Estimation of a *k*-monotone density: characterizations, consistency and minimax lower bounds

Fadoua Balabdaoui*

CEREMADE, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775, Paris, CEDEX 16, France

Jon A. Wellner[†]

Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322, USA

The classes of monotone or convex (and necessarily monotone) densities on $\mathbb{R}^+$ can be viewed as special cases of the classes of *k-monotone* densities on $\mathbb{R}^+$. These classes 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 (∞-monotone) densities on $\mathbb{R}^+$. In this paper we consider non-parametric maximum likelihood and least squares estimators of a *k*-monotone density g_0. We prove existence of the estimators and give characterizations. We also establish consistency properties, and show that the estimators are splines of degree $k-1$ with simple knots. We further provide asymptotic minimax risk lower bounds for estimating the derivatives $g_0^{(j)}(x_0), 0 = 1,\ldots,k-1$, at a fixed point x_0 under the assumption that $(-1)^k g_0^{(k)}(x_0) > 0$.

Keywords and Phrases: completely monotone, least squares, maximum likelihood, minimax risk, mixture models, multiply monotone, non-parametric estimation, rates of convergence, shape constraints

1 Introduction

Densities with monotone or convex shape are encountered in many non-parametric 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. Estimation of monotone densities on $(0, \infty)$ was initiated by GRENANDER (1956a,b) with related work by AYER *et al.* (1955), BRUNK (1958), and VAN EEDEN (1957a,b). Asymptotic theory of the maximum likelihood estimation

**fadoua@ceremade.dauphine.fr
†jaw@stat.washington.edu

(MLE) was developed by PRAKASA RAO (1969) with later contributions by GROENE-BOOM (1985, 1989), and KIM and POLLARD (1990).

Convex densities arise in connection with Poisson process models for bird migration and scale mixtures of triangular densities; see, for example, HAMPEL, (1987) and ANEVSKI, (2003). Estimation of convex densities on $(0, \infty)$ was apparently initiated by ANEVSKI (1994) (see also ANEVSKI, 2003), and was pursued by JONGBLOED (1995). The limit distribution theory for the MLE and least square (LS) estimators and their first derivative at a fixed point was obtained by GROENEBOOM, JONGBLOED, and WELLNER (2001). For consistent estimation of the estimators at the origin, see. BALABDAOUI (2007).

Estimation in the class of k-monotone densities on $\mathbb{R}^+$, denoted hereafter by $\mathcal{D}_k$, has been very recently considered in BALABDAOUI and WELLNER (2007) and has several motivating components. By definition, g is k-monotone on $(0, \infty)$ if g is non-negative and $(-1)^l g^{(l)}$ is non-increasing and convex for $l \in \{0, \ldots, k-2\}$ for $k \geq 2$, and simply non-negative and non-increasing 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 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 are scale mixtures of the triangular, or beta$(1, 2)$, densities. 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 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, p. 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, see JEWELL (1982) and the references therein, and BALABDAOUI and WELLNER (2007) for further motivation and references.

In BALABDAOUI and WELLNER (2007), the joint limit distribution theory for the MLE and LSE of a k-monotone density and their higher derivatives up to degree $k - 1$ at a fixed point is established modulo a spline conjecture. The rate of convergence of the j-th derivative, $j = 0, \ldots, k-1$ is shown to be $n^{(k-j)/(2k+1)}$. Note that these rates coincide with the minimax lower bounds obtained here. As for the joint limiting distribution, it depends on a Gaussian process H_k defined uniquely almost surely as follows:

- $H_k(t) \geq Y_k(t), t \in \mathbb{R}$.
- $(-1)^k H_k$ is $2k$-convex; that is, $(-1)^k H_k^{(2k-2)}$ exists and is convex.
- The process H_k satisfies

$$\int_{-\infty}^{\infty} (H_k(t) - Y_k(t)) \, dH_k^{(2k-1)}(t) = 0,$$

where Y_k is the $(k-1)$-th fold integral of a two-sided Brownian motion plus $(-1)^k k!/(2k)! t^{2k}, t \in \mathbb{R}$. JEWELL (1982) initiated the study of MLE 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 or the estimator of the mixing distribution function. As noted in BALABDAOUI and WELLNER (2007), it may be possible to obtain some insight into the asymptotics of the MLE of a completely monotone density by better understanding the behavior of the MLE of a k-monotone density for arbitrary $k > 2$. Indeed, as the class $\mathcal{D}_\infty$ is the intersection of all of the $\mathcal{D}_k$s, it can be well approximated by $\mathcal{D}_k$ with a large k.

Existence of the MLE and LSE of a k-monotone density, their characterization, their structure (splines of degree $k-1$ and with simple knots), and consistency of their derivatives up to degree $k-1$ are used in BALABDAOUI and WELLNER (2007). In this paper, we give proofs of those essential properties in sections 2 and 3. 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.

In the sequel, $X_1, \ldots, X_n$ are i.i.d. random variables with density $g_0 \in \mathcal{D}_k$, $\mathbb{G}_n$ is the corresponding empirical distribution function. We write $X_{(1)}, \ldots, X_{(n)}$ for the order statistics of $X_1, \ldots, X_n$, use the notation $z_+ = z1_{[z \geq 0]}$, and write λ for Lebesgue measure on $\mathbb{R}$.

2 Existence and characterizations

2.1 Mixture representation

Lemma 1 characterizing integrable k-monotone functions and giving an inversion formula follows from the results of WILLIAMSON (1956).

LEMMA 1. *(Integrable k-monotone characterization) A function g is an integrable k-monotone function if and only if it is of the form*

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

where F is non-decreasing and bounded on $(0, \infty)$. Thus g is a k-monotone density if and only if it is of the form of Equation 1 for some distribution function F on $(0, \infty)$. If F in Equation 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

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

where $G(t) = \int_0^t g(x)\mathrm{d}x$.

PROOF. The representation in equation 1 follows from theorem 5 of LÉVY (1962) by taking $k = n + 1$ and $f \equiv 0$ on $(-\infty, 0]$. The inversion formula 2 follows from lemma 1 in WILLIAMSON (1956) together with an integration by parts argument.

For $k = 1$ ($k = 2$), note that the characterization matches with the well-known fact that a density is non-decreasing (non-decreasing and convex) on $(0, \infty)$ if and only if it is a mixture of uniform densities (triangular densities). More generally, the characterization establishes a one-to-one correspondence between the class of k-monotone densities and the class of scale mixture of beta densities with parameters 1 and k. From the inversion formula in equation 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.

2.2 Existence and characterization of the estimators

We now consider the MLE and LSE of a k-monotone density g_0. We show that these estimators exist and give characterizations thereof. In the following, λ is the Lebesgue measure, $\mathcal{M}_k$ is the class of all k-monotone functions on $(0, \infty)$, and $\mathcal{M}_k \subset L_1(\lambda)$ is the class of all integrable k-monotone functions. Note that $\mathcal{D}_k \subset \mathcal{M}_k \cap L_1(\lambda)$.

Let

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

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 in SILVERMAN, 1982) as follows:

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

for $g \in \mathcal{M}_k \cap L_1(\lambda)$. Note that, using the one-to-one correspondence between the mixed k-monotone function $g = g_F$ and its corresponding mixing distribution F, maximizing ψ_n over the space $\mathcal{M}_k \cap L_1(\lambda)$ is equivalent to maximizing

$$\tilde{\psi}_n(F) = \int_0^\infty \log\left(\int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k}\, dF(t)\right) d\mathbb{G}_n(x) - \int_0^\infty \int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k}\, dF(t)\, dx$$

over the space of bounded and non-increasing functions F on $(0, \infty)$.

LEMMA 2. *The maximizer $\hat{g}_n$ of ψ_n over $\mathcal{M}_k \cap L_1(\lambda)$ exists and belongs to $\mathcal{D}_k$ (and hence is a density). Furthermore, $\hat{g}_n$ is of the form*

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

where $m \in \mathbb{N} \setminus \{0\}$, and $\hat{w}_1, \cdots, \hat{w}_m$ and $\hat{a}_1, \cdots, \hat{a}_m$ are respectively the weights and the support points of the maximizing (discrete) mixing distribution $\hat{F}_n$.

