Estimation of a $k-$monotone density, part 1: characterizations, consistency, and minimax lower bounds

Fadoua Balabdaoui[1]   and   Jon A. Wellner[2]

October 18, 2005

Abstract

Shape constrained densities are encountered in many nonparametric estimation problems. The classes of monotone or convex (and monotone) densities can be viewed as special cases of the classes of $k-$monotone densities. A density g is said to be $k-$monotone if $(-1)^l g^{(l)}$ is nonnegative, nonincreasing and convex for $l = 0, \ldots, k - 2$ if $k \geq 2$, and g is simply nonincreasing if $k = 1$. These classes of shaped constrained densities bridge the gap between the classes of monotone (1-monotone) and convex decreasing (2-monotone) densities for which asymptotic results are known, and the class of completely monotone ($\infty-$monotone) densities.

In this paper we consider both (nonparametric) Maximum Likelihood estimators and Least Squares estimators of a $k-$monotone density. We prove existence of the estimators and give characterizations. We also establish consistency properties, and show that the estimators are splines of order k (degree $k - 1$) with simple knots. We further provide asymptotic minimax risk lower bounds for estimating a $k-$monotone density $g_0(x_0)$ and its derivatives $g_0^{(j)}(x_0)$, $j = 1, \ldots, k - 1$, at a fixed point x_0 under the assumption that $(-1)^k g_0^{(k)}(x_0) > 0$.

1

1. Introduction

Shape constrained densities are encountered in many nonparametric estimation problems. Monotone densities arise naturally via connections with renewal theory and uniform mixing (see Vardi (1989) and Woodroofe and Sun (1993) for examples of the former, and Woodroofe and Sun (1993) for the latter in an astronomical context). Convex densities arise in connection with Poisson process models for bird migration and scale mixtures of triangular densities; see e.g. Hampel (1987), and Anevski (2003).

Estimation of monotone densities on the positive half-line $\mathbb{R}^+ = [0, \infty)$ was initiated by Grenander (1956) (with related work by Ayer, Brunk, Ewing, Reid, and Silverman (1955), Brunk (1958), and Van Eeden (1956), Van Eeden (1957)). Asymptotic theory of the maximum likelihood estimators was developed by Prakasa Rao (1969) with later contributions by Groeneboom (1985), Groeneboom (1989), and Kim and Pollard (1990).

Estimation of convex densities on $\mathbb{R}^+$ was apparently initiated by Anevski (1994) (see also Anevski (2003)), and was pursued by Jongbloed (1995). The limit distribution theory for the (nonparametric) maximum likelihood estimator and its first derivative at a fixed point was obtained by Groeneboom, Jongbloed, and Wellner (2001b).

Our goal is to develop nonparametric estimators and asymptotic theory for the classes of *k-monotone densities* on $[0, \infty)$ defined as follows: g is a $k-$monotone density on $(0, \infty)$ if g is nonnegative and $(-1)^l g^{(l)}$ is nonincreasing and convex for $l \in \{0, \ldots, k-2\}$ for $k \geq 2$, and simply nonnegative and nonincreasing when $k = 1$. As will be shown in section 2, it follows from the results of Williamson (1956), Lévy (1962), and Gneiting (1999) that g is a $k-$monotone density if and only if it can be represented as a scale mixture of $\mathrm{Beta}(1, k)$ densities. For $k = 1$ this recovers the well known facts that monotone densities are scale mixtures of uniform densities, and, for $k = 2$, that convex decreasing densities scale mixtures of the triangular, or $\mathrm{Beta}(1, 2)$, densities.

Our motivation for studying nonparametric estimation in the classes $\mathcal{D}_k$ has several components: besides the obvious goal of generalizing the existing theory for the $1-$monotone (i.e. monotone) and $2-$monotone (i.e. convex and decreasing) classes $\mathcal{D}_1$ and $\mathcal{D}_2$, these classes play an important role in several extensions of Hampel's bird migration problem which are discussed further in Balabdaoui and Wellner (2005a). They also provide a potential link to the important limiting case

of the $k-$monotone classes, namely the class $\mathcal{D}_\infty$ of completely monotone densities. Densities g in $\mathcal{D}_\infty$ have the property that $(-1)^l g^{(l)}(x) \geq 0$ for all $x \in (0,\infty)$ and $l \in \{0,1,\ldots\}$. It follows from Bernstein's theorem (see e.g. Feller (1971), page 439, or Gneiting (1998)) that $g \in \mathcal{D}_\infty$ if and only if it can be represented as a scale mixture of exponential densities. Completely monotone densities arise naturally in connection with mixtures of Poisson processes and have been used in reliability theory and empirical Bayes estimation. Jewell (1982) initiated the study of maximum likelihood estimation in the family $\mathcal{D}_\infty$ and succeeded in showing that the MLE $\hat{F}_n$ of the mixing distribution function is almost surely weakly consistent. Although consistency of the MLE follows now rather easily from the results of Pfanzagl (1988) and van de Geer (1993), little is known about rates of convergence or asymptotic distribution theory for either the estimator of the mixed density (the "forward" or "direct" problem) or the estimator of the mixing distribution function (the "inverse" problem). We hope that our development of methods and theory for general $k-$monotone densites may throw some light on the issues and problems.

In this paper we consider the Maximum Likelihood $\hat{g}_{n,k}$ and Least Squares $\tilde{g}_{n,k}$ estimators of a density $g_0 \in \mathcal{D}_k$ for a fixed integer $k \geq 2$ based on a sample $X_1,\ldots,X_n$ i.i.d. with density g_0. We show that the estimators exist, provide characterizations, and establish consistency of the estimators and their derivatives $\hat{g}_{n,k}^{(j)}$ and $\tilde{g}_{n,k}^{(j)}$ for $j \in \{1,\ldots,k-1\}$ (uniformly on closed sets bounded away from 0). In section 4 we establish asymptotic minimax lower bounds for estimation of $g_0^{(j)}(x_0)$, $j = 0,\ldots,k-1$ under the assumption that $g_0^{(k)}(x_0)$ exists and is non-zero. We also include statements of known results for estimation of a completely monotone density $g_0 \in \mathcal{D}_\infty$ whenever possible.

2. Existence and characterizations

The following lemma characterizing integrable $k-$monotone functions and giving an inversion formula follows from the results of Williamson (1956):

LEMMA 2.1. *(Integrable $k-$monotone characterization) A function g is an integrable k-monotone function if and only if it is of the form*

$$g(x) = \int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k} dF(t), \quad x > 0 \tag{2.1}$$

where F is nondecreasing and bounded on $(0,\infty)$. Thus g is a $k-$monotone density if and only if it is of the form (2.1) for some distribution function F on $(0,\infty)$. If

F in (2.1) satisfies $\lim_{t\to\infty} F(t) = \int_0^\infty g(x)dx$, then at a continuity point $t > 0$, F is given by

$$F(t) = G(t) - tg(t) + \cdots + \frac{(-1)^{k-1}}{(k-1)!}t^{k-1}g^{(k-2)}(t) + \frac{(-1)^k}{k!}t^k g^{(k-1)}(t), \quad (2.2)$$

where $G(t) = \int_0^t g(x)dx$.

Proof. The representation (2.1) follows from Theorem 5 of Lévy (1962) by taking $k = n+1$ and $f \equiv 0$ on $(-\infty, 0]$. The inversion formula (2.2) follows from Lemma 1 in Williamson (1956) together with an integration by parts argument. See Balabdaoui and Wellner (2005c) for details. ∎

For completeness and for comparison, we also give the corresponding characterization and inversion formula in the completely monotone case:

LEMMA 2.2. *(Integrable completely monotone characterization) A function g is an integrable completely monotone function if and only if it is of the form*

$$g(x) = \int_0^\infty \frac{1}{t}\exp(-x/t)dF(t), \quad x > 0 \qquad (2.3)$$

where F is nondecreasing and bounded on $(0,\infty)$. Thus g is a completely monotone density if and only if it is of the form (2.3) for some distribution function F on $(0,\infty)$. If F in (2.3) satisfies $\lim_{t\to\infty} F(t) = \int_0^\infty g(x)dx$, then at a continuity point $t > 0$, F is given by

$$F(t) = \lim_{k\to\infty} \sum_{j=0}^k \frac{(-1)^j}{j!}(kt)^j G^{(j)}(kt) \qquad (2.4)$$

where $G(t) = \int_0^t g(x)dx$.

