s^{-2} + m_1 s^{-1} + m_1^2 - m_2/2 + o(1)$ as $s \downarrow 0$. This expansion suggests that $p(x) \sim x$ as $x \to \infty$. However, it only implies $\int_0^x p(y)\,dy \sim \frac{1}{2}x^2$ as $x \to \infty$. When p is *monotone*, which holds if k is monotone, this asymptotic behavior of the integral of p implies $\lim_{x\to\infty} x^{-1} p(x) = 1$. These heuristics can be made rigorous by so-called Karamata theory; see, for instance, the Karamata theorems 1.7.1 (monotone form) and 1.7.6 (extended form) in Bingham, Goldie and Teugels (1987). An alternative proof for monotone k, based on a relation between (1) and the renewal equation with life time distribution $1 - k(x)/k(0)$, can be found in van Es, Jongbloed and van Zuijlen (1995), an earlier version of this paper.

Since the conditions needed to obtain consistency for $\tilde{F}_n^\infty$ are restrictive (k monotone) only to allow for a localization argument, it was decided to incorporate this localization in the definition of the estimator.

THEOREM 2. *Let p satisfy Condition 2, $0 < M < \infty$ and $x_0 \in (0, M)$ be fixed, and F be such that F has a continuous strictly positive derivative f in a neighborhood of x_0. Then, for $n \to \infty$,*

$$n^{1/3}\left(\frac{k(0)^2}{4f(x_0)g(x_0)}\right)^{1/3}\left(\tilde{F}_n^M(x_0) - F(x_0)\right) \to_{\mathscr{D}} Z,$$

where Z is the last time that the process $t \mapsto W(t) - t^2$ reaches its maximum. Here W is a standard two-sided Wiener process originating from zero and $\to_{\mathscr{D}}$ denotes convergence in distribution.

PROOF. Consider, for $a \in (0, 1)$ and $\tau \in [0, M)$, the event $T_n(a) > \tau$, where

$$T_n(a) = \inf\{t \in [0, M]\colon H_n(t) - at \text{ minimal}\}.$$

This event occurs if and only if the maximal affine function with slope a dominated by H_n on $[0, M)$ equals H_n at a point $t_0 \in (\tau, M]$, whereas for each $t \le t_0$, this affine function is *strictly* dominated by H_n. This is equivalent to $\tilde{F}_n^M(\tau) < a$. Therefore, for each $\tau \in [0, M)$ and $a \in (0, 1)$,

$$T_n(a) \le \tau \iff \tilde{F}_n^M(\tau) \ge a.$$

Fix $x_0 \in (0, M)$ meeting the requirements of the theorem. Then, for fixed $\alpha \in \mathbb{R}$, we have, for n sufficiently large,

$$\begin{aligned}
&n^{1/3}\big(\tilde{F}_n^M(x_0) - F(x_0)\big) < \alpha \\
&\quad\Longleftrightarrow\ \tilde{F}_n^M(x_0) < F(x_0) + \alpha n^{-1/3} \\
&\quad\Longleftrightarrow\ T_n(F(x_0) + \alpha n^{-1/3}) > x_0 \\
&\quad\Longleftrightarrow\ \inf\Big\{x_0 + tn^{-1/3} \in [0, M]\colon H_n(x_0 + tn^{-1/3}) - H_n(x_0) \\
&\qquad\qquad - F(x_0)tn^{-1/3} - \alpha t n^{-2/3} \text{ minimal}\Big\} > x_0 \qquad (6)\\
&\quad\Longleftrightarrow\ \inf\Big\{t \in [-x_0 n^{1/3}, (M - x_0)n^{1/3}]\colon \\
&\qquad\qquad n^{2/3}\big(H_n(x_0 + tn^{-1/3}) - H_n(x_0) - F(x_0)tn^{-1/3}\big) \\
&\qquad\qquad - \alpha t \text{ minimal}\Big\} > 0 \\
&\quad\Longleftrightarrow\ \inf\Big\{t \in [-x_0 n^{1/3}, (M - x_0)n^{1/3}]\colon Z_n(t) - \alpha t \text{ minimal}\Big\} > 0,
\end{aligned}$$

where

$$Z_n(t) = n^{2/3}\big(H_n(x_0 + tn^{-1/3}) - H_n(x_0) - F(x_0)tn^{-1/3}\big). \qquad (7)$$