REMARK 1. It follows from lemma 2.2 that the MLE $\hat{g}_n$ is a k-monotone spline of degree $k-1$ with m simple knots $\hat{a}_1, \cdots, \hat{a}_m$ (for a definition of splines and multiplicity of the knots, see, e.g., DE BOOR, 1978, and DE VORE and LORENTZ, 1993). Note that this is also equivalent to saying that $\hat{g}_n$ is a finite mixture of beta's with parameters 1 and k.

REMARK 2. It can be shown that the support points of the mixing distribution $\hat{F}_n$ fall stricty between the order statistics $X_{(1)}, \ldots, X_{(n)}$ with at most one support point in $(X_{(i)}, X_{(i+1)})$. Note also that by definition of the MLE, the last support point has to be strictly larger than $X_{(n)}$.

PROOF OF LEMMA 2. From LINDSAY (1983), we conclude that there exists a unique maximizer of l_n and the maximum is achieved by a finite mixture of at most n beta densities with parameters 1 and k. We denote this maximizer by $\hat{f}_n$.

By arguing as in GROENEBOOM *et al.* (2001, p. 1662), let $g \in \mathcal{M}_k \cap L_1(\lambda)$ such that $\int_0^\infty g(x)\, dx = c$. Then $g/c \in \mathcal{D}_k$, and we can write

$$\psi_n(\hat{f}_n) - \psi_n(g) = \int_0^\infty \log \hat{f}_n(x)\, d\mathbb{G}_n(x) - 1 - \int_0^\infty \log g(x)\, d\mathbb{G}_n(x) + c$$
$$= l_n(\hat{f}_n) - l_n(g/c) - \log c + c - 1 \geq -\log c + c - 1 \geq 0.$$

Hence, $\hat{g}_n = \hat{f}_n$.

REMARK 3. Considering maximization over the bigger set $\mathcal{M}_k \cap L_1(\lambda)$ is motivated by the fact that this set is a cone. Characterization of the MLE takes then a simpler form than the one we would obtain with $\mathcal{D}_k$.

Lemma 3 gives a necessary and sufficient condition for a function $\hat{g}_n \in \mathcal{M}_k \cap L_1(\lambda)$ to be the MLE.

For $k \geq 3$ it generalizes lemma 2.4 of GROENEBOOM *et al.* (2001).

LEMMA 3. *Let* $X_1, \cdots, X_n$ *be i.i.d. random variables from the true density* g_0. *A* k-*monotone spline* $\hat{g}_n$ *of degree* $k-1$ *and simple knots* $\hat{a}_1, \cdots, \hat{a}_m$ *is the MLE if and only if for all* $t > 0$

$$\hat{H}_n(t) \equiv \int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k \hat{g}_n(x)} \, d\mathbb{G}_n(x) \leq 1 \tag{3}$$

$$= 1, \text{if } t \in \{\hat{a}_1, \cdots, \hat{a}_m\}. \tag{4}$$

PROOF. See the Appendix.

REMARK 4. Note that t is a knot in $\{\hat{a}_1, \cdots, \hat{a}_m\}$ if and only if $(-1)^{k-1}\hat{g}_n^{(k-1)}(t-) < (-1)^{k-1}\hat{g}_n^{(k-1)}(t+)$. Thus, the equality condition in Equation 4 can re-expressed in terms of the left and right $(k-1)$-th derivative of $\hat{g}_n$ as in Lemma 2.4 of GROENE-BOOM *et al.* (2001) in the particular case of $k = 2$.

The MLE $\hat{g}_n$ can be computed by means of the support reduction algorithm of GROENEBOOM, JONGBLOED, and WELLNER (2008); also see BALADAOUI and WELLNER (2004) for further details.

Now, we briefly consider the LSE. The LS criterion is:

$$Q_n(g) = \frac{1}{2} \int_0^\infty g^2(x) \, dx - \int_0^\infty g(x) \, d\mathbb{G}_n(x). \tag{5}$$

We want to minimize this over $g \in \mathcal{D}_k \cap L_2(\lambda)$, the subset of square integrable k-monotone densities. 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 owing 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 $\mathcal{M}_k \cap L_2(\lambda)$. Here, one might wonder why consider the LSE when the MLE is a 'natural' density estimator. It turns out that the random processes involved in the characterization of the LSE for a finite sample size n gives a great insight into the limiting distribution of the estimator and its derivatives up to degree $k-1$. Thus, even though the MLE and LSE are asymptotically equivalent, it is easier to understand and establish the asymptotic theory of the MLE through the LSE: compare the characterization of the MLE in lemma 3 with the characterization of the LSE (over $\mathcal{M}_k \cap L_2(\lambda)$) given in lemma A.2. (For more details, see BALABDAOUI and WELLNER, 2007). We will not develop the characterization of the LSE further here, but postpone this study to the Appendix.

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 MLEs 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 1. Suppose that $\hat{g}_n$ is the MLE of g_0 in the class $\mathcal{D}_k$. Then,

$$H(\hat{g}_n, g_0) \to_{a.s.} 0 \text{ as } n \to \infty.$$

Furthermore, $\hat{F}_n \to_d F_0$ almost surely where $\hat{F}_n$ is the MLE of the mixing distribution function F_0

PROOF. Note that math $\mathcal{D}_k = \{g_F : F \text{ is a d.f. on } (0, \infty)\}$ with

$$g_F(x) = g(x) = \int_0^\infty k_x(t) \, dF(t)$$

and k_x is the the scaled beta$(1, k)$ kernel; that is, $k_x(t) = t(1 - tx/k)_+^{k-1}$. For all $x > 0$, the kernel k_x is bounded, continuous, and it is easy to see that it satisfies $\lim_{t \searrow 0} k_x(t) = \lim_{t \to \infty} k_x(t) = 0$. Hence, the map $F \mapsto g_F$ is continuous with respect to the vague topology for all $x > 0$. This implies that the class

$$\mathcal{G} = \left\{ \frac{g_F}{g_F + g_{F_0}}, F \text{ is a d.f. on } (0, \infty) \right\}$$

is continuous in F with respect to the vague topology for every $x > 0$. Now, as the family of sub-distributions F on $(0, \infty)$ is compact for the vague topology (see, e.g., BAUER, 1981), and the class $\mathcal{G}$ is uniformly bounded by 1, we conclude by lemma 5.1 of VAN DER GEER (1993) that g is P_0-Glivenko–Cantelli. It follows by corollary 1 of VAN DER VAART and WELLNER (2000) that $H(\hat{g}_n, g_0) \to_{a.s.} 0$. The second assertion of the proposition follows from lemma 5.2 of VAN DE GEER (1993).

Lemma 4 establishes a useful bound for k-monotone densities.

LEMMA 4. *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

$$g(x) = \int_x^\infty \frac{k}{y^k} (y - x)^{k-1} \, dF(y) = \frac{1}{x} \int_x^\infty \frac{kx}{y} \left(1 - \frac{x}{y}\right)^{k-1} dF(y)$$

$$\leq \frac{1}{x} \sup_{x \leq y < \infty} \frac{kx}{y} \left(1 - \frac{x}{y}\right)^{k-1} = \frac{k}{x} \sup_{0 < u \leq 1} u(1 - u)^{k-1} = \frac{1}{x} \left(1 - \frac{1}{k}\right)^{k-1}$$

by an easy calculation. [Note that when $k = 2$, this bound equals $1/(2x)$ which agrees with the bound given by JONGBLOED (1995, p. 117) and GROENEBOOM *et al.*, (2001, p. 1669) in this case.]

PROPOSITION 2. *Let $c > 0$. Then for $j = 0, 1, \cdots, k - 2$*

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

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

PROOF. Using the first part in the characterization of the MLE, we have

$$\int_0^\infty \frac{g_0(x)}{\hat{g}_n(x)} \, d\mathbb{G}_n(x) \leq 1. \tag{6}$$

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

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

The previous convergence is uniform on $[c, \infty), c > 0$. This follows as $\hat{g}_l$ and $\hat{g}$ are monotone and $\hat{g}$ is continuous.

