Estimation of a $k-$monotone density: limit distribution theory and the spline connection, with complete proofs

Fadoua Balabdaoui[*,‡]

Institute for Mathematical Stochastics
Georgia Augusta University Goettingen
Maschmuehlenweg 8-10
D-37073 Goettingen
Germany
e-mail: fadoua@math.uni-goettingen.de

and

Jon A Wellner[†]

Department of Statistics
Box 354322
University of Washington
Seattle, WA 98195-4322
e-mail: jaw@stat.washington.edu

University of Göttingen and University of Washington

Abstract: We study the asymptotic behavior of the Maximum Likelihood and Least Squares estimators of a k-monotone density g_0 at a fixed point x_0 when $k > 2$. In Balabdaoui and Wellner (2004a), it was proved that both estimators exist and are splines of degree $k-1$ with simple knots. These knots, which are also the jump points of the $(k-1)$-st derivative of the estimators, cluster around a point $x_0 > 0$ under the assumption that g_0 has a continuous k-th derivative in a neighborhood of x_0 and $(-1)^k g_0^{(k)}(x_0) > 0$. If τ_n^- and τ_n^+ are two successive knots,

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

[†]Research supported in part by National Science Foundation grant DMS-0203320, NIAID grant 2R01 AI291968-04, and an NWO Grant to the Vrije Universiteit, Amsterdam

[‡]Corresponding author

we prove that the random "gap" $\tau_n^+ - \tau_n^-$ is $O_p(n^{-1/(2k+1)})$ for any $k > 2$ if a conjecture about the upper bound on the error in a particular Hermite interpolation via odd-degree splines holds. Based on the order of the gap, the asymptotic distribution of the Maximum Likelihood and Least Squares estimators can be established. We find that the j-th derivative of the estimators at x_0 converges at the rate $n^{-(k-j)/(2k+1)}$ for $j = 0, \ldots, k-1$. The limiting distribution depends on an almost surely uniquely defined stochastic process H_k that stays above (below) the k-fold integral of Brownian motion plus a deterministic drift, when k is even (odd).

AMS 2000 subject classifications: Primary 62G05, 60G99; secondary 60G15, 62E20.
Keywords and phrases: asymptotic distribution, completely monotone, convex, Hermite interpolation, inversion, $k-$fold integral of Brownian motion, least squares, maximimum likelihood, minimax risk, mixture models, multiply monotone, nonparametric estimation, rates of convergence, shape constraints, splines.

1. Introduction

1.1. The estimation problem and motivation

A density function g on $\mathbb{R}^+$ is monotone (or $1-$monotone) if it is nonincreasing. It is $2-$monotone if it is nonincreasing and convex, and $k-$*monotone* for $k \geq 3$ if and only if $(-1)^j g^{(j)}$ is non-negative, nonincreasing, and convex for $j = 0, \ldots, k-2$.

We write $\mathcal{D}_k$ for the class of all $k-$monotone densities on $\mathbb{R}^+$, and $\mathcal{M}_k$ for the class of all $k-$monotone functions (without the density restriction). Suppose that $g_0 \in \mathcal{D}_k$ and that $X_1, \ldots, X_n$ are i.i.d. with density g_0. We write $\mathbb{G}_n$ for the empirical distribution function of $X_1, \ldots, X_n$. Our main interest is in the Maximum Likelihood Estimators (or MLE) $\hat{g}_n$ of $g_0 \in \mathcal{D}_k$.

When $k = 1$, it is well known that the maximum likelihood estimator $\hat{g}_n$ of $g_0 \in \mathcal{D}_1$ is the Grenander (1956) estimator; i.e. the left-derivative of the least concave majorant $\widehat{G}_n$ of $\mathbb{G}_n$, and if $g_0'(x_0) < 0$ with g_0' continuous in a neighborhood of x_0, then

$$n^{1/3}\left(\hat{g}_n(x_0) - g_0(x_0)\right) \to_d \left(\frac{1}{2} g_0(x_0)|g_0'(x_0)|\right)^{1/3} 2Z, \tag{1.1}$$

where $2Z$ is the slope at zero of the greatest convex minorant of two-sided Brownian motion $+t^2$, $t \in \mathbb{R}$; see PRAKASA RAO (1969), GROENEBOOM (1985), and KIM AND POLLARD (1990).

When $k = 2$, GROENEBOOM, JONGBLOED, AND WELLNER (2001B) considered both the MLE and LSE and established that if the true convex and nonincreasing density g_0 satisfies $g_0''(x_0) > 0$ (and g_0'' is continuous in a neighborhood of x_0), then

$$\begin{pmatrix} n^{2/5}\left(\bar{g}_n(x_0) - g_0(x_0)\right) \\ n^{1/5}\left(\bar{g}_n'(x_0) - g'(x_0)\right) \end{pmatrix} \to_d \begin{pmatrix} \left(\frac{1}{24} g_0^2(x_0) g_0''(x_0)\right)^{1/5} H^{(2)}(0) \\ \left(\frac{1}{24^3} g_0(x_0) g_0''(x_0)^3\right)^{1/5} H^{(3)}(0) \end{pmatrix} \tag{1.2}$$

where $\bar{g}_n$ is either the MLE or LSE and H is a random cubic spline function such that $H^{(2)}$ is convex and H stays above integrated two-sided Brownian motion $+t^4, t \in \mathbb{R}$, and touches exactly at those points where $H^{(2)}$ changes its slope (see GROENEBOOM, JONGBLOED, AND WELLNER (2001A)).

Our main interest in this paper is in establishing a generalization of the pointwise limit theory given in (1.1) and (1.2) for general $k \in \mathbb{N}$, $k \geq 1$.

Beyond the obvious motivation of extending the known results for $k = 1$ and $k = 2$ as listed above, there are several further reasons for considering such extensions:

(a) Pointwise limit distribution theory for natural nonparametric estimators of the piecewise smooth regression models of smoothness k considered by MAMMEN (1991) is only available for $k \in \{1, 2\}$. Similar models (with just one element in the partition) have been proposed for software reliability problems by MILLER AND SOFER (1986). Similarly, pointwise limit distribution theory is still lacking for the locally adaptive regression spline estimators considered by MAMMEN AND VAN DE GEER (1997).

(b) The classes of densities $\mathcal{D}_k$ have mixture representations as scale mixtures of Beta$(1, k)$ densities: as is known from WILLIAMSON (1956) (see also

Lévy (1962), Gneiting (1999), and Balabdaoui and Wellner (2004a)), $g \in \mathcal{D}_k$ if and only if there is a distribution function F on $(0, \infty)$ such that

$$g(x) = \int_0^\infty \frac{k}{y^k} (y-x)_+^{k-1} dF(y) = \int_0^\infty w \left(1 - \frac{wx}{k}\right)_+^{k-1} d\tilde{F}(w) \quad (1.3)$$

where $z_+ \equiv z 1\{z \geq 0\}$ and $\tilde{F} = F(k/\cdot)$. The second form of the mixture representation in the last display makes it clear that the limiting class of densities as $k \to \infty$, namely $\mathcal{D}_\infty$, is the class of scale mixtures of exponential distributions. In view of Feller (1971), pages 232-233, this is just the class of *completely monotone* densities; see also Widder (1941) and Gneiting (1998). To the best of our knowledge, there is no pointwise limit distribution theory available for the MLE in any class of mixed densities based on a smooth mixing kernel, including this particular case in which the kernel (or mixture density) is the exponential scale family as studied by Jewell (1982). On the other hand, maximum likelihood estimators in various classes of mixture models with smooth kernels have been proposed in a wide range of applications including pharmacokinetics (Mallet (1986), Mallet, Mentre, Steimer, and Lokiec (1988), and Davidian and Gallant (1992)), demography (Vaupel, Manton, and Stallard (1979)), and shock models and variations in hazard rates (Harris and Singpurwalla (1968), Doyle, Hansen, and McNolty (1980), Hill, Saunders, and Laud (1980)).

(c) The whole family of mixture models corresponding to $k \in (0, \infty)$ might be of some interest eventually, especially since the family of distributions corresponding to the classical Wicksell problem is contained in the class $\mathcal{D}_{1/2}$; see e.g. Groeneboom and Jongbloed (1995).

(d) The sub-class of $k-$monotone densities with mixing distribution F satisfying $g^{(k-1)}(0) = k! \int_0^\infty y^{-k} dF(y) < \infty$ can be regarded as the distributions arising in a generalization of Hampel's bird watching problem (Hampel (1987)) in which birds are captured $k-$ times, but only one "inter-catch" time is recorded. Based on those observed inter-catch times, the goal is to estimate the true distribution F of the resting times Y of the migrating birds, which we assume to have a density f with k-th moment $\mu_k(f) < \infty$. Furthermore, we assume that the time points of capture form the arrival time points of a Poisson process with rate λ, and given $Y = y$, the number of captures by time y is Poisson(λy) with λ small enough so that $\exp(-\lambda y) \approx 1$, and the probability of catching a bird more than k times is negligible (see also Hampel (1987) and Anevski (2003)). If $S_{k,1}$ denotes the elapsed time between the first and second captures (the only observed inter-catch), then it follows by a derivation analogous to Hampel's that the density of the time $S_{k,1}$ is given by

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

which is clearly k-monotone. We obtain F, the probability distribution of Y, by inverting the previous mixture representation; that is

$$F(t) = 1 - \frac{g^{(k-1)}(t)}{g^{(k-1)}(0+)}$$

at any point of continuity $t > 0$ of F.

In connection with (a), it is interesting to note that the definition of the family $\mathcal{D}_k$ is equivalent to $g \in \mathcal{D}_k$ if and only if $(-1)^{k-1} g^{(k-1)}$ (where $g^{(k-1)}$ is either the left- or right-derivative of $g^{(k-2)}$) is nonincreasing. This follows from Lemma 4.3 of Gneiting (1999) since Gneiting's condition $\lim_{x\to\infty} g(x) = 0$ is automatic for densities. Thus the equivalent definition of $\mathcal{D}_k$ has a natural connection with the work of Mammen (1991) in the nonparametric regression setting. In parallel to the treatment of convex regression estimation given by Groeneboom, Jongbloed, and Wellner (2001b), it seems clear that pointwise distribution theory for nonparametric least squares estimators for the regression problems in (a) could be developed if adequate theory were available for the the Maximum Likelihood and Least Squares estimators of densities in the class $\mathcal{D}_k$, so we focus exclusively on the density case in this paper. In Section 5 we comment further on the difficulties in obtaining corresponding limit theory for the smooth kernel cases discussed in (b).

1.2. Description of the key difficulty: the gap problem

The key result that Groeneboom, Jongbloed, and Wellner (2001b) used to establish (1.2) is that $\tau_n^+ - \tau_n^- = O_p(n^{-1/5})$ as $n \to \infty$, where τ_n^- and τ_n^+ are two successive jump points of the first derivative of $\bar{g}_n$ in the neighborhood of x_0. Such a result was already proved by Mammen (1991) (see Lemma 8) in the context of nonparametric regression, where the true regression curve, m, is piecewise concave/convex or convex/concave such that m is twice continuously differentiable in the neigborhood of x_0, and $m''(x_0) \neq 0$. Furthermore, Mammen (1991) conjectured the right form of the asymptotic distribution of his Least square estimator, which was later established by Groeneboom, Jongbloed, and Wellner (2001b).

To obtain the stochastic order $n^{-1/5}$ for the gap, Groeneboom, Jongbloed, and Wellner (2001b) used the characterizations of the estimators together with the "mid-point property" which we review in Section 4. For $k = 1$, the same property can be used to establish that $n^{-1/3}$ is the order of the gap. As a function of k, it is natural to conjecture that $n^{-1/(2k+1)}$ is the general form of the order of the gap. In the problem of nonparametric regression via splines, Mammen and van de Geer (1997) have conjectured that $n^{-1/(2k+1)}$ is the order of the distance between the knot points of their regression spline $\hat{m}$ under the assumption that the true regression curve m_0 satisfies our same working assumptions, but the question was left open (see Mammen and van de Geer (1997), page 400). In this manuscript, we refer to the problem of establishing the order of $\tau_n^+ - \tau_n^-$ as the *gap problem*.

In Section 4, we show that when $k > 2$, the gap problem is closely related to a "non-classical" Hermite interpolation problem via odd-degree splines. To put the interpolation problem encountered in the next section in context, it is useful to review briefly the related *complete interpolation problem* for odd-degree splines which is more "classical" and for which error bounds uniform in the knots are now available. Given a function $f \in C^{(k-1)}[0,1]$ and an increasing sequence $0 = y_0 < y_1 < \cdots < y_m < y_{m+1} = 1$ where

$m \ge 1$ is an integer, it is well-known that there exists a unique spline, called the *complete spline* and denoted here by Cf, of degree $2k-1$ with interior knots $y_1, \ldots, y_m$ that satisfies the $2k+m$ conditions

$$\begin{cases} (Cf)(y_i) = f(y_i), \quad i = 1, \ldots, m \\ (Cf)^{(l)}(y_0) = f^{(l)}(y_0), \ (Cf)^{(l)}(y_{m+1}) = f^{(l)}(y_{m+1}), \quad l = 0, \ldots, k-1; \end{cases}$$

see SCHOENBERG (1963), DE BOOR (1974), or NÜRNBERGER (1989), page 116, for further discussion. If $j \in \{0, \ldots, k\}$ and $f \in C^{(k+j)}[0,1]$, then there exists $c_{k,j} > 0$ such that

$$\sup_{0<y_1<\cdots<y_m<1} \|f - Cf\|_\infty \le c_{k,j} \|f^{(k+j)}\|_\infty. \tag{1.4}$$

For $j = k$, this "uniform in knots" bound in the complete interpolation problem was first conjectured by DE BOOR (1973) for $k > 4$ as a generalization that goes beyond $k = 2, 3$ and 4 for which the result was already established (see also DE BOOR (1974)). By a scaling argument, the bound (1.4) implies that, if $f \in C^{(2k)}[a,b], a < b \in \mathbb{R}$, the interpolation error in the complete interpolation problem is uniformly bounded in the knots, and that the bound is of the order of $(b-a)^{2k}$. One key property of the complete spline interpolant Cf is that $(Cf)^{(k)}$ is the Least Squares approximation of $f^{(k)}$ when $f^{(k)} \in L_2([0,1])$; i.e., if $\mathcal{S}_k(y_1, \cdots, y_m)$ denotes the space of splines of order k (degree $k-1$) and interior knots $y_1, \ldots, y_m$, then

$$\int_0^1 \Big((Cf)^{(k)} - f^{(k)}(x)\Big)^2 dx = \min_{S \in \mathcal{S}_k(y_1,\ldots,y_m)} \int_0^1 \Big(S(x) - f^{(k)}(x)\Big)^2 dx \tag{1.5}$$

(see e.g. SCHOENBERG (1963), DE BOOR (1974), NÜRNBERGER (1989)). Consequently, if L_∞ denotes the space of bounded functions on $[0,1]$, then the properly defined map

$$\begin{array}{rcl} C^{(k)}[0,1] & \to & \mathcal{S}_k(\underline{y}) \\ f^{(k)} & \to & (Cf)^{(k)} \end{array}$$

where $\underline{y} = (y_1, \ldots, y_m)$, is the restriction of the orthoprojector $P_{S_k(\underline{y})}$ from L_∞ to $\mathcal{S}_k(\underline{y})$ with respect to the inner product $\langle g, h \rangle = \int_0^1 g(x)h(x)dx$ which assigns to a function $g \in L_\infty$ the k-derivative of the complete spline interpolant of *any* primitive of g of order k (note that the difference between two primitives of g of order k is a polynomial of degree $k-1$).

de Boor (1974) pointed out that, in order to prove the conjecture, it is enough to prove that

$$\sup_{\underline{y}} \|P_{S_k(\underline{y})}\|_\infty = \sup_{\underline{y}} \sup_{g \in L_\infty} \frac{\|P_{S_k(\underline{y})}(g)\|_\infty}{\|g\|_\infty}$$

is bounded, and this was successfully achieved by Shadrin (2001).

