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

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

*University of Washington*

*November 1, 2004*

**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 and only 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. It is well-known that a density is completely monotone if and only if it is a scale mixture of exponential densities (Bernstein's theorem). Thus one motivation for studying the problem of estimation of a $k-$monotone density is to try to gain insight into the problem of estimating a completely monotone density.

In this series of four papers we consider both (nonparametric) Maximum Likelihood estimators and Least Squares estimators of a $k-$monotone estimator. In this first part (part 1), we prove existence of the estimators and give careful 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

[1] Research supported in part by National Science Foundation grant DMS-0203320

[2] Research supported in part by National Science Foundation grants DMS-0203320, and NIAID grant 2R01 AI291968-04

*AMS* 2000 *subject classifications.* Primary: 62G05; secondary 60G15, 62E20.

*Key words and phrases.* completely monotone densities, $k$-monotone densities, minimax risk, mixture models, nonparametric estimation, rates of convergence, shape constraints

bounds for estimating a $k-$monotone density $g_0(x_0)$ and its derivatives $g_0^{(j)}(x_0)$ at a fixed point $x_0$ under the assumption that $(-1)^k g_0^{(k)}(x_0) > 0$.

Part 2 of the series gives algorithms for computation of the estimators and an application of the methods to earthquake aftershock data. In part 3 we describe and establish existence of the limiting process $H_k$ which governs the asymptotic distribution theory modulo a certain conjecture involving a Hermite interpolation problem. In part 4 we give the limiting distribution theory in terms of $H_k$, again modulo the same Hermite interpolation problem.

# Outline

# 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), Anevski (2003), and Lavee, Safrie, and Meilijson (1991).

Estimation of monotone densities 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), Birgé (1987), Birgé (1989), and Kim and Pollard (1990).

Estimation of convex densities was apparently initiated by Anevski (1994) (see also Anevski (2003)), and was pursued by Wang (1994) and 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 here (and in the accompanying papers Balabdaoui and Wellner (2004a), Balabdaoui and Wellner (2004b), and Balabdaoui and Wellner (2004c)) 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 and only 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 Beta$(1,k)$ densities; i.e. with $x_+ \equiv x1\{x \geq 0\}$,

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

for some distribution function $F$ on $(0,\infty)$. Note that for $k=1$ this recovers the well known fact that monotone densities are in a one-to-one correspondence with scale mixtures of uniform densities, and, for $k=2$, the corresponding fact frequently used by Groeneboom, Jongbloed, and Wellner (2001b) that convex decreasing densities are in a one-to-one correspondence with scale mixtures of the triangular, or 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 will be discussed further in Balabdaoui and Wellner (2004d). 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 $(-1)^l g^{(l)}(x) \ge 0$ 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; i.e.

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

for some distribution function $F$ on $(0,\infty)$. Completely monotone densities arise naturally in connection with mixtures of Poisson processes and have been used in reliability theory (see e.g. Harris and Singpurwalla (1968), Doyle, Hansen, and McNolty (1980), Hill, Saunders, and Laud (1980)), and empirical Bayes estimation (see Robbins (1964) and Robbins (1980)). 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 $F$ is unique and 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 $\hat{g}_n$ of the mixed density $g$ in $\mathcal{D}_\infty$ (the "forward" or "direct" problem) or the estimator $\hat{F}_n$ of the mixing distribution function $F$ (the "inverse" problem). Although our present methods do not yield solutions of these difficult questions, the development of methods and theory for general $k-$monotone densites may throw some light on the issues and problems.

Now we briefly describe the contents of the four related papers of which the present manuscript is part 1. In this paper (part 1), we consider the Maximum Likelihood $\hat{g}_n$ and Least Squares $\tilde{g}_n$ estimators of a density $g_0 \in \mathcal{D}_k$ for a fixed integer $k \ge 2$ 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 $\hat{g}_n^{(j)}$ and $\tilde{g}_n^{(j)}$ 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. In part 1 we also include statements of known results for estimation of a completely monotone density $g_0 \in \mathcal{D}_\infty$.

In part 2 (Balabdaoui and Wellner (2004a)) we provide algorithms for computation of the estimators and for computation of (approximations to) the limit process $H_{c,k}$ defined in part 3 (Balabdaoui and Wellner (2004b)). We call the basic algorithm developed and used in part 2 an *iterative* $(2k-1)-$*spline algorithm* since it extends the "cubic spline algorithm" developed in Groeneboom, Jongbloed, and Wellner (2001a) and Groeneboom, Jongbloed, and Wellner (2003). Part 3 is devoted to a study of the corresponding canonical Gaussian problem and the "invelope" ($k$ even) or envelope ($k$ odd) processes $H_k = \lim_{c\to\infty} H_{c,k}$ which arise in the solution of the Gaussian version of the problem. Thus part 3 extends and is analogous to the treatment for the case $k = 2$ given by GROENEBOOM, JONGBLOED, AND WELLNER (2001A). Finally, part 4 (BALABDAOUI AND WELLNER (2004C)) gives joint asymptotic distribution theory at a fixed point $x_0 \in (0,\infty)$ of the vector of centered and scaled derivative estimators

$$\left(n^{(k-j)/(2k+1)}(\overline{g}_n^{(j)}(x_0) - g_0^{(j)}(x_0)), \quad j = 0,\dots,k-1\right)$$

where $\overline{g}_n$ is either the MLE $\hat{g}_n$ or the LSE $\tilde{g}_n$, under the assumption that $g_0^{(k)}(x_0)$ exists and is non-zero. This yields behavior of the corresponding estimators of the mixing distribution $F_0$ at fixed points (the inverse problem) as a corollary.

Thus the main outcome of parts 3 and 4 generalizes the asymptotic distribution theory for estimating a nondecreasing density, and a nondecreasing and convex density at a fixed point: If $x_0 > 0$ and $g_0$ is a $k$-monotone density defined on $(0,\infty)$ such that $g_0$ is $k$-times differentiable at $x_0$ with $(-1)^k g_0^{(k)}(x_0) > 0$, and $g_0^{(k)}$ is assumed to be continuous in a neighborhood of $x_0$, then our goal in parts 3 and 4 is to show that

$$\begin{pmatrix} n^{\frac{k}{2k+1}}(\bar{g}_n(x_0) - g_0(x_0)) \\ n^{\frac{k-1}{2k+1}}(\bar{g}_n^{(1)}(x_0) - g_0^{(1)}(x_0)) \\ \vdots \\ n^{\frac{1}{2k+1}}(\bar{g}_n^{(k-1)}(x_0) - g_0^{(k-1)}(x_0)) \end{pmatrix} \to_d \begin{pmatrix} c_0(g_0)H_k^{(k)}(0) \\ c_1(g_0)H_k^{(k+1)}(0) \\ \vdots \\ c_{k-1}(g_0)H_k^{(2k-1)}(0) \end{pmatrix}$$

and

$$n^{\frac{1}{2k+1}}(\bar{F}_n(x_0) - F(x_0)) \to_d \frac{(-1)^k x_0^k}{k!} c_{k-1}(g_0) H_k^{(2k-1)}(0),$$

where $\bar{g}_n$ is either the MLE of LSE, $\bar{F}_n$ is the corresponding estimator of the mixing distribution function $F_0$,

$$c_j(g_0) = \left\{ (g_0(x_0))^{k-j} \left( \frac{(-1)^k g_0^{(k)}(x_0)}{k!} \right)^{2j+1} \right\}^{\frac{1}{2k+1}},$$

for $j = 0,\cdots,k-1$, and $H_k$ is an almost surely uniquely defined stochastic process that is $(2k)$-convex (i.e., $H_k^{(2k-2)}$ exists and convex), and stays above (below) the $(k-1)$-fold integral of two-sided Brownian motion plus a polynomial drift of the form $t^{2k}/(2k)!$ if $k$ is even (odd). Only a change of scale is necessary to realize that $H_1$ and $H_2$ are very closely related to the greatest convex minorant of $W(t) + t^2, t \in \mathbb{R}$, where $W$ is two-sided Brownian motion, and the "invelope", $H$, of

$$\begin{cases} \int_0^t W(s)ds + t^4, & \text{if } t \ge 0 \\ \int_t^0 W(s)ds + t^4, & \text{if } t < 0. \end{cases}$$

Deriving the rate of convergence of both the estimators $\hat{g}_n$ and $\tilde{g}_n$ and their derivatives $\hat{g}_n^{(j)}$, $\tilde{g}_n^{(j)}$, $j = 1, \cdots, k-1$, and proving the existence of the stochastic processes $H_k$ for $k > 2$ involved in the joint asymptotic distribution still depends on a key conjecture: that the distance between two successive knots of the MLE or LSE that are in the neighborhood of $x_0$ is $O_p(n^{-1/(2k+1)})$ as the sample size $n \to \infty$, and that distance between two successive points of touch between the $(k-1)$-fold integral of two-sided Brownian motion plus $t^{2k}/(2k)!$ and $H_k$ is $O_p(1)$. Both problems are of the same nature and one can go from the first to the second one via a simple scaling argument. We refer to this common problem as the *gap problem*.