Using the inequality 6 we can show that the limit $\hat{g}$ and g_0 have to be the same, which implies the consistency result. The proof follows along the lines of GROENEBOOM *et al.* (2001, p. 1674–1675; see the Appendix). Consistency of the higher derivatives can be shown recursively using convexity of $(-1)^j \hat{g}_n^{(j)}$ for $j = 1, \ldots, k - 1$ in the same way as in the proof of lemma 3.1 of GROENEBOOM *et al.* (2001): for small $h > 0$, convexity of $(-1)^j \hat{g}_n^{(j)}$ allows us to write, for $j = 0, \ldots, k - 2$,

$$\frac{(-1)^j \hat{g}_n^{(j)}(x - h) - (-1)^j \hat{g}_n^{(j)}(x)}{-h}$$

$$\leq (-1)^j \hat{g}_n^{(j+1)}(x-) \leq (-1)^j \hat{g}_n^{(j+1)}(x+)$$

$$\leq \frac{(-1)^j \hat{g}_n^{(j)}(x + h) - (-1)^j \hat{g}_n^{(j)}(x)}{h}.$$

By letting $n \to \infty$, this implies that

$$\frac{(-1)^j g_0^{(j)}(x-h) - (-1)^j g_0^{(j)}(x)}{-h}$$
$$\leq \liminf_{n \to \infty} (-1)^j \hat{g}_n^{(j+1)}(x-) \leq \limsup_{n \to \infty} (-1)^j \hat{g}_n^{(j+1)}(x+)$$
$$\leq \frac{(-1)^j g_0^{(j)}(x+h) - (-1)^j g_0^{(j)}(x)}{h}.$$

By letting $h \searrow 0$, we conclude consistency of $\hat{g}_n^{(j)}(x), j = 0, \ldots, k-1$ for $x \in (0, \infty)$. Note that consistency of $(-1)^j \hat{g}_n^{(j)}, j = 1, \ldots, k-2$ is uniform on intervals of the form $[c, \infty)$ because of continuity of those derivatives. For $k-1$, only pointwise strong consistency of $(-1)^{k-1} \hat{g}_n^{(k-1)}$ can be claimed.

We also have strong and uniform consistency of the LSE $\tilde{g}_n$ on intervals of the form $[c, \infty), c > 0$. The relevant result and proof are deferred to the Appendix.

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 on the basic lemma 4.1 of GROENEBOOM (1996); see also JONGBLOED (2000). This basic method seems to go back to DONOHO and LIU (1987, 1991).

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 $g \in \mathcal{D}_k$ 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

$$\mathrm{MMR}_1(n, T, \mathcal{D}_{k,n}) = \inf_{t_n} \sup_{g \in \mathcal{D}_{k,n}} E_g |\hat{T}_n - Tg|.$$

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 (parameterized by $\tau > 0$) used here are Hellinger balls centered at g_0:

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

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. Lemma 5 is the basic tool for proving such a lower bound.

LEMMA 5. *Assume that there exists some subset $\{g_\epsilon : \epsilon > 0\}$ of densities in $\mathcal{D}_{k,n}$ such that, as $\epsilon \downarrow 0$,*

$$H^2(g_\epsilon, g_0) \leq \epsilon(1 + o(1)) \ \text{ and } \ |Tg_\epsilon - Tg_0| \geq (c\epsilon)^r (1 + o(1))$$

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

$$\sup_{\tau > 0} \liminf_{n \to \infty} n^r \mathrm{MMR}_1(n, T, \mathcal{D}_{k,n,\tau}) \geq \frac{1}{4} \left(\frac{cr}{2e} \right)^r.$$

PROOF. See GROENEBOOM (1996) and JONGBLOED (2000).

Here is the main result of this section.

PROPOSITION 3. *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 *(*$k \geq 2$*). 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:*

$$\sup_{\tau > 0} \liminf_{n \to \infty} n^{\frac{k-j}{2k+1}} \mathrm{MMR}_1(n, T_j, \mathcal{D}_{k,n,\tau}) \geq \left\{ |g_0^{(k)}(x_0)|^{2j+1} g_0(x_0)^{k-j} \right\}^{1/(2k+1)} d_{k,j},$$

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

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

where

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

Proposition 3 also yields lower bounds for estimation of the corresponding mixing distribution function F at a fixed point.

COROLLARY 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* $T g_0 = F_0(x_0)$ *where* F_0 *is given in terms of* g_0 *by 2,*

$$\sup_{\tau > 0} \liminf_{n \to \infty} n^{1/2k+1} \mathrm{MMR}_1(n, T, \mathcal{D}_{k,n,\tau})$$

$$\geq \left\{ |g_0^{(k)}(x_0)|^{2k-1} g_0(x_0) \right\}^{1/(2k+1)} \frac{x_0^k}{k!} d_{k,k-1}.$$

PROOF. See the Appendix.

Both the rates of convergence $n^{(k-j)/(2k+1)}$ and the dependence of our lower bound on the constants $g_0(x_0)$ and $g_0^{(k)}(x_0)$ match with the known results for $k = 1$ and $k = 2$ owing to GROENEBOOM (1985) and GROENEBOOM *et al.* (2001), and reappears in the limit distribution theory for $k \geq 3$ in BALABDAOUI and WELLNER (2007).

Appendix

PROOF OF LEMMA 3. The arguments generalize those in the proof of lemma 2.4 of GROENEBOOM *et al.* (2001). If $\hat{g}_n$ is the MLE, let $g_t(x) = k(t-x)_+^{k-1}/t^k$ for some $t > 0$. For all $\epsilon > 0$, we have that $\hat{g}_n + \epsilon g_t \in \mathcal{M}_k \cap L_1(\lambda)$, and hence

$$\lim_{\epsilon \searrow 0} \frac{1}{\epsilon}(\psi_n(\hat{g}_n + \epsilon g_t) - \psi_n(\hat{g}_n)) \leq 0 \iff \int_0^\infty \frac{k(t-x)_+^{k-1}/t^k}{\hat{g}_n(x)} \, \mathrm{d}\mathbb{G}_n(x) - 1 \leq 1$$

yielding the inequality in Equation 3. If $t \in \{\hat{a}_1, \cdots, \hat{a}_m\}$, then for $\epsilon \in \mathbb{R}$ such that $|\epsilon|$ is small enough, $\hat{g}_n + \epsilon g_t \in \mathcal{M}_k \cap L_1(\lambda)$, and hence

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon}(\psi_n(\hat{g}_n + \epsilon g_t) - \psi_n(\hat{g}_n)) = 0 \iff \int_0^\infty \frac{k(t-x)_+^{k-1}/t^k}{\hat{g}_n(x)} \, \mathrm{d}\mathbb{G}_n(x) - 1 = 0$$

yielding the identity in Equation 4.

Suppose now that $\hat{g}_n$ is a k-monotone spline of degree $k-1$ and with simple knots satisfying the condition in Equation 3, and let $g \in \mathcal{M}_k \cap L_1(\lambda)$. By lemma 1, there exists a non-decreasing and bounded function F on $(0, \infty)$ such that

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

We can write

$$\psi_n(\hat{g}_n) - \psi_n(g) = \int_0^\infty \log\left(\frac{\hat{g}_n(x)}{g(x)}\right) \mathrm{d}\mathbb{G}_n(x) - 1 + \int_0^\infty g(x) \, \mathrm{d}x$$

$$\geq \int_0^\infty \left(\frac{1 - g(x)}{\hat{g}_n(x)}\right) \mathrm{d}\mathbb{G}_n(x) - 1 + \int_0^\infty g(x) \, \mathrm{d}x,$$

using the inequality $\log z \geq 1 - 1/z, z > 0$

$$= -\int_0^\infty \frac{g(x)}{\hat{g}_n(x)} \, \mathrm{d}\mathbb{G}_n(x) + \int_0^\infty g(x) \, \mathrm{d}x$$

$$= -\int_0^\infty \int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k \hat{g}_n(x)} \, \mathrm{d}F(t) \, \mathrm{d}\mathbb{G}_n(x) + \int_0^\infty \mathrm{d}F(t)$$

$$= \int_0^\infty \left(-\int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k \hat{g}_n(x)} \, \mathrm{d}\mathbb{G}_n(x) + 1\right) \mathrm{d}F(t) \geq 0 \text{ by Equation 3.}$$

Hence, $\hat{g}_n$ is the MLE.