Proofs. Lemma 2.2 follows from the classical result of Bernstein; see Widder (1946), pages 141-163; Feller (1971), page 439; and Gneiting (1998). The inversion formula (2.4) follows from the development in Feller (1971), pages 232-233. For further details, see Balabdaoui and Wellner (2005a). ∎

It is easy to see that if g is a density, and F is chosen to be right-continuous and to satisfy the condition of the second part of Lemma 2.1, then F is a distribution function. For $k = 1$ ($k = 2$), note that the characterization matches with the well known fact that a density is nondecreasing (nondecreasing and convex) on $(0,\infty)$ if and only if it is a mixture of uniform densities (triangular densities). More generally,

4

the characterization establishes a one-to-one correspondance between the class of k-monotone densities and the class of scale mixture of Beta's with parameters 1 and k. From the inversion formula in (2.2), one can see that a natural estimator for the mixing distribution F is obtained by plugging in an estimator for the density g and it becomes clear that the rate of convergence of estimators of F will be controlled by the corresponding rate of convergence for estimators of the highest derivative $g^{(k-1)}$ of g. When k increases the densities become smoother, and therefore the inverse problem of estimating the mixing distribution F becomes harder.

We now consider the nonparametric Maximum Likelihood and Least Squares Estimators of a k-monotone density g_0. We show that these estimators exist and give characterizations thereof. In the following, $\mathcal{M}_k$ is the class of all k-monotone functions on $(0, \infty)$, $\mathcal{D}_k$ is the sub-class of k-monotone densities on $(0, \infty)$, $X_1, \cdots, X_n$ are i.i.d. from g_0, and $\mathbb{G}_n$ is their empirical distribution function, $\mathbb{G}_n(x) = n^{-1} \sum_1^n 1\{X_i \leq x\}$ for $x \geq 0$.

Let

$$l_n(g) = \int_0^\infty \log g(x) \, d\mathbb{G}_n(x)$$

be the log-likelihood function (really n^{-1} times the log-likelihood function). We want to maximize $l_n(g)$ over $g \in \mathcal{D}_k$. To do this, we change the optimization problem to one over the whole cone $\mathcal{M}_k \cap L_1(\lambda)$. This can be done by introducing the "adjusted likelihood function" $\psi_n(g)$ defined as follows:

$$\psi_n(g) = \int_0^\infty \log g(x) \, d\mathbb{G}_n(x) - \int_0^\infty g(x) dx,$$

for $g \in \mathcal{M}_k \cap L_1(\lambda)$. Then, as in GJW (2001a), Lemma 2.3, page 1661, the maximum likelihood estimator $\hat{g}_n$ also maximizes $\psi_n(g)$ over $\mathcal{M}_k \cap L_1(\lambda)$.

Using the integral representations established in the previous subsection, ψ_n can also be rewritten as

$$\psi_n(F) = \begin{cases} \int_0^\infty \log \left(\int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k} dF(t) \right) d\mathbb{G}_n(x) - \int_0^\infty \int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k} dF(t) dx, \\ \int_0^\infty \log \left(\int_0^\infty \frac{1}{t} \exp(-x/t) dF(t) \right) d\mathbb{G}_n(x) \\ \qquad - \int_0^\infty \int_0^\infty \frac{1}{t} \exp(-x/t) dF(t) dx, \end{cases}$$

where F is bounded and nondecreasing.

LEMMA 2.3. *The maximum likelihood estimator $\hat{g}_{n,k}$ in the classes $\mathcal{D}_k$, $k \in \{1, 2, \ldots, \infty\}$ exists. Furthermore, $\hat{g}_{n,k}$ is the maximizer of ψ_n over $\mathcal{M}_k \cap L_1(\lambda)$.*

5

Moreover, for $k \in \{1, 2, \ldots\}$ the density $\hat{g}_{n,k}$ is of the form

$$\hat{g}_{n,k}(x) = \hat{w}_1 \frac{k(\hat{a}_1 - x)_+^{k-1}}{\hat{a}_1^k} + \cdots + \hat{w}_m \frac{k(\hat{a}_m - x)_+^{k-1}}{\hat{a}_m^k},$$

for some $m = \hat{m}_k$, while for $k = \infty$, $\hat{g}_{n,\infty}$ is of the form

$$\hat{g}_{n,\infty}(x) = \frac{\hat{w}_1}{\hat{a}_1} \exp(-x/\hat{a}_1) + \cdots + \frac{\hat{w}_m}{\hat{a}_m} \exp(-x/\hat{a}_m)$$

for some $m = \hat{m}_\infty$ where $\hat{w}_1, \cdots, \hat{w}_m$ and $\hat{a}_1, \cdots, \hat{a}_m$ are respectively the weights and the support points of the maximizing mixing distribution $\hat{F}_{n,k}$.

Proof. From Lindsay (1983a) we conclude that there exists a unique maximizer of l_n and the maximum is achieved by a discrete mixing distribution function that has at most n support points.

By arguing as in Groeneboom, Jongbloed, and Wellner (2001b) page 1662 it follows that ψ_n is maximized over $\mathcal{M}_k \cap L_1(\lambda)$ by $\hat{g}_n \in \mathcal{D}_k$. In the case $k = \infty$, the assertions of the lemma are proved by Jewell (1982). ■

The following lemma gives a necessary and sufficient condition for a point t to be in the support of the maximizing distribution function $\hat{F}_{n,k}$. For $k \in \{3, \ldots\}$ it generalizes lemma 2.4, page 1662, Groeneboom, Jongbloed, and Wellner (2001b).

LEMMA 2.4. *Let $X_1, \cdots, X_n$ be i.i.d. random variables from the true density g_0, and let $\hat{F}_{n,k}$ and $\hat{g}_{n,k}$ be the MLE of the mixing and mixed distribution respectively. Then, for $k \in \{1, 2, \ldots\}$,*

$$\hat{H}_{n,k}(t) \equiv \mathbb{G}_n \left(\frac{k(t - X)_+^{k-1}/t^k}{\hat{g}_{n,k}(X)} \right) \leq 1, \tag{2.5}$$

with equality if and only if $t \in supp(\hat{F}_{n,k}) = \{\hat{a}_1, \cdots, \hat{a}_m\}$. In the case $k = \infty$

$$\hat{H}_{n,\infty}(t) \equiv \mathbb{G}_n \left(\frac{\exp(-X/t)}{t\hat{g}_{n,\infty}(X)} \right) \leq 1, \qquad \text{for all } \ t > 0 \tag{2.6}$$

with equality if and only if $t \in supp(\hat{F}_{n,\infty}) = \{\hat{a}_1, \cdots, \hat{a}_m\}$.

REMARK 2.1. *By factoring out t^{k-1} and replacing t by kv (say), it becomes clear that the function $\hat{H}_{n,\infty}$ on the right side of (2.6) is a natural limiting version as $k \to \infty$ of the functions $\hat{H}_{n,k}$ on the right side of (2.5).*

Proof. This follows the proof of proposition 2.3 of Groeneboom and Wellner (1992), page 58. The proof of the result for $k = \infty$ is given by Jewell (1982), page 481. ∎

Now we consider the Least Squares estimators. The least squares criterion is

$$Q_n(g) = \frac{1}{2} \int_0^\infty g^2(x)dx - \int_0^\infty g(x)d\mathbb{G}_n(x). \qquad (2.7)$$

We want to minimize this over $g \in \mathcal{D}_k \cap L_2(\lambda)$, the subset of square integrable $k-$monotone functions. Although existence of a minimizer of Q_n over $\mathcal{D}_k \cap L_2(\lambda)$ is quite easily established, the minimizer has a somewhat complicated characterization due to the density constraint $\int_0^\infty g(x)dx = 1$. Therefore we will actually consider the alternative optimization problem of minimizing $Q_n(g)$ over $\mathcal{M}_k \cap L_2(\lambda)$. In this optimization problem existence requires more work, but the resulting characterization of the estimator is considerably simpler. Further we will show that even though the resulting estimator does not necessarily have total mass one, it does have total mass converging almost surely to one and it consistently estimates $g_0 \in \mathcal{D}_k$.

Using arguments similar to those in the proof of Theorem 1 in Williamson (1956), one can show that $g \in \mathcal{M}_k$ if and only if

$$g(x) = \int_0^\infty (t - x)_+^{k-1}d\mu(t)$$

for a positive measure μ on $(0, \infty)$. Thus we can rewrite the criterion in terms of the corresponding measures μ: by Fubini's theorem

$$\int_0^\infty g^2(x)dx = \int_0^\infty \int_0^\infty r_k(t, t')d\mu(t)d\mu(t')$$

where $r_k(t, t') = \int_0^{t \wedge t'} (t - x)^{k-1}(t' - x)^{k-1}dx$, and

$$\int_0^\infty g(x)d\mathbb{G}_n(x) = \int_0^\infty \int_0^\infty (t - x)_+^{k-1}d\mu(t)d\mathbb{G}_n(x) = \int_0^\infty s_{n,k}(t)d\mu(t)$$

where $s_{n,k}(t) \equiv \mathbb{G}_n((t - X)_+^{k-1})$. Hence it follows that, with $g = g_\mu$

$$Q_n(g) = \frac{1}{2} \int_0^\infty \int_0^\infty r_k(t, t')d\mu(t)d\mu(t') - \int_0^\infty s_{n,k}(t)d\mu(t) \equiv \Phi_n(\mu)$$

Now we want to minimize Φ_n over the set $\mathcal{X}$ of all non-negative measures μ on R^+. Since Φ_n is convex and can be restricted to a subset $\mathcal{C}$ of $\mathcal{X}$ on which it is lower semicontinuous, a solution exists and is unique.