We will show in parts 3 and 4 that the gap problem can be reduced to the solution of a certain problem related to Hermite interpolation. That is, the gap problem has a solution if the following conjecture involving Hermite interpolation is true: Consider Hermite interpolation (as described for example in Nürnberger (1989), pages 108-109 or DeVore and Lorentz (1993) pages 161 - 162) of some smooth function $f$ via splines of odd-degree. More specifically, if $f$ is some real-valued function in $C^{(j)}[0,1]$ for some $j \geq 2$, $0 = t_0 < t_1 < \cdots < t_{2k-4} < t_{2k-3} = 1$ is a given increasing sequence, then the uniquely defined spline $Hf$ of degree $2k-1$ and interior knots $t_1, \ldots, t_{2k-4}$ satisfying the $4k-4$ conditions

$$(Hf)(t_i) = f(t_i), \quad \text{and} \quad (Hf)'(t_i) = f'(t_i), \quad i = 0, \ldots, 2k-3,$$

then we conjecture that there exists a constant $c_{k,j}$ depending only on $k$ and $j$ such that

$$\sup_{t_1, \cdots, t_{2k-4} \in (0,1)} \|f - Hf\|_\infty \leq c_{k,j} \, \|f^{(j)}\|_\infty,$$

where $\|\cdot\|_\infty$ is the supremum norm over $[0,1]$.

This Hermite interpolation problem has apparently not been investigated in detail in the spline or approximation theory literature, and hence an analysis of the corresponding interpolation error is yet to be developed. It is, however, precisely the interpolation problem involved in understanding our least squares estimators, both for finite sample sizes and in the limiting Gaussian problem: as will be shown in parts 3 and 4, the connecting link is the classical theorem of Schoenberg and Whitney (Schoenberg and Whitney (1953)) and its generalization by Karlin and Ziegler (Karlin and Ziegler (1966)); see Nürnberger (1989) page 109, or DeVore and Lorentz (1993), page 162.

However, the approximation theory literature has considered a related conjecture for another Hermite problem whose solution is a different odd-degree spline, also called a *complete spline*. Given a function $f \in C^{(k-1)}[0,1]$, and an increasing sequence $0 = t_0 < t_1 < \cdots < t_m < t_{m+1} = 1$, the complete spline interpolant, $Cf$, of degree $2k-1$ with interior knots $t_1, \cdots, t_m$ satisfies the $2k+m$ conditions

$$\begin{cases} (Cf)(t_i) = f(t_i), \quad i = 1, \cdots, m \\ (Cf)^{(l)}(t_0) = f^{(l)}(t_0), \ (Cf)^{(l)}(t_{m+1}) = f^{(l)}(t_{m+1}), \quad l = 0, \cdots, k-1. \end{cases}$$

When $f$ is in $C^{(2k)}[0,1]$, the error in this more "common" Hermite problem is known to be uniformly bounded independently of the location of the knots. This result was first conjectured by DE BOOR (1973) in (1972) for $k > 4$ since it was known to be true for $k = 2, 3, 4$ (DE BOOR (1974)), DE BOOR (1976)). The problem remained unsolved for more than 25 years: SHADRIN (2001) presents a proof of de Boor's conjecture. Thus there is a closely related interpolation problem in which the interpolation error does hold uniformly in the knots, and this gives some hope that "uniformity in the knots" will hold in our problem as well.

In our Hermite interpolation problem, the spline interpolant matches not only the value of the function at the knots but also the value of its first derivative. So intuitively, one should expect our spline to "behave better" than the complete spline, and the interpolation error to be smaller. On the other hand, our conjecture is supported by numerical evidence suggesting that $c_{k,2k} = 1/(2k)!$ for $k = 3, 4$ and 5. This evidence as well as other details related to both Hermite problems can be found in parts 3 and 4.

# 2 The Maximum Likelihood and Least Squares estimators: Existence and characterization

## 2.1 Mixture representation of a $k$-monotone density

WILLIAMSON (1956) gave the following characterization of a $k$-monotone function on $(0, \infty)$:

**Theorem 2.1** *(Williamson, 1956) A function $g$ is $k$-monotone on $(0, \infty)$ if and only if there exists a nondecreasing function $\gamma$ bounded at 0 such that*

$$g(x) = \int_0^\infty (1 - tx)_+^{k-1} d\gamma(t), \qquad x > 0 \tag{2.1}$$

*where $y_+ = y1_{(0,\infty)}(y)$.*

The next theorem gives an inversion formula for the function $\gamma$:

**Theorem 2.2** *(Williamson, 1956) If $g$ is of the form (2.1) with $\gamma(0) = 0$, then at a continuity point $t > 0$, $\gamma$ is given by*

$$\gamma(t) = \sum_{j=0}^{k-1} \frac{(-1)^{k-l} g^{(j)}(1/u)}{j!} \left(\frac{1}{u}\right)^j.$$

For proofs of Theorems 2.2.1 and 2.2.2, see WILLIAMSON (1956). ■

From the characterization given in (2.1), we can easily derive another integral representation for $k$-monotone functions that are Lebesgue integrable on $(0,\infty)$; i.e., $\int_0^\infty g(x)dx < \infty$.

**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*

$$g(x) = \int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k} dF(t), \qquad x > 0 \tag{2.2}$$

*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 (2.2) for some distribution function $F$ on $(0,\infty)$.*

**Proof.** This follows from Theorem 5 of Lévy (1962) by taking $k = n+1$ and $f \equiv 0$ on $(-\infty, 0]$. ■

**Lemma 2.2** *(k-monotone inversion formula) If $F$ in (2.2) satisfies $\lim_{t\to\infty} F(t) = \int_0^\infty g(x)dx$, then at a continuity point $t > 0$, $F$ is given by*

$$F(t) = G(t) - tg(t) + \cdots + \frac{(-1)^{k-1}}{(k-1)!} t^{k-1} g^{(k-2)}(t) + \frac{(-1)^k}{k!} t^k g^{(k-1)}(t), \tag{2.3}$$

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

**Proof.** By the mixture form in (2.2), we have for all $t > 0$

$$F(\infty) - F(t) = \frac{(-1)^k}{k!} \int_t^\infty x^k dg^{(k-1)}(x).$$

But, for $j = 1, \cdots, k$, $t^j G^{(j)}(t) \searrow 0$ as $t \to \infty$. This follows from Lemma 1 in Williamson (1956) applied to the $(k+1)$-monotone function $G(\infty) - G(t)$. Therefore, for $j = 1, \cdots, k$, $t^j g^{(j-1)}(t) \searrow 0$ as $t \to \infty$.