Establishing the characterization and structure of the LSE

In this optimization problem, existence requires more work because there is no available theory as in the case of the MLE. However, 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 Q_n 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 $s_{n,k}(t) \equiv \int_0^\infty (t - x)_+^{k-1} \, d\mathbb{G}_n(x)$. 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 $(0, \infty)$.

PROPOSITION A.1. *The functional Φ_n admits a unique minimizer $\tilde{\mu}$, and hence the LSE $\tilde{g}_n$ exists and is unique.*

PROOF. Uniqueness follows from strict convexity of Φ_n. To prove existence, it can be shown that Φ_n can be restricted to a subset $\mathcal{C}$ of $\mathcal{X}$ on which it is lower semicontinuous, and hence the minimization problem admits a solution by applying theorem 38.B of ZEIDLER (1985, p. 152). In the following, we will exhibit the subset $\mathcal{C}$, and show that the conditions of Zeidler's theorem are satisfied.

We begin by checking the hypotheses of Zeidler's theorem 38.B (ZEIDLER, p. 152 1985). We identity X of Zeidler's theorem with the space $\mathcal{X}$ of non-negative measures on $[0, \infty)$, and we show that we can take M of Zeidler's theorem to be

$$\mathcal{C} \equiv \{\mu \in \mathcal{X} : \mu(t, \infty) \leq D t^{-(k-1/2)}\}$$

for some constant $D < \infty$.

First, we can, without loss, restrict the minimization to the space of non-negative measures on $[X_{(1)}, \infty)$, where $X_{(1)} > 0$ is the first-order statistic of the data. To see this, note that we can decompose any measure μ as $\mu = \mu_1 + \mu_2$, where μ_1 is concentrated on $[0, X_{(1)})$ and μ_2 is concentrated on $[X_{(1)}, \infty)$. As the second term of Q_n is zero for μ_1, the contribution of the μ_1 component to $Q_n(\mu)$ is always non-negative, so we make $\inf Q_n(\mu)$ no larger by restricting to measures on $[X_{(1)}, \infty)$.

We can restrict further to measures μ with $\int_0^\infty t^{k-1} \, d\mu(t) \leq D$ for some finite $D = D_\omega$. To show this, we first give a lower bound for $r_k(s, t)$. For $s, t \geq t_0 > 0$ we have

$$r_k(s,t) \geq \frac{(1 - e^{-v_0})t_0}{2k} s^{k-1} t^{k-1}, \tag{A.1}$$

where $v_0 \approx 1.59$. To prove Equation A.1, we will use the inequality

$$(1 - v/k)^{k-1} \geq e^{-v}, \qquad 0 \leq v \leq v_0, \ k \geq 2. \tag{A.2}$$

[This inequality holds by straightforward computation; see HALL and WELLNER (1979), especially their proposition 2.]

Thus, we compute

$$
\begin{aligned}
r_k(s,t) &= \int_0^\infty (s-x)_+^{k-1} (t-x)_+^{k-1} \, \mathrm{d}x \\
&= \frac{1}{k} s^{k-1} t^{k-1} \int_0^\infty \left(1 - \frac{y}{sk}\right)_+^{k-1} \left(1 - \frac{y}{tk}\right)_+^{k-1} \, \mathrm{d}y \\
&\geq \frac{1}{k} s^{k-1} t^{k-1} \int_0^{v_0(t \wedge s)} e^{-y/s} e^{-y/t} \, \mathrm{d}y \\
&= \frac{1}{k} s^{k-1} t^{k-1} \frac{1}{c} \int_0^{v_0(t \wedge s)} c e^{-cy} \, \mathrm{d}y, \quad c \equiv 1/s + 1/t \\
&= \frac{1}{k} s^{k-1} t^{k-1} \frac{1}{c} \left(1 - \exp(-c(t \wedge s)v_0)\right) \\
&\geq \frac{1}{k} s^{k-1} t^{k-1} \frac{1}{c} \left(1 - \exp(-v_0)\right)
\end{aligned}
$$

as

$$c(s \wedge t) = \frac{s+t}{st}(s \wedge t) = \left\{ \begin{array}{ll} (t+s)/t, & s \leq t \\ (t+s)/s, & s \geq t \end{array} \right\} \geq 1.$$

But, we also have

$$\frac{1}{c} = \frac{1}{(1/s) + (1/t)} = \frac{st}{s+t} \geq \frac{1}{2} s \wedge t \geq \frac{1}{2} t_0$$

for $s, t \geq t_0$, so we conclude that Equation A.1 holds.

From the inequality A.1, we conclude that for measures μ concentrated on $[X_{(1)}, \infty)$ we have

$$\int \int r_k(s,t) \, \mathrm{d}\mu(s) \, \mathrm{d}\mu(t) \geq \frac{(1 - e^{-v_0})X_{(1)}}{2k} \left(\int_0^\infty t^{k-1} \, \mathrm{d}\mu(t) \right)^2.$$

In contrast,

$$\int_0^\infty s_{n,k}(t) \, \mathrm{d}\mu(t) \leq \int_0^\infty t^{k-1} \, \mathrm{d}\mu(t).$$

Combining these two inequalities it follows that for any measure μ concentrated on $[X_{(1)}, \infty)$ we have

$$
\begin{aligned}
\Phi_n(\mu) &= \frac{1}{2} \int \int r_k(t,s) \, \mathrm{d}\mu(t) \, \mathrm{d}\mu(s) - \int_0^\infty s_{n,k}(t) \, \mathrm{d}\mu(t) \\
&\geq \frac{(1-e^{-v_0})X_{(1)}}{4k} \left(\int_0^\infty t^{k-1} \, \mathrm{d}\mu(t) \right)^2 - \int_0^\infty t^{k-1} \, \mathrm{d}\mu(t) \\
&\equiv A m_{k-1}^2 - m_{k-1}.
\end{aligned}
$$

This lower bound is strictly positive if

$$
m_{k-1} > 1/A = \frac{4k}{(1-e^{-v_0})X_{(1)}}.
$$

But for such measures μ we can make Φ smaller by taking the zero measure. Thus, we may restrict the minimization problem to the collection of measures μ satisfying

$$
m_{k-1} \leq 1/A. \tag{A.3}
$$

Now we decompose any measure μ on $[X_{(1)}, \infty)$ as $\mu = \mu_1 + \mu_2$ where μ_1 is concentrated on $[X_{(1)}, MX_{(n)}]$ and μ_2 is concentrated on $(MX_{(n)}, \infty)$ for some (large) $M > 0$. Then, it follows that

$$
\begin{aligned}
\Phi_n(\mu) &\geq \frac{1}{2} \int \int r_k(t,s) \, \mathrm{d}\mu_2(t) \, \mathrm{d}\mu_2(s) - \int_0^\infty t^{k-1} \, \mathrm{d}\mu(t) \\
&\geq \frac{(1-e^{v_0})MX_{(n)}}{4k} (MX_{(n)})^{2k-2} \mu(MX_{(n)}, \infty)^2 - 1/A \\
&\equiv B\mu(MX_{(n)}, \infty)^2 - 1/A > 0
\end{aligned}
$$

if

$$
\mu(MX_{(n)}, \infty)^2 > \frac{1}{AB} = \frac{4k}{(1-e^{-v_0})X_{(1)}} \frac{4k}{(1-e^{-v_0})(MX_{(n)})^{2k-1}},
$$

and hence we can restrict to measures μ with

$$
\mu(MX_{(n)}, \infty) \leq \frac{4k}{(1-e^{-v_0})X_{(1)}^{1/2} X_{(n)}^{k-1/2}} \frac{1}{M^{k-1/2}}
$$

for every $M \geq 1$.

But this implies that μ satisfies

$$
\int_0^\infty t^{k-3/4} \, \mathrm{d}\mu(t) \leq D
$$

for some $0 < D = D_\omega < \infty$, and this implies that t^{k-1} is uniformly integrable over $\mu \in \mathcal{C}$.