7

PROPOSITION 2.1. *The problem of minimizing $\Phi_n(\mu)$ over all non-negative measures μ has a unique solution $\tilde{\mu}$.*

Proof. Existence follows from Zeidler (1985), Theorem 38.B, page 152. See Balabdaoui and Wellner (2005c) for the verification of the hypotheses of required hypotheses. Uniqueness follows from the strict convexity of Φ_n. ∎

The following proposition characterizes the least squares estimators.

PROPOSITION 2.2. *For $k \in \{1, 2, \ldots\}$ define $\mathbb{Y}_{n,k}$ and $\tilde{H}_{n,k}$ respectively by*

$$\mathbb{Y}_{n,k}(t) = \int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_2} \mathbb{G}_n(t_1)dt_1 dt_2 \cdots dt_{k-1} = \int_0^t \frac{(t-x)^{k-1}}{(k-1)!}d\mathbb{G}_n(x)$$

and

$$\tilde{H}_{n,k}(t) = \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} \tilde{g}_n(t_1)dt_1 dt_2 \cdots dt_k = \int_0^t \frac{(t-x)^{k-1}}{(k-1)!}\tilde{g}_n(x)dx$$

for $t \geq 0$. Then $\tilde{g}_{n,k}$ is the LS estimator over $\mathcal{M}_k \cap L_2(\lambda)$ if and only if the following conditions are satisfied for $\tilde{g}_{n,k}$ and $\tilde{H}_{n,k}$:

$$\begin{cases} \tilde{H}_{n,k}(t) \geq \mathbb{Y}_{n,k}(t), & \text{for } t \geq 0, \quad \text{and} \\ \int_0^\infty (\tilde{H}_{n,k} - \mathbb{Y}_{n,k})d\tilde{g}_{n,k}^{(k-1)} = 0. \end{cases} \tag{2.8}$$

Proof. This follows along the lines of the proof of lemma 2.2 of Groeneboom, Jongbloed, and Wellner (2001b) using integration by parts in the sufficiency part of the proof, and by taking the perturbations $\tilde{g}_{n,k} + \epsilon g_t$ and $\tilde{g}_{n,k}(1+\epsilon)$ for ϵ sufficiently small where $g_t(x) \equiv (t-x)_+^{k-1}/(k-1)!$ for $x \geq 0$ and $t > 0$ in the necessity half. ∎

In order to prove that the LSE is a spline of degree $k-1$, we need the following result.

LEMMA 2.5. *Let $[a, b] \subseteq (0, \infty)$ and let g be a nonnegative and nonincreasing function on $[a, b]$. For any polynomial P_{k-1} of degree $\leq k-1$ on $[a, b]$, if the function*

$$\Delta(t) = \int_0^t (t-s)^{k-1}g(s)ds - P_{k-1}(s), \quad t \in [a, b]$$

admits infinitely many zeros in $[a, b]$, then there exists $t_0 \in [a, b]$ such that $g \equiv 0$ on $[t_0, b]$ and $g > 0$ on $[a, t_0)$ if $t_0 > a$.

Proof. By applying the mean value theorem k times, it follows that $(k-1)!g = \Delta^{(k)}$ admits infinitely many zeros in $[a,b]$. But since g is assumed to be nonnegative and nonincreasing, this implies that if t_0 is the smallest zero of g in $[a,b]$, then $g \equiv 0$ on $[t_0, b]$. By definition of t_0, $g > 0$ on $[a, t_0)$ if $t_0 > a$. ∎

Now we will use the characterization of the LSE $\tilde{g}_n$ together with the previous lemma to show that it is a finite mixture of $Beta(1,k)$'s. We know from Proposition 2.2 that $\tilde{g}_{n,k}$ is the LSE if and only if (2.8) holds. The equality condition in the second part of (2.8) implies that $\tilde{H}_{n,k}$ and $\mathbb{Y}_{n,k}$ have to be equal at any point of increase of the monotone function $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$. Therefore, the set of points of increase of $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ is included in the set of zeros of the function $\tilde{\Delta}_{n,k} = \tilde{H}_{n,k} - \mathbb{Y}_{n,k}$. Now, note that $\mathbb{Y}_{n,k}$ can be given by the explicit expression:

$$\mathbb{Y}_{n,k}(t) = \frac{1}{(k-1)!} \frac{1}{n} \sum_{j=1}^{n} (t - X_{(j)})_+^{k-1}, \quad \text{for } t > 0.$$

In other words, $\mathbb{Y}_{n,k}$ is a spline of degree $k-1$ with simple knots $X_{(1)}, \cdots, X_{(n)}$ (for a definition of the multiplicity of knots, see e.g. de Boor (1978), page 96, or DeVore and Lorentz (1993), page 140). Also note that the function $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ cannot have a positive density with respect to Lebesgue measure λ. Indeed, if we assume otherwise, then we can find $0 \leq j \leq n$ and an interval $I \subset (X_{(j)}, X_{(j+1)})$ (with $X_{(0)} = 0$ and $X_{(n+1)} = \infty$) such that I has a nonempty interior, and $\tilde{H}_{n,k} \equiv \mathbb{Y}_{n,k}$ on I. This implies that $\tilde{H}_{n,k}^{(k)} \equiv \mathbb{Y}_{n,k}^{(k)} \equiv 0$, since $\mathbb{Y}_{n,k}$ is a polynomial of degree $k-1$ on I, and hence $\tilde{g}_{n,k} \equiv 0$ on I. But the latter is impossible since it was assumed that $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ was strictly increasing on I. Thus the monotone function $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ can have only two components: discrete and singular. In the following theorem, we will prove that it is actually discrete with finitely many points of jump.

PROPOSITION 2.3. *There exists $m \in \mathbb{N}\backslash\{0\}$, $\tilde{a}_1, \cdots, \tilde{a}_m$ and $\tilde{w}_1, \cdots, \tilde{w}_m$ such that for all $x > 0$, the LSE $\tilde{g}_{n,k}$ is given by*

$$\tilde{g}_{n,k}(x) = \tilde{w}_1 \frac{k(\tilde{a}_1 - x)_+^{k-1}}{\tilde{a}_1^k} + \cdots + \tilde{w}_m \frac{k(\tilde{a}_m - x)_+^{k-1}}{\tilde{a}_m^k}. \tag{2.9}$$

Consequently, the equality part of proposition 2.2 can be re-expressed as $\tilde{H}_{n,k}(t) = \mathbb{Y}_{n,k}(t)$ if $t \in supp\{\tilde{F}_{n,k}\}$.

Proof. We need to consider two cases:

(i) The number of zeros of $\tilde{\Delta}_{n,k} = \tilde{H}_{n,k} - \mathbb{Y}_{n,k}$ is finite. This implies by (??) that the number of points of increase of $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ is also finite. Therefore, $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ is discrete with finitely many jumps and hence $\tilde{g}_{n,k}$ is of the form given in (2.9).

(ii) Now, suppose that $\tilde{\Delta}_{n,k}$ has infinitely many zeros. Let j be the smallest integer in $\{0, \cdots, n-1\}$ such that $[X_{(j)}, X_{(j+1)}]$ contains infinitely many zeros of $\tilde{\Delta}_{n,k}$ (with $X_{(0)} = 0$ and $X_{(n+1)} = \infty$). By Lemma 2.5, if t_j is the smallest zero of $\tilde{g}_n$ in $[X_{(j)}, X_{(j+1)}]$, then $\tilde{g}_{n,k} \equiv 0$ on $[t_j, X_{(j+1)}]$ and $\tilde{g}_{n,k} > 0$ on $[X_{(j)}, t_j)$ if $t_j > X_{(j)}$. Note that from the proof of Proposition 2.1, we know that the minimizing measure $\tilde{\mu}_n$ does not put any mass on $(0, X_{(1)}]$, and hence the integer j has to be strictly greater than 0.