Now, using integration by parts, we can write

$$\begin{aligned}
F(\infty) - F(t) &= \frac{(-1)^k}{k!} \Big[ x^k g^{(k-1)}(x) \Big]_t^\infty + \frac{(-1)^{(k-1)}}{(k-1)!} \int_t^\infty x^{k-1} g^{(k-1)}(x) dx \\
&= -\frac{(-1)^k}{k!} t^k g^{(k-1)}(t) - \frac{(-1)^{k-1}}{(k-1)!} t^{k-1} g^{(k-2)}(t) \\
&\quad + \frac{(-1)^{k-2}}{(k-2)!} \int_t^\infty x^{k-2} g^{(k-2)}(x) dx \\
&\ \vdots \\
&= -\frac{(-1)^k}{k!} t^k g^{(k-1)}(t) - \frac{(-1)^{k-1}}{(k-1)!} t^{k-1} g^{(k-2)}(x) + \cdots - \int_t^\infty g(x) dx,
\end{aligned}$$

Using the fact that $F(\infty) = \int_0^\infty g(x)dx$, the result follows. ■

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

**Lemma 2.3** *(Integrable completely monotone characterization) A function $g$ is an integrable $k$-monotone function if and only if it is of the form*

$$g(x) = \int_0^\infty \frac{1}{t} \exp(-x/t) dF(t), \qquad x > 0 \tag{2.4}$$

*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 (2.4) for some distribution function $F$ on $(0, \infty)$.*

**Lemma 2.4** *(Completely-monotone inversion formula) If $F$ in (2.4) satisfies $\lim_{t\to\infty} F(t) = \int_0^\infty g(x)dx$, then at a continuity point $t > 0$, $F$ is given by*

$$F(t) = \lim_{k\to\infty} \sum_{j=0}^{k} \frac{(-1)^j}{j!} (kt)^j G^{(j)}(kt) \tag{2.5}$$

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

The characterization in (2.2) is more relevant for us since we are dealing with $k$-monotone densities. 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 Lemma 2.2, 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, 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 (2.3), 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.

In the next section, we 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, $\mathbb{G}_n(x) = n^{-1} \sum_1^n 1\{X_i \le x\}$ for $x \ge 0$

## 2.2 Maximum likelihood estimation of a $k$-monotone density

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, but we will abuse notation slightly in this same way throughout). We want to minimize $l_n(g)$ over $g \in \mathcal{D}_k$. To do this, it is frequently of help to 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, $\psi_n$ can also be rewritten as

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

where $F$ is bounded and nondecreasing.

**Lemma 2.5** *The maximum likelihood estimator $\hat{g}_{n,k}$ in the classes $\mathcal{D}_k$, $k \in \{1, 2, \ldots, \infty\}$ exists and is unique. Furthermore, $\hat{g}_{n,k}$ is the maximizer of $\psi_n$ over $\mathcal{M}_k \cap L_1(\lambda)$. 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.** First, we prove that there exists a density $\hat{g}_n$ that maximizes the "usual" log-likelihood $l_n = \int_0^\infty \log g(x) d\mathbb{G}_n(x)$ over the class $\mathcal{D}_k$ with $k$ finite. For $g$ in $\mathcal{D}_k$, let $F$ be the distribution function such that

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

The unicomponent likelihood curve $\Gamma$ as defined by LINDSAY (1983A) (see also LINDSAY (1995)) is then

$$\Gamma = \left\{ \left( \frac{k(y-X_1)_+^{k-1}}{y^k}, \frac{k(y-X_2)_+^{k-1}}{y^k}, \ldots, \frac{k(y-X_n)_+^{k-1}}{y^k} \right) : \quad y \in [0,\infty) \right\}.$$

It is easy to see that $\Gamma$ is bounded (notice that the $i$-th component is equal to 0 whenever $y < X_i$). Also, $\Gamma$ is closed. By Theorems 18 and 22 of LINDSAY (1995), there exists a unique maximizer of $l_n$ and the maximum is achieved by a discrete distribution function that has at most $n$ support points.

Now, let $g$ be a $k$-monotone function in $\mathcal{M}_k \cap L_1(\lambda)$ and let $\int_0^\infty g(x)dx = c$ so that $g/c \in \mathcal{D}_k$. We have

$$\begin{aligned}
\psi_n(g) - \psi_n(\hat{g}_n) &= \int_0^\infty \log\left(\frac{g(x)}{c}\right) d\mathbb{G}_n(x) + \log(c) - c + 1 - \int_0^\infty \log\left(\hat{g}_n(x)\right) d\mathbb{G}_n(x) \\
&\leq \int_0^\infty \log\left(\frac{g(x)}{c}\right) d\mathbb{G}_n(x) - \int_0^\infty \log\left(\hat{g}_n(x)\right) d\mathbb{G}_n(x) \\
&\leq 0
\end{aligned}$$

since $\log(c) \leq c - 1$. Thus $\psi_n$ 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 2.6** *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 $k \in \{1, 2, \ldots\}$,*

$$\hat{H}_{n,k}(t) \equiv \mathbb{G}_n \left( \frac{k(t-X)_+^{k-1}/t^k}{\hat{g}_{n,k}(X)} \right) \leq 1, \tag{2.6}$$

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

$$\hat{H}_{n,\infty}(t) \equiv \mathbb{G}_n \left( \frac{\exp(-X/t)}{t\hat{g}_{n,\infty}(X)} \right) \leq 1, \qquad \textit{for all} \;\; t > 0 \tag{2.7}$$

*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 (2.7) is a natural limiting version as $k \to \infty$ of the functions $\hat{H}_{n,k}$ on the right side of (2.6).*

**Proof.** Since $\hat{F}_n$ maximizes the log-likelihood

$$l_n(F) = \frac{1}{n}\sum_{j=1}^n \log\left(\int_0^\infty \frac{k(y-X_j)_+^{k-1}}{y^k}dF(y)\right),$$

it follows that for all $t > 0$

$$\lim_{\epsilon \searrow 0} \frac{l_n((1-\epsilon)\hat{F}_n + \epsilon\delta_t) - l_n(\hat{F}_n)}{\epsilon} \le 0.$$

This yields

$$\frac{1}{n}\sum_{j=1}^n \frac{k(t-X_j)_+^{k-1}/t^k - \hat{g}_n(X_j)}{\hat{g}_n(X_j)} \le 0$$

or

$$\frac{1}{n}\sum_{j=1}^n \frac{k(t-X_j)_+^{k-1}/t^k}{\hat{g}_n(X_j)} \le 1. \tag{2.8}$$

Now, let $M_n$ be the set defined by

$$M_n = \left\{t > 0 \ : \ \frac{1}{n}\sum_{j=1}^n \frac{k(t-X_j)_+^{k-1}/t^k}{\hat{g}_n(X_j)} = 1\right\}.$$

We will prove now that $M_n = \mathrm{supp}(\hat{F}_n)$. We write $P_{\hat{F}_n}$ for the probability measure associated with $\hat{F}_n$. Integrating the left hand side of (2.8) with respect to $\hat{F}_n$, we have

$$\frac{1}{n}\sum_{j=1}^n \frac{\int_0^\infty \left(k(t-X_j)_+^{k-1}/t^k\right)d\hat{F}_n(t)}{\hat{g}_n(X_j)} \quad = \quad \frac{1}{n}\sum_{j=1}^n \frac{\hat{g}_n(X_j)}{\hat{g}_n(X_j)} = 1.$$

But, using the definition of $M_n$, we can write,

$$\begin{aligned} 1 \quad &= \quad \frac{1}{n}\sum_{j=1}^n \frac{\int_0^\infty \left(k(t-X_j)_+^{k-1}/t^k\right)d\hat{F}_n(t)}{\hat{g}_n(X_j)} \\ &= \quad P_{\hat{F}_n}(M_n) + \frac{1}{n}\sum_{j=1}^n \int_{\mathbb{R}^+\setminus M_n} \frac{\left(k(t-X_j)_+^{k-1}/t^k\right)}{\hat{g}_n(X_j)}d\hat{F}_n(t), \end{aligned}$$

and so

$$\begin{aligned} P_{\hat{F}_n}(\mathbb{R}^+ \setminus M_n) \quad &= \quad \int_{\mathbb{R}^+\setminus M_n} \frac{1}{n}\sum_{j=1}^n \frac{\left(k(t-X_j)_+^{k-1}/t^k\right)}{\hat{g}_n(X_j)}d\hat{F}_n(t) \\ &< \quad P_{\hat{F}_n}(\mathbb{R}^+ \setminus M_n), \ \text{ if } P_{\hat{F}_n}(\mathbb{R}^+ \setminus M_n) > 0. \end{aligned}$$

This is a contradiction and we conclude that $P_{\hat{F}_n}(\mathbb{R}^+ \setminus M_n) = 0$.