Alternatively, for $\lambda \geq 1$ we have

$$\int_{t>\lambda} t^{k-1}\,\mathrm{d}\mu(t) = \lambda^{k-1}\mu(\lambda,\infty) + (k-1)\int_\lambda^\infty s^{k-2}\mu(s,\infty)-\,\mathrm{d}s$$

$$\leq \lambda^{k-1}\frac{K}{\lambda^{k-1/2}} + (k-1)\int_\lambda^\infty s^{k-2}Ks^{-(k-1/2)}\,\mathrm{d}s$$

$$= K\lambda^{-1/2} + (k-1)K\int_\lambda^\infty s^{-3/2}\,\mathrm{d}s$$

$$\leq K\lambda^{-1/2} + (k-1)2K\lambda^{-1/2}$$

$$\to 0 \quad \text{as } \lambda\to\infty$$

uniformly in $\mu\in\mathcal{C}$.

This implies that for $\{\mu_m\}\subset\mathcal{C}$ satisfying $\mu_m\Rightarrow\mu_0$ we have

$$\limsup_{m\to\infty}\int_0^\infty s_{n,k}(t)\,\mathrm{d}\mu_m(t) \leq \int_0^\infty s_{n,k}(t)\,\mathrm{d}\mu_0(t),$$

and hence Φ_n is lower semicontinuous on $\mathcal{C}$:

$$\liminf_{m\to\infty}\Phi_n(\mu_m)\geq\Phi_n(\mu_0).$$

As Q_n is lower semi-compact (i.e., the sets $\mathcal{C}_r\equiv\{\mu\in\mathcal{C}:\Phi_n(\mu)\leq r\}$ are compact for $r\in\mathbb{R}$), the existence of a minimum follows from ZEIDLER (1985, theorem 38.B, p. 152).

Lemma A.2 characterizes the LSE.

LEMMA A.1. *For $k\geq 1$ define $\mathbb{Y}_n$ and $\tilde{H}_n$ respectively by*

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

and

$$\tilde{H}_n(t) = \int_0^t\int_0^{t_k}\cdots\int_0^{t_2}\tilde{g}_n(t_1)\mathrm{d}t_1\mathrm{d}t_2\cdots dt_k = \int_0^t\frac{(t-x)^{k-1}}{(k-1)!}\tilde{g}_n(x)\mathrm{d}x$$

for $t\geq 0$. Then $\tilde{g}_n$ is the LSE over $\mathcal{M}_k\cap L_2(\lambda)$ if and only if the following conditions are satisfied:

$$\begin{cases} \tilde{H}_n(t)\geq\mathbb{Y}_n(t), & \text{for } t\geq 0, \quad \text{and} \\ \int_0^\infty(\tilde{H}_n-\mathbb{Y}_n)\,d\tilde{g}_n^{(k-1)}=0. \end{cases} \tag{A.4}$$

PROOF. The arguments are very similar to those used in the proof of lemma 3.

Now, to prove that the LSE is a spline of degree $k-1$ with simple knots, we need the following intermediate result.

PROPOSITION A.2. *Let $[a,b]]\subseteq(0,\infty)$ and let g be a non-negative and non-increasing function on $[a,b]$. For any polynomial P_{k-1} of degree $\leq k-1$ on $[a,b]$, if the function*

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

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 as g is assumed to be non-negative and non-increasing, 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 proposition to show that it is a finite mixture of beta$(1, k)$s. We know from lemma A.1 that $\tilde{g}_n$ is the LSE if and only if Equation A.4 holds. The equality condition in the second part of equation A.4 implies that $\tilde{H}_n$ and $\mathbb{Y}_n$ have to be equal at any point of increase of the monotone function $(-1)^{k-1}\tilde{g}_n^{(k-1)}$. Therefore, the set of points of increase of $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ is included in the set of zeros of the function $\tilde{\Delta}_n = \tilde{H}_n - \mathbb{Y}_n$.

Now, note that $\mathbb{Y}_n$ can be given by the explicit expression:

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

In other words, $\mathbb{Y}_n$ is a spline of degree $k-1$ with simple knots at $X_{(1)}, \ldots, X_{(n)}$, the order statistics of $X_1, \ldots, X_n$. Also note that the function $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ 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 $\mathcal{I} \subset (X_{(j)}, X_{(j+1)})$ (with $X_{(0)} = 0$ and $X_{(n+1)} = \infty$) such that $\mathcal{I}$ has a non-empty interior, and $\tilde{H}_n \equiv \mathbb{Y}_n$ on $\mathcal{I}$. This implies that $\tilde{H}_n^{(k)} \equiv \mathbb{Y}_n^{(k)} \equiv 0$, as $\mathbb{Y}_n$ is a polynomial of degree $k-1$ on $\mathcal{I}$, and hence $\tilde{g}_n \equiv 0$ on $\mathcal{I}$. But the latter is impossible as it was assumed that $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ was strictly increasing on I. Thus, the monotone function $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ can have only two components: discrete and singular. In the following, we will prove that it is actually discrete with finitely many points of jump.

PROPOSITION A.2. There exists $m \in \mathbb{N} \setminus \{0\}, \tilde{a}_1, \cdots, \tilde{a}_m$ and $\tilde{w}_1, \cdots, \tilde{w}_m$ such that for all $x > 0$, the LSE $\tilde{g}_n$ is given by

$$\tilde{g}_n(x) = \tilde{w}_1 \frac{k(\tilde{a}_1 - x)_+^{k-1}}{\tilde{a}_1^k} + \cdots + \tilde{w}_m \frac{k(\tilde{a}_m - x)_+^{k-1}}{\tilde{a}_m^k}. \tag{A.5}$$

Consequently, the equality part in Equation A.4 can be re-expressed as $\tilde{H}_n(t) = \mathbb{Y}_n(t)$ if t is a point in the support of the minimizing (mixing) measure $\tilde{F}_n$ (or a knot of $\tilde{g}_n$).

PROOF. We need to consider two cases:

(i) The number of zeros of $\tilde{\Delta}_n = \tilde{H}_n - \mathbb{Y}_n$ is finite. This implies by the equality condition in Equation A.4 that the number of points of increase of $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ is

also finite. Therefore, $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ is discrete with finitely many jumps and hence $\tilde{g}_n$ is of the form given in Equation A.5.

(ii) Now, suppose that $\tilde{\Delta}_n$ 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 proposition A.2, if t_j is the smallest zero of $\tilde{g}_n$ in $[X_{(j)}, X_{(j+1)}]$, then $\tilde{g}_n \equiv 0$ on $[t_j, X_{(j+1)}]$ and $\tilde{g}_n > 0$ on $[X_{(j)}, t_j)$ if $t_j > X_{(j)}$. Note that from the proof of proposition A.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 $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ has finitely many points of increase in $(0, X_{(j)})$. We also know that $\tilde{g}_n \equiv 0$ on $[t_j, \infty)$. Thus we only need to show that the number of points of increase of $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ 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 $\tilde{\Delta}_n$ that is strictly smaller than z_j. If $z_j < t_j$, then for the same reason, $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ has no point of increase in $[X_{(j)}, z_j)$. Finally, $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ cannot have infinitely many points of increase in $[z_j, t_j)$ because that would imply that $\tilde{\Delta}_n$ has infinitely many zeros in (z_j, t_j), and hence by lemma A.1, we can find $t'_j \in (z_j, t_j)$ such that $\tilde{g}_n \equiv 0$ on $[t'_j, t_j]$. But, this is impossible as $\tilde{g}_n > 0$ on $[X_{(j)}, t_j)$.

PROOF OF THE IDENTITY $\hat{g} = g_0$ (proof of Proposition 2). For $0 < \alpha < 1$ define $\eta_\alpha = G_0^{-1}(1-\alpha)$. Let $\epsilon > 0$ be small so that $\epsilon < \eta_\epsilon$.