Now, by definition of j, $\tilde{\Delta}_n$ has finitely many zeros to the left of $X_{(j)}$, which implies that $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ has finitely many points of increase in $(0, X_{(j)})$. We also know that $\tilde{g}_{n,k} \equiv 0$ on $[t_j, \infty)$. Thus we only need to show that the number of points of increase of $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ in $[X_{(j)}, t_j)$ is finite, when $t_j > X_{(j)}$. This can be argued as follows: Consider z_j to be the smallest zero of $\tilde{\Delta}_n$ in $[X_{(j)}, X_{(j+1)})$. If $z_j \geq t_j$, then we cannot possibly have any point of increase of $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ in $[X_{(j)}, t_j)$ because it would imply that we have a zero of $\tilde{\Delta}_{n,k}$ that is strictly smaller than z_j. If $z_j < t_j$, then for the same reason, $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ has no point of increase in $[X_{(j)}, z_j)$. Finally, $(-1)^{k-1}\tilde{g}_{n,k}^{(k-1)}$ cannot have infinitely many points of increase in $[z_j, t_j)$ because that would imply that $\tilde{\Delta}_{n,k}$ has infinitely zeros in (z_j, t_j), and hence by Lemma 2.5, we can find $t'_j \in (z_j, t_j)$ such that $\tilde{g}_n \equiv 0$ on $[t'_j, t_j]$. But this impossible since $\tilde{g}_{n,k} > 0$ on $[X_{(j)}, t_j)$. ∎

3. Consistency

In this section, we will prove that both the MLE and LSE are strongly consistent. Furthermore, we will show that this consistency is uniform on intervals of the form $[c, \infty)$, where $c > 0$.

Consistency of the maximum likelihood estimators for the classes $\mathcal{D}_k$ in the sense of Hellinger convergence of the mixed density is a relatively simple straightforward consequence of the methods of Pfanzagl (1988) and van de Geer (1993). As usual, the Hellinger distance H is given by $H^2(p, q) = (1/2) \int \{\sqrt{p} - \sqrt{q}\}^2 d\mu$ for any common dominating measure μ.

PROPOSITION 3.1. *Suppose that $\hat{g}_{n,k}$ is the MLE of g_0 in the class $\mathcal{D}_k$, $k \in$*

$\{1, \ldots, \infty\}$. *Then*

$$H(\hat{g}_{n,k}, g_0) \to_{a.s.} 0 \qquad as \quad n \to \infty.$$

Furthermore $\hat{F}_{n,k} \to_d F_0$ *almost surely where* $\hat{F}_{n,k}$ *is the MLE of the mixing distribution function* F_0.

Proof. This follows from the methods of Pfanzagl (1988) and van de Geer (1993) by using the Glivenko-Cantelli preservation theorems of van der Vaart and Wellner (2000). ∎

The following lemma establishes a useful bound for k-monotone densities.

LEMMA 3.1. *If g is a k-monotone density function for $k \geq 2$, then*

$$g(x) \leq \frac{1}{x}\left(1 - \frac{1}{k}\right)^{k-1}$$

for all $x > 0$.

Proof. We have

$$
\begin{aligned}
g(x) &= \int_x^\infty \frac{k}{y^k}(y-x)^{k-1}dF(y) = \frac{1}{x}\int_x^\infty \frac{kx}{y}(1 - \frac{x}{y})^{k-1}dF(y) \\
&\leq \frac{1}{x}\sup_{x \leq y < \infty} \frac{kx}{y}\left(1 - \frac{x}{y}\right)^{k-1} = \frac{k}{x}\sup_{0 < u \leq 1} u(1-u)^{k-1} = \frac{1}{x}\left(1 - \frac{1}{k}\right)^{k-1}
\end{aligned}
$$

by an easy calculation. (Note that when $k = 2$, this bound equals $1/(2x)$ which agrees with the bound given by Jongbloed (1995), page 117, and Groeneboom, Jongbloed, and Wellner (2001b), page 1669 in this case.) ∎

PROPOSITION 3.2. *Let g_0 be a k-monotone density on $(0, \infty)$ and fix $c > 0$. Then for $j = 0, 1, \cdots, k - 2$ (with $\hat{g}_{n,k}^{(0)} = \hat{g}_{n,k}$, $g_0^{(0)} = g_0$)*

$$\sup_{x \in [c, \infty)} |\hat{g}_{n,k}^{(j)}(x) - g_0^{(j)}(x)| \to_{a.s.} 0, \quad as \ n \to \infty,$$

and for each $x > 0$ at which g_0 is $(k - 1)$-times differentiable, $\hat{g}_{n,k}^{(k-1)}(x) \to_{a.s.} g_0^{(k-1)}(x)$.

Proof. Let F_0 be the mixing distribution function associated with g_0. Then for all $x > 0$, we have $g_0(x) = \int_0^\infty k(t - x)_+^{k-1}/t^k dF_0(t)$. Now, let $Y_1, Y_2, \ldots$ be i.i.d. from F_0. Let $\mathbb{F}_n$ be the corresponding empirical distribution and g_n the mixed density $g_n(x) = \int_0^\infty k(t - x)_+^{k-1}/t^k d\mathbb{F}_n(t)$, for $x > 0$.

11

Let $d > 0$. Using integration by parts, we have for all $x > d$

$$
\begin{aligned}
&|g_n(x) - g_0(x)| \\
&= \left| \int_x^\infty k \frac{(t-x)^{k-1}}{t^k} d(\mathbb{F}_n - F_0)(t) \right| \\
&= \left| \int_x^\infty k \frac{(k-1)t^k(t-x)^{k-2} - kt^{k-1}(t-x)^{k-1}}{t^{2k}} (\mathbb{F}_n - F_0)(t)dt \right| \\
&\leq \left(\int_d^\infty k \frac{(t-d)^{k-2}}{t^k} dt + k^2 \int_d^\infty \frac{(t-d)^{k-2}}{t^k} dt \right) \|\mathbb{F}_n - F_0\|_\infty \\
&\leq \left(2k^2 \int_d^\infty \frac{(t-d)^{k-2}}{t^k} dt \right) \|\mathbb{F}_n - F_0\|_\infty \\
&= C_d \|\mathbb{F}_n - F_0\|_\infty.
\end{aligned}
$$

By the Glivenko-Cantelli theorem, the sequence of k-monotone densities $(g_n)_n$ satisfies $\sup_{x \in [d,\infty)} |g_n(x) - g_0(x)| \to_{a.s.} 0$ as $n \to \infty$. Since the MLE $\hat{g}_{n,k}$ maximizes the criterion function over the class $\mathcal{M}_k \cap L_1(\lambda)$, we have

$$
\lim_{\epsilon \searrow 0} \frac{1}{\epsilon} \left(\psi_n((1-\epsilon)\hat{g}_{n,k} + \epsilon g_n) - \psi_n(\hat{g}_{n,k}) \right) \leq 0,
$$

and this is equivalent to

$$
\int_0^\infty \frac{g_n(x)}{\hat{g}_{n,k}(x)} d\mathbb{G}_n(x) \leq 1. \tag{3.1}
$$

Let $\hat{F}_{n,k}$ denote again the MLE of the mixing distribution. By the Helly-Bray theorem, there exists a subsequence $\{\hat{F}_{l,k}\}$ that converges weakly to some distribution function $\hat{F}$ and hence for all $x > 0$ $\hat{g}_{l,k}(x) \to \hat{g}(x)$ as $l \to \infty$ where

$$
\hat{g}(x) = \int_0^\infty k \frac{(t-x)_+^{k-1}}{t^k} d\hat{F}(t), \qquad x > 0.
$$

The previous convergence is uniform on intervals of the form $[d, \infty)$, $d > 0$. This follows since $\hat{g}_{l,k}$ and $\hat{g}$ are monotone and $\hat{g}$ is continuous. Using convergence of g_n to g_0 and the inequality (3.1) we can show that the limit $\hat{g}$ and g_0 have to be the same, which implies the consistency result. Consistency of the higher derivatives can be shown recursively using convexity of $(-1)^j \hat{g}_{n,k}^{(j)}$ for $j = 1, \ldots, k-2$. The proof follows along the lines of Jongbloed (1995), pages 117-119, and Groeneboom, Jongbloed, and Wellner (2001b), pages 1674-1675; see Balabdaoui and Wellner (2005c) for complete details. ∎

We also have strong and uniform consistency of the LSE $\tilde{g}_{n,k}$ on intervals of the form $[c, \infty), c > 0$.

PROPOSITION 3.3. *Fix $c > 0$ and suppose that the true k-monotone density g_0 satisfies $\int_0^\infty x^{-1/2} dG_0(x) < \infty$. Then $\|\tilde{g}_{n,k} - g_0\|_2 \to_{a.s.} 0$,*

$$\sup_{x \in [c,\infty)} |\tilde{g}_{n,k}^{(j)}(x) - g_0^{(j)}(x)| \to_{a.s.} 0, \quad as \ n \to \infty,$$

for $j = 0, 1, \cdots, k-2$, and, for each $x > 0$ at which g_0 is $(k-1)$-times differentiable, $\tilde{g}_{n,k}^{(k-1)}(x) \to_{a.s.} g_0^{(k-1)}(x)$.