The Hermite interpolation problem which arises naturally in Section 4 appears to be another variant of interpolation problems via odd-degree splines which has not yet been studied in the approximation theory or spline literature. More specifically, if f is some real-valued function in $C^{(j)}[0,1]$ for some $j \geq 1$, $0 = y_0 < y_1 < \cdots < y_{2k-4} < y_{2k-3} = 1$ is a given increasing sequence, then there exists a unique spline $\mathcal{H}_k f$ of degree $2k-1$ and interior knots $y_1, \ldots, y_{2k-4}$ satisfying the $4k-4$ conditions

$$(\mathcal{H}_k f)(y_i) = f(y_i), \quad \text{and} \quad (\mathcal{H}_k f)'(y_i) = f'(y_i), \quad i = 0, \ldots, 2k-3. \tag{1.6}$$

It turns out that deriving the stochastic order of the distance between two successive knots of the MLE and LSE in the neighborhood of the point of estimation is very closely linked to bounding the error in this new Hermite interpolation independently of the locations of the knots of the spline interpolant. More precisely, if $g_t(x) = (x-t)_+^{k-1}/(k-1)!$ is the power truncated function of degree $k-1$ with unique knot t, then we conjecture that there is a constant $d_k > 0$ such that

$$\sup_{t \in (0,1)} \sup_{0 < y_1 < \cdots < y_{2k-4} < 1} \|g_t - \mathcal{H}_k g_t\|_\infty \leq d_k. \tag{1.7}$$

As shown in Balabdaoui and Wellner (2005), the preceding formulation implies that boundedness of the error independently of the knots of the spline interpolant holds true for any $f \in C^{(k+j)}$, that is

$$\sup_{0 < y_1 < \cdots < y_{2k-4} < 1} \|f - \mathcal{H}_k f\|_\infty \leq d_{k,j} \|f^{(k+j)}\|_\infty.$$

If $j = k$ and $\|f^{(2k)}\|_\infty \le 1$, it follows from Proposition 1 of Balabdaoui and Wellner (2005) that the interpolation error must be bounded above by the error for interpolating the perfect spline

$$S^*(t) = \frac{1}{(2k)!}\left(t^{2k} + 2\sum_{i=1}^{2k-4}(-1)^i (t-\tau_j)_+^{2k}\right)$$

For a definition of perfect splines, see e.g. Bojanov, Hakopian and Sahakian (1993), Chapter 6. Based on large number of simulations, we found that

$$\sup_{0<y_1<\cdots<y_{2k-4}<1} \|S^* - \mathcal{H}_k S^*\|_\infty \le \frac{2}{(2k)!}$$

for fairly large values of k (see the last column in Table 2 in Balabdaoui and Wellner (2005)). The latter strongly suggests that for $f \in C^{(2k)}[0,1]$ we have

$$\sup_{0<y_1<\cdots<y_{2k-4}<1} \|f - \mathcal{H}_k f\|_\infty \le \frac{2}{(2k)!}\,\|f^{(2k)}\|_\infty. \tag{1.8}$$

Based on Conjecture (1.7), we will prove that the distance between two consecutive knots in a neighborhood of x_0 is $O_p(n^{-1/(2k+1)})$.

After a brief introduction of the MLE and LSE and their respective characterizations, we give in Section 3 a statement of our main result which gives the joint asymptotic distribution of the successive derivatives of the MLE and LSE. The obtained convergence rate $n^{-(k-j)/(2k+1)}$ for the j-th derivative of any of the estimators was found by Balabdaoui and Wellner (2004a) to be the asymptotic minimax lower bound for estimating $g_0^{(j)}(x_0)$, $j = 0, \ldots, k-1$ under the same working assumptions. The limiting distribution depends on the higher derivatives of H_k, an almost surely uniquely defined process that stays above (below) the $(k-1)$-fold integral of Brownian motion plus the drift $(k!/(2k)!)\,t^{2k}$, when k is even (odd), and is $(2k-2)-$convex; i.e. the $2k-2$ derivative of H_k is convex. The process H_k is studied separately in Balabdaoui and Wellner (2004c). Proving the existence of H_k relies also on our conjecture in (1.7) since the key problem, also referred to as the *gap problem*, depends on a very similar Hermite interpolation problem, except that the knots of the estimators are replaced by the points of touch between the $(k-1)$-fold integral of Brownian motion plus the drift $(k!/(2k)!)\, t^{2k}$ and H_k. For more discussion of the background and related problems, see BALABDAOUI AND WELLNER (2004A). For a discussion of algorithms and computational issues, see BALABDAOUI AND WELLNER (2004B).

2. The estimators and their characterization

Let $X_1, \cdots, X_n$ be n independent observations from a common k-monotone density g_0. We consider nonpametric estimation of g_0 via the Least Squares and Maximum Likelihood methods, and that of its mixture distribution F_0, that is the distribution function on $(0, \infty)$ such that

$$g_0(x) = \int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k} dF_0(t), \quad x > 0.$$

In other words, g_0 is a scale mixture of Beta$(1, k)$ densities. The mixing distribution is furthermore given at any point of continuity t by the inversion formula

$$F_0(t) = \sum_{j=0}^{k} (-1)^j \frac{t^j}{j!} G_0^{(j)}(t) \tag{2.9}$$

where $G_0(t) = \int_0^t g_0(x)dx$. AAn estimator for F_0 can be obtained by simply plugging in estimators of $G_0^{(j)} = g_0^{(j-1)}$, $j = 0, \ldots, k$, in the inversion formula (2.9). We call estimation of the (mixed) k-monotone density g_0 the *direct problem*, and estimation of the mixing distribution function F_0 the *inverse problem*. For more technical details on the mixture representation and the inversion formula, see Lemma 2.1 of BALABDAOUI AND WELLNER (2004A).

Now, we give the definition of the Least Squares and Maximum Likelihood estimators; these were already considered in the case $k = 2$ by GROENEBOOM, Jongbloed, and Wellner (2001b). The LSE, $\tilde{g}_n$, is the minimizer of the criterion function

$$\Phi_n(g) = \frac{1}{2}\int_0^\infty g^2(t)dt - \int_0^\infty g(t)d\mathbb{G}_n(t)$$

over the class $\mathcal{M}_k$, whereas the MLE, $\hat{g}_n$, maximizes the "adjusted" log-likelihood function, i.e.

$$l_n(g) = \int_0^\infty \log g(t)d\mathbb{G}_n(t) - \int_0^\infty g(t)dt$$

over the same class. In Balabdaoui and Wellner (2004a), we find that both estimators exist and are splines of degree $k-1$, i.e., their $(k-1)$-st derivative is stepwise. Furthermore, as shown in Balabdaoui and Wellner (2004a), the LSE's and MLE's are characterized as follows: Let $\tilde{H}_n$ and $\mathbb{Y}_n$ be the processes defined for all $x \ge 0$ by

$$\begin{aligned} \mathbb{Y}_n(x) &= \int_0^x \int_0^{t_{k-1}} \cdots \int_0^{t_2} \mathbb{G}_n(t_1)dt_1dt_2\ldots dt_{k-1} \\ &= \int_0^x \frac{(x-t)^{k-1}}{(k-1)!}d\mathbb{G}_n(t), \end{aligned} \tag{2.10}$$

and

$$\begin{aligned} \tilde{H}_n(x) &= \int_0^x \int_0^{t_k} \cdots \int_0^{t_2} \tilde{g}_n(t_1)dt_1dt_2\ldots dt_k \\ &= \int_0^x \frac{(x-t)^{k-1}}{(k-1)!}\tilde{g}_n(t)dt. \end{aligned} \tag{2.11}$$

Then the k-monotone function $\tilde{g}_n$ is the LSE if and only if

$$\tilde{H}_n(x) \begin{cases} \ge \mathbb{Y}_n(x), & \text{for all } x \ge 0 \\ = \mathbb{Y}_n(x), & \text{if} \quad (-1)^{k-1}\tilde{g}_n^{(k-1)}(x-) < (-1)^{k-1}\tilde{g}_n^{(k-1)}(x+). \end{cases} \tag{2.12}$$

For the MLE we define the process

$$\widehat{H}_n(x,g) = \int_0^x \frac{k(x-t)^{k-1}}{x^k\hat{g}_n(t)}d\mathbb{G}_n(t) \tag{2.13}$$

for all $x \geq 0$ and $g \in \mathcal{D}_k$. Then, a necessary and sufficient condition for the k-monotone function $\hat{g}_n$ to be the MLE is given by

$$\widehat{H}_n(x, \hat{g}_n) \begin{cases} \leq 1, & \text{for all } x \geq 0 \\ = 1, & \text{if } \ (-1)^{k-1}\hat{g}_n^{(k-1)}(x-) < (-1)^{k-1}\hat{g}_n^{(k-1)}(x+). \end{cases} \tag{2.14}$$

These characterizations are crucial for understanding the local asymptotique behavior of the LSE and MLE. They were exploited in Balabdaoui and Wellner (2004a) to show uniform strong consistency of the estimators on intervals of the form $[c, \infty), c > 0$. Here, they prove to be once again very useful for establishing their limit theory in both the direct and inverse problems.

3. The asymptotic distribution

3.1. The main convergence theorem

To prepare for a statement of the main result, we first recall the following theorem from Balabdaoui and Wellner (2004c) giving existence of the processes H_k.

Theorem 3.1 *For all $k \geq 1$, let Y_k denote the stochastic process defined by*

$$Y_k(t) = \begin{cases} \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} dW(s) + \frac{(-1)^k k!}{(2k)!} t^{2k}, & t \geq 0 \\ \int_t^0 \frac{(t-s)^{k-1}}{(k-1)!} dW(s) + \frac{(-1)^k k!}{(2k)!} t^{2k}, & t < 0. \end{cases}$$

If Conjecture (1.7) holds (also see the discussion in Balabdaoui and Wellner (2004c)*), then there exists an almost surely uniquely defined stochastic process H_k characterized by the following four conditions:*

(i) The process H_k stays everywhere above the process Y_k:

$$H_k(t) \geq Y_k(t), \qquad t \in \mathbb{R}.$$

(ii) $(-1)^k H_k$ is $2k$-convex; i.e. $(-1)^k H_k^{(2k-2)}$ exists and is convex.

(iii) *The process H_k satisfies*

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

(iv) *If k is even, $\lim_{|t|\to\infty}(H_k^{(2j)}(t) - Y_k^{(2j)}(t)) = 0$ for $j = 0, \ldots, (k-2)/2$; if k is odd, $\lim_{t\to\infty}(H_k(t) - Y_k(t)) = 0$ and $\lim_{|t|\to\infty}(H_k^{(2j+1)}(t) - Y_k^{(2j+1)}(t)) = 0$ for $j = 0, \ldots, (k-3)/2$.*

Now we are able to state the main result of this paper which generalizes Theorem 6.2 of Groeneboom, Jongbloed, and Wellner (2001b) for estimating convex (2-monotone) densities:

Theorem 3.2 *Let $x_0 > 0$ and g_0 be a k-monotone density such that g_0 is k-times differentiable at x_0 with $(-1)^k g_0^{(k)}(x_0) > 0$ and assume that $g_0^{(k)}$ is continuous in a neighborhood of x_0. Let $\bar{g}_n$ denote either the LSE, $\tilde{g}_n$ or the MLE $\hat{g}_n$ and let $\bar{F}_n$ be the corresponding mixing measure. If Conjecture (1.7) holds, then*

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

and

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

where

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

for $j = 0, \ldots, k-1$.

3.2. The key results and outline of the proofs

Our proof of Theorem 3.2 proceeds by solving the key gap problem assuming that our Conjecture (1.7) holds. This is carried out in Section 4 in which the main result is:

Lemma 3.1 *Let $k \ge 3$ and $\bar{g}_n$ denote either the LSE $\tilde{g}_n$ or the MLE $\hat{g}_n$. If $g_0 \in \mathcal{D}_k$ satisfies $g_0^{(k)}(x_0) \neq 0$ and Conjecture (1.7) holds, then $\tau_{2k-3} - \tau_0 = O_p(n^{-1/(2k+1)})$ where $\tau_0 < \cdots < \tau_{2k-3}$ are $2k-2$ successive jump points of $\bar{g}_n^{(k-1)}$ in a neighborhood of x_0.*

Using Lemma 3.1 we can establish the rate(s) of convergence of the estimators $\tilde{g}_n$ and $\hat{g}_n$ and their derivatives viewed as local processes in $n^{-1/(2k+1)}$ neighborhoods of the fixed point x_0. This is accomplished in Proposition 3.1 (which depends in turn on a preliminary "existence of points" result given in Proposition 6.1). Once the rates have been established, we define for the LSE localized versions $\mathbb{Y}_n^{loc}$, $\tilde{H}_n^{loc}$ of the processes $\mathbb{Y}_n$, $\tilde{H}_n$ given in (2.10) and (2.11) respectively, and $\widehat{\mathbb{Y}}_n^{loc}$, $\widehat{H}_n^{loc}$ related to the process $\widehat{H}_n$ given in (2.13) in the case of the MLE. The proof then proceeds by showing that:

- The localized processes $\mathbb{Y}_n^{loc}$ and $\widehat{\mathbb{Y}}_n^{loc}$ converge weakly to $\mathbb{Y}_{a,\sigma}$ where

$$Y_{a,\sigma}(t) = \begin{cases} \sigma \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1) ds_1 \dots ds_{k-1} + a(-1)^k \frac{k!}{(2k)!} t^{2k}, & t \ge 0 \\ \sigma \int_t^0 \int_{s_{k-1}}^0 \cdots \int_{s_2}^0 W(s_1) ds_1 \dots ds_{k-1} + a(-1)^k \frac{k!}{(2k)!} t^{2k}, & t \le 0 \end{cases}$$

  with $\sigma = \sqrt{g(x_0)}$, $a = (-1)^k g_0^{(k)}(x_0)/k!$ and W a two-sided Brownian motion process starting from 0.
- The localized processes $\tilde{H}_n^{loc}$ and $\widehat{H}_n^{loc}$ satisfy Fenchel (inequality and equality) relations relative to the localized processes $\mathbb{Y}_n^{loc}$ and $\widehat{\mathbb{Y}}_n^{loc}$ respectively.
- We then show via tightness that the localized processes $\tilde{H}_n^{loc}$ and $\widehat{H}_n^{loc}$ (and all their derivatives up to order $2k-1$) converge to a limit process satisfying the conditions (i) - (iv) of Theorem 3.1, and hence the limit

process in both cases is just H_k (up to scaling by constants). When specialized to $t = 0$ this gives the conclusion of Theorem 3.2.

Here is the key rates of convergence proposition.

Proposition 3.1 *Fix $x_0 > 0$ and let g_0 be a k-monotone density such that $(-1)^k g_0^{(k)}(x_0) > 0$. Let $\bar{g}_n$ denote either the MLE $\hat{g}_n$ or the LSE $\tilde{g}_n$. If Conjecture (1.7) holds, then for each $M > 0$ we have,*

$$\begin{aligned} &\sup_{|t|\le M} \left| \bar{g}_n^{(j)}(x_0 + n^{-1/(2k+1)}t) - \sum_{i=j}^{k-1} \frac{n^{-(i-j)/(2k+1)} g_0^{(i)}(x_0)}{(i-j)!} t^{i-j} \right| \\ &\quad = O_p(n^{-(k-j)/(2k+1)}) \end{aligned} \tag{3.1}$$

for $j = 0, \ldots, k-1$.

For the LSE, we define the local $\mathbb{Y}_n$ and $\tilde{H}_n$-processes respectively by

$$\begin{aligned} \mathbb{Y}_n^{loc}(t) &= n^{\frac{2k}{2k+1}} \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \\ &\qquad \left\{ \mathbb{G}_n(v_1) - \mathbb{G}_n(x_0) - \int_{x_0}^{v_1} \sum_{j=0}^{k-1} \frac{(u-x_0)^j}{j!} g_0^{(j)}(x_0) du \right\} \Pi_{i=1}^{k-1} dv_i, \end{aligned}$$

and

$$\begin{aligned} \tilde{H}_n^{loc}(t) &= n^{\frac{2k}{2k+1}} \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_k} \cdots \int_{x_0}^{v_2} \\ &\qquad \left\{ \tilde{g}_n(v_1) - \sum_{j=0}^{k-1} \frac{(v_1-x_0)^j}{j!} g_0^{(j)}(x_0) \right\} dv_1 \ldots dv_k \\ &\qquad + \tilde{A}_{k-1,n} t^{k-1} + \tilde{A}_{k-2,n} t^{k-2} + \cdots + \tilde{A}_{1,n} t + \tilde{A}_{0,n}, \end{aligned}$$

where

$$\tilde{A}_{j,n} = \frac{n^{(2k-j)/(2k+1)}}{j!} \left(\tilde{H}_n^{(j)}(x_0) - \mathbb{Y}_n^{(j)}(x_0) \right), \qquad j = 0, \ldots, k-1.$$

Let $r_k \equiv 1/(2k+1)$. In the case of the MLE, the local processes $\widehat{\mathbb{Y}}_n^{loc}$ and $\widehat{H}_n^{loc}$ are defined as

$$\begin{aligned}\frac{\widehat{\mathbb{Y}}_n^{loc}(t)}{g_0(x_0)} &= n^{2kr_k}\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_1}\frac{g_0(v)-\sum_{j=0}^{k-1}\frac{(v-x_0)^j}{j!}g_0^{(j)}(x_0)}{\hat{g}_n(v)} \\ &\qquad dv dv_1\ldots dv_{k-1} \\ &\quad + n^{2kr_k}\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_1}\frac{1}{\hat{g}_n(v)}d(\mathbb{G}_n-G_0)(v) \\ &\qquad dv_1\ldots dv_{k-1}\end{aligned}$$

and

$$\begin{aligned}\frac{\widehat{H}_n^{loc}(t)}{g_0(x_0)} &= n^{2kr_k}\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_1}\frac{\hat{g}_n(v)-\sum_{j=0}^{k-1}\frac{(v-x_0)^j}{j!}g_0^{(j)}(x_0)}{\hat{g}_n(v)} \\ &\qquad dv dv_1\ldots dv_{k-1}+\widehat{A}_{(k-1)n}t^{k-1}+\cdots+\widehat{A}_{0n}\end{aligned}$$

where

$$\widehat{A}_{jn} = -\frac{n^{(2k-j)r_k}}{(k-1)!j!}g_0(x_0)\left(\widehat{H}_n^{(j)}(x_0)-\frac{(k-1)!}{(k-j)!}x_0^{k-j}\right), \qquad j=0,\ldots,k-1.$$

In the following lemma, we will give the asymptotic distribution of the local process $\mathbb{Y}_n^{loc}$ and $\widehat{\mathbb{Y}}_n^{loc}$ in terms of the $(k-1)$-fold integral of two-sided Brownian motion, $g_0(x_0)$, and $g_0^{(k)}(x_0)$ assuming that the true density g_0 is k-times continuously differentiable at x_0. We denote by $\bar{\mathbb{Y}}_n^{loc}$ either $\mathbb{Y}_n^{loc}$ or $\widehat{\mathbb{Y}}_n^{loc}$.