By Equation 6, there exists a number $D_\epsilon > 0$ such that $\hat{g}_l(\eta_\epsilon) \geq D_\epsilon$ for sufficiently large l. To see this, note that Equation 6, implies that

$$1 \geq \int_0^\infty \frac{g_0(x)}{\hat{g}_l(x)}\, d\mathbb{G}_l(x) \geq \int_{\eta_\epsilon}^\infty g_0(x)\hat{g}_l(x)\, d\mathbb{G}_l(x) \geq \frac{1}{\hat{g}_l(\eta_\epsilon)} \int_{\eta_\epsilon}^\infty g_0(x)\, d\mathbb{G}_l(x),$$

and hence

$$\liminf_l \hat{g}_l(\eta_\epsilon) \geq \liminf_l \int_{\eta_\epsilon}^\infty g_0(x)\, d\mathbb{G}_l(x) = \int_{\eta_\epsilon}^\infty g_0(x)\, dG_0(x) > 0,$$

by the choice of η_ϵ, and the claim follows by taking $D_\epsilon = \int_{\eta_\epsilon}^\infty g_0(x)\, dG_0(x)/2$. Hence, by the bound in lemma 4, we have

$$\hat{g}_l(z) \leq \frac{1}{z}\left(1 - \frac{1}{k}\right)^{k-1} \equiv \frac{e_k}{z}, \quad g_0(z) \leq \frac{1}{z}\left(1 - \frac{1}{k}\right)^{k-1} \equiv \frac{e_k}{z}.$$

It follows that $g_0/\hat{g}_l$ is uniformly bounded on the interval $[\epsilon, \eta_\epsilon]$; that is, there exist two constants $\underline{c}_\epsilon$ and $\overline{c}_\epsilon$ such that for all $x \in [\epsilon, \eta_\epsilon]$

$$\underline{c}_\epsilon \leq \frac{g_0(x)}{\hat{g}_l(x)} \leq \overline{c}_\epsilon.$$

In fact,

$$\frac{g_0(x)}{\hat{g}_l(x)} \le \frac{g_0(\epsilon)}{\hat{g}_l(\eta_\epsilon)} \le \frac{\epsilon^{-1}e_k}{D_\epsilon},$$

while

$$\frac{g_0(x)}{\hat{g}_l(x)} \ge \frac{g_0(\eta_\epsilon)}{\hat{g}_l(\epsilon)} \ge \frac{g_0(\eta_\epsilon)}{\epsilon^{-1}e_k}$$

Therefore,

$$\frac{g_0(x)}{\hat{g}_l(x)} \to \frac{g_0(x)}{\hat{g}(x)}$$

uniformly on $[\epsilon, \eta_\epsilon]$. Using Equation 6, we have for sufficiently large l and

$$\int_\epsilon^{\eta_\epsilon} \frac{g_0(x)}{\hat{g}(x)} \, \mathrm{d}\mathbb{G}_l(x) \le \int_\epsilon^{\eta_\epsilon} \left(\frac{g_0(x)}{\hat{g}_l(x)} + \epsilon \right) \mathrm{d}\mathbb{G}_l(x) \le 1 + \epsilon.$$

But as $\mathbb{G}_l$ converges weakly to G_0 the distribution function of g_0 and $g_0/\hat{g}$ is continuous and bounded on $[\epsilon, \eta_\epsilon]$; we conclude that

$$\int_\epsilon^{\eta_\epsilon} \frac{g_0(x)}{\hat{g}(x)} \, \mathrm{d}G_0(x) \le 1 + \epsilon.$$

Now, by Lebesgue's monotone convergence theorem, we conclude that

$$\int_0^\infty \frac{g_0(x)}{\hat{g}(x)} \, \mathrm{d}G_0(x) \le 1,$$

which is equivalent to

$$\int_0^\infty \frac{g_0^2(x)}{\hat{g}(x)} \, \mathrm{d}x \le 1. \tag{A.6}$$

Define $\tau = \int_0^\infty \hat{g}(x) \, \mathrm{d}x$. Then, $\hat{h} = \tau^{-1}\hat{g}$ is a k-monotone density. By Equation A.6, we have that

$$\int_0^\infty \frac{g_0^2(x)}{\hat{h}(x)} \, \mathrm{d}x = \tau \int_0^\infty \frac{g_0^2(x)}{\hat{g}(x)} \, \mathrm{d}x \le \tau.$$

Now, consider the function

$$K(g) = \int_0^\infty \frac{g_0^2(x)}{g(x)} \, \mathrm{d}x$$

defined on the class $\mathcal{C}_d$ of all continuous densities g on $[0, \infty)$. Minimizing K is equivalent to minimizing

$$\int_0^\infty \left(\frac{g_0^2(x)}{g(x)} + g(x) \right) \mathrm{d}x.$$

It is easy to see that the integrand is minimized pointwise by taking $g(x) = g_0(x)$. Hence, $\inf_{\mathcal{C}_d} K(g) \ge 1$. In particular, $K(\hat{h}) \ge 1$ which implies that $\tau = 1$.

Now, if $g \ne g_0$ at a point x, it follows that $g \ne g_0$ on an interval of positive length. Hence, $g_0 \ne g \Rightarrow K(g) > 1$. We conclude that we have necessarily $\hat{h} = \hat{g} = g_0$.

PROPOSITION A.4. Fix $c > 0$ and suppose that the true k-monotone density g_0 satisfies $\int_0^\infty x^{-1/2} g_0(x)\,\mathrm{d}x < \infty$. Then, $\| \tilde{g}_n - g_0 \|_2 \to_{\text{a.s.}} 0$,

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

for $j = 0, 1, \cdots, k-2$, and, for each $x > 0$ at which g_0 is $(k-1)$-times differentiable, $\tilde{g}_n^{(k-1)}(x) \to_{\text{a.s.}} g_0^{(k-1)}(x)$. Here, $\| \cdot \|_2$ denotes the L_2-norm.

PROOF. The main difficulty here is that the LSE $\tilde{g}_n$ is not necessarily a density in that it may integrate to more than one; indeed it can be shown that $\int_0^\infty \tilde{g}_1(x)\,\mathrm{d}x = ((2k-2)/k)(1 - 1/(2k-1))^{k-2} > 1$ 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 $k = 2$; that is, consistency of the LSE of a convex and decreasing density (see GROENEBOOM *et al.*, 2001). 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 $\int_0^\infty \tilde{g}_n(x)\,\mathrm{d}x = O_p(1)$.

Here we will first show that $\int_0^\infty \tilde{g}_n^2 \,\mathrm{d}\lambda = O_p(1)$. Note that the equality part in Equation (A.4) can be re-written as $\int_0^\infty \tilde{g}_n^2(x)\,\mathrm{d}x = \int_0^\infty \tilde{g}_n(x)\,\mathrm{d}\mathbb{G}_n(x)$ and hence

$$\sqrt{\int_0^\infty \tilde{g}_n^2(x)\,\mathrm{d}x} = \int_0^\infty \tilde{u}_n(x)\,\mathrm{d}\mathbb{G}_n(x), \tag{A.7}$$

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

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

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

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

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

$$\int_0^\infty \tilde{u}_n(x)\,\mathrm{d}\mathbb{G}_n(x) \leq \int_0^\infty G(x)\,\mathrm{d}\mathbb{G}_n(x) \to_{\text{a.s.}} \int_0^\infty G(x)\,\mathrm{d}G_0(x), \quad \text{as } n \to \infty$$

and hence by Equation A.7 the integral $\int_0^\infty \tilde{g}_n^2 \,\mathrm{d}\lambda$ 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 $\int_0^\infty \tilde{g}_n^2 \mathrm{d}\lambda \leq M_k$. We have

$$\tilde{g}_n(\tau_n) \leq \tilde{g}_n(\delta/2) = (2/\delta)(\delta/2)\tilde{g}_n(\delta/2) \leq (2/\delta)\int_0^{\delta/2}\tilde{g}_n(x)\mathrm{d}x$$

$$\leq (2/\delta)\sqrt{\delta/2}\sqrt{\int_0^{\delta/2}\tilde{g}_n^2(x)\mathrm{d}x} \leq \sqrt{2/\delta}\sqrt{\int_0^\infty\tilde{g}_n^2(x)\mathrm{d}x}$$

$$= \sqrt{2M_k/\delta}. \tag{A.8}$$

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

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

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

$$\sqrt{\delta M_k} \geq \int_{\tau_n}^\delta \tilde{g}_n(x)\mathrm{d}x = \tilde{g}_n(\delta)(\delta - \tau_n) - \frac{\tilde{g}_n'(\delta)}{2}(\delta - \tau_n)^2 + \cdots + (-1)^{k-1}\frac{\tilde{g}_n^{(k-1)}(\delta)}{k!}(\delta - \tau_n)^k$$

$$\geq (\delta - \tau_n)\left(\tilde{g}_n(\delta) + \frac{1}{k}(-1)\tilde{g}_n'(\delta)(\delta - \tau_n) + \cdots + (-1)^{k-1}\frac{\tilde{g}_n^{(k-1)}(\delta)}{(k-1)!}(\delta - \tau_n)^{k-1}\right)$$

$$= (\delta - \tau_n)\left(\tilde{g}_n(\delta)\left(1 - \frac{1}{k}\right) + \frac{1}{k}\tilde{g}_n(\tau_n)\right) \geq \frac{\delta}{2k}\tilde{g}_n(\tau_n)$$