Proof. We write $\tilde{g}_n$ for $\tilde{g}_{n,k}$ throughout the proof. The main difficulty here is that the least squares estimator $\tilde{g}_n$ is not necessarily a density in that it may integrate to more than one; indeed it can be shown that $\int_0^\infty \tilde{g}_1(x)dx = ((2k-2)/k)(1 - 1/(2k-1))^{k-2} > 1$ for $k \geq 3$. However, once we show that $\tilde{g}_n$ stays bounded in L_2 with high probability, the proof of consistency will be much like the one used for $k = 2$; i.e., consistency of the LSE of a convex and decreasing density (see Groeneboom, Jongbloed, and Wellner (2001b)). The proof for $k = 2$ is based on the very important fact that the LSE is a density, which helps in showing that $\tilde{g}_n$ at the last jump point $\tau_n \in [0, \delta]$ of $\tilde{g}_n'$ for a fixed $\delta > 0$ is uniformly bounded. The proof would have been similar if we only knew that $\int_0^\infty \tilde{g}_n(x)dx = O_p(1)$.

Here we will first show that $\int_0^\infty \tilde{g}_n^2 d\lambda = O_p(1)$. From the last display in the proof of Proposition 2.2 $\int_0^\infty \tilde{g}_n^2(x)dx = \int_0^\infty \tilde{g}_n(x)d\mathbb{G}_n(x)$ and hence

$$\sqrt{\int_0^\infty \tilde{g}_n^2(x)dx} = \int_0^\infty \tilde{u}_n(x)d\mathbb{G}_n(x), \tag{3.2}$$

where $\tilde{u}_n \equiv \tilde{g}_n/\|\tilde{g}_n\|_2$ satisfies $\|\tilde{u}_n\|_2 = 1$. Take $\mathcal{F}_k$ to be the class of functions

$$\mathcal{F}_k = \left\{ g \in \mathcal{M}_k, \int_0^\infty g^2 d\lambda = 1 \right\}.$$

In the following, we show that $\mathcal{F}_k$ has an envelope $G \in L_1(G_0)$.

Note that for $g \in \mathcal{F}_k$ we have

$$1 = \int_0^\infty g^2 d\lambda \geq \int_0^x g^2 d\lambda \geq xg^2(x),$$

since g is decreasing. Therefore $g(x) \leq 1/\sqrt{x} \equiv G(x)$ for all $x > 0$ and $g \in \mathcal{F}_k$; i.e. G is an envelope for the class $\mathcal{F}_k$. Since $G \in L_1(G_0)$ (by our hypothesis) it follows from the strong law that

$$\int_0^\infty \tilde{u}_n(x)d\mathbb{G}_n(x) \leq \int_0^\infty G(x)d\mathbb{G}_n(x) \to_{a.s.} \int_0^\infty G(x)dG_0(x), \quad as \ n \to \infty$$

13

and hence by (3.2) the integral $\int_0^\infty \tilde{g}_n^2 d\lambda$ is bounded (almost surely) by some constant M_k.

Now we are ready to complete the proof. Let $\delta > 0$ and τ_n be the last jump point of $\tilde{g}_n^{(k-1)}$ if there are jump points in the interval $(0, \delta]$, otherwise we take τ_n to be 0. To show that the sequence $(\tilde{g}_n(\tau_n))_n$ stays bounded, we consider two cases:

1. $\tau_n \geq \delta/2$. Let n be large enough so that $\int_0^\infty \tilde{g}_n^2 d\lambda \leq M_k$. We have

$$
\begin{aligned}
\tilde{g}_n(\tau_n) &\leq \tilde{g}_n(\delta/2) = (2/\delta)(\delta/2)\tilde{g}_n(\delta/2) \leq (2/\delta)\int_0^{\delta/2} \tilde{g}_n(x)dx \\
&\leq (2/\delta)\sqrt{\delta/2}\sqrt{\int_0^{\delta/2} \tilde{g}_n^2(x)dx} \leq \sqrt{2/\delta}\sqrt{\int_0^\infty \tilde{g}_n^2(x)dx} \\
&= \sqrt{2M_k/\delta}.
\end{aligned}
\tag{3.3}
$$

2. $\tau_n < \delta/2$. We have

$$
\int_{\tau_n}^{\delta} \tilde{g}_n(x)dx \leq \sqrt{\delta - \tau_n}\sqrt{\int_{\tau_n}^{\delta} \tilde{g}_n^2(x)dx} \leq \sqrt{\delta}\sqrt{\int_0^\infty \tilde{g}_n^2(x)dx} = \sqrt{\delta M_k}.
$$

Using that $\tilde{g}_n$ is a polynomial of degree $k-1$ on the interval $[\tau_n, \delta]$ we have

$$
\begin{aligned}
\sqrt{\delta M_k} &\geq \int_{\tau_n}^{\delta} \tilde{g}_n(x)dx \\
&= \tilde{g}_n(\delta)(\delta - \tau_n) - \frac{\tilde{g}_n'(\delta)}{2}(\delta - \tau_n)^2 + \cdots + (-1)^{k-1}\frac{\tilde{g}_n^{(k-1)}(\delta)}{k!}(\delta - \tau_n)^k \\
&\geq (\delta - \tau_n)\left(\tilde{g}_n(\delta) + \frac{1}{k}(-1)\tilde{g}_n'(\delta)(\delta - \tau_n)\right. \\
&\qquad\qquad \left. + \cdots + (-1)^{k-1}\frac{\tilde{g}_n^{(k-1)}(\delta)}{(k-1)!}(\delta - \tau_n)^{k-1}\right) \\
&= (\delta - \tau_n)\left(\tilde{g}_n(\delta)\left(1 - \frac{1}{k}\right) + \frac{1}{k}\tilde{g}_n(\tau_n)\right) \geq \frac{\delta}{2k}\tilde{g}_n(\tau_n)
\end{aligned}
$$

and hence $\tilde{g}_n(\tau_n) \leq 2k\sqrt{M_k/\delta}$. By combining the bounds, we have for large n, $\tilde{g}_n(\tau_n) \leq 2k\sqrt{M_k/\delta} = C_k$. Now, since $\tilde{g}_n(\delta) \leq \tilde{g}_n(\tau_n)$, the sequence $\tilde{g}_n(x)$ is uniformly bounded almost surely for all $x \geq \delta$. Using a Cantor diagonalization argument, we can find a subsequence $\{n_l\}$ so that, for each $x \geq \delta$, $g_{n_l}(x) \to \tilde{g}(x)$, as $l \to \infty$. By Fatou's lemma, we have

$$
\int_\delta^\infty (\tilde{g}(x) - g_0(x))^2 dx \leq \liminf_{l \to \infty} \int_\delta^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2 dx.
\tag{3.4}
$$

14

On the other hand, the characterization of $\tilde{g}_n$ implies that $Q_n(\tilde{g}_n) \le Q_n(g_0)$, and this yields

$$\int_0^\infty (\tilde{g}_n(x) - g_0(x))^2 dx \le 2 \int_0^\infty (\tilde{g}_n(x) - g_0(x)) d(\mathbb{G}_n(x) - G_0(x)).$$

Thus we can write

$$\int_\delta^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2 dx \le \int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2 dx$$

$$\le 2 \int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x)) d(\mathbb{G}_{n_l}(x) - G_0(x)) \to_{a.s.} 0, \qquad (3.5)$$

as $l \to \infty$. The last convergence is justified as follows: since $\int_0^\infty \tilde{g}_{n_l}^2 d\lambda$ is bounded almost surely, we can find a constant $C > 0$ such that $\tilde{g}_{n_l} - g_0$ admits $G(x) = C/\sqrt{x}, x > 0$, as an envelope. Since $G \in L_1(G_0)$ by hypothesis and since the class of functions $\{(g - g_0)1_{[G \le M]} : g \in \mathcal{M}_k \cap L_2(\lambda)\}$ is a Glivenko-Cantelli class for every $M > 0$ (each element is a difference of two bounded monotone functions) (3.5) holds. From (3.4), we conclude that $\int_\delta^\infty (\tilde{g}(x) - g_0(x))^2 dx \le 0$, and therefore, $\tilde{g} \equiv g_0$ on $(0, \infty)$ since $\delta > 0$ can be chosen arbitrarily small. We have proved that there exists Ω_0 with $P(\Omega_0) = 1$ and such that for each $\omega \in \Omega_0$ and any given subsequence $\tilde{g}_{n_k}(\cdot, \omega)$, we can extract a further subsequence $\tilde{g}_{n_l}(\cdot, \omega)$ that converges to g_0 on $(0, \infty)$. It follows that $\tilde{g}_n$ converges to g_0 on $(0, \infty)$, and this convergence is uniform on intervals of the form $[c, \infty)$, $c > 0$ by the monotonicity and continuity of g_0. As for the MLE, consistency of the higher derivatives can be shown recursively using convexity of $(-1)^j \tilde{g}_n^{(j)}$ for $j = 1, \ldots, k-2$. ■

4. Asymptotic minimax risk lower bounds for the rates of convergence

In this section our goal is to derive minimax lower bounds for the behavior of *any estimator* of a $k-$monotone density g and its first $k - 1$ derivatives at a point x_0 for which the $k-$th derivative exists and is non-zero. The proof will rely upon the basic Lemma 4.1 of Groeneboom (1996); see also Jongbloed (2000). This basic method seems to go back to Donoho and Liu (1987) and Donoho and Liu (1991)). The relationship of our results to other rate results due to Fan (1991), and Zhang (1990) will be discussed later in the section.