Lemma 3.2 *Let x_0 be a point where g_0 is continuously $k-$times differentiable in a neighborhood of x_0 with $(-1)^k g_0^{(k)}(x_0) > 0$. Then $\bar{\mathbb{Y}}_n^{loc} \Rightarrow Y_{a,\sigma}$ in $C[-K,K]$ for each $K>0$ where*

$$Y_{a,\sigma}(t) = \begin{cases}\sqrt{g_0(x_0)}\int_0^t\int_0^{s_{k-1}}\cdots\int_0^{s_2}W(s_1)ds_1\ldots ds_{k-1}+a(-1)^k\frac{k!}{2k!}t^{2k}, & t\geq 0\\ \sqrt{g_0(x_0)}\int_t^0\int_{s_{k-1}}^0\cdots\int_{s_2}^0 W(s_1)ds_1\ldots ds_{k-1}+a(-1)^k\frac{k!}{2k!}t^{2k}, & t<0\end{cases}$$

where W is standard two-sided Brownian motion starting at 0, $\sigma=\sqrt{g_0(x_0)}$, and $a=(-1)^k g_0^{(k)}(x_0)/k!$.

Now, let $\bar{H}_n^{loc}$ denote either $\tilde{H}_n^{loc}$ or $\widehat{H}_n^{loc}$.

Lemma 3.3 *The localized processes $\bar{\mathbb{Y}}_n^{loc}$ and $\bar{H}_n^{loc}$, satisfy*

$$\bar{H}_n^{loc}(t) - \bar{\mathbb{Y}}_n^{loc}(t) \geq 0 \qquad \textit{for all} \;\; t \geq 0,$$

with equality if $x_0 + tn^{-1/(2k+1)}$ is a jump point of $\bar{g}_n^{(k-1)}$.

Lemma 3.4 *The limit process $Y_{a,\sigma}$ in Lemma 3.2 satisfies*

$$Y_{a,\sigma}(t) \stackrel{d}{=} \frac{1}{s_1} Y_k\left(\frac{t}{s_2}\right)$$

where $Y_k \equiv Y_{1,1}$ and

$$s_1 = \frac{1}{\sqrt{g_0(x_0)}}\left(\frac{(-1)^k g_0^{(k)}(x_0)}{k!\sqrt{g_0(x_0)}}\right)^{(2k-1)/(2k+1)} \tag{3.2}$$

$$s_2 = \left(\frac{\sqrt{g_0(x_0)}}{\frac{(-1)^k g_0^{(k)}(x_0)}{k!}}\right)^{2/(2k+1)}. \tag{3.3}$$

To show that the derivatives of $\bar{H}_n^{loc}$ are tight, we need the following lemma.

Lemma 3.5 *For all $j \in \{0, \ldots, k-1\}$, let $\bar{A}_{jn}$ denote either $\tilde{A}_{jn}$ or $\widehat{A}_{jn}$. If Conjecture (1.7) holds, then*

$$\bar{A}_{jn} = O_p(1). \tag{3.4}$$

Now we rescale the processes $\bar{\mathbb{Y}}_n^{loc}$ and $\bar{H}_n^{loc}$ so that the rescaled $\bar{\mathbb{Y}}_n^{loc}$ converges to the canonical limit process Y_k defined in Lemma 3.4. Since the scaling of $\bar{\mathbb{Y}}_n^{loc}$ will be exactly the same as the one we used for Y_k, we define $\bar{H}_n^l$ and $\bar{\mathbb{Y}}_n^l$ by

$$\bar{H}_n^l(t) = s_1 \bar{H}_n^{loc}(s_2 t), \qquad \bar{\mathbb{Y}}_n^l(t) = s_1 \bar{\mathbb{Y}}_n^{loc}(s_2 t)$$

where s_1 and s_2 are given by (3.2) and (3.3) respectively.

Lemma 3.6 *Let $c > 0$. Then*

$$((\bar{H}_n^l)^{(0)}, (\bar{H}_n^l)^{(1)}, \ldots, (\bar{H}_n^l)^{(2k-1)}) \Rightarrow (H_k^{(0)}, H_k^{(1)}, \ldots, H_k^{(2k-1)})$$

in $(D[-c, c])^{2k}$ where H_k is the stochastic process defined in Theorem 3.1.

Proofs of Theorem 3.2 and the results given in Subsection 3.2 can be found in Appendix 1.

4. The gap problem - Spline connection

Recall that it was assumed that g_0 is k-times continuously differentiable at x_0 and that $(-1)^k g_0^{(k)}(x_0) > 0$. Under a weaker assumption, Balabdaoui and Wellner (2004a) proved strong consistency of the $(k-1)$-st derivative of the MLE and LSE. This consistency result together with the above assumption imply that the number of jump points of this derivative, in a small neighborhood of x_0, diverges to infinity almost surely as the sample size $n \to \infty$. This "clustering" phenomenon is one of the most crucial elements in studying the local asymptotics of the estimators. The jump points form then a sequence that converges to x_0 almost surely and therefore the distance between two successive jump points, for example located just before and after x_0, converges to 0 as $n \to \infty$. But it is not enough to know that the "gap" between these points converges to 0: an upper bound for this rate of convergence is needed.

To prove Lemma 3.1, we will focus first on the LSE because it is somewhat easier to handle through the simple form of its characterization. The arguments for the MLE could be built upon those used for the LSE, but in this case, one has to deal with some extra difficulties due to the non-linear nature of its characterization.

We start first by describing the difficulties of establishing this result for the general case $k > 2$.

4.1. Fundamental differences

Let τ_n^- and τ_n^+ be the last and first jump points of the $(k-1)$-st derivative of the LSE $\tilde{g}_n$, located before and after x_0 respectively. To obtain a better understanding of the gap problem, we describe the reasoning used by Groeneboom, Jongbloed, and Wellner (2001b) in order to prove that $\tau_n^+ - \tau_n^- = O_p(n^{-1/5})$ for the special case $k = 2$. The LSE $\tilde{g}_n$ is characterized by

$$\tilde{H}_n(x) \begin{cases} \geq \mathbb{Y}_n(x), & x \geq 0 \\ = \mathbb{Y}_n(x), & \text{if } x \text{ is a jump point of } \tilde{g}_n' \end{cases} \tag{4.1}$$

where $\tilde{H}_n(x) = \int_0^x (x-t)\tilde{g}_n(t)dt$ and $\mathbb{Y}_n(x) = \int_0^x \mathbb{G}_n(t)dt$. On the interval $[\tau_n^-, \tau_n^+)$, the function $\tilde{g}_n'$ is constant since there are no more jump points in this interval. This implies that $\tilde{H}_n$ is polynomial of degree 3 on $[\tau_n^-, \tau_n^+)$. But, from the characterization in (4.1), it follows that

$$\tilde{H}_n(\tau_n^{\pm}) = \mathbb{Y}_n(\tau_n^{\pm}), \quad \tilde{H}_n'(\tau_n^{\pm}) = \mathbb{Y}_n'(\tau_n^{\pm}).$$

These four boundary conditions allow us to fully determine the cubic polynomial $\tilde{H}_n$ on $[\tau_n^-, \tau_n^+]$. Using the explicit expression for $\tilde{H}_n$ and evaluating it at the mid-point $\bar{\tau} = (\tau_n^- + \tau_n^+)/2$, Groeneboom, Jongbloed, and Wellner (2001b) established that

$$\tilde{H}_n(\bar{\tau}_n) = \frac{\mathbb{Y}_n(\tau_n^-) + \mathbb{Y}_n(\tau_n^+)}{2} - \frac{\mathbb{G}_n(\tau_n^+) - \mathbb{G}_n(\tau_n^-)}{8}\,(\tau_n^+ - \tau_n^-).$$

Groeneboom, Jongbloed and Wellner refer to this as the "mid-point property". By applying the first condition (the inequality condition) in (4.1), it follows that

$$\frac{\mathbb{Y}_n(\tau_n^-) + \mathbb{Y}_n(\tau_n^+)}{2} - \frac{\mathbb{G}_n(\tau_n^+) - \mathbb{G}_n(\tau_n^-)}{8}\,(\tau_n^+ - \tau_n^-) \geq \mathbb{Y}_n(\bar{\tau}_n).$$

The inequality in the last display can be rewritten as

$$\frac{Y_0(\tau_n^-) + Y_0(\tau_n^+)}{2} - \frac{G_0(\tau_n^+) - G_0(\tau_n^-)}{8}\,(\tau_n^+ - \tau_n^-) \geq \mathbb{E}_n$$

where G_0 and Y_0 are the true counterparts of $\mathbb{G}_n$ and $\mathbb{Y}_n$ respectively, and $\mathbb{E}_n$ is a random error. Using empirical process theory, Groeneboom, Jongbloed, and Wellner (2001b) showed that

$$|\mathbb{E}_n| = O_p(n^{-4/5}) + o_p((\tau_n^+ - \tau_n^-)^4). \tag{4.2}$$

On the other hand, Groeneboom, Jongbloed, and Wellner (2001b) established that there exists a universal constant $C > 0$ such that

$$\begin{aligned} &\frac{Y_0(\tau_n^-) + Y_0(\tau_n^+)}{2} - \frac{G_0(\tau_n^+) - G_0(\tau_n^-)}{8}\,(\tau_n^+ - \tau_n^-) \\ &= -C g_0''(x_0)(\tau_n^+ - \tau_n^-)^4 + \ o_p((\tau_n^+ - \tau_n^-)^4). \end{aligned} \tag{4.3}$$

Combining the results in (4.2) and (4.3), it follows that

$$\tau_n^+ - \tau_n^- = O_p(n^{-1/5}).$$

The problem has two main features that make the above arguments work. First of all, the polynomial $\tilde{H}_n$ can be fully determined on $[\tau_n^-, \tau_n^+]$ and therefore it can be evaluated at any point between τ_n^- and τ_n^+. Second of all, it can be expressed via the empirical process $\mathbb{Y}_n$ and that enables us to "get rid of" terms depending on $\tilde{g}_n$ whose rate of convergence is still unknown at this stage. We should also add that the problem is symmetric around $\bar{\tau}_n$, a property that helps establishing the formula derived in (4.3).

When $k > 2$, it follows from the characterization of the LSE given in (2.12), that for any two successive jump points of $\tilde{g}_n^{(k-1)}$, τ_n^-, τ_n^+, the four equalities

$$\tilde{H}_n(\tau_n^\pm) = \mathbb{Y}_n(\tau_n^\pm), \quad \text{and} \quad \tilde{H}_n'(\tau_n^\pm) = \mathbb{Y}_n'(\tau_n^\pm)$$

still hold. However, these equations are not enough to determine the polynomial $\tilde{H}_n$, now of degree $2k-1$, on the interval $[\tau_n^-, \tau_n^+]$. One would need $2k$ conditions to be able to achieve this. [We would be in this situation if we had equality of the higher derivatives of $\tilde{H}_n$ and $\mathbb{Y}_n$ at τ_n^- and τ_n^+, that is

$$\tilde{H}_n^{(j)}(\tau_n^-) = \mathbb{Y}_n^{(j)}(\tau_n^-), \qquad \tilde{H}_n^{(j)}(\tau_n^+) = \mathbb{Y}_n^{(j)}(\tau_n^+) \tag{4.4}$$

for $j = 0, \ldots, k-1$, but the characterization (2.12) does not give this much.] Thus it becomes clear that two jump points are not sufficient to determine the piecewise polynomial $\tilde{H}_n$. However, if we consider $p > 2$ jump points $\tau_{n,0} < \cdots < \tau_{n,p-1}$ (all located e.g. after x_0), then $\tilde{H}_n$ is a spline of degree $2k-1$ with interior knots $\tau_{n,1}, \cdots, \tau_{n,p-2}$; that is, $\tilde{H}_n$ is a polynomial of degree $2k-1$ on $(\tau_{n,j}, \tau_{n,j+1})$ for $j = 0, \ldots, p-2$ and is $(2k-2)$-times differentiable at its knot points $\tau_{n,0}, \ldots, \tau_{n,p-1}$. In the next subsection, we prove that if $p = 2k-2$, the spline $\tilde{H}_n$ is completely determined on $[\tau_{n,0}, \tau_{n,2k-3}]$ by the conditions

$$\tilde{H}_n(\tau_{n,i}) \;=\; \mathbb{Y}_n(\tau_{n,i}), \;\text{ and }\; \tilde{H}_n'(\tau_{n,i}) = \mathbb{Y}_n'(\tau_{n,i}), \qquad i = 0, \ldots, 2k-3\,. \tag{4.5}$$

This result proves to be very useful for determining the stochastic order of the distance between two successive jump points in a small neighborhood of x_0 if our Conjecture (1.7) on the uniform boundedness of the error in the "non-classical" Hermite interpolation problem via splines of odd-degree defined in (1.6) holds.

4.2. The gap problem for the LSE - Hermite interpolation

In the next lemma, we prove that given $2k-2$ successive jump points $\tau_{n,0} < \cdots < \tau_{n,2k-3}$ of $\tilde{g}_n^{(k-1)}$, $\tilde{H}_n$ is the unique solution of the Hermite problem given by (4.5). In the following, we will omit writing the subscript n explictly in the knots, but their dependence on the sample size should be kept in mind.

Lemma 4.1 *The function $\tilde{H}_n$ characterized by (2.12) is a spline of degree $2k-1$. Moreover, given any $2k-2$ successive jump points of $\tilde{H}_n^{(2k-1)}$, $\tau_0 < \ldots < \tau_{2k-3}$, the $(2k-1)$-th spline $\tilde{H}_n$ is uniquely determined on $[\tau_0, \tau_{2k-3}]$ by the values of the process $\mathbb{Y}_n$ and of its derivative $\mathbb{Y}_n'$ at $\tau_0, \ldots, \tau_{2k-3}$.*

Proof. We know that for any jump point τ of $\tilde{H}_n^{(2k-1)}$, we have

$$\tilde{H}_n(\tau) = \mathbb{Y}_n(\tau) \quad \text{ and } \quad \tilde{H}_n'(\tau) = \mathbb{Y}_n'(\tau).$$

This can be viewed as a *Hermite interpolation problem* if we consider that the *interpolated function* is the process $\mathbb{Y}_n$ and that the *interpolating spline* is $\tilde{H}_n$ (see e.g. Nürnberger (1989), Definition 3.6, pages 108 and 109). Existence and uniqueness of the spline interpolant follows easily from the Schoenberg-Whitney-Karlin-Ziegler Theorem (Schoenberg and Whitney (1953); Theorem 3, page 529, Karlin and Ziegler (1966); or see Theorem 3.7, page 109, Nürnberger (1989); or Theorem 9.2, page 162, DeVore and Lorentz (1993)).

In the following lemma, we prove a preparatory result that will be used later for deriving the stochastic order of the distance between successive knots of $\tilde{g}_n$ in a neighborhood of x_0. Given a fixed set of points $\tau_0, \cdots, \tau_{2k-3}$, let $\mathcal{H}_k$ denote again the spline interpolation operator which assigns to each differentiable function f the unique spline $\mathcal{H}_k[f]$ with interior knots $\tau_1, \cdots,$ τ_{2k-4} and degree $2k-1$, and satisfying the boundary conditions given in (1.6).