The proof of the result for $k = \infty$ is given by JEWELL (1982), page 481. $\blacksquare$

## 2.3 The Least Squares estimator of a $k$-monotone density

The least squares criterion is

$$Q_n(g) = \frac{1}{2}\int_0^\infty g^2(x)dx - \int_0^\infty g(x)d\mathbb{G}_n(x)\,. \tag{2.9}$$

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

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

where

$$r_k(t,t') \equiv \int_0^\infty (t-x)_+^{k-1}(t'-x)_+^{k-1}dx = \int_0^{t\wedge t'} (t-x)^{k-1}(t'-x)^{k-1}dx\,,$$

and

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

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 $\Phi_n$ over the set $\mathcal{X}$ of all non-negative measures $\mu$ on $R^+$. Since $\Phi_n$ 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.

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

**Proof.** Existence follows from Zeidler (1985), Theorem 38.B, page 152. Here we verify the hypotheses of that theorem.

We identity $X$ of Zeidler's theorem with the space $\mathcal{X}$ of nonnegative 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 $\mu$ as $\mu = \mu_1 + \mu_2$ where $\mu_1$ is concentrated on $[0, X_{(1)})$ and $\mu_2$ is concentrated on $[X_{(1)}, \infty)$. Since the second term of $\Phi_n$ is zero for $\mu_1$, the contribution of the $\mu_1$ component to $\Phi_n(\mu)$ is always non-negative, so we make $\inf \Phi_n(\mu)$ no larger by restricting to measures on $[X_{(1)}, \infty)$.

We can restrict further to measures $\mu$ 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

$$r_k(s,t) \geq \frac{(1 - e^{-v_0}) t_0}{2k} s^{k-1} t^{k-1} \tag{2.10}$$

where $v_0 \approx 1.59$. To prove (2.10) we will use the inequality

$$(1 - v/k)^{k-1} \geq e^{-v}, \qquad 0 \leq v \leq v_0, \quad k \geq 2\,. \tag{2.11}$$

(This inequality holds by straightforward computation; see Hall and Wellner (1979), especially their Proposition 2.) Thus we compute

$$\begin{aligned}
r_k(s,t) &= \int_0^\infty (s-x)_+^{k-1} (t-x)_+^{k-1}\, dx \\
&= s^{k-1} t^{k-1} \int_0^\infty (1 - x/s)_+^{k-1} (1 - x/t)_+^{k-1} dx \\
&= \frac{1}{k} s^{k-1} t^{k-1} \int_0^\infty \left(1 - \frac{y}{sk}\right)_+^{k-1} \left(1 - \frac{y}{tk}\right)_+^{k-1} dy \\
&\geq \frac{1}{k} s^{k-1} t^{k-1} \int_0^{v_0(t \wedge s)} e^{-y/s} e^{-y/t} dy
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{k} s^{k-1} t^{k-1} \int_0^{v_0(t\wedge s)} e^{-cy} dy, \qquad c \equiv 1/s + 1/t \\
&= \frac{1}{k} s^{k-1} t^{k-1} \frac{1}{c} \int_0^{v_0(t\wedge s)} c e^{-cy} dy, \\
&= \frac{1}{k} s^{k-1} t^{k-1} \frac{1}{c} \left(1 - \exp(-c(t \wedge s) v_0)\right) \\
&\geq \frac{1}{k} s^{k-1} t^{k-1} \frac{1}{c} \left(1 - \exp(-v_0)\right)
\end{aligned}
$$

since

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

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 (2.10) holds.

From the inequality (2.10) we conclude that for measures $\mu$ concentrated on $[X_{(1)}, \infty)$ we have

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

On the other hand,

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

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

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

This lower bound is strictly positive if

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

But for such measures $\mu$ we can make $\Phi$ smaller by taking the zero measure. Thus we may restrict the minimization problem to the collection of measures $\mu$ satisfying

$$
m_{k-1} \leq 1/A\,. \tag{2.12}
$$

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

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

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 $\mu$ 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 $\mu$ satisfies

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

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

$$\begin{aligned}
\int_{t>\lambda} t^{k-1}d\mu(t) &= \lambda^{k-1}\mu(\lambda, \infty) + (k-1)\int_\lambda^\infty s^{k-2}\mu(s,\infty)ds \\
&\leq \lambda^{k-1}\frac{K}{\lambda^{k-1/2}} + (k-1)\int_\lambda^\infty s^{k-2}Ks^{-(k-1/2)}ds \\
&= K\lambda^{-1/2} + (k-1)K\int_\lambda^\infty s^{-3/2}ds \\
&\leq K\lambda^{-1/2} + (k-1)2K\lambda^{-1/2} \\
&\to 0 \quad \text{ as } \lambda \to \infty
\end{aligned}$$

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

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

$$\limsup \int_0^\infty s_{n,k}(t)d\mu_m(t) \leq \int_0^\infty s_{n,k}(t)d\mu_0(t),$$

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

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

Since $\Phi_n$ is lower semi-compact (i.e. the sets $\mathcal{C}_r \equiv \{\mu \in \mathcal{C} : \; \Phi_n(\mu) \le r\}$ are compact for $r \in \mathbb{R}$), the existence of a minimum follows from ZEIDLER (1985), Theorem 38.B, page 152. Uniqueness follows from the strict convexity of $\Phi_n$. ■

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*

$$\mathbb{Y}_{n,k}(t) = \int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_2} \mathbb{G}_n(t_1) dt_1 dt_2 \cdots dt_{k-1}, \qquad x \ge 0,$$

*and*

$$\tilde{H}_{n,k}(t) = \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} \tilde{g}_n(t_1) dt_1 dt_2 \cdots dt_k, \qquad x \ge 0 .$$

*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 $\tilde{H}_{n,k}$:*

$$\begin{cases} \tilde{H}_{n,k}(t) \ge \mathbb{Y}_{n,k}(t), & \textit{for } \; t \ge 0, \\ \textit{and} & \\ \tilde{H}_{n,k}(t) = \mathbb{Y}_{n,k}(t) & \textit{for } \; t \in supp\{\tilde{F}_{n,k}\} . \end{cases} \tag{2.13}$$

**Remark 2.2** *Note that for $k \in \{1, 2, \ldots\}$ the processes $\mathbb{Y}_{n,k}$ and $\tilde{H}_{n,k}$ can be written in the more compact forms*

$$\mathbb{Y}_{n,k}(t) = \int_0^t \frac{(t-x)^{k-1}}{(k-1)!} d\mathbb{G}_n(x)$$

*and*

$$\tilde{H}_{n,k}(t) = \int_0^t \frac{(t-x)^{k-1}}{(k-1)!} \tilde{g}_n(x) dx.$$

**Proof.** Let $\tilde{g}_n \in \mathcal{M}_k \cap L_2(\lambda)$ satisfy (2.13), and let $g$ be an arbitrary function in $\mathcal{M}_k \cap L_2(\lambda)$. Then

$$Q_n(g) - Q_n(\tilde{g}_n) = \frac{1}{2} \int g^2(x) dx - \frac{1}{2} \int \tilde{g}_n^2(x) dx - \int g(x) d\mathbb{G}_n(x) + \int \tilde{g}_n(x) d\mathbb{G}_n(x).$$