As before, let $\mathcal{D}_k$ denote the class of $k-$monotone densities on $[0, \infty)$. Here is the notation we will need. Consider estimation of the $j-$th derivative of $g \in \mathcal{D}_k$

15

at x_0 for $j \in \{0, 1, \ldots, k-1\}$. If $\hat{T}_n$ is an arbitrary estimator of the real-valued functional T of g, then the (L_1-)minimax risk based on a sample $X_1, \ldots, X_n$ of size n from g which is known to be in a suitable subset $\mathcal{D}_{k,n}$ of $\mathcal{D}_k$ is defined by

$$MMR_1(n, T, \mathcal{D}_{k,n}) = \inf_{t_n} \sup_{g \in \mathcal{D}_{k,n}} E_g |\hat{T}_n - Tg| \,.$$

Here the infimum ranges over all possible measurable functions $t_n : \mathbb{R}^n \to \mathbb{R}$, and $\hat{T}_n = t_n(X_1, \ldots, X_n)$. When the subclasses $\mathcal{D}_{k,n}$ are taken to be shrinking to one fixed $g_0 \in \mathcal{D}_k$, the minimax risk is called *local* at g_0. The shrinking classes (parametrized by $\tau > 0$) used here are Hellinger balls centered at g_0:

$$\mathcal{D}_{k,n} \equiv \mathcal{D}_{k,n,\tau} = \left\{ g \in \mathcal{D}_k : \ H^2(g, g_0) = \frac{1}{2} \int_0^\infty (\sqrt{g(x)} - \sqrt{g_0(x)})^2 dx \le \tau/n \right\},$$

The behavior, for $n \to \infty$ of such a local minimax risk MMR_1 will depend on n (rate of convergence to zero) and the density g_0 toward which the subclasses shrink. The following lemma is the basic tool for proving such a lower bound.

LEMMA 4.1. *Assume that there exists some subset $\{g_\epsilon : \epsilon > 0\}$ of densities in $\mathcal{D}_{k,n}$ such that, as $\epsilon \downarrow 0$,*

$$H^2(g_\epsilon, g_0) \le \epsilon(1 + o(1)) \quad and \quad |Tg_\epsilon - Tg_0| \ge (c\epsilon)^r (1 + o(1))$$

for some $c > 0$ and $r > 0$. Then

$$\sup_{\tau > 0} \liminf_{n \to \infty} n^r MMR_1(n, T, \mathcal{D}_{k,n}) \ge \frac{1}{4} \left(\frac{cr}{2e} \right)^r \,.$$

Proof. See Jongbloed (1995) and Jongbloed (2000). ∎

Here is the main result of this section:

PROPOSITION 4.1. *Let $g_0 \in \mathcal{D}_k$ and x_0 be a fixed point in $(0, \infty)$ such that g_0 is $k-$times continuously differentiable at x_0 ($k \ge 2$). An asymptotic lower bound for the local minimax risk of any estimator $\hat{T}_{n,j}$ for estimating the functional $T_j g_0 = g_0^{(j)}(x_0)$, is given by:*

$$\sup_{\tau > 0} \liminf_{n \to \infty} n^{\frac{k-j}{2k+1}} MMR_1(n, T_j, \mathcal{D}_{k,n,\tau}) \ge \left\{ |g_0^{(k)}(x_0)|^{2j+1} g_0(x_0)^{k-j} \right\}^{1/(2k+1)} d_{k,j},$$

16

where $d_{k,j} > 0$, $j \in \{0, \ldots, k-1\}$. Here

$$d_{k,j} = \frac{1}{4}\left(4\frac{k-j}{2k+1}e^{-1}\right)^{\frac{k-j}{2k+1}}\frac{\lambda_{k,1}^{(j)}}{(\lambda_{k,2})^{\frac{k-j}{2k+1}}}$$

where

$$\lambda_{k,2} = \begin{cases} 2^{4(k+1)}\frac{(2k+3)(k+2)}{(k+1)^2}\frac{((2(k+1))!)^2}{(4k+7)!((k-1)!)^2\left(\binom{k}{k/2-1}\right)^2}, & k \quad even \\[4mm] 2^{4(k+2)}(2k+3)(k+2)\frac{((2(k+1))!)^2}{(4k+7)!(k!)^2\left(\binom{k+1}{(k-1)/2}\right)^2}, & k \quad odd. \end{cases}$$

Proposition 4.1 also yields lower bounds for estimation of the corresponding mixing distribution function F at a fixed point.

COROLLARY 4.1. *Let $g_0 \in \mathcal{D}_k$ and let x_0 be a fixed point in $(0, \infty)$ such that g_0 is $k-$times continuously differentiable at x_0, $k \geq 2$. Then, for estimating $Tg_0 = F(x_0)$ where F_0 is given in terms of g_0 by (2.2),*

$$\sup_{\tau > 0} \liminf_{n \to \infty} n^{\frac{1}{2k+1}} MMR_1(n, T, \mathcal{D}_{k,n,\tau}) \geq \left\{|g_0^{(k)}(x_0)|^{2k-1}g_0(x_0)\right\}^{1/(2k+1)}\frac{x_0^k}{k!}d_{k,k-1},$$

The dependence of our lower bound on the constants $g_0(x_0)$ and $g_0^{(k)}(x_0)$ matches with the known results for $k = 1$ and $k = 2$ due to Groeneboom (1985) and Groeneboom, Jongbloed, and Wellner (2001b), and reappears in the limit distribution theory for $k \geq 3$ in Balabdaoui and Wellner (2004c).

The result of Corollary 4.1 is consistent with the lower bound results of Zhang (1990) and Fan (1991) in the deconvolution setting as we now explain.

To link up with the deconvolution literature we transform our scale mixture problem to a location mixture or deconvolution problem. To do this we will reparametrize our $k-$monotone densities so that the beta kernels converge to the limiting exponential kernels: Note that if

$$g(x) = \int_0^\infty \frac{1}{y}\left(1 - \frac{x}{ky}\right)_+^{k-1} dF(y),$$

then for $X \sim g$, $Z = Z_k \sim k \times \text{Beta}(1, k)$, and $Y \sim F$ with Y and Z independent, we have $X \stackrel{d}{=} ZY$. Thus $X^* \equiv \log X = \log Y + \log Z \equiv Y^* + Z^*$. Hence the density g^* of X^* is given by

$$g^*(x) = \int_{-\infty}^\infty \left(1 - \frac{1}{k}e^{x-y}\right)_+^{k-1}e^{x-y}dF^*(y) = \int_{-\infty}^\infty f_{Z^*}(x-y)dF^*(y)$$

where $F^*(y) = F(e^y)$ is the distribution function of Y^*.

For the completely monotone case corresponding to $k = \infty$, the corresponding formulas for g and g^* are given by

$$g(x) = \int_0^\infty \frac{1}{y} \exp(-x/y) dF(y)\,,$$

and

$$g^*(x) = \int_{-\infty}^\infty \exp(-e^{x-y}) e^{x-y} dF^*(y) = \int_{-\infty}^\infty f_{Z_\infty^*}(x-y) dF^*(y)\,.$$

According to Fan (1991), we need to compute the characteristic function ϕ_{Z^*} and bound its modulus above and below for large arguments. Thus we calculate first for Z_∞^*: from Abramowitz and Stegun (1964), page 930,

$$\phi_{Z_\infty^*}(t) = \int_{-\infty}^\infty e^{itz} e^{-e^z} e^z dz = \int_0^\infty e^{it \log v} e^{-v} dv = \Gamma(1 + it)\,.$$

Thus by Abramowitz and Stegun (1964), page 256,

$$|\phi_{Z_\infty^*}(t)|^2 = \Gamma(1 + it)\Gamma(1 - it) = \frac{\pi t}{\sinh(\pi t)} = \frac{2\pi t}{e^{\pi t} - e^{-\pi t}}\,,$$

and it follows that

$$\sqrt{2\pi|t|} \exp(-\pi|t|/2) \leq |\phi_{Z_\infty^*}(t)| \leq \sqrt{3\pi|t|} \exp(-\pi|t|/2)$$

for $|t| \geq 1$. Thus the hypothesis (1.3) of Fan (1991) holds with $\beta = 1$ and $\beta_0 = \beta_1 = 1/2$. This implies the first hypothesis of Fan's theorem 4, page 1263, and thus we are in the case of a "super-smooth" convolution kernel. Fan's second hypothesis is easily satisfied by the current extreme value distribution function since $f_{Z_\infty^*}(y) = O(|y|^{-2})$ as $y \to \pm\infty$. It therefore follows in the completely monotone case ($k = \infty$) that for estimation of $F_0^*(y_0) = F(e^{y_0})$ the resulting minimax lower bound yields the rate of convergence $(\log n)^{-1}$. This rate could also be deduced from Zhang (1990), Corollary 3, page 824. (Note that the tail behavior of the characteristic function of our extreme value kernel coincides with the tail behavior of the characteristic function of the Cauchy kernel and that Zhang's example 2 yields the rate $(\log n)^{-1}$ in the case of the Cauchy kernel.)