Lemma 4.2 *Let $\bar{\tau} \in \cup_{i=0}^{2k-4}(\tau_i, \tau_{i+1})$. If $e_k(t)$ denotes the error at t of the Hermite interpolation of the function $x^{2k}/(2k)!$; i.e.,*

$$e_k(t) = \frac{t^{2k}}{(2k)!} - \mathcal{H}_k\left[\frac{x^{2k}}{(2k)!}\right](t)$$

then

$$g_0^{(k)}(\bar{\tau})e_k(\bar{\tau}) \le \mathbb{E}_n + \mathbb{R}_n \tag{4.6}$$

where $\mathbb{E}_n$ defined in (4.8) is a random error and $\mathbb{R}_n$ defined in (4.9) is a remainder that both depend on the knots $\tau_0, \ldots, \tau_{2k-3}$ and the point $\bar{\tau}$.

Proof. Let $\bar{\tau} \in \cup_{i=0}^{2k-4}(\tau_i, \tau_{i+1})$. From the characterization in (2.12) and the fact that $\tilde{H}_n = \mathcal{H}_k[\mathbb{Y}_n]$ on $[\tau_0, \tau_{2k-3}]$, it follows that

$$\mathcal{H}_k[\mathbb{Y}_n](\bar{\tau}) \ge \mathbb{Y}_n(\bar{\tau}).$$

Let Y_0 the true counterpart of $\mathbb{Y}_n$; i.e., $Y_0(x) = \int_0^x (x-t)^{k-1}/(k-1)!\, g_0(t)dt$. Then, we can rewrite the previous inequality as

$$\mathcal{H}_k[Y_0](\bar{\tau}) - Y_0(\bar{\tau}) \geq -\mathbb{E}_n(\bar{\tau}) \tag{4.7}$$

where

$$\mathbb{E}_n = \mathcal{H}_k[\mathbb{Y}_n - Y_0](\bar{\tau}) - [\mathbb{Y}_n - Y_0](\bar{\tau}). \tag{4.8}$$

Based on the working assumptions, the function Y_0 is $(2k)$-times continuously differentiable in a small neighborhood of x_0. Now, Taylor expansion of $Y_0(t)$, with integral remainder, around $\bar{\tau}$ up to the order $2k$ yields

$$Y_0(t) = \sum_{j=0}^{2k-1} \frac{(t-\bar{\tau})^j}{j!} Y_0^{(j)}(\bar{\tau}) + \int_{\bar{\tau}}^{\tau_{2k-3}} \frac{(t-u)_+^{2k-1}}{(2k-1)!} g_0^{(k)}(u)du$$

for all $t \in [\tau_0, \tau_{2k-3}]$. Using this expansion along with the fact that the operator $\mathcal{H}_k$ is linear and does preserve polynomials of degree $2k-1$, we can rewrite the inequality in (4.7) as

$$\frac{1}{(2k-1)!} \int_{\bar{\tau}}^{\tau_{2k-3}} \mathcal{H}_k[(t-u)_+^{2k-1}](\bar{\tau})\ g_0^{(k)}(u)du \geq -\mathbb{E}_n.$$

In the previous display, $\mathcal{H}_k[(t-u)_+^{2k-1}](\bar{\tau})$ is the Hermite spline interpolant of the truncated power function $t \mapsto (t-u)_+^{2k-1}$ (u is fixed), evaluated at the point $\bar{\tau}$. Now, we can rewrite the left side of the previous inequality as

$$\begin{aligned}
&\int_{\bar{\tau}}^{\tau_{2k-3}} \frac{1}{(2k-1)!} \mathcal{H}_k[(t-u)_+^{2k-1}](\bar{\tau})\ g_0^{(k)}(u)du \\
&= g_0^{(k)}(\bar{\tau}) \frac{1}{(2k-1)!} \int_{\bar{\tau}}^{\tau_{2k-3}} \mathcal{H}_k[(t-u)_+^{2k-1}](\bar{\tau})du \\
&\quad + \frac{1}{(2k-1)!} \int_{\bar{\tau}}^{\tau_{2k-3}} \mathcal{H}_k[(t-u)_+^{2k-1}](\bar{\tau}) \left(g_0^{(k)}(u) - g_0^{(k)}(\bar{\tau})\right) du \\
&= g_0^{(k)}(\bar{\tau}) \frac{1}{(2k-1)!} \mathcal{H}_k \left[\int_{\bar{\tau}}^{\tau_{2k-3}} [(t-u)_+^{2k-1}]du\right](\bar{\tau}) + \mathbb{R}_n, \qquad (4.9)
\end{aligned}$$

using once again linearity of the operator $\mathcal{H}_k$. The remainder $\mathbb{R}_n$ is equal to the Hermite interpolant of the function

$$t \mapsto \frac{1}{(2k-1)!} \int_{\bar{\tau}}^{t} \frac{(t-u)^{2k-1}}{(2k-1)!} (g_0^{(k)}(u) - g_0^{(k)}(\bar{\tau}))du$$

at the point $\bar{\tau}$. On the other hand, we can further rewrite the integral term in (4.9) as

$$\frac{1}{(2k-1)!}\mathcal{H}_k\left[\int_{\bar{\tau}}^{\tau_{2k-3}}(t-u)_+^{2k-1}du\right](\bar{\tau})$$
$$= \frac{1}{(2k-1)!}\mathcal{H}_k\left[\int_{\bar{\tau}}^{t}(t-u)^{2k-1}du\right](\bar{\tau}) = \frac{1}{(2k)!}\mathcal{H}_k\left[(t-\bar{\tau})^{2k}\right](\bar{\tau}).$$

In other words, the integral term in (4.9) is nothing but the value of the Hermite spline interpolant of the function $t \mapsto (t-\bar{\tau})^{2k}/(2k)!$ at the point $\bar{\tau}$. As claimed in the lemma, this value is also equal to $-e_k(\bar{\tau})$,where e_k the error of the Hermite interpolation of the function $x^{2k}/(2k)!$. Indeed, let $P_{2k-1}(t) = (t-\bar{\tau})^{2k}/(2k)! - t^{2k}/(2k)!$. Since P_{2k-1} is a polynomial of degree $2k-1$, we have

$$\mathcal{H}_k\left[\frac{(x-\bar{\tau})^{2k}}{(2k)!}\right](t) = \mathcal{H}_k\left[\frac{x^{2k}}{(2k)!}\right](t) + P_{2k-1}(t).$$

If $t = \bar{\tau}$, $P_{2k-1}(\bar{\tau}) = 0 - \bar{\tau}^{2k}/(2k)! = -\bar{\tau}^{2k}/(2k)!$, which implies that

$$\mathcal{H}_k\left[\frac{(x-\bar{\tau})^{2k}}{(2k)!}\right](\bar{\tau}) = \mathcal{H}_k\left[\frac{x^{2k}}{(2k)!}\right](\bar{\tau}) - \frac{\bar{\tau}^{2k}}{(2k)!} = -e_k(\bar{\tau}).$$

■

The error e_k defined in Lemma 4.2 can be recognized as a monospline of degree $2k$ with $2k-2$ simple knots $\tau_0, \cdots, \tau_{2k-3}$. For a definition of monosplines, see e.g. Michelli (1972), Bojanov, Hakopian and Sahakian (1993), Nürnberger (1989), page 194 or DeVore and Lorentz (1993), page 136. In the next lemma, we state an important property of e_k.

Lemma 4.3 *The function $x \mapsto e_k(x)$ has no other zeros than $\tau_0, \ldots, \tau_{2k-3}$ in $[\tau_0, \tau_{2k-3}]$. Furthermore, $(-1)^k e_k \geq 0$ on $[\tau_0, \tau_{2k-3}]$.*

Proof. See Appendix 3. ■

In Lemma 4.2, the key inequality in (4.6) can be rewritten as

$$(-1)^k g_0^{(k)}(\bar{\tau}) \cdot (-1)^k e_k(\bar{\tau}) \leq \mathbb{E}_n + \mathbb{R}_n, \tag{4.10}$$

where the first factor on the right side is already known to be positive by k-monotonicity of g_0. Lemma 4.4 and Lemma 4.5 are the final steps toward establishing the order of the gap for the LSE based on the Conjecture (1.7).

Lemma 4.4 *If Conjecture (1.7) holds, then* $\mathbb{E}_n$ *in (4.6) of Lemma 4.2 satisfies*

$$|\mathbb{E}_n| = O_p(n^{-k/(2k+1)}) + o_p((\tau_{2k-3} - \tau_0)^{2k}).$$

Proof. We have

$$\mathbb{E}_n = \mathcal{H}_k[\mathbb{Y}_n - Y_0](\bar{\tau}) - [\mathbb{Y}_n - Y_0](\bar{\tau}).$$

Using the (generalized) Taylor expansion of $\mathbb{Y}_n(t)$ and $Y_0(t)$ around the point $\bar{\tau}$ up to the order $k-1$ yields

$$\begin{aligned} \mathbb{Y}_n(t) - Y_0(t) &= \sum_{j=0}^{k-1} \frac{(t-\bar{\tau})^j}{j!}[\mathbb{Y}_n^{(j)}(\bar{\tau}) - Y_0^{(j)}(\bar{\tau})] \\ &\quad + \int_{\bar{\tau}}^{t} \frac{(x-u)^{k-1}}{(k-1)!} d(\mathbb{G}_n - G_0)(x), \end{aligned}$$

and therefore,

$$\begin{aligned} \mathbb{E}_n &= \mathcal{H}_k\left[\int_{\bar{\tau}}^{t} \frac{1}{(k-1)!}(t-x)^{k-1} d(\mathbb{G}_n - G_0)(x)\right](\bar{\tau}) \\ &= \mathcal{H}_k\left[\int_{\bar{\tau}}^{\tau_{2k-3}} g_t(x) d(\mathbb{G}_n - G_0)(x)\right](\bar{\tau}), \quad \text{where } g_t(x) = \tfrac{(t-x)_+^{k-1}}{(k-1)!} \\ &= \int_{\bar{\tau}}^{\tau_{2k-3}} \mathcal{H}_k[g_t(x)](\bar{\tau}) d(\mathbb{G}_n - G_0)(x), \quad \text{by linearity of } \mathcal{H}_k \\ &= \int_{\tau_0}^{\tau_{2k-3}} f_{\bar{\tau}}(x) d(\mathbb{G}_n - G_0)(x). \end{aligned}$$

Given $x \in [\bar{\tau}, \tau_{2k-3}]$, $f_{\bar{\tau}}(x) = \mathcal{H}_k[g_t(x)](\bar{\tau}) 1_{[\bar{\tau}, \tau_{2k-3}]}(x)$, where $\mathcal{H}_k[g_t(x)](\bar{\tau})$ is the value at $\bar{\tau}$ of the Hermite spline interpolant of the function $t \mapsto g_t(x) = (t-x)_+^{k-1}/(k-1)!$. Thus, $f_{\bar{\tau}}(x)$ depends on the knots $\tau_0, \cdots, \tau_{2k-3}$, and the point $s = \bar{\tau} \in [\tau_0, \tau_{2k-3}]$, and can be viewed as an element of the class of functions

$$\mathcal{F}^{(1)}_{y_0,R} = \{f_s(x) = f_{s,y_0,\ldots,y_{2k-3}}(x) : x \in [y_0, y_{2k-3}],\ s \in [y_0, y_{2k-3}], \\ x_0 - \delta \le y_0 < y_1 < \cdots < y_{2k-3} \le y_0 + R\}. \quad (4.11)$$

In view of Conjecture (1.7) together with the triangle inequality, there exists a constant $C > 0$ depending only on k such that

$$|f_s(x)| \le C(y_{2k-3} - y_0)^{k-1} 1_{[y_0, y_{2k-3}]}(x)$$

and hence the collection $\mathcal{F}^{(1)}_{y_0,R}$ has envelope function $F_{y_0,R}$ given by

$$F_{y_0,R}(x) = CR^{k-1} 1_{[y_0, y_0+R]}(x).$$

Furthermore, $\mathcal{F}^{(1)}_{y_0,R}$ is a VC-subgraph collection of functions (see Lemma 7.1 in Appendix 2, for a detailed argument), and hence by van der Vaart and Wellner (1996), Theorem 2.6.7, page 141

$$\sup_Q N(\epsilon \|F\|_{Q,2}, \mathcal{F}^{(1)}_{y_0,R}, L_2(Q)) \le \left(\frac{K}{\epsilon}\right)^{V_k}$$

for $0 < \epsilon < 1$ where $V_k = 2(V(\mathcal{F}_{y_0,R}) - 1)$ with $V(\mathcal{F}_{y_0,R})$ the VC-dimension of the collection of subgraphs and the constant K depends only on $V(\mathcal{F}_{y_0,R})$. It follows that

$$\sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon \|F_{y_0,R}\|_{Q,2}, \mathcal{F}^{(1)}_{y_0,R}, L_2(Q))} d\epsilon < \infty.$$

On the other hand, if $y_0 \in [x_0 - \delta, x_0 + \delta]$ (an event which occurs with increasing probability) for some small $\delta > 0$, then we can find a constant $M > 0$ depending only on δ, and g_0 such that $0 < \sup_{t \in [y_0, y_0+R]} g_0(t) < M$. Therefore,

$$EF^2_{y_0,R}(X_1) = C^2 R^{2(k-1)} \int_{y_0}^{y_0+R} g_0(x) dx \le C^2 M R^{2k-1}.$$

Application of Lemma 7.1 with $d = k$ and $\alpha = k$ yields

$$|\mathbb{E}_n| = o_p((\tau_{2k-3} - \tau_0)^{2k}) + O_p(n^{-2k/(2k+1)}).$$

■

Lemma 4.5 *If the bound in (1.8) holds, then $\mathbb{R}_n$ of Lemma 4.2 satisfies*

$$|\mathbb{R}_n| = o_p((\tau_{2k-3} - \tau_0)^{2k}).$$

Proof. By definition, $\mathbb{R}_n$ is the value at $\bar{\tau}$ of the Hermite spline interpolant of the function

$$t \mapsto \int_{\bar{\tau}}^{t} \frac{(t-u)^{2k-1}}{(2k-1)!}(g_0^{(k)}(u) - g_0^{(k)}(\bar{\tau}))du \tag{4.12}$$

By (1.8), there exists a constant $D > 0$ depending only on k such that

$$|\mathbb{R}_n| \le D \sup_{t \in [\tau_0, \tau_{2k-3}]} |g_0^{(k)}(t) - g_0^{(k)}(\bar{\tau})| \, (\tau_{2k-3} - \tau_0)^{2k}.$$

In the previous bound, we used the fact that the $(2k)$-times derivative of the function in (4.12) is $g_0^{(k)}(t) - g_0^{(k)}(\bar{\tau})$. But, note that this derivative is $o_p(1)$, which follows from uniform continuity of $g_0^{(k)}$ on compacts. This in turn implies the claimed bound. ■

Proof of Lemma 3.1 for the LSE. Let $j_0 \in \{0, \ldots, 2k-4\}$ be such that $[\tau_{j_0}, \tau_{j_0+1}]$ is the largest knot interval; i.e., $\tau_{j_0+1} - \tau_{j_0} = \max_{0 \le j \le 2k-4}(\tau_{j+1} - \tau_j)$. Let $a = \tau_0$, $b = \tau_{2k-3}$. Using the inequality in (4.10) and since the bounds on $\mathbb{R}_n$ and $\mathbb{E}_n$ are independent of the choice of $\bar{\tau}$ in $\cup_{j=0}^{2k-4}(\tau_j, \tau_{j+1})$, it follows that

$$\sup_{\bar{\tau} \in (\tau_{j_0}, \tau_{j_0+1})} (-1)^k e_k(\bar{\tau}) \le O_p(n^{-2k/(2k+1)}) + o_p((\tau_{2k-3} - \tau_0)^{2k}).$$

Now, on the interval $[\tau_{j_0}, \tau_{j_0+1}]$, the Hermite spline interpolant of the function $x^{2k}/(2k)!$ reduces to a polynomial of degree $2k-1$. On the other hand, the best uniform approximation of the function x^{2k} on $[\tau_{j_0}, \tau_{j_0+1}]$ from the space of polynomials of degree $\leq 2k-1$ is given by the polynomial

$$x \mapsto x^{2k} - \left(\frac{\tau_{j_0+1} - \tau_{j_0}}{2}\right)^{2k} \frac{1}{2^{2k-1}} T_{2k}\left(\frac{2x - (\tau_{j_0} + \tau_{j_0+1})}{\tau_{j_0+1} - \tau_{j_0}}\right), \quad (4.13)$$

where T_{2k} is the Chebyshev polynomial of degree $2k$ (defined on $[-1, 1]$), see, e.g., Nürnberger (1989), Theorem 3.23, page 46 or DeVore and Lorentz (1993), Theorem 6.1, page 75. It follows that

$$\begin{aligned} \sup_{\bar{\tau} \in (\tau_{j_0}, \tau_{j_0+1})} (-1)^k e_k(\bar{\tau}) &\geq \left\| \frac{T_{2k}}{2^{4k-1}(2k)!} \right\|_\infty (\tau_{j_0+1} - \tau_{j_0})^{2k} \quad (4.14) \\ &= \frac{1}{2^{4k-1}(2k)!} (\tau_{j_0+1} - \tau_{j_0})^{2k} \end{aligned}$$

since $\|T_{2k}\|_\infty = 1$. But,

$$\tau_{2k-3} - \tau_0 = \sum_{j=0}^{2k-4} (\tau_{j+1} - \tau_j) \leq (2k-3)(\tau_{j_0+1} - \tau_{j_0}).$$

Hence,

$$\sup_{\bar{\tau} \in (\tau_{j_0}, \tau_{j_0+1})} (-1)^k e_k(\bar{\tau}) \geq \frac{1}{(2k-3)^{2k} 2^{4k-1}(2k)!} (\tau_{2k-3} - \tau_0)^{2k}.$$

Combining the results obtained above, we conclude that

$$\frac{(-1)^k g_0^{(k)}(x_0)}{(2k-3)^{2k} 2^{4k-1}(2k)!} (\tau_{2k-3} - \tau_0)^{2k} \leq O_p(n^{-2k/(2k+1)}) + o_p((\tau_{2k-3} - \tau_0)^{2k})$$

which implies that $\tau_{2k-3} - \tau_0 = O_p(n^{-1/(2k+1)})$. ■

4.3. The gap problem for the MLE

To show Lemma 3.1 for the MLE, one needs to deal with an extra diffucutly posed by the nonlinear form of the characterization of this estimator, given in (2.14). In the following, we show how one can get around this difficulty. The main idea is to "linearize" the characterization of the MLE, and hence be able to re-use the arguments developed for the LSE in the previous subsetion.