This process Z_n can be decomposed as

$$Z_n(t) = n^{2/3}\big(H(x_0 + tn^{-1/3}) - H(x_0) - F(x_0)tn^{-1/3}\big) + W_n(t) + R_n(t). \qquad (8)$$

Here

$$W_n(t) = n^{2/3} p(0) \int_0^\infty \big(1_{[0,\, x_0 + n^{-1/3}t)}(z) - 1_{[0,\, x_0)}(z)\big)\, d(G_n - G)(z)$$

and, defining the (α-Hölder continuous, $\alpha > 1/2$) function $\tilde{p} = p - p(0)1_{[0,\infty)}$,

$$R_n(t) = n^{2/3} \int_0^\infty \big(\tilde{p}(x_0 + n^{-1/3}t - z) - \tilde{p}(x_0 - z)\big)\, d(G_n - G)(z).$$

We will show in the Appendix that $\sup_{|t|\le K} |R_n(t)| \to 0$ in probability as $n \to \infty$ for $K \in (0, \infty)$. The asymptotics of W_n is well known. This process also plays an important role in the distribution theory of the maximum likelihood estimator of a decreasing density. For example, Example 3.2.14 in van der Vaart and Wellner (1996) immediately gives that $k(0)g(x_0)^{-1/2}W_n$ converges in distribution in $l^\infty([-K, K])$, for each $0 < K < \infty$, to a standard two-sided Brownian motion W. Therefore, also using a Taylor expansion for the "deterministic part" of Z_n, for each $0 < K < \infty$, $Z_n \to Z$ in distribution in $l^\infty([-K, K])$, where

$$Z(t) = \frac{1}{2} f(x_0) t^2 + \frac{\sqrt{g(x_0)}}{k(0)} W(t).$$

Moreover, applying Corollary 3.2.6 in van der Vaart and Wellner (1996) to the class of functions $\{m_x - m_{x_0}: |x - x_0| < \delta\}$ $(\delta > 0)$, where

$$m_x(z) = p(x - z) - xF(x_0),$$

and using Theorem 1, we obtain

$$\inf\{t \in [-x_0 n^{1/3}, (M - x_0)n^{1/3}]: Z_n(t) - \alpha t \text{ minimal}\} = O_p(1).$$

Hence, by Theorem 3.2.2 in van der Vaart and Wellner (1996),

$$\inf\{t \in [-x_0 n^{1/3}, (M - x_0)n^{1/3}]: Z_n(t) - \alpha t \text{ minimal}\} \to_{\mathscr{D}} \operatorname*{argmin}_{t\in\mathbb{R}}\left(\frac{1}{2}f(x_0)t^2 + \frac{\sqrt{g(x_0)}}{k(0)}W(t) - \alpha t\right).$$

Finally, using the property of W that, for each $a > 0$ and $b \in \mathbb{R}$,

$$\operatorname*{argmin}_{t\in\mathbb{R}}(aW(t) + (t - b)^2) =_{\mathscr{D}} a^{2/3} \operatorname*{argmin}_{t\in\mathbb{R}}(W(t) + t^2) + b,$$

we obtain

$$\operatorname*{argmin}_{t\in\mathbb{R}}\left(\frac{1}{2}f(x_0)t^2 + \frac{\sqrt{g(x_0)}}{k(0)}W(t) - \alpha t\right) =_{\mathscr{D}} \frac{2^{2/3}g(x_0)^{1/3}}{f(x_0)^{2/3}k(0)^{2/3}} \operatorname*{argmin}_{t\in\mathbb{R}}(W(t) + t^2) + \frac{\alpha}{f(x_0)},$$

so that

$$\begin{aligned}
&\lim_{n\to\infty} P\big(n^{1/3}(\tilde{F}_n^M(x_0) - F(x_0)) < \alpha\big)\\
&\quad = P\left(\operatorname*{argmin}_{t\in\mathbb{R}}\left(\frac{1}{2}f(x_0)t^2 + \frac{\sqrt{g(x_0)}}{k(0)}W(t) - \alpha t\right) > 0\right)\\
&\quad = P\Big(2\operatorname*{argmin}_{t\in\mathbb{R}}(W(t) + t^2) > -\alpha 2^{1/3} f(x_0)^{-1/3} g(x_0)^{-1/3} k(0)^{2/3}\Big)\\
&\quad = P\Big(2\operatorname*{argmin}_{t\in\mathbb{R}}(W(t) + t^2) < \alpha 2^{1/3} f(x_0)^{-1/3} g(x_0)^{-1/3} k(0)^{2/3}\Big),
\end{aligned}$$

from which the theorem follows. □

THEOREM 3. *Let k be the uniform density on $[0, 1)$, $x_0 \in (0, \infty)$ be fixed and F be such that F has a continuous strictly positive derivative f in a neighborhood of x_0. Then, for $n \to \infty$,*

$$n^{1/3}(4f(x_0)F(x_0))^{-1/3}\big(\tilde{F}_n^{\infty}(x_0) - F(x_0)\big) \to_{\mathscr{D}} Z,$$

where Z is defined as in Theorem 2.