Now, using integration by parts

$$\begin{aligned} & \int_0^\infty (g(x) - \tilde{g}_n(x)) d\mathbb{G}_n(x) \\ = & \; - \int_0^\infty \mathbb{G}_n(x) (g'(x) - \tilde{g}_n'(x)) dx \\ = & \; \int_0^\infty \left( \int_0^x \mathbb{G}_n(y) dy \right) (g''(x) - \tilde{g}_n''(x)) dx \\ \vdots & \\ = & \; (-1)^k \int_0^\infty \mathbb{Y}_n(x) (dg^{(k-1)}(x) - d\tilde{g}_n^{(k-1)}(x)), \end{aligned}$$

and

$$\begin{aligned}
&\int_0^\infty (g^2(x) - \tilde{g}_n^2(x))dx \\
&= \int_0^\infty (g(x) + \tilde{g}_n(x))(g(x) - \tilde{g}_n(x))dx \\
&= -\int_0^\infty \left( \int_0^x g(y)dy + \int_0^x \tilde{g}_n(y)dy \right)(g'(x) - \tilde{g}_n'(x))dx \\
&\vdots \\
&= (-1)^k \int_0^\infty (G_k(x) + \tilde{H}_n(x))(dg^{(k-1)}(x) - d\tilde{g}_n^{(k-1)}(x)),
\end{aligned}$$

where $G_k$ is the $k$-th order integral of $g$. Hence,

$$\begin{aligned}
Q_n(g) - Q_n(\tilde{g}_n) &= \frac{1}{2}(-1)^k \int_0^\infty (G_k(x) + \tilde{H}_n(x))(dg^{(k-1)}(x) - d\tilde{g}_n^{(k-1)}(x)) \\
&\quad - (-1)^k \int_0^\infty \mathbb{Y}_n(x)(dg^{(k-1)}(x) - d\tilde{g}_n^{(k-1)}(x)) \\
&= \frac{1}{2}(-1)^k \int_0^\infty (G_k(x) - \tilde{H}_n(x))(dg^{(k-1)}(x) - d\tilde{g}_n^{(k-1)}(x)) \\
&\quad + (-1)^k \int_0^\infty (\tilde{H}_n(x) - \mathbb{Y}_n(x))(dg^{(k-1)}(x) - d\tilde{g}_n^{(k-1)}(x)) \\
&\geq (-1)^k \int_0^\infty (\tilde{H}_n(x) - \mathbb{Y}_n(x))(dg^{(k-1)}(x) - d\tilde{g}_n^{(k-1)}(x)).
\end{aligned}$$

To see that, we notice (using integration by parts) that

$$(-1)^k \int_0^k (G_k(x) - \tilde{H}_n(x))(dg^{(k-1)}(x) - d\tilde{g}_n^{(k-1)}(x)) = \int_0^\infty (g(x) - \tilde{g}_n(x))^2 dx.$$

But condition (2.13) implies that

$$\int_0^\infty (\tilde{H}_n(x) - \mathbb{Y}_n(x))d\tilde{g}_n^{(k-1)}(x) = 0.$$

Therefore,

$$Q_n(g) - Q_n(\tilde{g}_n) \geq \int_0^\infty (\tilde{H}_n(x) - \mathbb{Y}_n(x))(-1)^k dg^{(k-1)}(x) \geq 0,$$

since $\tilde{H}_n \geq \mathbb{Y}_n$ and $(-1)^{k-2} dg^{(k-1)}(x) = (-1)^k dg^{(k-1)}(x) \geq 0$ because $(-1)^{k-2} g^{(k-2)}$ is convex.

Conversely, take $g_t \in \mathcal{M}_k$ to be

$$g_t(x) = \frac{(t-x)_+^{k-1}}{(k-1)!}, \qquad x \geq 0.$$

We have:

$$\lim_{\epsilon \to 0} \frac{Q_n(\tilde{g}_n + \epsilon g_t) - Q_n(\tilde{g}_n)}{\epsilon} = \int_0^t \frac{(t-x)^{k-1}}{(k-1)!} \tilde{g}_n(x) dx - \int_0^t \frac{(t-x)^{k-1}}{(k-1)!} d\mathbb{G}_n(x).$$

Using integration by parts, we obtain

$$0 \leq \lim_{\epsilon \to 0} \frac{Q_n(\tilde{g}_n + \epsilon g_t) - Q_n(\tilde{g}_n)}{\epsilon} = \tilde{H}_n(t) - \mathbb{Y}_n(t)\,.$$

Finally, since $\tilde{g}_n$ maximizes $Q_n$ it follows that

$$\begin{aligned} 0 &= \lim_{\epsilon \to 0} \frac{Q_n((1+\epsilon)\tilde{g}_n) - Q_n(\tilde{g}_n)}{\epsilon} = \int_0^\infty \tilde{g}_n^2(x) dx - \int_0^\infty \tilde{g}_n(x) d\mathbb{G}_n(x) \\ &= \int_0^\infty (\tilde{H}_n(x) - \mathbb{Y}_n(x))(-1)^{k-1} d\tilde{g}_n^{(k-1)}(x), \end{aligned}$$

which holds if and only if the equality in (2.13) holds. (Additional argument needed for the $k = \infty$ case...?!) ■

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

**Lemma 2.7** *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 $t_0 > a$. ■

**Remark 2.3** *In the previous lemma, the assumption that $\Delta$ has infinitely many zeros can be weakened. Indeed, we obtain the same conclusion if we assume that $\Delta$ has $k+1$ distinct zeros in $[a,b]$.*

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 2.13 that $\tilde{g}_n$ is the LSE if and only if

$$\tilde{H}_n(t) \geq \mathbb{Y}_n(t), \quad \text{for } t > 0, \tag{2.14}$$

and

$$\int_0^\infty \left(\tilde{H}_n(t) - \mathbb{Y}_n(t)\right) d\tilde{g}_n^{(k-1)}(t) = 0 \tag{2.15}$$

where

$$\tilde{H}_n(t) = \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} \tilde{g}_n(t) dt,$$

and

$$\mathbb{Y}_n(t) = \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} d\mathbb{G}_n(t).$$

The condition in (2.15) 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 $X_{(1)}, \cdots, X_{(n)}$ (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 $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ cannot have a positive density with respect to Lebesgue measure $\lambda$. Indeed, if we assume otherwise, then we can find $0 \le j \le n$ 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 $\tilde{H}_n \equiv \mathbb{Y}_n$ on $I$. This implies that $\tilde{H}_n^{(k)} \equiv \mathbb{Y}_n^{(k)} \equiv 0$, since $\mathbb{Y}_n$ is a polynomial of degree $k-1$ on $I$, and hence $\tilde{g}_n \equiv 0$ on I. But the latter is impossible since 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 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 $\tilde{g}_n$ is given by*

$$\tilde{g}_n(x) = \tilde{w}_1 \frac{k(\tilde{a}_1 - x)_+^{k-1}}{\tilde{a}_1^k} + \cdots + \tilde{w}_m \frac{k(\tilde{a}_m - x)_+^{k-1}}{\tilde{a}_m^k}. \tag{2.16}$$

**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 (2.15) 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 (2.16).

(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 $\tilde{\Delta}_n$ (with $X_{(0)} = 0$ and $X_{(n+1)} = \infty$). By Lemma 2.7, 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 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 $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ 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 zeros in $(z_j, t_j)$, and hence by Lemma 2.7, 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 $\tilde{g}_n > 0$ 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 $c > 0$.

## 3.1 Consistency of the maximum likelihood estimator

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), van de Geer (1993), and van de Geer (1996). 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 $\mu$.

**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*

$$H(\hat{g}_{n,k}, g_0) \rightarrow_{a.s.} 0 \qquad as \;\; n \rightarrow \infty\,.$$

*Furthermore $\hat{F}_{n,k} \rightarrow_d F_0$ 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), van de Geer (1993), and van de Geer (1996), by using the Glivenko-Cantelli preservation theorems of

van der Vaart and Wellner (2000). See also van de Geer (1999), page 54, example 4.2.4, and Wellner (2003b), pages 98 to 99.
■

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

**Lemma 3.1** *If $g$ is a $k$-monotone density function then*

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

*for all $x > 0$.*

**Proof.** We have

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

since, with $g_k(u) = u(1-u)^{k-1}$ we have

$$g_k'(u) = (1-u)^{k-1} - u(k-1)(1-u)^{k-2} = (1-u)^{k-2}(1-ku)$$

which equals zero if $u = 1/k$ and this yields a maximum. (Note that when $k = 2$, this bound equals $1/(2x)$ which agrees with the bound given by Jongbloed (1995), page 117 in this case.) ■

**Proposition 3.2** *Let $g_0$ be a $k$-monotone density on $(0, \infty)$ and fix $c > 0$. Then*

$$\sup_{x \geq c} |\hat{g}_n(x) - g_0(x)| \to_{a.s.} 0, \qquad \text{as} \quad n \to \infty.$$

**Proof.** Let $F_0$ be the mixing distribution function associated with $g_0$. Then for all $x > 0$, we have

$$g_0(x) = \int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k} dF_0(t).$$

Now, let $Y_1, \cdots, Y_m$ be i.i.d. from $F_0$. Taking $m = n$, let $\mathbb{F}_n$ be the corresponding empirical distribution and $g_n$ the mixed density

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

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

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

By the Glivenko-Cantelli theorem, the sequence of $k$-monotone densities $(g_n)_n$ satisfies

$$\sup_{x \in [d,\infty)} |g_n(x) - g_0(x)| \rightarrow_{a.s.} 0, \qquad \text{as} \quad n \to \infty.$$

Since the MLE $\hat{g}_n$ maximizes the criterion function over the class $\mathcal{M}_k \cap L_1(\lambda)$, we have

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

and this is equivalent to

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

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 $x > 0$

$$\hat{g}_l(x) \to \hat{g}(x), \quad \text{as} \quad l \to \infty,$$

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 $\hat{g}_l$ and $\hat{g}$ are monotone and $\hat{g}$ is continuous.