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, as $\tilde{g}_n(\delta) \leq \tilde{g}_n(\tau_n)$, 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

$$\int_\delta^\infty (\tilde{g}(x) - g_0(x))^2\mathrm{d}x \leq \liminf_{l\to\infty}\int_\delta^\infty(\tilde{g}_{n_l}(x) - g_0(x))^2\mathrm{d}x. \tag{A.9}$$

However, the characterization of $\tilde{g}_n$ implies that $Q_n(\tilde{g}_n) \leq Q_n(g_0)$, and this yields

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

Thus, we can write

$$\int_\delta^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2\mathrm{d}x \leq \int_0^\infty(\tilde{g}_{n_l}(x) - g_0(x))^2\mathrm{d}x$$

$$\leq 2\int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x))\mathrm{d}(\mathbb{G}_{n_l}(x) - G_0(x)) \to_{\text{a.s.}} 0, \tag{A.10}$$

as $l \to \infty$. The last convergence is justified as follows: as $\int_0^\infty \tilde{g}_n^2\mathrm{d}\lambda$ 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 as the class of functions $\{(g - g_0)1_{[G \leq M]} : g \in \mathcal{M}_k \cap L_2(\lambda)\}$ is a Glivenko-Cantelli class for every $M > 0$ (each

element is a difference of two bounded monotone functions) Equation A.10 holds. From Equation A.9, we conclude that $\int_\delta^\infty (\tilde{g}(x) - g_0(x))^2 \mathrm{d}x \le 0$, and therefore, $\tilde{g} \equiv g_0$ on $(0, \infty)$ as $\delta > 0$ 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 the convexity of $(-1)^j \tilde{g}_n^{(j)}$ for $j = 1, \dots, k - 2$.

PROOF OF PROPOSITION 3. Let μ be a positive number and consider the function $\tilde{g}_\mu$ defined by

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

Now, consider the perturbation

$$g_\mu(x) = g_0(x) + s(\mu)\tilde{g}_\mu(x), \quad x \in (0, \infty),$$

where $s(\mu)$ is a scale to be determined later. 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

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

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

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

To compute $r^{(j)}(0)$, note that for $m \ge 2$ and $2n \ge m$ we have

$$\begin{aligned}
\left((1 - x^2)^n\right)^{(m)} &= \left(((1 - x^2)^n)'\right)^{(m-1)} \\
&= \left(-2nx(1 - x^2)^{n-1}\right)^{(m-1)} \\
&= -2n \left(x((1 - x^2)^{n-1})^{(m-1)} + (m-1)((1 - x^2)^{n-1})^{m-2}\right),
\end{aligned}$$

where in the last equality we used Leibniz's formula for the derivatives of a product; see, for example, APOSTOL (1957, p. 99). Evaluating the last expression at $x = 0$ yields

$$x_{n,m} \equiv \left((1-x^2)^n\right)^{(m)} \Big|_{x=0} = -2n(m-1)x_{n-1,m-2}.$$

If m is even, we find that

$$x_{n,m} = (-2)^{m/2} \prod_{i=0}^{m/2-1}(n-i) \times \prod_{i=0}^{m/2-1}(m-2i-1) \times x_{n-m/2,0}$$

$$= (-2)^{m/2} \prod_{i=0}^{m/2-1}(n-i) \times \prod_{i=0}^{m/2-1}(m-2i-1)$$

as $x_{n-m/2,0} = 1$. Similarly, when m is odd,

$$x_{n,m} = (-2)^{(m-1)/2} \prod_{i=0}^{(m-1)/2-1}(n-i) \times \prod_{i=0}^{(m-1)/2-1}(m-2i-1) \times x_{n-(m-1)/2,1}$$

$$= 0$$

as $x_{n-(m-1)/2,1} = 0$. Now we have, for $1 \le j \le k$,

$$r^{(j)}(x) = \left((1-x^2)^{k+1}(1+x)\right)^{(j)}$$

$$= (x+1)\left((1-x^2)^{k+1}\right)^{(j)} + j\left((1-x^2)^{k+1}\right)^{(j-1)}$$

and hence

$$r^{(j)}(0) = \left((1-x^2)^{k+1}\right)^{(j)} \Big|_{x=0} + j\left((1-x^2)^{k+1}\right)^{(j-1)} \Big|_{x=0}.$$

Therefore, when j is even the second term vanishes and

$$r^{(j)}(0) = (-2)^{j/2} \prod_{i=0}^{j/2-1}(k+1-i) \times \prod_{j=0}^{j/2-1}(j-2i-1) \ne 0.$$

When j is odd, the first term vanishes and

$$r^{(j)}(0) = (-2)^{(j-1)/2} \prod_{i=0}^{(j-1)/2-1}(k+1-i) \times \prod_{i=0}^{(j-1)/2}(j-2i) \ne 0.$$

Summarizing, we have shown 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

$$C_{k,k} = \begin{cases} 2 \times (-1)^{k/2}(k+1)(k-1)!\binom{k}{k/2-1} & \text{if } k \text{ is even} \\ (-1)^{(k-1)/2}k!\binom{k+1}{(k-1)/2} & \text{if } k \text{ is odd.} \end{cases} \tag{A.11}$$

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

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

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,

$$\int_0^\infty \frac{(g_\mu(x) - g_0(x))^2}{g_0(x)} dx = \left(\frac{\left(g_0^{(k)}(x_0) \right)^2}{g_0(x_0)} \frac{\int_{-1}^1 (1-z^2)^{2(k+1)}(z+1)^2 dz}{(C_{k,k})^2} \right) \mu^{2k+1} + o(\mu^{2k+2})$$

as $\mu \searrow 0$. This gives control of the Hellinger distance as well in view of JONGBLOED (2000, lemma 2, p. 282), or JONGBLOED (1995, corollary 3.2, pp. 30–31). We set

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

$$= \begin{cases} 2^{4(k+1)} \frac{(2k+3)(k+2)}{(k+1)^2} \frac{((2(k+1)))^2}{(4k+7)((k-1))^2 \left(\binom{k}{k/2-1}l \right)^2} & k \text{ even} \\ 2^{4(k+2)}(2k+3)(k+2) \frac{((2(k+1)))^2}{(4k+7)(k)^2 \left(\binom{k+1}{(k-1)/2}l \right)^2} & k \text{ odd.} \end{cases}$$

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 $0 \leq j \leq k-1$, the modulus of continuity, m_j, of the functional T_j satisfies

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

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

$$\sup_{\tau>0} \liminf_{n\to\infty} n^{(k-j)/(2k+1)} \mathrm{MMR}_1(n, T_j, \mathcal{D}_{k,n,\tau})$$

$$\geq \frac{1}{4} \left(4\frac{k-j}{2k+1}e^{-1} \right)^{(k-j)/(2k+1)} (r_{k,j})^{(k-j)/(2k+1)}, \tag{A.12}$$

which can be rewritten as

$$\sup_{\tau>0} \liminf_{n\to\infty} n^{(k-j)/(2k+1)} \mathrm{MMR}_1(n, T_j, \mathcal{D}_{k,n,\tau})$$

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

for $j = 0, \cdots, k - 1$. Finally, note that the fact that the function g_μ is not exactly a density will not affect the obtained constants its integral converges to 1 as $\mu \to 0$.