PROOF. Note that the function p associated with k is given by

$$p(x) = (1 + \lfloor x \rfloor) 1_{[0,\infty)}(x),$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x, and that p satisfies Condition 1 contrary to Condition 2. Note also that $F(x) = \sum_{j=0}^{\lfloor x \rfloor} g(x-j)$. Let $x_0 \in [i, i+1)$ for some $i \in \mathbb{N}$. The proof of Theorem 2 can be copied until the decomposition of Z_n in a deterministic part, W_n and R_n as given in (8). In this case, however, $R_n(t) = 0$ for all $|t| \le K$ and n sufficiently large. The process W_n in this situation is given by

$$\begin{aligned} W_n(t) &= \sum_{j=0}^{i} n^{2/3} \int_0^\infty \big(1_{[0,\, x_0+n^{-1/3}t-j)}(z) - 1_{[0,\, x_0-j)}(z)\big)\, d(G_n - G)(z) \\ &= \sum_{j=0}^{i} W_n^{(j)}(t). \end{aligned}$$

Using that $W_n^{(j)} \to_{\mathscr{D}} \sqrt{g(x_0 - j)}W^{(j)}$ for independent standard two-sided Brownian motions $W^{(j)}$, we get that $W_n \to_{\mathscr{D}} \sqrt{F(x_0)}W$ for a standard two-sided Brownian motion W. The result can be obtained along the same lines as in Theorem 2. □

REMARK 2. Theorem 3 can be adapted to cover situations where the kernel k has more than one, but a finite number of jumps. Denoting by $0 = a_1 < a_2 < \cdots < a_m$ the discontinuity points of p and assuming k to satisfy a Hölder condition of order $\alpha > 1/2$ between its discontinuity points, the following asymptotic result can be derived. If $x_0 \in [a_i, a_{i+1})$ for some i, then

$$n^{1/3}\left(4f(x_0)\sum_{j=0}^{i} g(x_0 - a_j)(p(a_j) - p(a_j-))^2\right)^{-1/3} \big(\tilde{F}_n^M(x_0) - F(x_0)\big) \to_{\mathscr{D}} Z.$$

REMARK 3. In deriving our results we have assumed that the distribution F is concentrated on $[0,\infty)$. This can be generalized to the condition that F has a finite left threshold. If the support of F extends to minus infinity, there is a need to control the “right tail” of p. See (3).

5. Discussion. The asymptotic behavior established by Theorem 2 coincides with the asymptotic behavior conjectured in Groeneboom and Wellner (1992) for the NPMLE in case of decreasing kernels on $[0,\infty)$. Apart from a universal constant, the asymptotic variance of $\tilde{F}_n^M(x_0)$ depends on k and F in exactly the same way as the lower bound on the minimax risk for estimating $F(x_0)$ as derived in van Es (1991a).

For certain choices of kernels k, (1) has a simple solution. For instance, if $k(x) = e^{-x}1_{[0,\infty)}(x)$, then $p(x) = (1 + x)1_{[0,\infty)}(x)$. For this *exponential deconvolution problem*, the asymptotic inverse estimator $\tilde{F}_n^\infty$ is compared to the

NPMLE $\hat{F}_n$ in Jongbloed (1998). It turns out that these estimators are first order asymptotically equivalent, in the sense that, for each $x_0 \geq 0$,

$$n^{1/3}\big(\tilde{F}_n^{\infty}(x_0) - \hat{F}_n(x_0)\big) \to_P 0 \quad \text{for } n \to \infty.$$

Also for the uniform deconvolution problem, we saw in the previous section that $p(x) = (1 + \lfloor x \rfloor)1_{[0,\infty)}(x)$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. In this case the asymptotic distribution of $\tilde{F}_n^{\infty}(x_0)$ is given in Theorem 3. However, in van Es and van Zuijlen (1996), an estimator is introduced which has an asymptotic variance which is strictly smaller than the asymptotic variance of $\tilde{F}_n(x_0)$. Under the restriction $F(1) = 1$, this estimator coincides with the NPMLE. See also van Es (1991b).