Much of the following is along the lines of Jongbloed (1995), pages 117-119, and Groeneboom, Jongbloed, and Wellner (2001b), pages 1674-1675. We are going to show that $\hat{g}$ and the true density $g_0$ have to be the same. For $0 < \alpha < 1$ define

$\eta_\alpha = G_0^{-1}(1-\alpha)$. Fix $\epsilon$ so small that $\epsilon < \eta_\epsilon$. By (3.1) there is a number $D_\epsilon > 0$ such that $\hat{g}_l(1/\epsilon) \geq D_\epsilon$ for sufficiently large $l$. To see this, note that (3.1) implies that

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

and hence

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

by the choice of $\eta_\epsilon$ and hence we can certainly take $D_\epsilon = \int_{\eta_\epsilon}^\infty g_0(x)dG_0(x)/2$.

Hence, by continuity of $g_l$ and the bound in Lemma 3.1

$$\hat{g}_l(z) \leq \frac{1}{z}(1-\frac{1}{k})^{k-1} \equiv \frac{e_k}{z}\,, \qquad g_l(z) \leq \frac{1}{z}(1-\frac{1}{k})^{k-1} \equiv \frac{e_k}{z}\,,$$

$g_l/\hat{g}_l$ 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 $x \in [\epsilon, \eta_\epsilon]$

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

In fact,

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

while

$$\frac{g_l(x)}{\hat{g}_l(x)} \geq \frac{g_l(\eta_\epsilon)}{\hat{g}_l(\epsilon)} \geq \frac{g_0(\eta_\epsilon)/2}{\epsilon^{-1}e_k}$$

using the (uniform) convergence of $g_l$ to $g_0$. Therefore

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

uniformly on $[\epsilon, \eta_\epsilon]$. For sufficiently large $l$, we have using (3.1)

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

But since $\mathbb{G}_l$ 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

$$\int_\epsilon^{\eta_\epsilon} \frac{g_0(x)}{\hat{g}(x)} dG_0(x) \leq 1+\epsilon.$$

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

$$\int_0^\infty \frac{g_0(x)}{\hat{g}(x)} dG_0(x) \le 1,$$

which is equivalent to

$$\int_0^\infty \frac{g_0^2(x)}{\hat{g}(x)} dx \le 1. \tag{3.2}$$

Define $\tau = \int_0^\infty \hat{g}(x)dx$. Then $\hat{h} = \tau^{-1}\hat{g}$ is a $k$-monotone density. By (3.2), we have that

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

Now consider the function

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

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

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

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 $\hat{h} = \hat{g} = g_0$.

We have proved that from each subsequence of $\hat{g}_n$, we can extract a further subsequence that converges to $g_0$ almost surely. The convergence is again uniform on intervals of the form $[c, \infty)$, $c > 0$ by monotonicity of $\hat{g}_n$ and $\hat{g}$ and continuity of $g_0$. ■

**Corollary 3.1** *Let $c > 0$. For $j = 1, \cdots, k-2$,*

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

*and for each $x > 0$ at which $g_0$ is $k-1$-times differentiable,*

$$\hat{g}_n^{(k-1)}(x) \to_{a.s.} g_0^{(k-1)}(x) \, .$$

**Proof.** This follows along the lines of the proof in Jongbloed (1995), page 119, and Groeneboom, Jongbloed, and Wellner (2001b), Lemma 3.1, page 1675. ■

## 3.2 The Least Squares estimator

We also have strong and uniform consistency of the LSE $\tilde{g}$ 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 $\|\tilde{g}_n - g_0\|_2 \to_{a.s.} 0$, and*

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

**Proof.** The main difficulty here is that we don't know whether the LSE $\tilde{g}_n$ is a genuine density; i.e. $\tilde{g}_n \in \mathcal{M}_k$ but not necessarily $\tilde{g}_n \in \mathcal{D}_k$. 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 $\int_0^\infty \tilde{g}_n^2 d\lambda = O(1)$ almost surely. 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

$$\sqrt{\int_0^\infty \tilde{g}_n^2(x) dx} = \int_0^\infty \tilde{u}_n(x) d\mathbb{G}_n(x), \tag{3.3}$$

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

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

since $g$ is decreasing. Therefore

$$g(x) \le \frac{1}{\sqrt{x}} \equiv G(x)$$

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

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

and hence by (3.3) 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. Most of the following arguments are similar to those of proof of consistency of the LSE when $k = 2$ as given in Groeneboom, Jongbloed, and Wellner (2001b).

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

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

$$\begin{aligned} \tilde{g}_n(\tau_n) &\le \tilde{g}_n(\delta/2) \le (2/\delta)(\delta/2)\tilde{g}_n(\delta/2) \le (2/\delta) \int_0^{\delta/2} \tilde{g}_n(x)dx \\ &\le (2/\delta)\sqrt{\delta/2}\sqrt{\int_0^{\delta/2} \tilde{g}_n^2(x)dx} \le \sqrt{2/\delta}\sqrt{\int_0^\infty \tilde{g}_n^2(x)dx} \\ &= \sqrt{2M_k/\delta}. \end{aligned} \tag{3.4}$$

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

$$\begin{aligned} \int_{\tau_n}^\delta \tilde{g}_n(x)dx &\le \sqrt{\delta - \tau_n}\sqrt{\int_{\tau_n}^\delta \tilde{g}_n^2(x)dx} \\ &\le \sqrt{\delta}\sqrt{\int_0^\infty \tilde{g}_n^2(x)dx} = \sqrt{\delta M_k}. \end{aligned}$$

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

$$\begin{aligned} \sqrt{\delta M_k} &\ge \int_{\tau_n}^\delta \tilde{g}_n(x)dx \\ &= \tilde{g}_n(\delta)(\delta - \tau_n) - \frac{\tilde{g}_n'(\delta)}{2}(\delta - \tau_n)^2 \end{aligned}$$

$$
\begin{aligned}
& \quad + \cdots + (-1)^{k-1} \frac{\tilde{g}_n^{(k-1)}(\delta)}{k!} (\delta - \tau_n)^k \\
\geq & \quad (\delta - \tau_n) \Bigg( \tilde{g}_n(\delta) + \frac{1}{k}(-1)\tilde{g}_n'(\delta)(\delta - \tau_n) \\
& \qquad + \cdots + (-1)^{k-1} \frac{\tilde{g}_n^{(k-1)}(\delta)}{(k-1)!} (\delta - \tau_n)^{k-1} \Bigg) \\
= & \quad (\delta - \tau_n) \left( \tilde{g}_n(\delta) \left(1 - \frac{1}{k}\right) + \frac{1}{k} \tilde{g}_n(\tau_n) \right) \\
\geq & \quad \frac{\delta}{2k} \tilde{g}_n(\tau_n)
\end{aligned}
$$

and hence

$$
\tilde{g}_n(\tau_n) \leq 2k\sqrt{M_k/\delta}.
$$

Therefore, combining the obtained bounds, we have for large $n$

$$
\tilde{g}_n(\tau_n) \leq 2k\sqrt{M_k/\delta} = C_k. \tag{3.5}
$$

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

$$
\int_\delta^\infty (\tilde{g}(x) - g_0(x))^2 dx \leq \liminf_{l \to \infty} \int_\delta^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2 dx. \tag{3.6}
$$

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

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

Thus we can write

$$
\begin{aligned}
& \int_\delta^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2 dx \\
& \quad \leq \int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2 dx \\
& \quad \leq 2 \int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x)) d(\mathbb{G}_{n_l}(x) - G_0(x)) \to_{a.s.} 0,
\end{aligned} \tag{3.7}
$$