We can also follow the deconvolution approach to obtain a minimax lower bound for estimation of the mixing distribution in the $k-$monotone case: the characteristic

18

function of $Z_k^* = \log Z_k$ is given by

$$
\begin{aligned}
\phi_{Z_k^*}(t) &= \int_{-\infty}^{\infty} e^{itz} \left(1 - \frac{1}{k}e^z\right)_+^{k-1} e^z dz = \int_0^k e^{it\log v}(1 - v/k)_+^{k-1} dv \\
&= \frac{k^{it}\Gamma(k+1)\Gamma(1+it)}{\Gamma(k+1+it)}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
|\phi_{Z_k^*}(t)|^2 &= \frac{k^{it}\Gamma(k+1)\Gamma(1+it)}{\Gamma(k+1+it)} \frac{k^{-it}\Gamma(k+1)\Gamma(1-it)}{\Gamma(k+1-it)} \\
&= \frac{(k!)^2}{(k^2+t^2)\cdots(1+t^2)} \sim \frac{(k!)^2}{t^{2k}} \qquad \text{as } t \to \infty.
\end{aligned}
$$

It should also be noted that

$$
\lim_{k\to\infty} |\phi_{Z_k^*}(t)|^2 = \lim_{k\to\infty} \frac{(k!)^2}{(k^2+t^2)\cdots(1+t^2)} = \frac{\pi t}{\sinh(\pi t)} = |\phi_{Z_\infty^*}(t)|^2.
$$

Thus $|\phi_{Z_k^*}(t)| \sim k!/t^k$ as $\to \infty$, and we are in the situation of a smooth convolution kernel of hypothesis (1.4) of Fan (1991), page 1263, with Fan's $\beta = k$ in our setting. Thus Fan's theorem (extended to negative values of l) gives our rate of convergence for estimating $F^*(y_0) = F(e^{y_0})$ or $g^{(k-1)}$ by taking $l = -1$, $\alpha + m = 0$, and $\beta = k$. By "extending" Fan's theorem further and taking $l = -(k-j)$, we get the rate of convergence $n^{-(k-j)/(2k+1)}$, $j = 1, \ldots, k-1$ for estimation of $g_0^{(j)}(x_0)$.

Proof of Proposition 4.1. Let μ be a positive number and consider the function g_μ defined by:

$$
g_\mu(x) = g_0(x) + s(\mu)(x_0 + \mu - x)^{k+1}(x - x_0 + \mu)^{k+2} 1_{[x_0-\mu,x_0+\mu]}(x), \quad x \in (0,\infty)
$$

where $s(\mu)$ is a scale to be determined later. We denote the unscaled perturbation function by $\tilde{g}_\mu$; i.e.,

$$
\tilde{g}_\mu(x) = (x_0 + \mu - x)^{k+1}(x - x_0 + \mu)^{k+2} 1_{[x_0-\mu,x_0+\mu]}(x).
$$

If μ is chosen small enough so that the true density g_0 is k-times continuously differentiable on $[x_0 - \mu, x_0 + \mu]$, the perturbed function g_μ is also k-times differentiable on $[x_0 - \mu, x_0 + \mu]$ with a continuous k-th derivative. Now, let r be the function defined on $(0,\infty)$ by

$$
r(x) = (1-x)^{k+1}(1+x)^{k+2} 1_{[-1,1]}(x) = (1-x^2)^{k+1}(1+x) 1_{[-1,1]}(x).
$$

19

Then, we can write $\tilde{g}_\mu$ as $\tilde{g}_\mu(x) = \mu^{2k+3} r((x - x_0)/\mu)$. Then for $0 \le j \le k$

$$g_\mu^{(j)}(x_0) - g_0^{(j)}(x_0) = s(\mu)\mu^{2k+3-j} r^{(j)}(0).$$

The scale $s(\mu)$ should be chosen so that $(-1)^j g_\mu^{(j)}(x) > 0$ for all $0 \le j \le k$, for $x \in [x_0 - \mu, x_0 + \mu]$. But for μ small enough, the sign of $(-1)^j g_\mu^{(j)}$ will be that of $(-1)^j g_0^{(j)}(x_0)$, and hence g_μ is k–monotone. For $j = k$, $g_\mu^{(k)}(x_0) = g_0^{(k)}(x_0) + s(\mu)\mu^{k+3} r^{(k)}(0)$. Assume that $r^{(k)}(0) \ne 0$. Set $s(\mu) = g_0^{(k)}(x_0)/(\mu^{k+3} r^{(k)}(0))$. Then for $0 \le j \le k - 1$

$$g_\mu^{(j)}(x_0) = g_0^{(j)}(x_0) + \mu^{k-j} \frac{g_0^{(k)}(x_0) r^{(j)}(0)}{r^{(k)}(0)} = g_0^{(j)}(x_0) + o(\mu),$$

as $\mu \to 0$, and

$$(-1)^k g_\mu^{(k)}(x_0) = 2(-1)^k g_0^{(k)}(x_0) > 0.$$

for $j = k$. It can be shown (after some algebra; see Balabdaoui and Wellner (2005c) for the details) that

$$r^{(j)}(0) = \begin{cases} (-2)^{j/2} \prod_{i=0}^{j/2-1}(k + 1 - i) \times \prod_{i=0}^{j/2-1}(j - 2i - 1) \ne 0, & j \text{ even} \\ (-2)^{(j-1)/2} \prod_{i=0}^{(j-1)/2-1}(k + 1 - i) \times \prod_{i=0}^{(j-1)/2}(j - 2i) \ne 0, & j \text{ odd.} \end{cases}$$

We set $C_{k,j} = r^{(j)}(0)$ for $1 \le j \le k$. Then $C_{k,k}$ becomes

$$C_{k,k} = \begin{cases} (-2)^{k/2} \prod_{i=0}^{k/2-1}(k + 1 - i) \times \prod_{i=0}^{k/2-1}(k - 2i - 1), & \text{if } k \text{ is even} \\ (-2)^{(k-1)/2} \prod_{i=0}^{(k-1)/2-1}(k + 1 - i) \times \prod_{i=0}^{(k-1)/2}(k - 2i), & \text{if } k \text{ is odd.} \end{cases}$$

The previous expressions can be given in a more compact form. After some algebra, we find that

$$C_{k,k} = \begin{cases} 2 \times (-1)^{k/2}(k + 1)(k - 1)!\binom{k}{k/2-1}, & \text{if } k \text{ is even} \\ (-1)^{(k-1)/2} k!\binom{k+1}{(k-1)/2}, & \text{if } k \text{ is odd.} \end{cases} \tag{4.1}$$

We have for $0 \le j \le k - 1$,

$$|T_j(g_\mu) - T_j(g_0)| = \left| \frac{C_{k,j}}{C_{k,k}} g_0^{(k)}(x_0) \right| \mu^{k-j} \equiv \lambda_{k,1}^{(j)} \left| g_0^{(k)}(x_0) \right| \mu^{k-j}$$

where we defined $\lambda_{k,1}^{(j)} = |C_{k,j}/C_{k,k}|$ for $j \in \{0, \ldots, k - 1\}$. Furthermore, by computation and change of variables,

$$\int_0^\infty \frac{(g_\mu(x) - g_0(x))^2}{g_0(x)} dx$$

$$= \left(\frac{\left(g_0^{(k)}(x_0)\right)^2}{g_0(x_0)} \frac{\int_{-1}^1 (1 - z^2)^{2(k+1)}(z + 1)^2 dz}{(C_{k,k})^2} \right) \mu^{2k+1} + o(\mu^{2k+2})$$