APPENDIX

We will show that $\sup_{|t|\leq K} |R_n(t)| \to 0$ in probability as $n \to \infty$, where $R_n(t)$ is defined in the proof of Theorem 2.

Observe that

$$\begin{aligned}\sup_{|t|\leq K} |R_n(t)| &= \max_{-k_n+1\leq i\leq k_n} \sup_{t\in[t_{i-1},\, t_i)} |R_n(t)| \\ &\leq \max_{-k_n+1\leq i\leq k_n} \Big(|R_n(t_i)| + \sup_{t\in[t_{i-1},\, t_i)} |R_n(t) - R_n(t_i)|\Big),\end{aligned}$$

where $0 = t_0 < t_1 < \cdots < t_{k_n} = K$ and $t_{-i} := -t_i$, $i = 1, \ldots, k_n$. Using Markov's inequality, we obtain

$$\begin{aligned}\varepsilon P\Big(\sup_{|t|\leq K} |R_n(t)| > \varepsilon\Big) &\leq E \sup_{|t|\leq K} |R_n(t)| \\ &\leq E \max_{-k_n+1\leq i\leq k_n} |R_n(t_i)| \\ &\quad + E\Big(\max_{-k_n+1\leq i\leq k_n} \sup_{t\in[t_{i-1},\, t_i]} |R_n(t) - R_n(t_i)|\Big).\end{aligned} \tag{9}$$

If we now consider the second expectation in (9), we see that, for each $t \in [t_{i-1}, t_i]$,

$$\begin{aligned}&|R_n(t) - R_n(t_i)| \\ &\quad = n^{2/3}\left|\int_0^{\infty} \big(\tilde{p}(x_0 + n^{-1/3}t - z) - \tilde{p}(x_0 + n^{-1/3}t_i - z)\big)\, d(G_n - G)(z)\right| \\ &\quad \leq n^{2/3} \int_0^{\infty} \big|\tilde{p}(x_0 + n^{-1/3}t - z) - \tilde{p}(x_0 + n^{-1/3}t_i - z)\big|\, d(G_n + G)(z) \\ &\quad \leq 2n^{2/3} L n^{-\alpha/3} |t - t_i|^{\alpha} \\ &\quad \leq 2n^{(2-\alpha)/3} L |t_i - t_{i-1}|^{\alpha},\end{aligned}$$

where α and L are the Hölder index and constant of $\tilde{p}$, respectively. If we take the grid of t_i's equally spaced such that $|t_i - t_{i-1}|^\alpha = \delta n^{-(2-\alpha)/3}$, we see that the random variables $\sup_{t\in[t_{i-1}, t_i)} |R_n(t) - R_n(t_i)|$ are bounded uniformly in i by the nonrandom quantity $2L\delta$ that can be made arbitrarily small just by taking δ small. Hence,

$$E\Big(\max_{-k_n+1\le i\le k_n} \sup_{t\in[t_{i-1}, t_i]} |R_n(t) - R_n(t_i)|\Big) \le 2L\delta.$$

Note that $k_n = O(n^{(2-\alpha)/3\alpha})$.

To bound the first expectation on the right-hand side of (9), we can use the two lemmas 2.2.9 (Bernstein's inequality) and 2.2.10 in van der Vaart and Wellner (1996). Denote by $Z_1, Z_2, \dots, Z_n$ a sample from g and write, for fixed t, $R_n(t) = \sum_{i=1}^n Y_i$, where

$$\begin{aligned} Y_i = n^{-1/3}\Big(&\tilde{p}(x_0 + n^{-1/3}t - Z_i) - \tilde{p}(x_0 - Z_i) - H(x_0 + n^{-1/3}t) + H(x_0) \\ &+ p(0)G(x_0 + n^{-1/3}t) - p(0)G(x_0)\Big). \end{aligned}$$

Note that Y_i has expectation zero and bounded range. Indeed,

$$|Y_i| \le n^{-1/3}(Ln^{-\alpha/3}|t|^\alpha + Cn^{-1/3}|t|) = C'n^{-(1+\alpha)/3} + CKn^{-2/3}.$$