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) (3.7) holds. From (3.6), 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 $\Omega_0$ 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 $g_0$. ■

**Corollary 3.2** *Let $c > 0$. Under the assumption of Proposition 3.3, we have for $j = 1, \cdots, k-2$,*

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

*and for each $x > 0$ at which $g_0$ is $k-1$-times differentiable,*

$$\tilde{g}_n^{(k-1)}(x) \to_{a.s.} g_0^{(k-1)}(x)\,.$$

**Proof.** See the proof of Corollary 3.1. ■

# 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 Kiefer (1982), Stone (1980), 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 $g \in \mathcal{D}_k$ 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 $\epsilon \downarrow 0$,*

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

*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 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:*

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

*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*

$$\lambda_{k,2} = 2^{4(k+1)} \frac{(2k+3)(k+2)}{(k+1)^2} \frac{((2(k+1))!)^2}{(4k+7)!((k-1)!)^2 \left(\binom{k}{k/2-1}\right)^2}, \quad \textit{when } k \textit{ is even}$$

*and*

$$\lambda_{k,2} = 2^{4(k+2)}(2k+3)(k+2) \frac{((2(k+1))!)^2}{(4k+7)!(k!)^2 \left(\binom{k+1}{(k-1)/2}\right)^2} \quad \textit{when } k \textit{ is odd}$$

*and, with* $r(x) \equiv (1-x^2)^{k+1}(1+x)$ *for* $-1 \le x \le 1$ *and* $C_{k,j} \equiv r^{(j)}(0)$*,*

$$\lambda_{k,1}^{(j)} = \left| \frac{C_{k,j}}{C_{k,k}} \right|, \qquad 0 \le j \le k-1.$$

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 differentiable at* $x_0$*,* $k \ge 2$*. Then, for estimating* $Tg_0 = F(x_0)$ *where* $F_0$ *is given in terms of* $g_0$ *by (2.3),*

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

The lower bound results in Proposition 4.1 are consistent with the results of Kiefer (1982) and Stone (1980) (although our result involves a slightly stronger lower bound since the supremum is over just a local neighborhood of the truth). In particular, Kiefer showed that rates of convergence in estimation cannot be improved by order restrictions, but that order restrictions might result in improvements of the constants. This latter suggestion has been investigated in detail in the case of monotone densities by Birgé (1987), Birgé (1989). 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 will reappear in the limit distribution theory for $k \ge 3$ 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

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

then for $X \sim g$, $Z = Z_k \sim k \times \mathrm{Beta}(1, k)$, 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 $\phi_{Z^*}$ and bound its modulus above and below for large arguments. Thus we calculate first for $Z_\infty^*$: 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

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

for $|t| \ge 1$. Thus the hypothesis (1.3) of Fan (1991) holds with $\beta = 1$, $\beta_1 = 1/2$ and $\beta_0 = 1/2$. 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 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

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

It should also be noted that

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

Thus

$$|\phi_{Z_k^*}(t)| \sim \frac{k!}{t^k} \qquad \text{as } \ t\to\infty\,,$$

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 $\mu$ be a positive number and consider the function $g_\mu$ 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 $\mu$ is chosen small enough so that the true density $g_0$ is $k$-times differentiable on $[x_0 - \mu, x_0 + \mu]$ and $g_0^{(k)}$ is continuous on the latter interval, the perturbed function $g_\mu$ 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

$$\tilde{g}_\mu(x) = \mu^{2k+3} r\left(\frac{x - x_0}{\mu}\right).$$

Then for $0 \leq j \leq k$

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

The scale $s(\mu)$ should be chosen so that for all $0 \leq j \leq k$

$$(-1)^j g_\mu^{(j)}(x) > 0, \quad \text{for} \quad x \in [x_0 - \mu, x_0 + \mu].$$

But for $\mu$ small enough, the sign of $(-1)^j g_\mu^{(j)}$ will be that of $(-1)^j g_0^{(j)}(x_0)$, and hence $g_\mu$ 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 $r^{(k)}(0) \neq 0$. Set

$$s(\mu) = \frac{g_0^{(k)}(x_0)}{r^{(k)}(0)} \times \frac{1}{\mu^{k+3}}.$$

Then for $0 \leq j \leq k-1$

$$\begin{aligned} g_\mu^{(j)}(x_0) &= g_0^{(j)}(x_0) + \mu^{k-j} \frac{g_0^{(k)}(x_0) r^{(j)}(0)}{r^{(k)}(0)} \\ &= g_0^{(j)}(x_0) + o(\mu), \quad \text{as} \quad \mu \searrow 0 \end{aligned}$$

and so we can choose $\mu$ small enough so that $(-1)^j g_\mu^{(j)}(x_0) > 0$. For $j = k$

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

To show that $r^{(j)}(0) \neq 0$ for $0 \leq j \leq k$, we define

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

Let $m \geq 2$ and $2n \geq m$. We have

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

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

$$x_{n,m} = -2n(m-1)x_{n-1,m-2}.$$

If $m$ is even, we obtain

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

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

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

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

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

and hence

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

Therefore, when $j$ is even, the second term vanishes and

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

When $j$ is odd, the first term vanishes and

$$\begin{aligned} r^{(j)}(0) &= (-2)^{(j-1)/2} \prod_{i=0}^{(j-1)/2-1} (k+1-i) \times j \times \prod_{i=0}^{(j-1)/2-1} (j-2i-2) \\ &= (-2)^{(j-1)/2} \prod_{i=0}^{(j-1)/2-1} (k+1-i) \times \prod_{i=0}^{(j-1)/2} (j-2i) \neq 0. \end{aligned}$$

We set

$$C_{k,j} = r^{(j)}(0), \quad \text{for} \quad 1 \leq j \leq k\,.$$

Then $C_{k,k}$ specializes to

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

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 $0 \leq j \leq k-1$,

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

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

$$\begin{aligned} &\int_0^\infty \frac{(g_\mu(x) - g_0(x))^2}{g_0(x)} dx \\ &= \frac{\left(g_0^{(k)}(x_0)\right)^2}{\mu^{2(k+3)}(C_{k,k})^2} \int_{x_0-\mu}^{x_0+\mu} \frac{(x_0+\mu-x)^{2(k+1)}(x-x_0+\mu)^{2(k+2)}}{g_0(x)} dx \\ &= \frac{\left(g_0^{(k)}(x_0)\right)^2}{\mu^{2(k+3)}(C_{k,k})^2} \int_{-\mu}^{\mu} \frac{(\mu^2-y^2)^{2(k+1)}(y+\mu)^2}{g_0(x_0+y)} dy \\ &= \frac{\left(g_0^{(k)}(x_0)\right)^2}{\mu^{2(k+3)}(C_{k,k})^2} \times \mu^{4(k+1)+3} \int_{-1}^{1} \frac{(1-z^2)^{2(k+1)}(z+1)^2}{g_0(x_0+\mu z)} dz \end{aligned}$$

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

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

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

The constants $\lambda_{k,2}$ can be given more explicitly using the formula

$$
I_{n,2p} = \int_0^1 (1-x^2)^n x^{2p} dx = 2^{2n+1} \frac{n!(n+1)!}{(2n+2)!} \frac{\binom{n+p}{n+1}}{\binom{2(n+p)+1}{2(n+1)}},
$$

for any integers $n$ and $p$, using the convention

$$
\binom{n+p}{n+1} = \binom{2(n+p)+1}{2(n+1)} = 1
$$

when $p = 0$. We have,

$$
\int_{-1}^{1} (1-x^2)^{2(k+1)}(x+1)^2 dx = \int_{-1}^{1} (1-x^2)^{2(k+1)} x^2 dx + \int_{-1}^{1} (1-x^2)^{2(k+1)} dx,
$$

since

$$
\int_{-1}^{1} (1-x^2)^{2(k+1)} x dx = 0,
$$

and hence

$$
\begin{aligned}
\int_{-1}^{1} (1-x^2)^{2(k+1)}(x+1)^2 dx &= 2(I_{2(k+1),2} + I_{2(k+1),0}) \\
&= 2^{4k+6} \frac{(2(k+1))!(2k+3)!}{(4k+6)!} \frac{\binom{2k+3}{2k+3}}{\binom{4k+7}{4k+6}} + \frac{2^{4k+5}\,((2(k+1))!)^2}{(4k+5)!} \\
&= 2^{4k+5} \frac{((2(k+1))!)^2}{(4k+6)!} \left( \frac{2(2k+3)}{4k+7} + (4k+6) \right) \\
&= 2^{4k+5} \frac{((2(k+1))!)^2}{(4k+7)!} \left( (4k+6) + (4k+6)(4k+7) \right)
\end{aligned}
$$

$$
\begin{aligned}
&= 2^{4k+5}\frac{((2(k+1))!)^2}{(4k+7)!}(4k+6)(4k+8) \\
&= 2^{4(k+2)}(2k+3)(k+2)\frac{((2(k+1))!)^2}{(4k+7)!}. \qquad (4.2)
\end{aligned}
$$