Lemma 4.6 *Let $\tau_0, \cdots, \tau_{2k-3}$ be $2k-2$ sucessive jump points of $\hat{g}_n^{(k-1)}$. Then,*

$$\mathcal{H}_k[\mathbb{Y}_n] - \mathbb{Y}_n \geq g_0(\tau_0)\left(\check{f}_n - \mathcal{H}_k[\check{f}_n] + \Delta_n - \mathcal{H}_k\left[\Delta_n\right]\right)$$

on $[\tau_0, \tau_{2k-3}]$, where $\mathbb{Y}_n$ is the same empirical process introduced in (2.10),

$$\check{f}_n(x) \equiv -\int_{\tau_0}^{t} \frac{(x-t)^{k-1}}{(k-1)!}\left(\frac{1}{\hat{g}_n(t)} - \frac{1}{g_0(\tau_0)}\right) d(\hat{G}_n(t) - G_0(t))$$

and

$$\Delta_n(x) \equiv \int_{\tau_0}^{x} \frac{(x-t)^{k-1}}{(k-1)!}\left(\frac{1}{\hat{g}_n(t)} - \frac{1}{g_0(\tau_0)}\right) d(\mathbb{G}_n(t) - G_0(t)).$$

Proof. Let $\hat{G}_n(x) = \int_0^x \hat{g}_n(s)ds$. The characterization in (2.14) can be rewritten as

$$\int_0^x \frac{(x-t)^{k-1}}{\hat{g}_n(t)}\, d(\hat{G}_n(t) - \mathbb{G}_n(t)) \begin{cases} \geq 0, & \text{for } x > 0 \\ = 0, & \text{if } x \text{ is a jump point of } \hat{g}_n^{(k-1)}. \end{cases} \tag{4.15}$$

Note that when x is a jump point of $\hat{g}_n^{(k-1)}$, the two parts of (4.15) imply that the first derivative of the function on the right side is equal to 0 at the jump point x; i.e.,

$$\int_0^x \frac{(x-t)^{k-2}}{\hat{g}_n(t)}\, d(\hat{G}_n(t) - \mathbb{G}_n(t)) = 0. \tag{4.16}$$

For $x > 0$, let

$$\hat{H}_n(x) = \int_0^x \frac{(x-t)^{k-1}}{(k-1)!}\, d\hat{G}_n(t).$$

Note that $\hat{H}_n \neq \widehat{H}_n$ defined in (2.13), and on $[\tau_0, \tau_{2k-3}]$, $\hat{H}_n$ is a spline of degree $2k-1$ with knots $\tau_0, \cdots, \tau_{2k-3}$. For $x \in [\tau_0, \tau_{2k-3}]$, we can write

$$\int_0^x \frac{(x-t)^{k-1}}{\hat{g}_n(t)}\, d(\hat{G}_n(t) - \mathbb{G}_n(t))$$

$$
\begin{aligned}
&= \frac{1}{g_0(\tau_0)} \int_0^x (x-t)^{k-1}\, d(\hat{G}_n(t) - \mathbb{G}_n(t)) \\
&\quad + \int_0^x (x-t)^{k-1} \left(\frac{1}{\hat{g}_n(t)} - \frac{1}{g_0(\tau_0)} \right) d(\hat{G}_n(t) - \mathbb{G}_n(t)) \\
&= \frac{(\hat{H}_n(x) - \mathbb{Y}_n(x))}{g_0(\tau_0)} + \int_0^{\tau_0} (x-t)^{k-1} \left(\frac{1}{\hat{g}_n(t)} - \frac{1}{g_0(\tau_0)} \right) d(\hat{G}_n(t) - \mathbb{G}_n(t)) \\
&\quad + \int_{\tau_0}^x (x-t)^{k-1} \left(\frac{1}{\hat{g}_n(t)} - \frac{1}{g_0(\tau_0)} \right) d(\hat{G}_n(t) - G_0(t)) \\
&\quad + \int_{\tau_0}^x (x-t)^{k-1} \left(\frac{1}{\hat{g}_n(t)} - \frac{1}{g_0(\tau_0)} \right) d(G_0(t) - \mathbb{G}_n(t)) \\
&= \frac{1}{g_0(\tau_0)} (\hat{H}_n(x) - \mathbb{Y}_n(x)) + p_n(x) - \check{f}_n(x) - \Delta_n(x).
\end{aligned}
$$

Note that

$$
p_n(x) \equiv \int_0^{\tau_0} (x-t)^{k-1} \left(\frac{1}{\hat{g}_n(t)} - \frac{1}{g_0(\tau_0)} \right) d(\hat{G}_n(t) - \mathbb{G}_n(t))
$$

is a polynomial of degree $k-1$. From (4.15) and (4.16), it follows that $\hat{H}_n$ is the Hermite spline interpolant of the function

$$
\mathbb{Y}_n + g_0(\tau_0) \left\{ -p_n + \check{f}_n + \Delta_n \right\}
$$

such that

$$
\hat{H}_n \geq \mathbb{Y}_n + g_0(\tau_0)(-p_n + \check{f}_n + \Delta_n).
$$

Hence,

$$
\mathcal{H}_k \left[\mathbb{Y}_n + g_0(\tau_0) \left\{ -p_n + \check{f}_n + \Delta_n \right\} \right] \geq \mathbb{Y}_n + g_0(\tau_0) \left\{ -p_n + \check{f}_n + \Delta_n \right\}
$$

on $[\tau_0, \tau_{2k-3}]$, or equivalently

$$
\mathcal{H}_k[\mathbb{Y}_n] - \mathbb{Y}_n \geq g_0(\tau_0) \left(\check{f}_n - \mathcal{H}_k[\check{f}_n] + \Delta_n - \mathcal{H}_k\left[\Delta_n\right] \right).
$$

■

As $\mathcal{H}_k[\mathbb{Y}_n] - \mathbb{Y}_n$ has been already studied for proving the order of the gap in the case of the LSE, the final step is to evaluate each of the interpolation errors

$$
\mathcal{E}_1 = \check{f}_n - \mathcal{H}_k[\check{f}_n] \text{ and } \mathcal{E}_2 = \Delta_n - \mathcal{H}_k[\Delta_n]. \tag{4.17}
$$

Lemma 4.7 *Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be the interpolation errors defined in (4.17). Then,*

$$\|\mathcal{E}_1\|_\infty = o_p((\tau_{2k-3} - \tau_0)^{2k}) \text{ and } \|\mathcal{E}_2\|_\infty = o_p((\tau_{2k-3} - \tau_0)^{2k}) + O_p\left(n^{-\frac{2k}{2k+1}}\right).$$

Proof. See Appendix 3.

Proof of Lemma 3.1 for the MLE. From our study of the distance between the knots of the LSE, and using very similar caluculations we can show that for all $\bar{\tau} \in \cup_{j=0}^{2k-4}(\tau_j, \tau_{j+1})$

$$(-1)^k g_0^{(k)}(\bar{\tau})(-1)^k e_k(\bar{\tau}) \le \mathbb{E}_n + \mathbb{R}_n - g_0(\tau_0)(\mathcal{E}_1(\bar{\tau}) + \mathcal{E}_2(\bar{\tau})),$$

which implies that by the results obtained for the LSE

$$D(\tau_{2k-3} - \tau_0)^{2k}(1 + o_p(1)) \quad \le \quad O_p\left(n^{-\frac{2k}{2k+1}}\right) + g_0(\tau_0)\left(\|\mathcal{E}_1\|_\infty + \|\mathcal{E}_2\|_\infty\right)$$

for some constant $D > 0$ depending on k and x_0. Hence, it follows from Lemma 4.7 that

$$D(\tau_{2k-3} - \tau_0)^{2k}(1 + o_p(1)) \le O_p\left(n^{-\frac{2k}{2k+1}}\right)$$

which yields the order $n^{-1/(2k+1)}$ for the distance between the knots of the MLE in the neighborhood of x_0. ∎

5. Conclusions and discussion

As noted in Section 1, one of the motivations for this work was to try to approach the problem of pointwise limit theory for the MLE's in both the forward and inverse problems for the family of completely monotone densities on $\mathbb{R}^+$. This is one very important special case of the family of nonparametric mixture models with a smooth kernel as was mentioned in part (b) of our discussion in Section 1. Jewell (1982) established consistency of the MLE's of $g \in \mathcal{D}_\infty$ and the corresponding mixing distribution function F in this setting, but local rates of convergence and limiting distribution theory remain unknown. Our initial hope was that we might be able to learn about the problem with $k = \infty$ by studying the problem for fixed k, and then taking limits as $k \to \infty$. Unfortunately, we now believe that new tools and methods will be needed. Here is the state of affairs as we understand it now.

In terms of the rates of convergence, and localization properties, our development here shows that the local behavior of the estimators near a fixed point $x_0 > 0$ becomes dependent on an increasing number of jump points or knots in the spline problem. In other words, one needs to consider $2k-2$ consecutive jump points (knots) $\tau_{0,n} < \cdots < \tau_{n,2k-3}$ of the $(k-1)$-st derivative of the estimators in a neighborhood of x_0 in order to be able to find a bound on $\tau_{n,j+1} - \tau_{n,j}, j = 0, \ldots, 2k-4$ as $n \to \infty$. Thus the problem becomes increasingly "less local" with increasing k, and this leads us to suspect that the situation in the $k = \infty$ (or completely monotone) problem might be only "weakly local" or perhaps even "completely non-local" in senses yet to be precisely defined.

Another aspect of this problem is that although the MLE is asymptotically equivalent to the (mass unconstrained) LSE for each fixed k if our conjecture (1.7) holds, they seem to differ increasingly as k increases. For $k = 1$, the MLE and the LSE are identical; for $k = 2$, the MLE differs from the (mass unconstrained), but the LSE always has total mass 1. For $k \geq 3$, the MLE and LSE differ, and, moreover, the total amount of mass in the unconstrained LSE for $n = 1$ is $M_k = ((2k-1)/k)(1 - 1/(2k-1))^{k-1} \nearrow 2e^{-1/2} \approx 1.21306\ldots \neq 1$ as $k \to \infty$. We do know know how the mass of the unconstrained LSE behaves jointly in n and k, even though (by consistency) the mass of the LSE converges to 1 as $n \to \infty$ for fixed k. We also do not even know if the unconstrained LSE exists for the scale mixture of exponentials, even though it is clear that the constrained estimator (defined by the least squares criterion minimized over $\mathcal{D}_k$ rather than $\mathcal{M}_k$) with mass 1 does exist. Since our current proof techniques rely so heavily on showing equivalence between the MLE and the (unconstrained) LSE, it seems likely that new methods will be required. We do not know if the (mass) constrained LSE's and the MLE's are asymptotically equivalent either for finite k or for $k = \infty$. Our current plan is to study the constrained LSE's with total mass constrained to be 1 for finite sample sizes, to investigate the asymptotic equivalence of these mass-constrained LSE's and the MLE's, and to (perhaps) extend this study to $k = \infty$ via limits on k. We do not yet know the "right" Gaussian version of the estimation problem in the completely monotone case.

Another way to view these difficulties might be to take the following perspective: since more knowledge is available concerning the MLE's for the families $\mathcal{D}_k$ with k finite, and since $\mathcal{D}_\infty$ is the intersection of all the $\mathcal{D}_k$'s (and hence well-approximated by $\mathcal{D}_k$ with k large), we can fruitfully consider estimation via model selection, choosing k based on the data, over the collection $\cup_{k=1}^{\infty} \mathcal{D}_k$.

In summary, we have tried to shed some more light on the local behavior of two nonparametric estimators of a k-monotone density, the Maximum Likelihood and Least Squares estimators. We have shown that they are both adaptive splines of degree $k - 1$, with knots determined by the data and their corresponding criterion functions. When $(-1)^k g_0^{(k)}(x_0) > 0$, the distance between their knots in a neighborhood of a point $x_0 > 0$ was shown to be $n^{-1/(2k+1)}$ if a conjecture concerning the uniform boundedness of the interpolation error in a new Hermite interpolation problem holds, and once this control of the distance between the knots is available, pointwise limit distribution theory follows via a route paralleling previous results for $k = 1, 2$. Although we do not exclude the possibility that this order could be established via other different approaches, we hope that the techniques developed here demonstrate that there could still be many interesting and powerful connections between statistics and approximation theory.

6. Appendix 1 - Proofs for Subsection 3.2, and proof of Theorem 3.2

The proof of Proposition 3.1 will rely crucially on Proposition 6.1. Consider the event $J_n = J_n^{(1)} \cap J_n^{(2)}$ where $J_n^{(i)}$, $i = 1, 2$, are defined by

$$\begin{aligned} J_n^{(1)} &\equiv J_n^{(1)}(x_0, k, M) \\ &= \Big\{ \text{there exist } (k+1) \text{ jump points } \ \tau_{n,1}, \ldots, \tau_{n,k+1} \\ &\qquad \text{(not necessarily successive) satisfying} \\ &\qquad x_0 - n^{-1/(2k+1)} \le \tau_{n,1} < \cdots < \tau_{n,k+1} \le x_0 + M n^{-1/(2k+1)} \\ &\qquad k n^{-1/(2k+1)} \le \tau_{n,k+1} - \tau_{n,1} \le M n^{-1/(2k+1)} \Big\}, \end{aligned}$$

and

$$J_n^{(2)} \equiv J_n^{(2)}(j, k, c_j) = \left\{ \inf_{t \in [\tau_{n,1}, \tau_{n,k+1}]} \left| \bar{g}_n^{(j)}(t) - g_0^{(j)}(t) \right| \le c_j n^{-(k-j)/(2k+1)} \right\}.$$

Proposition 6.1 *Suppose that $(-1)^k g_0^{(k)}(x_0) > 0$ and $g_0^{(k)}$ is continuous in a neighborhood of x_0. Let $\bar{g}_n$ be either the MLE $\hat{g}_n$ or the LSE $\tilde{g}_n$ and let $0 \le j \le k-1$. Suppose also that $\int_0^\infty y^{-1/2} dG_0(y) < \infty$ holds. Then, if Conjecture (1.7) holds, for any $\epsilon > 0$, there exists $M > 0$ and $c_j > 0$ such that $P(J_n) > 1 - \epsilon$ for all sufficiently large n.*

Proof. Fix $\epsilon > 0$. In what follows, we consider only the LSE since the result in the case of the MLE can be proved similarly by using the same perturbation functions and uniform consistency of the estimator. We will start with $j = 0$. For ease of notation, we will write the jump points of $\tilde{g}_n^{(k-1)}$ without the subscript n. Let τ_1 be the first jump point of $\tilde{g}_n^{(k-1)}$ after $x_0 - n^{-1/(2k+1)}$, τ_2 the first jump point after $\tau_1 + n^{-1/(2k+1)}$, $\ldots$, τ_{k+1} the first jump point after $\tau_k + n^{-1/(2k+1)}$. By Lemma 3.1, there exists $M > 0$ such that

$$0 \le \tau_{k+1} - \tau_1 \le M n^{-1/(2k+1)}$$

with probability $> 1-\epsilon$. Note that by construction $\tau_{k+1} - \tau_1 \ge kn^{-1/(2k+1)}$. Fix $c > 0$ and consider the event

$$\inf_{t \in [\tau_1, \tau_{k+1}]} |\tilde{g}_n(t) - g_0(t)| > cn^{-k/(2k+1)}. \tag{6.18}$$