Note also that

$$\begin{aligned} \mathrm{Var}(R_n(t)) = n\mathrm{Var}(Y_1) &\le n\cdot n^{-2/3}E(\tilde{p}(x_0 + n^{-1/3}t - Z_1) - \tilde{p}(x_0 - Z_1))^2 \\ &\le L^2K^{2\alpha}n^{(1-2\alpha)/3}. \end{aligned}$$

Using Bernstein's inequality, we obtain the following bound on the tail of $R_n(t)$:

$$\begin{aligned} P(|R_n(t)| > x) &= P(|Y_1 + Y_2 + \cdots + Y_n| > x) \\ &\le 2\exp\bigg\{-\frac{1}{2}\frac{x^2}{C_1n^{(1-2\alpha)/3} + (C_2n^{-(1+\alpha)/3} + C_3n^{-2/3})x}\bigg\}. \end{aligned}$$

Applying Lemma 2.2.10 in van der Vaart and Wellner (1996) to $R_n(t_{-k_n+1})$, $\dots, R_n(t_{k_n})$, and using that $\|\cdot\|_1 \le \|\cdot\|_{\Psi_1}$, we get

$$\begin{aligned} E\max_{-k_n+1\le i\le k_n} |R_n(t_i)| \le C\big(&C_1^{1/2}n^{(1-2\alpha)/6}\sqrt{\log(1+2k_n)} \\ &+ (C_2n^{-(1+\alpha)/3} + C_3n^{-2/3})\log(1+2k_n)\big). \end{aligned}$$

Using $k_n = O(n^{(2-\alpha)/3\alpha})$ and $\alpha > 1/2$, the result follows.

Acknowledgment. The authors thank Piet Groeneboom, Aad van der Vaart and two referees for their advice and comments on an earlier version of this manuscript that only covered the case of decreasing densities k.

REFERENCES

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation.* Cambridge Univ. Press.

FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.

GRIPENBERG, G., LONDEN, S.-O. and STAFFANS, O. (1990). *Volterra Integral and Functional Equations.* Cambridge Univ. Press.

GROENEBOOM, P. (1996). Lectures on inverse problems. *Ecole d'Eté de Probabilités de Saint-Flour XXIV. Lectures on Probability Theory and Statistics* 67–164. Springer, Berlin.

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

HALL, P. and DIGGLE, P. J. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55** 523–531.

JONGBLOED, G. (1995). Three statistical inverse problems. Ph.D. dissertation, Delft Univ. Technology.

JONGBLOED, G. (1998). Exponential deconvolution: two asymptotically equivalent estimators. *Statist. Neerlandica* **52**, 6–17.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, New York.

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

VAN ES, A. J. (1991a). Aspects of nonparametric density estimation. *CWI Tract* **77**. Centre for Mathematics and Computer Science, Amsterdam.

VAN ES, A. J. (1991b). Uniform deconvolution: nonparametric maximum likelihood and inverse estimation. In *Nonparametric Functional Estimation and Related Topics* (G. Roussas, ed.) 191–198. Kluwer, Dordrecht.

VAN ES, B. and KOK, A. (1997). Simple kernel estimators for certain nonparametric deconvolution problems. *Statist. Probab. Lett.* To appear.

VAN ES, A. J. and VAN ZUIJLEN, M. C. A. (1996). Convex minorant estimators of distributions in nonparametric deconvolution problems. *Scand. J. Statist.* **23** 85–104.

VAN ES, B., JONGBLOED, G. and VAN ZUIJLEN, M. (1995). Nonparametric deconvolution for decreasing kernels. Report 95–77, Delft Univ. Technology.

BERT VAN ES
KORTEWEG DE VRIES
INSTITUTE FOR MATHEMATICS
UNIVERSITY OF AMSTERDAM
PLANTAGE MUIDERGRACHT 24
1018 TV AMSTERDAM
THE NETHERLANDS
E-MAIL: vanes@fwi.uva.nl

GEURT JONGBLOED
DEPARTMENT OF MATHEMATICS
VRIJE UNIVERSITEIT
DE BOELELAAN 1081A
1081 HV AMSTERDAM
THE NETHERLANDS
E-MAIL: geurt@cs.vu.nl

MARTIEN VAN ZUIJLEN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NIJMEGEN
TOERNOOIVELD
6525 ED NIJMEGEN
THE NETHERLANDS
E-MAIL: zuijlen@sci.kun.nl