Combining and (4.1) and (4.2), we find that $\lambda_{k,2}$ is given by

$$
\lambda_{k,2} = 2^{4(k+1)}\frac{(2k+3)(k+2)}{(k+1)^2}\frac{((2(k+1))!)^2}{(4k+7)!((k-1)!)^2\left(\binom{k}{k/2-1}\right)^2}, \quad \text{when } k \text{ is even,}
$$

and

$$
\lambda_{k,2} = 2^{4(k+2)}(2k+3)(k+2)\frac{((2(k+1))!)^2}{(4k+7)!(k!)^2\left(\binom{k+1}{(k-1)/2}\right)^2}, \quad \text{when } k \text{ is odd.}
$$

Now, by using the change of variable $\epsilon = \mu^{2k+1}(b_k + o(1))$, where

$$
b_k = \lambda_{k,2}\frac{\left(g_0^{(k)}(x_0)\right)^2}{g_0(x_0)}
$$

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

$$
r_{k,j} = \frac{\left(\lambda_{k,1}^{(j)} g_0^{(k)}(x_0)\right)^{(2k+1)/(k-j)}}{b_k}
$$

and hence

$$
\sup_{\tau>0}\lim_{n\to\infty}\inf n^{\frac{k-j}{2k+1}} MMR_1(n, T_j, \mathcal{D}_{k,n,\tau}) \ge \frac{1}{4}\left(4\frac{k-j}{2k+1}e^{-1}\right)^{\frac{k-j}{2k+1}}(r_{k,j})^{\frac{k-j}{2k+1}}, \qquad (4.3)
$$

which can be rewritten as

$$\sup_{\tau>0} \lim_{n\to\infty} \inf n^{\frac{k-j}{2k+1}} MMR_1(n, T_j, \mathcal{D}_{k,n,\tau})$$
$$\geq \frac{1}{4}\left(4\frac{k-j}{2k+1}e^{-1}\right)^{\frac{k-j}{2k+1}} \frac{\lambda_{k,1}^{(j)}}{(\lambda_{k,2})^{\frac{k-j}{2k+1}}} \left\{ \left|g_0^{(k)}(x_0)\right|^{\frac{2j+1}{2k+1}} g_0(x_0)^{\frac{k-j}{2k+1}} \right\}$$

for $j = 0, \cdots, k-1$. ■

**Remark 4.1** *It might seem that a more natural choice for a perturbation would have been*

$$g_\mu(x) = g_0(x) + s(\mu)(x_0 + \mu - x)^{k+1}(x - x_0 + \mu)^{k+1} 1_{[x_0-\mu, x_0+\mu]}(x).$$

*The scale $s(\mu)$ can be chosen such that the perturbed function is $k$-monotone and $k$-times differentiable with a continuous $k$-th derivative in the neighborhood $[x_0-\mu, x_0+\mu]$. However, using this perturbation, asymptotic lower bounds can only be derived for estimating the functionals $T_j(g)$ when $j$ is even since $g_\mu^{(2l+1)}(x_0) = g_0^{(2l+1)}(x_0)$ for $l \in \mathbb{N}$.*

**Acknowledgements:** We gratefully acknowledge helpful conversations with Carl de Boor, Nira Dyn, Tilmann Gneiting, and Piet Groeneboom.

# 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.

Apostol, T. (1957). *Mathematical Analysis*, Addison-Wesley, Reading.

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. (2004d). Generalizations of Hampel's problem via mixtures of Poisson processes. Manuscript in progress.

Birgé, L. (1987). Estimating a density under order restrictions: nonasymptotic minimax risk. *Ann. Statist.* **15**, 995 - 1012.

Birgé, L. (1989). The Grenander estimator: a nonasymptotic approach. *Ann. Statist.* **17**, 1532 - 1549.

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

de Boor, C. (1973). The quasi-interpolant as a tool in elementary polynomial spline theory. In *Approximation Theory* (Austin, TX, 1973), 269 - 276. Academic Press, New York.

de Boor, C. (1974). Bounding the error in spline interpolation. *SIAM Review* **16**, 531 - 544.

de Boor, C. (1976). A bound on the $L_\infty$ norm of $L_2$ approximation by splines in terms of global mesh ratio. *Math. Comp.* **30**, 765 - 771.

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. Geometrizing rates of convergence, II, III. *Ann. Statist.* **19**, 633 - 667, 668 - 701.

Doyle, J., Hansen, E. and McNolty, F. (1980). Properties of the mixed exponential failure process. *Technometrics* **22**, 555-565.

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

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. (2001a). A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.* **29**, 1620 - 1652.

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

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2003). The support reduction algorithm for computing nonparametric function estimates in mixture models. Available in Math. ArXiv. at: *http://front.math/ucdavis.edu/math.ST/0405511*.

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

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.

Harris, C. M. and Singpurwalla, N. (1968). Life distributions derived from stochastic hazard functions. *IEEE Trans. on Reliability* **R-17**, 70- 79.

Hill, D. L., Saunders, R. and Laud, P. W. (1980). Maximum likelihood estimation for mixtures. *Canad. J. Statist.* **8**, 87 - 93.

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.

Karlin, S. and Ziegler, Z. (1966). Tchebysheffian spline functions. *SIAM Num.* **3**, 514 - 543.

Kiefer, J. (1982). Optimum rates for non-parametric density and regression estimates under order restrictions. In *Statistics and Probability: Essays in Honor of C. R. Rao*, 419 - 428. G. Kallianpur, P. R. Krishnaiah, and J.K. Ghosh, editors. North Holland, Amsterdam.

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

Lavee, D., Safrie, U. N., and Meilijson, I. (1991). For how long do trans-Saharan migrants stop over at an oasis? *Ornis Scandinavica* **22**, 33 - 44.

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.

Lindsay, B. (1983b). The geometry of mixture likelihoods, Part II: the exponential family. *Ann. Statist.* **11**, 783 - 792.

Lindsay, B. (1995). *Mixture Models: Theory, Geometry and Applications.* IMS, Hayward.

Nürnberger, G. (1989). *Approximation by Spline Functions.* Springer-Verlag, New York.

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.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1 - 20.

Robbins, H. (1980). Estimation and prediction for mixtures of the exponential distribution. *Proc. nat. Acad. Sciences* **77**, 2382-2383.

Shadrin, A. Yu. (2001). The $L_\infty$ norm of the $L_2-$spline projector is bounded independently of the knot sequence: a proof of de Boor's conjecture. *Acta Math.* **187**, 59 - 137.

Schoenberg, I. J. and Whitney, A. (1953). On Pólya frequency functions III. the positivity of translation determinants with applications to the interpolation problem by spline curves. *TAMS* **74**, 246 - 259.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348 - 1360.

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

van de Geer, S. (1996). Rates of convergence for the maximum likelihood estimator in mixture models. *J. Nonparametric Statist.* **6**, 293 - 310.

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.

Wang, Y. (1994). The limiting distribution in concave regression. Preprint, Univ. Missouri, Columbia.

Wellner, J. A. (2003b). Empirical processes: theory and applications. Lectures presented at Torgnon, Italy, July 2003. Available at:
http://www.stat.washington.edu/jaw/RESEARCH/TALKS/talks.html.

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.

University of Washington
Statistics
Box 354322
Seattle, Washington 98195-4322
U.S.A.
*e-mail: balabda@stat.washington.edu*

University of Washington
Statistics
Box 354322
Seattle, Washington 98195-4322
U.S.A.
*e-mail: jaw@stat.washington.edu*