On this set and for any nonnegative function g on $[\tau_1, \tau_{k+1}]$, we have

$$\left| \int_{\tau_1}^{\tau_{k+1}} (\tilde{g}_n(t) - g_0(t))\, g(t) dt \right| \ge cn^{-k/(2k+1)} \int_{\tau_1}^{\tau_{k+1}} g(t) dt. \tag{6.19}$$

Now, let B be the B-spline of degree $k-1$ and with support $[x_1, x_{k+1}]$ (for definitions and basic material on B-splines, see e.g. NÜRNBERGER (1989), Theorems 2.6 - 29.9, pages 98 - 99). The B-spline is given by

$$B(t) = [x_1, \cdots, x_{k+1}](-1)^k k (t - \cdot)_+^{k-1}$$

where $[t_0, \cdots, t_m]g$ is the divided difference of degree m at the points $t_0, \cdots, t_m$; i.e.,

$$[t_0]g = g(t_0) \quad \text{and} \quad [t_0, \cdots, t_{m+1}]g = \frac{[t_1, \cdots, t_{m+1}]g - [t_0, \cdots, t_m]g}{t_{m+1} - t_0},$$

(see e.g. DEVORE AND LORENTZ (1993), pages 120 - 123, or DE BOOR (2001), pages 3-12)). After some algebra, we find that B can be given more explictly by

$$B(t) = (-1)^k k \left(\frac{(t - \tau_1)_+^{k-1}}{\prod_{j \ne 1} (\tau_j - \tau_1)} + \cdots + \frac{(t - \tau_k)_+^{k-1}}{\prod_{j \ne k} (\tau_j - \tau_k)} \right).$$

for all $t \in [\tau_1, \tau_{k+1}]$. Let $|\eta| > 0$ and consider the perturbation function

$$p(t) = \prod_{1 \le i < j \le k+1} (\tau_j - \tau_i) \times B(t).$$

It is easy to check that for $|\eta|$ small enough, the perturbed function

$$\tilde{g}_{\eta,n}(t) = \tilde{g}_n(t) + \eta p(t)$$

is k-monotone on $(0, \infty)$. Indeed, p was chosen so that it satisfies $p^{(j)}(\tau_1) = p^{(j)}(\tau_{k+1}) = 0$ for $0 \le j \le k-2$, which guarantees that the perturbed function $\tilde{g}_{\eta,n}$ belongs to $C^{k-2}(0,\infty)$. For $0 \le j \le k-3$, the properties of strict convexity and monotonicity of $(-1)^j \tilde{g}_n^{(j)}$ on $(0,\infty)$ are preserved by $\tilde{g}_{\eta,n}^{(j)}$ as long as $|\eta|$ is small enough. For $k-2$, $(-1)^{k-2}\tilde{g}_n^{(k-2)}$ is a convex and nonincreasing on $(0,\infty)$ piecewise linear function. Now note that p is a spline of degree $k-1$ whose knots are included in the set of knots of $\tilde{g}_n$. Moreover, for small values of $|\eta|$ it can be easily checked that $(-1)^{k-2}\tilde{g}_{\eta,n}^{(k-2)}$ is nonincreasing and convex on $(0,\infty)$.

It follows that

$$\lim_{\eta\to 0} \frac{Q_n(\tilde{g}_{\eta,n}) - Q_n(\tilde{g}_n)}{\eta} = 0\,.$$

This implies that

$$\int_{\tau_1}^{\tau_{k+1}} p(t)d(\tilde{G}_n - \mathbb{G}_n)(t) = 0.$$

The previous equality can be rewritten as

$$\int_{\tau_1}^{\tau_{k+1}} p(t)\left(\tilde{g}_n(t) - g_0(t)\right)dt = \int_{\tau_1}^{\tau_{k+1}} p(t)d(\mathbb{G}_n(t) - G_0(t)).$$

Taking $g \equiv p$ in (6.19), we obtain

$$\begin{aligned} \left|\int_{\tau_1}^{\tau_{k+1}} p(t)d(\mathbb{G}_n(t) - G_0(t))\right| &\ge cn^{-k/(2k+1)}\int_{\tau_1}^{\tau_{k+1}} p(t)dt \\ &= cn^{-k/(2k+1)}\prod_{1\le i<j\le k+1}(\tau_j - \tau_i) \quad (6.20)\\ &\ge cn^{-k/(2k+1)}\left(n^{-1/(2k+1)}\right)^{k(k+1)/2} \quad (6.21)\\ &= cn^{-(3+k)k/(2(2k+1))} \end{aligned}$$

where in (6.20), we used the fact that B-splines integrate to 1, whereas in (6.21) we used the facts that there are $k(k+1)/2$ terms in the product $\prod_{1\le i<j\le k+1}(\tau_j - \tau_i)$ and that $\tau_j - \tau_i \ge n^{-1/(2k+1)}, 1 \le i < j \le k+1$. Let $0 < y_0 < y_1 < \cdots < y_{k-1} < y_k$ be $(k+1)$ points in $(0,\infty)$ and consider the function $f_{y_0,y_1,\ldots,y_{k-1},y_k}$ defined by

$$f_{y_0,\ldots,y_k}(t) = (-1)^k k \prod_{0\le i<j\le k}(y_j - y_i)\sum_{l=0}^{k-1}\frac{(y_l - t)_+^{k-1}}{\prod_{j\ne l}(y_j - y_l)}$$

$$= \sum_{j=0}^{k-1} \alpha_j (y_j - t)_+^{k-1} \tag{6.22}$$

where

$$\alpha_j = (-1)^k k \frac{\prod_{0 \le l < l' \le k} (y_{l'} - y_l)}{\prod_{j' \ne j} (y_j - y_{j'})}.$$

Let $R > 0$ and consider the collection of functions

$$\mathcal{F}^{(2)}_{y_0,R} = \{ f_{y_0,\ldots,y_k}(t) : y_0 < y_1 < \cdots < y_{k-1} \le y_0 + R \} \tag{6.23}$$

where $f_{y_0,\ldots,y_{k-1}}$ is as defined in (6.22). Here the components of $y = (y_0, \ldots, y_{2k-3})$ play the role of the τ's. We first find an envelope function for the class $\mathcal{F}^{(2)}_{y_0,R}$. Note that for $j = 0, \ldots, k$, the product $\prod_{j' \ne j}(y_{j'} - y_j)$ contains k terms and hence α_j is a product of $k(k+1)/2 - k = k(k-1)/2$ terms that are at most R distant from one another. It follows that

$$\alpha_j \le k R^{k(k-1)/2}, \quad \text{for } j = 0, \ldots, k.$$

Thus the functions being summed in (6.22) have common envelope $kR^{k(k-1)/2}(y_0 + R - t)_+^{k-1} 1_{[y_0, y_0+R]}(t)$, and this yields the envelope

$$F_{y_0,R}(t) = k^2 R^{k(k-1)/2} (y_0 + R - t)_+^{k-1} 1_{[y_0, y_0+R]}(t)$$

for the class $\mathcal{F}^{(2)}_{y_0,R}$. Furthermore, $\mathcal{F}^{(2)}_{y_0,R}$ is a VC-subgraph collection of functions (see Lemma 7.1 in Appendix 2 for details of the argument), and hence by van der Vaart and Wellner (1996), Theorem 2.6.7, page 141,

$$\sup_Q N\left(\epsilon \| F_{y_0,R} \|_{Q,2}, \mathcal{F}^{(2)}_{y_0,R}, L_2(Q) \right) \le \left(\frac{K}{\epsilon} \right)^{V_k}.$$

for $0 < \epsilon < 1$ where $V_k = 2(V(\mathcal{F}) - 1)$ with $V(\mathcal{F}) = V(\mathcal{F}_{y_0,R})$ the VC - dimension of the collection of subgraphs. Therefore

$$\sup_Q \int_0^1 \sqrt{1 + \log(N \left(\epsilon \| F_{y_0,R} \|_{Q,2}, \mathcal{F}^{(2)}_{y_0,R}, L_2(Q) \right)} d\epsilon < \infty.$$

On the other hand, if y_0 is in a small neighborhood $[x_0 - \delta, x_0 + \delta]$ for some small $\delta > 0$, there exists some constant $C > 0$ depending only on δ, R and $g_0(x_0)$ such that $0 < g_0 < C$ on $[y_0, y_0 + R]$ for all $y_0 \in [x_0 - \delta, x_0 + \delta]$. It follows that

$$\begin{aligned} EF^2_{y_0,R}(X_1) &\leq k^4 R^{k(k-1)} \int_{y_0}^{y_0+R} (y_0 + R - x)^{2k-2} g_0(x)dx \\ &\leq \frac{k^4 C}{2k-1} R^{k(k-1)} R^{2k-1} = \frac{k^4 C}{2k-1} R^{k(k+1)-1}. \end{aligned}$$

Therefore, by van der Vaart and Wellner (1996), Theorem 2.14.1, we have

$$\begin{aligned} & E\left\{ \left(\sup_{f_{y_0,y_1,\ldots,y_k} \in \mathcal{F}^{(2)}_{y_0,R}} \left| (\mathbb{G}_n - G_0)(f_{y_0,y_1,\ldots,y_k}) \right| \right)^2 \right\} \\ & \leq \frac{K'}{n} EF^2_{y_0,R}(X_1) = O(n^{-1} R^{k(k+1)-1}), \end{aligned} \tag{6.24}$$

for some constant K' depending only on k, x_0, and δ. Application of Lemma 7.1 in Appendix 1 with $d = k(k+1)/2$ and $\alpha = k$ yields

$$\left| (\mathbb{P}_n - P_0)(f_{y_0,y_1,\ldots,y_k}) \right| \leq \epsilon (y_k - y_0)^{(3+k)k/2} + O_p\left(n^{-(3+k)k/(2(2k+1))} \right)$$

uniformly in $y_0, \ldots, y_k$. It follows that

$$\left| \int_{\tau_1}^{\tau_{k+1}} p(t) d(\mathbb{G}_n - G_0)(t) \right| = O_p\left(n^{-(3+k)k/(2(2k+1))} \right)$$

and we can choose $c_0 = c$ to be large enough so that the probability of the event (6.18) is arbitrarily small. This proves the result for $j = 0$.

Now let $1 \leq j \leq k - 1$. This time we will need $(k + 1 + j)$ jump points $\tau_1 < \cdots < \tau_{k+1+j}$. As for $j = 0$, τ_1 is taken to be the first jump point of $\tilde{g}_n^{(k-1)}$ after $x_0 - n^{-1/(2k+1)}$, τ_2 the first jump point after $\tau_1 + n^{-1/(2k+1)}$ and so on. Notice that the existence of at least $k + 1 + j$ jump points is guaranteed by the fact that $g_0^{(k)}(x_0) \neq 0$ which implies that with probability 1, the number of jump points tends to infinity with increasing sample size

n. Consider the function

$$q_j(t) = \prod_{1\le i<j\le k+j+1} (\tau_j - \tau_i) \times B_j(t)$$

where B_j is the B-spline of degree $k+j-1$ with support $[\tau_1, \tau_{k+1+j}]$; i.e.,

$$B_j(t) \;=\; (-1)^{k+j}(k+j)\left(\frac{(\tau_1 - t)_+^{k+j-1}}{\prod_{j\neq 1}(\tau_j-\tau_1)} + \cdots + \frac{(\tau_{k+j} - t)_+^{k+j-1}}{\prod_{j\neq k+j}(\tau_j-\tau_{k+j})}\right).$$

It is easy to check that $p_j = q_j^{(j)}$ is a valid perturbation function (it is a spline of degree $k-1$) since for $|\eta|$ small enough, the function

$$\tilde{g}_{\eta,n,j} = \tilde{g}_n + \eta p_j$$

is k-monotone. It follows that

$$\lim_{\eta\to 0} \frac{Q_n(\tilde{g}_{\eta,n,j}) - Q_n(\tilde{g}_n)}{\eta} = 0$$

which implies that

$$\int_{\tau_1}^{\tau_{k+1+j}} p_j(t)(\tilde{g}_n(t) - g_0(t))dt = \int_{\tau_1}^{\tau_{k+1+j}} p_j(t)d(\mathbb{G}_n(t) - G_0(t))dt.$$

By successive integrations by parts and using the fact that $q_j^{(i)}(\tau_1) = q_j^{(i)}(\tau_{k+1+j}) = 0$ for $i = 0, \ldots, k+j-2$, we obtain

$$\int_{\tau_1}^{\tau_{k+1+j}} (-1)^j q_j(t)(\tilde{g}_n^{(j)}(t) - g_0^{(j)}(t))dt = \int_{\tau_1}^{\tau_{k+1+j}} p_j(t)d(\mathbb{G}_n(t) - G_0(t))dt.$$

Therefore, if we assume that there exists $c > 0$ such that

$$\inf_{t\in[\tau_1,\tau_{k+1+j}]} \left|\tilde{g}_n^{(j)}(t) - g_0^{(j)}(t)\right| > c\, n^{-(k-j)/(2k+1)} \tag{6.25}$$

then

$$\begin{aligned}
&\left|\int_{\tau_1}^{\tau_{k+1+j}} p_j(t)d(\mathbb{G}_n(t) - G_0(t))dt\right| \\
&\quad\ge\; c\, n^{-(k-j)/(2k+1)} \int_{\tau_1}^{\tau_{k+1+j}} q_j(t)dt \\
&\quad\ge\; c\,(k+j)\, n^{-(k-j)/(2k+1)} \left(n^{-1/(2k+1)}\right)^{(k+1+j)(k+2+j)/2} \\
&\quad=\; c\,(k+j)\, n^{-((2(k-j)+(k+j)(k+j+1))/(2(2k+1))} \\
&\quad=\; c\,(k+j)\, n^{-(3k-j+(k+j)^2)/(2(2k+1))}.
\end{aligned}$$

Using similar empirical process arguments as in the proof for $j = 0$ together with an application of Lemma 7.1 with $2d = 3k - j + (k + j)^2$ and $\alpha = k$, it follows that

$$\left| \int_{\tau_1}^{\tau_{k+1+j}} p_j(t) d(\mathbb{G}_n(t) - G_0(t)) dt \right| = O_p \left(n^{-(3k-j+(k+j)^2)/(2(2k+1))} \right)$$

and the result for $1 \leq j \leq k - 1$ follows. ■

With Proposition 6.1 in hand we are prepared for the proof of the rates Proposition 3.1.