20

as $\mu \searrow 0$. This gives control of the Hellinger distance as well in view of Jongbloed (2000), Lemma 2, page 282, or Jongbloed (1995), Corollary 3.2, pages 30 and 31. We set

$$
\lambda_{k,2} = \frac{\int_{-1}^{1}(1-z^2)^{2(k+1)}(z+1)^2 dz}{(C_{k,k})^2}
$$

$$
= \begin{cases} 2^{4(k+1)}\dfrac{(2k+3)(k+2)}{(k+1)^2}\dfrac{((2(k+1))!)^2}{(4k+7)!((k-1)!)^2\left(\binom{k}{k/2-1}\right)^2}, & k \text{ even} \\[3ex] 2^{4(k+2)}(2k+3)(k+2)\dfrac{((2(k+1))!)^2}{(4k+7)!(k!)^2\left(\binom{k+1}{(k-1)/2}\right)^2}, & k \text{ odd} \end{cases}
$$

Now, by using the change of variable $\epsilon = \mu^{2k+1}(b_k + o(1))$, where $b_k = \lambda_{k,2}(g_0^{(k)}(x_0))^2/g_0(x_0))$ so that $\mu = (\epsilon/b_k)^{1/(2k+1)}(1+o(1))$, then for $0 \le j \le k-1$, the modulus of continuity, m_j, of the functional T_j satisfies

$$
m_j(\epsilon) \ge \lambda_{k,1}^{(j)} g_0^{(k)}(x_0) \left(\frac{\epsilon}{b_k}\right)^{(k-j)/(2k+1)} (1+o(1)).
$$

The result is that $m_j(\epsilon) \ge (r_{k,j}\epsilon)^{\frac{k-j}{2k+1}}(1+o(1))$, where $r_{k,j} = (\lambda_{k,1}^{(j)} g_0^{(k)}(x_0))^{(2k+1)/(k-j)}/b_k$ and hence

$$
\sup_{\tau>0} \lim_{n\to\infty} \inf n^{\frac{k-j}{2k+1}} MMR_1(n,T_j,\mathcal{D}_{k,n,\tau}) \ge \frac{1}{4}\left(4\frac{k-j}{2k+1}e^{-1}\right)^{\frac{k-j}{2k+1}} (r_{k,j})^{\frac{k-j}{2k+1}}, \quad (4.2)
$$

which can be rewritten as

$$
\sup_{\tau>0} \lim_{n\to\infty} \inf n^{\frac{k-j}{2k+1}} MMR_1(n,T_j,\mathcal{D}_{k,n,\tau})
$$

$$
\ge \frac{1}{4}\left(4\frac{k-j}{2k+1}e^{-1}\right)^{\frac{k-j}{2k+1}} \frac{\lambda_{k,1}^{(j)}}{(\lambda_{k,2})^{\frac{k-j}{2k+1}}}\left\{ \left|g_0^{(k)}(x_0)\right|^{\frac{2j+1}{2k+1}} g_0(x_0)^{\frac{k-j}{2k+1}}\right\}
$$

for $j = 0, \cdots, k-1$. ∎

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions.* National Bureau of Standards, Washington. (Seventh reprinting by Dover, New York, 1970).

Anevski, D. (1994). Estimating the derivative of a convex density. *Technical Report* **1994:8**, Dept. of Math. Statistics, Univ. of Lund.

Anevski, D. (2003). Estimating the derivative of a convex density. *Statistica Neerl.* **57**, 245 - 257.

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. Statist.* **26**, 641 - 647.

Balabdaoui, F. (2004). Nonparametric estimation of k-monotone density: A new asymptotic distribution theory. Ph.D. dissertation, University of Washington, Department of Statistics.

Balabdaoui, F. and Wellner, J. A.   (2004a). Estimation of a k-monotone density, part 2: algorithms for computation and numerical results. Technical Report 460, Department of Statistics, University of Washington. Available at: *http://www.stat.washington.edu/www/research/reports/2004/.*

Balabdaoui, F. and Wellner, J. A. (2004b). Estimation of a k-monotone density, part 3: limiting Gaussian versions of the problem; invelopes and envelopes Technical Report 461, Department of Statistics, University of Washington. Available at: *http://www.stat.washington.edu/www/research/reports/2004/.*

Balabdaoui, F. and Wellner, J. A.   (2004c). Estimation of a k-monotone density, part 4: limit distribution theory and the spline connection. Technical Report 462, Department of Statistics, University of Washington. Available at: *http://www.stat.washington.edu/www/research/reports/2004/.*

Balabdaoui, F. and Wellner, J. A. (2005a). Mixtures and monotonicity: a review of estimation under monotonicity constraints. To appear, Kjell Doksum Festschrift volume.

Balabdaoui, F. and Wellner, J. A. (2005b). Conjecture of error boundedness in a new Hermite interpolation problem via splines of odd-degree. *Technical Report* **480**, Department of Statistics, University of Washington.

Balabdaoui, F. and Wellner, J. A. (2005c). Estimation of a k-monotone density, part 1: characterizations, consistency, and minimax lower bounds. *Technical Report* **459**, Department of Statistics, University of Washington.

Brunk, H. D. (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.* **29**, 437 - 454.

de Boor, C. (1978). *A Practical Guide to Splines.* Springer-Verlag, New York.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation.* Springer-Verlag, New York.

Donoho, D. L. and Liu, R. C. (1987). Geometrizing rates of convergence, I. Technical Report 137, Dept. of Statistics, Univ. California, Berkeley.

Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence, II, III. *Ann. Statist.* **19**, 633 - 667, 668 - 701.

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

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications.* Vol. 2, 2nd ed. Wiley, New York.

Gneiting, T. (1998). On the Bernstein-Hausdorff-Widder conditions for completely monotone functions. *Exposition. Math.* **16**, 181 - 183.

Gneiting, T. (1999). Radial positive definite functions generated by Euclid's hat. *J. Multivariate Analysis* **69**, 88 - 119.

Grenander, U. (1956). On the theory of mortality measurement, Part II. *Skand. Actuar.* **39**, 125 - 153.

Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer,* Vol. II. Lucien M. LeCam and Richard A. Olshen eds. Wadsworth, New York. 529 - 555.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Th. Rel. Fields* **81**, 79 - 109.

Groeneboom, P. (1996). Inverse problems in statistics. *Proceedings of the St. Flour Summer School in Probability. Lecture Notes in Math.* **1648**, 67 - 164. Springer, Berlin.

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001b). Estimation of convex functions: characterizations and asymptotic theory. *Ann. Statist.* **29**, 1653 - 1698.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, Boston.

Hampel, F.R. (1987). Design, modelling and analysis of some biological datasets. In *Design, data and analysis, by some friends of Cuthbert Daniel*, C.L. Mallows, editor, 111 - 115. Wiley, New York.

Jewell, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10**, 479 - 484.

Jongbloed, G. (1995). *Three Statistical Inverse Problems; estimators-algorithms-asymptotics.* Ph.D. dissertation, Delft University of Technology, Department of Mathematics.

Jongbloed, G. (2000). Minimax lower bounds and moduli of continuity. *Statist. Probab. Lett.* **50**, 279 - 284.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18**, 191-219.

Lévy, P. (1962). Extensions d'un théorème de D. Dugué et M. Girault. *Z. Wahrscheinlichkeitstheorie* **1**, 159 - 173.

Lindsay, B. (1983a). The geometry of mixture likelihoods: a general theory. *Ann. Statist.* **11**, 86 - 94.

Pfanzagl, J. (1988). Consistency of maximum likelihood estimators for certain nonparametric families, in particular: mixtures: *J. Statist. Planning and Inference* **19**, 293 - 310.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankya Series A* **31**, 23 - 36.

van de Geer, S. (1993). Hellinger consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21**, 14 - 44.

van de Geer, S. (1999). *Empirical Processes in M-Estimation.* Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (2000). Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes, pp. 115 - 134 In *High Dimensional Probability II*, Evarist Giné, David Mason, and Jon A. Wellner, editors, Birkhäuser, Boston.

Van Eeden, C. (1956). Maximum likelihood estimation of ordered probabilities. *Proc. Konink. Nederl. Akad. Wetensch. A* **59**, 444 - 455.

Van Eeden, C. (1957). Maximum likelihood estimation of partially ordered probabilities. *Proc. Konink. Nederl. Akad. Wetensch. A* **60**, 128 - 136.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76**, 751 - 761.

Widder, D. V. (1941). *The Laplace Transform.* Second printing. Princeton University Press, Princeton.

Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.* **23**, 189 - 207.

Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statistica Sinica* **3**, 501 - 515.

Zeidler, E. (1985). *Nonlinear Functional Analysis and its Applications III: Variational Methods and Optimization.* Springer-Verlag, New York.

Zhang, C-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806 - 831.

INSTITUTE FOR MATHEMATICAL STOCHASTICS
GEORGIA AUGUSTA UNIVERSITY GOETTINGEN
MASCHMUEHLENWEG 8 - 10
D-37073 GOETTINGEN
GERMANY
e-mail: fadoua@math.uni-goettingen.de

UNIVERSITY OF WASHINGTON
STATISTICS
BOX 354322
SEATTLE, WASHINGTON 98195-4322
U.S.A.
e-mail: jaw@stat.washington.edu