PROOF OF COROLLARY 1. Let $G_\mu(x) = \int_0^x g_\mu(t) \mathrm{d}t$. Using the inversion formula in Equation 2, we have

$$T(g_\mu) - T(g_0) = G_\mu(x_0) - G_0(x_0) + \sum_{j=1}^{k} (-1)^j \frac{x_0^j}{j!} (T_{j-1}(g_\mu) - T_{j-1}(g_0)).$$

For $j = 1, \ldots, k$, we have already established before that $|T_{j-1}(g_\mu) - T_{j-1}(g_0)| = \lambda_{k,1}^{(j)} |g_0^{(k)}(x_0)| \mu^{k-j}$. In constrast, we have for $\mu > 0$ small enough

$$\begin{aligned}
G_\mu(x_0) &= G_0(x_0) + s(\mu) \int_{x_0-\mu}^{x_0+\mu} (x_0 + \mu - x)^{k+1} (x - x_0 + \mu)^{k+2} \mathrm{d}x \\
&= G_0(x_0) + s(\mu) \int_{x_0-\mu}^{x_0+\mu} (x - x_0 + \mu)^{k+1} (x - x_0 - \mu)^{k+1} \mathrm{d}x \\
&= G_0(x_0) + s(\mu) \int_{-1}^{1} r(x) \mathrm{d}x \mu^{2k+4} \\
&= G_0(x_0) + \frac{g_0^{(k)}(x_0)}{\mu^{k+3} r^{(k)}(0)} \left(\int_{-1}^{1} r(x) \mathrm{d}x \right) \mu^{2k+4} \\
&= G_0(x_0) + O(\mu^{k+1}).
\end{aligned}$$

Hence,

$$|T(g_\mu) - T(g_0)| = \frac{x_0^k}{k!} \lambda_{k,1}^{(k-1)} |g_0^{(k)}(x_0)| \mu + o(\mu).$$

Using again the change of variable $\epsilon = \mu^{2k+1}(b_k + o(1))$, we obtain the claimed lower bound in the same way as in proproition 3.

Acknowledgements

References

ANEVSKI, D. (1994), *Estimating the derivative of a convex density*, Technical Report 1994:8, Department of Mathematical Statistics, University of Lund.

ANEVSKI, D. (2003), Estimating the derivative of a convex density, *Statistica Neerlandica* **57**, 245–257.

Apostol, T. M. (1957), *Mathematical analysis: a modern approach to advanced calculus*, Addision-Wesley Publishing Company, Inc., Reading, MA.

Ayer M., H. D. Brunk, G. M. Ewing, W. T. Reid and E. Silverman (1955), An empirical distribution function for sampling with incomplete information, *The Annals of Mathematcial Statistics* **26**, 641–647.

Balabdaoui, F. (2007), Consistent estimation of a convex density at the origin, *Mathematical Methods of Statistics* **16**, 77–95.

Balabdaoui, F. and J. A. Wellner (2004), *Estimation of a K-monotone density, part 2: algorithm for computation and numerical results*, Technical Report 460, Department of Statistics, University of Washington.

Balabdaoui, F. and J. A. Wellner (2007), Estimation of a K-monotone density: Limit distribution theory and the spline connection, *The Annals of Statistics* **35**, 2536–2564.

Brunk H. D. (1981), *Probability theory and elements of measure theory*, Academic Press Inc. [Harcourt Brace Jovanovich Publisher], London. Second edition of the translation by R. B. Burckel from the third German edition, Probability and Mathematical Statistics.

Brunk H. D. (1958), On the estimation of parameters restricted by inequalities. *Annals of Mathematical Statistics* **29**, 437–454.

De Boor, C. (1978), *A practical guide to splines, vol. 27 of Applied Mathematical Science*, Springer-Verlag, New York.

DeVore, R. A. and G. G. Lorentz (1993), *Constructive approximation, vol. 303 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, Springer-Verlag Berlin.

Donoho, D. L. and R. C. Liu (1987), *Geometrizing rates of convergence. i*, Technical Report. 137, Department of Statistics, University of Califonia, Berkeley.

Donoho, D. L. and R. C. Liu (1991), Geometrizing rates of convergence. II III, *The Annals of Statistics* **19**, 633–667, 668–701.

van Eeden, C. (1957a), Maximum likelihood estimation of partially or completely ordered parameters. I, *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings. Series A Mathematical Sciences* **19**, 128–136.

van Eeden, C. (1957b), Maximum likelihood estimation of partially or completely ordered parameters. II, *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings. Series A Mathematical Sciences* **19**, 201–211.

Feller, W. (1971), *An introduction to probability theory and its applications*, vol. II, 2nd edition, John Wiley & Sons Inc., New york.

van de Geer, S. (1993), Hellinger-consistency of certain nonparametric maximum likelihood estimators, *The Annals of Statistics* **21**, 14–44.

Gneiting, T. (1998), On the Bernstein–Hausdorff–Widder conditions for completely monotone function. *Expositionces Mathaticae* **16**, 88–119.

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

Grenander, U. (1956a), On the theory of mortality measurement. I *Skandinavisk Aktuarietidskrift* **39**, 70–96.

Grenander, U. (1956b), On the theory of mortality measurement. II, *Skandinavisk Aktuarietidskrift* **39**, 125–153.

Groeneboom, P. (1985), Estimating a monotone density, in: *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer Vol. II* (Berkeley, CA. 1983), Wadsworth Statist./Probab. Ser., Wadsworth, Belmont, CA.

Groeneboom, P. (1989), Brownian motion with a parabolic drift and Airy functions, *Probability Theory and Related Fields* **81**, 79–109.

Groeneboom, P. (1996), Lectures on inverse problems, in: *Lectures on probability theory and statistics (Saint-Flour 1994)*, vol. 1648 *of Lecture Notes in Mathematics*, Springer, Berlin, pp. 67–164.

Groeneboom, P., G. Jongbloed and J. A. Wellner (2001), Estimation of a convex function: characterizations and asymptotic theory, *The Annals of Statistics* **29**, 1653–1698.

GROENEBOOM, P., G. JONGBLOED and J. A. WELLNER (2008), The support reduction algorithm for computing non-parametric function estimates in mixture models, *Scandinavian Journal of Statistics* **35**, 385–399.

HALL, W. J. and J. A. WELLNER (1979), The rate of convergence in law of the maximum of an exponential sample, *Statistica Neerlandica* **33**, 151–154.

HAMPEL, F. R. (1987). Design modelling and anlysis of some biological datasets, in: *design, data and analysis, by some friends of Cuthbert Daniel* C. MALLOWS (ed.), Wiley, New york, pp. 111–115.

JEWELL, N. P. (1982), Mixtures of exponential distributions, *The Annals of Statistics* **10**, 479–484.

JONGBLOED, G. (1995), *Three statistical inverse problems*, Ph.D. Thesis, Delft University of Technology.

JONGBLOED, G. (2000), Minimax lower bounds and moduli of continuity, *Statistics & Probability Letters* **50**, 279–284.

KIM, J. and D. POLLARD (1990), Cube root asymptotics, *The Annals of Statistics* **18**, 191–219.

LÉVY, P. (1962) Extensions d'un théorème de D. Dugué et M, Girault. *Probability Theory and Related Fields* **1**, 159–173.

LINDSAY, B. G. (1983), The geometry of mixture likelihoods: a general theory, *The Annals of Statistics* **11**, 86–94.

PFANZAGL, J. (1988), Consistency of maximum likelihood estimators for certain nonparametric families in particular: mixtures, *Journal of Statistical Planning and Inference* **19**, 137–158.

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

SILVERMAN, B. W. (1982), On the estimation of a probability density function by the maximum penalized likelihood method, *The Annals of Statistics* **10**, 795–810.

VAN DER VAART, A. and J. A. WELLNER (2000), Preservation theorems for Glivenko-Cantelli and unifrom Gilvenko-Cantelli classes, in: *Hign dimensional probability, II (Seattle, WA, 1999), vol. 47 of Progr. Probab.*, BirKhäuser Boston, Boston, MA, pp. 115–133.

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

WILLIAMSON R. E. (1956), Multiply monotone functions and their Laplace transforms, *Duke Mathematical Journal* **23**, 189–207.

WOODROOFE, M. and J. SUN (1993), A penalized maximum likelihood estimate of f(0+) when f is nonincreasing, *Statistics Sinica* **3**, 501–515.

ZEIDLER, E. (1985), *Nonlinear functional analysis and its applications. III*, Springer-Verlag, New york: Variational methods and optimization Translated from German by Leo F. Boron.