Proof of Proposition 3.1. We will use induction starting from the highest order of differentiation $k - 1$. The techniques used here are very much analogous to the ones used in the case $k = 2$ in Groeneboom, Jongbloed, and Wellner (2001b). But this was possible mainly because of the result established in the previous lemma. We begin by establishing the rate for $j = k-1$. Let $M > 0$ and $0 < \epsilon < 1$. We consider two sequences of $(k+1)$ jump points $\tau_{1,1}, \ldots, \tau_{k+1,1}$ and $\tau_{1,2}, \ldots, \tau_{k+1,2}$ as described in the previous proposition, where $\tau_{1,1}$ is the first jump point of $\bar{g}_n^{(k-1)}$ after $x_0 + Mn^{-1/(2k+1)}$ and $\tau_{1,2}$ is the first jump after $\tau_{k+1,1} + n^{-1/(2k+1)}$. Similarly, we define two other sequences $\tau_{1,-1} \ldots, \tau_{k+1,-1}$ and $\tau_{1,-2}, \ldots, \tau_{k+1,-2}$ to the left of x_0. By the previous theorem, we can find $c > 0$ so that,

$$\inf_{t \in [\tau_{1,i}, \tau_{k+1,i}]} |\bar{g}_n^{(k-2)}(t) - g_0^{(k-2)}(t)| < cn^{-2/(2k+1)}$$

for $i = -2, -1, 1, 2$ with probability greater than $1 - \epsilon$. Let ξ_1 and ξ_2 be the minimizer of $|\bar{g}_n^{(k-2)} - g_0^{(k-2)}|$ on $[\tau_{1,1}, \tau_{k+1,1}]$ and $[\tau_{1,2}, \tau_{k+1,2}]$ respectively. Define ξ_{-1} and ξ_{-2} similarly to the left of x_0. For all $t \in [x_0 - Mn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$, we have with probability greater than $1 - \epsilon$

$$\begin{aligned} (-1)^{k-2}\bar{g}_n^{(k-1)}(t-) &\leq (-1)^{k-2}\bar{g}_n^{(k-1)}(t+) \\ &\leq \frac{(-1)^{k-2}\bar{g}_n^{(k-2)}(\xi_2) - (-1)^{k-2}\bar{g}_n^{(k-2)}(\xi_1)}{\xi_2 - \xi_1} \end{aligned}$$

$$\leq \frac{(-1)^{k-2} g_0^{(k-2)}(\xi_2) - (-1)^{k-2} g_0^{(k-2)}(\xi_1) + 2cn^{-2/(2k+1)}}{\xi_2 - \xi_1}$$
$$\leq (-1)^{k-2} g_0^{(k-1)}(\xi_2) + 2cn^{-1/(2k+1)}$$

since $\xi_2 - \xi_1 \geq n^{-1/(2k+1)}$. Similarly, with probability greater than $1-\epsilon$, we have that

$$(-1)^{k-2}\bar{g}_n^{(k-1)}(t+) \geq (-1)^{k-2}\bar{g}_n^{(k-1)}(t-) \geq (-1)^{k-2} g_0^{(k-1)}(\xi_{-2}) - 2cn^{-1/(2k+1)}.$$

Now, using the fact that $\xi_{\pm 2} = x_0 + O_p(n^{-1/(2k+1)})$ and differentiability of $g_0^{(k-1)}$ at the point x_0, we obtain (3.1) for $j = 0$. Using similar arguments as in the proof of Lemma 4.4 in Groeneboom, Jongbloed, and Wellner (2001b), we can show (3.1) for $j = k-2$ which specializes to

$$\sup_{|t| \leq M} \left| \bar{g}_n^{(k-2)}(x_0 + n^{-1/(2k+1)}t) - g_0^{(k-2)}(x_0) - n^{-1/(2k+1)} t g_0^{(k-1)}(x_0) \right|$$
$$= O_p(n^{-2/(2k+1)})$$

for all $M > 0$. Indeed, since the jump points $\tau_{j,i}, j = 1, \ldots, k+1, i = -2, -1, 1, 2$ are at distance from x_0 that is $O_p(n^{-1/(2k+1)})$, we can find with probability exceeding $1-\epsilon$, $K > M$ such that ξ_1 and ξ_2 are in $[x_0 - Mn^{-1/(2k+1)}, x_0 + Kn^{-1/(2k+1)}]$, ξ_{-2} and ξ_{-1} in $[x_0 - Kn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$. But we know that, with probability greater than $1-\epsilon$, we can find $c > 0$ such that

$$|\bar{g}_n^{(k-2)}(\xi_{\pm 1}) - g_0^{(k-2)}(\xi_{\pm 1})| \leq cn^{-2/(2k+1)}.$$

Also, with probability greater than $1-\epsilon$, we can find $c' > 0$ such that

$$\sup_{t \in [x_0 - Kn^{-1/(2k+1)}, x_0 + Kn^{-1/(2k+1)}]} \left| \bar{g}_n^{(k-1)}(t) - g_0^{(k-1)}(x_0) \right| \leq c' n^{-1/(2k+1)}.$$

Hence, with probability greater than $1-3\epsilon$, we have for any $t \in [x_0 - Mn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$

$$(-1)^{k-2}\bar{g}_n^{(k-2)}(t)$$

$$
\begin{aligned}
&\geq\ (-1)^{k-2}\bar{g}_n^{(k-2)}(\xi_1)+(-1)^{k-2}\bar{g}_n^{(k-1)}(\xi_1)(t-\xi_1)\\
&\geq\ (-1)^{k-2}g_0^{(k-2)}(\xi_1)-cn^{-2/(2k+1)}+((-1)^{k-2}g_0^{(k-1)}(x_0)\\
&\qquad +c'n^{-1/(2k+1)})(t-\xi_1)\\
&\geq\ (-1)^{k-2}g_0^{(k-2)}(x_0)+(\xi_1-x_0)(-1)^{k-2}g_0^{(k-1)}(x_0)\\
&\qquad +(t-\xi_1)(-1)^{k-2}g_0^{(k-1)}(x_0)\\
&\qquad -cn^{-2/(2k+1)}-c'n^{-1/(2k+1)}(\xi_1-t) \qquad (6.26)\\
&\geq\ (-1)^{k-2}g_0^{(k-2)}(x_0)+(t-x_0)(-1)^{k-2}g_0^{(k-1)}(x_0)-(c+2Kc')n^{-2/(2k+1)}.
\end{aligned}
$$

where in (6.26), we used convexity of $(-1)^{k-2}g_0^{(k-2)}$ "from below". On the other hand, using convexity of $(-1)^{k-2}g_0^{(k-2)}$ but this time "from above", we have

$$
\begin{aligned}
&(-1)^{k-2}\bar{g}_n^{(k-2)}(t)\\
&\leq\ (-1)^{k-2}\bar{g}_n^{(k-2)}(\xi_{-1})+\frac{(-1)^{k-2}\bar{g}_n^{(k-2)}(\xi_1)-(-1)^{k-2}\bar{g}_n^{(k-2)}(\xi_{-1})}{\xi_1-\xi_{-1}}(t-\xi_{-1})\\
&\leq\ (-1)^{k-2}\bar{g}_0^{(k-2)}(\xi_{-1})+cn^{-2/(2k+1)}\\
&\qquad +\frac{(-1)^{k-2}g_0^{(k-2)}(\xi_1)-(-1)^{k-2}g_0^{(k-2)}(\xi_{-1})+2cn^{-2/(2k+1)}}{\xi_1-\xi_{-1}}(t-\xi_{-1})\\
&\leq\ (-1)^{k-2}g_0^{(k-2)}(x_0)+(\xi_{-1}-x_0)(-1)^{k-2}g_0^{(k-2)}(x_0)\\
&\qquad +\frac{1}{2}(\xi_{-1}-x_0)^2(-1)^{k-2}g_0^{(k)}(\nu)\\
&\qquad +(-1)^{k-2}g_0^{(k-1)}(\xi_1)(t-\xi_{-1})+2cn^{-2/(2k+1)}\frac{(t-\xi_{-1})}{\xi_1-\xi_{-1}}\\
&\leq\ (-1)^{k-2}g_0^{(k-2)}(x_0)+(\xi_{-1}-x_0)(-1)^{k-2}g_0^{(k-2)}(x_0)\\
&\qquad +\frac{1}{2}(\xi_{-1}-x_0)^2(-1)^{k-2}g_0^{(k)}(\nu)\\
&\qquad +\left((-1)^{k-2}g_0^{(k-1)}(x_0)+c'n^{-1/(2k+1)}\right)(t-\xi_{-1})+2cn^{-2/(2k+1)}\frac{(t-\xi_{-1})}{\xi_1-\xi_{-1}}\\
&\leq\ (-1)^{k-2}g_0^{(k-2)}(x_0)+(t-x_0)(-1)^{k-2}g_0^{(k-1)}(x_0)\\
&\qquad +\left(\frac{D_1}{2}+2c+2Kc'\right)n^{-2/(2k+1)}
\end{aligned}
$$

where $\nu\in(\xi_{-1},x_0)$, $D_1=\sup_{x\in[x_0-\delta,x_0+\delta]}|g_0^{(k)}(x)|$ and $[x_0-\delta,x_0+\delta]$ can be taken to be the largest neighborhood where $g_0^{(k)}$ exists and is continuous.

In all the previous calculations, n is taken sufficiently large so that $[x_0 - Kn^{-1/(2k+1)}, x_0 + Kn^{-1/(2k+1)}] \subseteq [x_0 - \delta, x_0 + \delta]$. We conclude that (3.1) holds for $j = k - 2$. Now, suppose that (3.1) is true for all $j' > j - 1$; i.e., for all $M > 0$

$$\sup_{|t|<M} \left| \bar{g}_n^{(j')}(x_0 + n^{-1/(2k+1)}t) - \sum_{i=j'}^{k-1} \frac{n^{-(i-j')/(2k+1)} g_0^{(i)}(x_0)}{(i-j')!} t^{i-j'} \right| \\ = O_p(n^{-(k-j')/(2k+1)}).$$

We are going to prove (3.1) for $j - 1$. We assume without loss of generality that k and $j - 1$ are even. In what follows, $\xi_{\pm 1}$ denotes the same numbers introduced before but this time they are associated with $\bar{g}_n^{(j-1)}$; i.e., for any $0 < \epsilon < 1$, there exist $c > 0$ and $K > M$ such that

$$|\bar{g}_n^{(j-1)}(\xi_{\pm 1}) - g_0^{(j-1)}(\xi_{\pm 1})| \le cn^{-(k-j+1)/(2k+1)}$$

with probability greater than $1 - \epsilon$ and where $\xi_1 \in [x_0 + Mn^{-1/(2k+1)}, x_0 + Kn^{-1/(2k+1)}]$ and $\xi_{-1} \in [x_0 - Kn^{-1/(2k+1)}, x_0 - Mn^{-1/(2k+1)}]$. Now, using the induction assumption, we know that we can find $c' > 0$ such that, with probability greater than $1 - \epsilon$,

$$\begin{aligned} & -c'n^{-(k-j')/(2k+1)} \\ \le \quad & \bar{g}_n^{(j')}(x_0 + n^{-1/(2k+1)}t) - \sum_{i=j'}^{k-1} \frac{n^{-(i-j')/(2k+1)} g_0^{(i)}(x_0)}{(i-j')!} t^{i-j'} \\ \le \quad & c'n^{-(k-j')/(2k+1)} \end{aligned} \tag{6.27}$$

for all $|t| \le M$ and $j' > j - 1$. Using convexity of $\bar{g}_n^{(j-1)}$ "from below", we have for all $|t - x_0| \le Mn^{-1/(2k+1)}$ with probability greater than $1 - 2\epsilon$,

$$\begin{aligned} & \bar{g}_n^{(j-1)}(t) \\ \ge \quad & \bar{g}_n^{(j-1)}(\xi_1) + \bar{g}_n^{(j)}(\xi_1)(t - \xi_1) + \cdots + \frac{1}{(k-j)!} \bar{g}_n^{(k-1)}(\xi_1)(t - \xi_1)^{k-j} \\ \ge \quad & g_0^{(j-1)}(\xi_1) - cn^{-(k-j+1)/(2k+1)} + \left(\sum_{i=j}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j)!} (\xi_1 - x_0)^{i-j} (t - \xi_1) \right) \end{aligned}$$

$$
\begin{aligned}
&+ \left(\sum_{i=j+1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j-1)!} (\xi_1 - x_0)^{i-j-1} \right) \frac{(t-\xi_1)^2}{2!} \\
&+ \cdots + g_0^{(k-1)}(x_0) \frac{(t-\xi_1)^{k-j}}{(k-j)!} \\
&+ c' n^{-(k-j)/(2k+1)} (t-\xi_1) - c' n^{-(k-j-1)/(2k+1)} \frac{(t-\xi_1)^2}{2!} \\
&+ \cdots - c' n^{-1/(2k+1)} \frac{(t-\xi_1)^{k-j}}{(k-j)!}. \qquad (6.28)
\end{aligned}
$$

Using Taylor expansion of $g_0^{(j-1)}(\xi_1)$ around $g_0^{(j-1)}(x_0)$, we can write

$$
\begin{aligned}
g_0^{(j-1)}(\xi_1) &= g_0^{(j-1)}(x_0) + g_0^{(j)}(x_0)(\xi_1 - x_0) + \cdots + \frac{g_0^{(k-1)}(x_0)}{(k-j)!}(\xi_1 - x_0)^{k-j} \\
&\quad + \frac{g_0^{(k)}(\nu)}{(k-j+1)!}(\xi_1 - x_0)^{k-j+1}
\end{aligned}
$$

where $\nu \in (x_0, \xi_1)$. Using this expansion and the fact that $|t - \xi_1| \le K n^{-1/(2k+1)}$, the right side of (6.28) can be bounded below by

$$
\begin{aligned}
&\sum_{i=j-1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j+1)!}(\xi_1 - x_0)^{i-j+1} + \sum_{i=j}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j)!}(\xi_1 - x_0)^{i-j}(t-\xi_1) \\
&+ \sum_{i=j+1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j-1)!}(\xi_1 - x_0)^{i-j-1} \frac{(t-\xi_1)^2}{2!} + \cdots + g_0^{(k-1)}(x_0)\frac{(t-\xi_1)^{k-j}}{(k-j)!} \\
&- \left(c + c' \sum_{p=1}^{k-j} \frac{K^p}{p!} \right) n^{-(k-j+1)/(2k+1)} + \frac{g_0^{(k)}(\nu)}{(k-j+1)!}(\xi_1 - x_0)^{k-j+1} \\
= \;& g_0^{(j-1)}(x_0) + g_0^{(j)}(x_0)(t - x_0) \\
&+ \frac{g_0^{(j+1)}(x_0)}{2!} \left((\xi_1 - x_0)^2 + 2(\xi_1 - x_0)(t-\xi_1) + (t-\xi_1)^2 \right) \\
&+ \cdots + \frac{g_0^{(k-1)}(x_0)}{(k-j)!} \sum_{p=0}^{k-j} \frac{(k-j)!}{(k-j-p)!p!} (\xi_1 - x_0)^{k-j-p}(t-\xi_1)^p \\
&- \left(c + c' \sum_{p=1}^{k-j} \frac{K^p}{p!} \right) n^{-(k-j+1)/(2k+1)} + \frac{g_0^{(k)}(\nu)}{(k-j+1)!}(\xi_1 - x_0)^{k-j+1}
\end{aligned}
$$

$$
\begin{aligned}
&= g_0^{(j-1)}(x_0) + g_0^{(j)}(x_0)(t-x_0) + \cdots + \frac{g^{(k-1)}(x_0)}{(k-j)!}(t-x_0)^{k-j} \\
&\quad - \left(c + c' \sum_{p=1}^{k-j} \frac{K^p}{p!}\right) n^{-(k-j+1)/(2k+1)} - \frac{D_1 K^{k-j+1}}{(k-j+1)!} n^{-(k-j+1)/(2k+1)}
\end{aligned}
$$

since $0 \le \xi_1 - x_0 \le K n^{-1/(2k+1)}$. Now, we use convexity of $\bar{g}_n^{(j-1)}$ "from above". We first need to establish a useful inequality. Since $\bar{g}_n^{(k-2)}$ is convex, we have for all $t' \in [x_0 - Mn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$ and

$$
\bar{g}_n^{(k-2)}(t') \le \bar{g}_n^{(k-2)}(\xi_{-1}) + \frac{\bar{g}_n^{(k-2)}(\xi_1) - \bar{g}_n^{(k-2)}(\xi_{-1})}{\xi_{n,1} - \xi_{-1}}(t' - \xi_{-1}).
$$

By successive integrations of the last inequality between ξ_{-1} and t, we obtain

$$
\begin{aligned}
\bar{g}_n^{(j-1)}(t) - \bar{g}_n^{(j-1)}(\xi_{-1}) \quad &\le \quad \bar{g}_n^{(j)}(\xi_{-1})(t-\xi_1) + \bar{g}_n^{(j+1)}(\xi_{-1}) \frac{(t-\xi_{-1})^2}{2!} \\
&\quad + \cdots + \frac{\bar{g}_n^{(k-2)}(\xi_1) - \bar{g}_n^{(k-2)}(\xi_{-1})}{\xi_1 - \xi_{-1}} \frac{(t-\xi_{-1})^{k-j}}{(k-j)!}.
\end{aligned}
$$

It follows that with probability greater than $1 - 2\epsilon$, we have

$$
\begin{aligned}
&\bar{g}_n^{(j-1)}(t) \\
&\le \bar{g}_n^{(j-1)}(\xi_{-1}) + \bar{g}_n^{(j)}(\xi_{-1})(t-\xi_{-1}) + \bar{g}_n^{(j+1)}(\xi_{-1}) \frac{(t-\xi_{-1})^2}{2!} \\
&\quad + \cdots + \frac{g_0^{(k-2)}(\xi_1) - g_0^{(k-2)}(\xi_{-1}) + 2cn^{-2/(2k+1)}}{\xi_1 - \xi_{-1}} \frac{(t-\xi_{-1})^{k-j}}{(k-j)!} \\
&\le g_0^{(j-1)}(\xi_{-1}) + cn^{-(k-j+1)/(2k+1)} \\
&\quad + \left(\sum_{i=j}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j)!}(\xi_{-1} - x_0)^{i-j} + c' n^{-(k-j)/(2k+1)}\right)(t - \xi_{-1}) \\
&\quad + \left(\sum_{i=j+1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j-1)!}(\xi_{-1} - x_0)^{i-j-1} + c' n^{-(k-j-1)/(2k+1)}\right) \frac{(t-\xi_{-1})^2}{2!} \\
&\quad + \cdots + \left(g_0^{(k-1)}(\xi_1) + \frac{c}{K} n^{-1/(2k+1)}\right) \frac{(t-\xi_{-1})^{k-j}}{(k-j)!} \\
&\le \sum_{i=j-1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j+1)!}(\xi_{-1} - x_0)^{i-j+1} + \frac{g^{(k)}(\nu)}{k!}(\xi_{-1} - x_0)^{k-j+1}
\end{aligned}
$$

$$
\begin{aligned}
& + \left(\sum_{i=j}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j)!}(\xi_{-1}-x_0)^{i-j} \right)(t-\xi_{-1}) \\
& + \cdots + \left(\sum_{i=j+1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j-1)!}(\xi_{-1}-x_0)^{i-j-1} \right) \frac{(t-\xi_{-1})^2}{2!} \\
& + \left(g_0^{(k-1)}(x_0) + cn^{-1/(2k+1)} \right) \frac{(t-\xi_{-1})^{k-j}}{(k-j)!} \\
& + \left(c(1+K^{k-j}) + c' \sum_{p=1}^{k-j} \frac{K^p}{p!} + \frac{D_1 K^{k-j+1}}{k!} \right) n^{-(k-j+1)/(2k+1)} \\
= \;& g_0^{(j-1)}(x_0) + g_0^{(j)}(x_0)(t-x_0) + \cdots + g_0^{(k-j)}(x_0)\frac{(t-x_0)^{k-j}}{(k-j)!} \\
& + K' n^{-(k-j+1)/(2k+1)}
\end{aligned}
$$

with $K' = c(1+K^{k-j}) + c' \sum_{p=1}^{k-j} \frac{K^p}{p!} + \frac{D_1 K^{k-j+1}}{k!}$. It follows that (3.1) holds for $j-1$. ■

Proof of Lemma 3.2. Fix $K > 0$. Recall that $r_k \equiv 1/(2k+1)$. We will prove the lemma for $t \geq 0$; similar arguments can be used for $t \in [-K, 0)$. In the Least Squares case, we have

$$
\begin{aligned}
\mathbb{Y}_n^{loc}(t) &= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Big\{ \mathbb{G}_n(v_1) - \mathbb{G}_n(x_0) \\
& \quad - \int_{x_0}^{v_1} \Big(g_0(x_0) + (u-x_0)g_0'(x_0) \\
& \quad\quad + \cdots + \frac{1}{(k-1)!}(u-x_0)^{k-1} g_0^{(k-1)}(x_0) \Big) du \Big\} \\
& \quad\quad dv_1 dv_2 \ldots dv_{k-1} \\
&= A_n + B_n,
\end{aligned}
$$

where

$$
\begin{aligned}
A_n &= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \\
& \quad \Big\{ \mathbb{G}_n(v_1) - \mathbb{G}_n(x_0) - (G_0(v_1) - G_0(x_0)) \Big\} dv_1 dv_2 \ldots dv_{k-1},
\end{aligned}
$$

and

$$
\begin{aligned}
B_n &= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \\
&\qquad \Bigg\{ G_0(v_1) - G_0(x_0) - \int_{x_0}^{v_1} \Bigg(g_0(x_0) + (u-x_0) g_0'(x_0) \\
&\qquad + \cdots + \frac{1}{(k-1)!}(u-x_0)^{k-1} g_0^{(k-1)}(x_0) \Bigg) du \Bigg\} dv_1 dv_2 \ldots dv_{k-1}.
\end{aligned}
$$

But, with $\mathbb{U}_n$ denoting $\sqrt{n}(\Gamma_n - I)$, $\Gamma_n(t) = n^{-1}\sum_{i=1}^n 1_{[\xi_i \le t]}$ where $\xi_1, \ldots, \xi_n$ are i.i.d. $U(0,1)$ random variables, we have

$$
\begin{aligned}
A_n &\stackrel{d}{=} n^{2kr_k - 1/2} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \\
&\qquad \cdots \int_{x_0}^{v_2} \Big(\mathbb{U}_n(G_0(v_1)) - \mathbb{U}_n(G_0(x_0)) \Big) \Big) \prod_{j=1}^{k-1} dv_j \\
&= n^{(k-1/2)r_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \\
&\qquad \cdots \int_{x_0}^{v_2} \Big(\mathbb{U}_n(G_0(v_1)) - \mathbb{U}_n(G_0(x_0)) \Big) \prod_{j=1}^{k-1} dv_j,
\end{aligned}
$$

and using Taylor expansion of $G_0(v_1)$ in the neighborhood of x_0,

$$
\begin{aligned}
B_n &= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \frac{(v_1-x_0)^{k+1}}{(k+1)!} \Big(g_0^{(k)}(v_1^*) - g_0^{(k)}(x_0) \Big) \\
&\qquad \cdot \prod_{i=1}^{k-1} dv_i \\
&\quad + n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \frac{(v_1-x_0)^{k+1}}{(k+1)!} g_0^{(k)}(x_0) \prod_{i=1}^{k-1} dv_i \\
&= B_{n1} + B_{n2},
\end{aligned}
$$

where $|v_1^* - x_0| \le |v_1 - x_0|$. Now,

$$
\frac{B_{n2}}{g_0^{(k)}(x_0)} = \frac{n^{2kr_k}}{(k+2)!} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_3} (v_2 - x_0)^{k+2} dv_2 \cdots dv_{k-1}
$$

$$
\begin{aligned}
&= \frac{n^{2kr_k}}{(k+3)!}\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_4}(v_3-x_0)^{k+3}dv_3\cdots dv_{k-1}\\
&\vdots\\
&= \frac{n^{2kr_k}}{(2k-1)!}\int_{x_0}^{x_0+tn^{-r_k}}(v_{k-1}-x_0)^{2k-1}dv_{k-1}\\
&= n^{2kr_k}\frac{1}{(2k)!}\left(\frac{t}{n^{r_k}}\right)^{2k} = \frac{1}{(2k)!}g_0^{(k)}(x_0)t^{2k}.
\end{aligned}
$$

Furthermore, by continuity of $g_0^{(k)}$ at x_0, we deduce that $B_{n1}(t) = o(1)$ uniformly in $0 \le t \le K$ and hence

$$B_n \to \frac{1}{(2k)!}g_0^{(k)}(x_0)t^{2k}, \tag{6.29}$$

as $n \to \infty$ uniformly in $0 \le t \le K$. Using the identity

$$\mathbb{U}(G_0(v)) - \mathbb{U}(G_0(x_0)) \overset{d}{=} W(G_0(v)) - W(G_0(x_0)) - (G_0(v) - G_0(x_0))W(1),$$

where W is two-sided Brownian motion process, we have

$$
\begin{aligned}
A_n \overset{d}{=}\;& n^{(k-1/2)r_k}\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_2}\\
&\qquad\Big(\mathbb{U}_n(v_1)-\mathbb{U}(v_1)-(\mathbb{U}_n(x_0)-\mathbb{U}(x_0))\Big)dv_1\ldots dv_{k-1}\\
&+ n^{(k-1/2)r_k}\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_2}\Big(W(G_0(v))-W(G_0(x_0))\Big)\\
&- W(1)n^{(k-1/2)r_k}\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_2}\\
&\qquad(G_0(v_1)-G_0(x_0))dv_1\ldots dv_{k-1}\\
=\;& A_{n1}+A_{n2}+A_{n3}.
\end{aligned}
$$

But,

$$
\begin{aligned}
A_{n1} &\le 2n^{(k-1/2)r_k}\|\mathbb{U}_n-\mathbb{U}\|_\infty\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_2}dv_1\ldots dv_{k-1}\\
&= 2n^{(k-1/2)r_k}\|\mathbb{U}_n-\mathbb{U}\|_\infty\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_3}(v_2-x_0)dv_2\ldots dv_{k-1}
\end{aligned}
$$

$$
\begin{aligned}
&= 2n^{(k-1/2)r_k}\|\mathbb{U}_n - \mathbb{U}\|_\infty \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_4} \frac{1}{2}(v_3 - x_0)^2 dv_3 \\
&\vdots \\
&= 2n^{(k-1/2)r_k}\|\mathbb{U}_n - \mathbb{U}\|_\infty \frac{1}{(k-2)!} \int_{x_0}^{x_0+tn^{-r_k}} (v_{k-1} - x_0)^{k-2} dv_{k-1} \\
&= 2n^{(k-1/2)r_k}\|\mathbb{U}_n - \mathbb{U}\|_\infty \frac{1}{(k-1)!} \left(\frac{t}{n^{r_k}}\right)^{k-1} \\
&= 2t^{k-1} n^{r_k/2} O\left(\frac{\log(n)^2}{n^{1/2}}\right) \\
&= 2t^{k-1} O\left(\frac{\log(n)^2}{n^{kr_k}}\right)
\end{aligned} \tag{6.30}
$$

since $\|\mathbb{U}_n - \mathbb{U}\|_\infty = O\left(n^{-1/2}\left(\log(n)\right)^2\right)$ via Komlós, Major and Tusnády (1975); see e.g. Shorack and Wellner (1986), page 494. On the other hand, using the fact that g_0 is nonincreasing, we have

$$
\begin{aligned}
A_{n3} &\le |W(1)| g_0(x_0) n^{(k-1/2)r_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} (v_1 - x_0) dv_1 \ldots dv_{k-1} \\
&= |W(1)| g_0(x_0) n^{(k-1/2)r_k} \frac{1}{k!} \left(\frac{t}{n^{r_k}}\right)^k \\
&= |W(1)| g_0(x_0) t^k n^{-r_k/2} \to_p 0,
\end{aligned} \tag{6.31}
$$

as $n \to \infty$ uniformly in $0 \le t \le K$. Finally, using the change of variables $s_j = n^{1/(2k+1)}(v_j - x_0) = n^{r_k}(v_j - x_0)$ for $j = 1, \ldots, k-1$, we have

$$
\begin{aligned}
A_{n2} &= n^{(k-1/2)r_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Big(W(G_0(v_1)) - W(G_0(x_0)) \Big) \\
&\qquad dv_1 \ldots dv_{k-1} \\
&= n^{(k-1/2)r_k} n^{-(k-1)r_k} \int_0^t \int_0^{s_{k-1}} \\
&\qquad \cdots \int_0^{s_2} \Big(W(G_0(n^{-r_k}s_1 + x_0)) - W(G_0(x_0)) \Big)
\end{aligned}
$$

$$
\begin{aligned}
& ds_1 \dots ds_{k-1} \\
\stackrel{d}{=} \; & n^{r_k/2} \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{v_2} W\Big(G_0(n^{-r_k} s_1 + x_0) - G_0(x_0) \Big) ds_1 \dots ds_{k-1} \\
\stackrel{d}{=} \; & \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W\Big(n^{r_k} (G_0(n^{-r_k} s_1 + x_0) - G_0(x_0)) \Big) ds_1 \dots ds_{k-1} \\
\to \; & \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1 g_0(x_0)) ds_1 \dots ds_{k-1} \quad \text{as } n \to \infty \\
\stackrel{d}{=} \; & \sqrt{g_0(x_0)} \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1) ds_1 \dots ds_{k-1}.
\end{aligned}
\tag{6.32}
$$

Therefore, combining (6.29), (6.30), (6.31) and (6.32) yields

$$
\begin{aligned}
\mathbb{Y}_n^{loc}(t) \; & \Rightarrow \; \sqrt{g_0(x_0)} \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1) ds_1 \dots ds_{k-1} + \frac{1}{(2k)!} t^{2k} g_0^{(k)}(x_0) \\
& \equiv \; Y_{a,\sigma}(t)
\end{aligned}
$$

for $0 \le t \le K$. A similar argument for $-K \le t < 0$ yields the conclusion. In the Maximum Likelihood case, we apply very similar argumuments along with uniform consistency of $\hat{g}_n$. ■

Proof of Lemma 3.3. We now consider the difference of the two local processes $\mathbb{Y}_n^{loc}$ and $\tilde{H}_n^{loc}$. We have

$$
\begin{aligned}
& \tilde{H}_n^{loc}(t) - \mathbb{Y}_n^{loc}(t) \\
& \quad = \; n^{2kr_k} \int_{x_0}^{x_0 + tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Bigg\{ \Bigg((\tilde{G}_n(v_1) - \tilde{G}_n(x_0)) \\
& \qquad\qquad - (\mathbb{G}_n(v_1) - \mathbb{G}_n(x_0)) \Bigg) \\
& \qquad\qquad dv_1 \cdots dv_{k-1} \Bigg\} + \tilde{A}_{(k-1)n} t^{k-1} + \cdots + \tilde{A}_{1n} t + \tilde{A}_{0n} \\
& \quad = \; n^{2kr_k} \int_{x_0}^{x_0 + tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Big(\tilde{G}_n(v_1) - \mathbb{G}_n(v_1) \Big) dv_1 \cdots dv_{k-1} \\
& \qquad - \frac{n^{(k+1)r_k}}{(k-1)!} \Big(\tilde{G}_n(x_0) - \mathbb{G}_n(x_0) \Big) t^{k-1} \\
& \qquad + \tilde{A}_{(k-1)n} t^{k-1} + \cdots + \tilde{A}_{1n} t + \tilde{A}_{0n}
\end{aligned}
$$

$$
\begin{aligned}
&= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Big(\tilde{G}_n(v_1) - \mathbb{G}_n(v_1)\Big) dv_1 \dots dv_{k-1} \\
&\quad - \tilde{A}_{(k-1)n} t^{k-1} + \tilde{A}_{(k-1)n} t^{k-1} + \cdots + \tilde{A}_{1n} t + \tilde{A}_{0n} \\
&= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Big(\tilde{G}_n(v_1) - \mathbb{G}_n(v_1)\Big) dv_1 \dots dv_{k-1} \\
&\quad + \tilde{A}_{(k-2)n} t^{k-2} + \cdots + \tilde{A}_{1n} t + \tilde{A}_{0n} \\
&= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{0}^{v_2} \Big(\tilde{G}_n(v_1) - \mathbb{G}_n(v_1)\Big) dv_1 \dots dv_{k-1} \\
&\quad - n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_3} dv_2 \dots dv_{k-1} \\
&\qquad \cdot \int_0^{x_0} \Big(\tilde{G}_n(v_1) - \mathbb{G}_n(v_1)\Big) dv_1 \\
&\quad + \tilde{A}_{(k-2)n} t^{k-2} + \cdots + \tilde{A}_{1n} t + \tilde{A}_{0n} \\
&= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{0}^{v_2} \Big(\tilde{G}_n(v_1) - \mathbb{G}_n(v_1)\Big) dv_1 \dots dv_{k-1} \\
&\quad - n^{(k+2)r_k} \frac{t^{k-2}}{(k-2)!} \times \int_0^{x_0} \Big(\tilde{G}_n(v_1) - \mathbb{G}_n(v_1)\Big) dv_1 + \tilde{A}_{(k-2)n} t^{k-2} \\
&\quad + \tilde{A}_{(k-3)n} t^{k-3} + \cdots + \tilde{A}_{1n} t + \tilde{A}_{0n} \\
&= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{0}^{v_2} \Big(\tilde{G}_n(v_1) - \mathbb{G}_n(v_1)\Big) dv_1 \dots dv_{k-1} \\
&\quad - \tilde{A}_{(k-2)n} t^{k-2} + \tilde{A}_{(k-2)n} t^{k-2} + \cdots + \tilde{A}_{1n} t + \tilde{A}_{0n} \\
&= n^{2kr_k} \int_{x_0}^{x_0+tn^{-r_k}} \int_{x_0}^{v_{k-1}} \cdots \int_{0}^{v_2} \Big(\tilde{G}_n(v_1) - \mathbb{G}_n(v_1)\Big) dv_1 \dots dv_{k-1} \\
&\quad + \tilde{A}_{(k-3)n} t^{k-2} + \cdots + \tilde{A}_{1n} t + \tilde{A}_{0n} \\
&\vdots \\
&= n^{2kr_k} \Big(\tilde{H}_n(x_0 + tn^{-r_k}) - \mathbb{Y}_n(x_0 + tn^{-r_k})\Big) \geq 0,
\end{aligned}
$$

by the first Fenchel condition satisfied by the LSE. Very similar calculations yield the second Fenchel condition satsified by $\widehat{H}_n^{loc}$ in the case of the MLE.

Similarly, for the localized processes $\widehat{\mathbb{Y}}_n^{loc}$ and $\widehat{H}_n^{loc}$, by the particular choice of $\widehat{A}_{jn}$, $0 \leq j \leq k-1$, we have

$$(\widehat{H}_n^{loc}(t) - \widehat{\mathbb{Y}}_n^{loc}(t))/g_0(x_0)$$

$$
\begin{aligned}
&= n^{2kr_k}\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_1}\frac{\hat{g}_n(v)-g_0(v)}{\hat{g}_n(v)}dvdv_1\dots dv_{k-1}\\
&\quad - n^{2kr_k}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_1}\frac{1}{\hat{g}_n(v)}d(\mathbb{G}_n-G_0)(v)dv_1\dots dv_{k-1}\\
&\quad + \widehat{A}_{(k-1)n}t^{k-1}+\cdots+\widehat{A}_{0n}\\
&= n^{2kr_k}\left(\frac{t^k}{k!}n^{-kr_k}-\int_{x_0}^{x_0+n^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_1}\frac{1}{\hat{g}_n(v)}d\mathbb{G}_n(v)\Pi_{i=1}^{k-1}dv_i\right)\\
&\quad + \widehat{A}_{(k-1)n}t^{k-1}+\cdots+\widehat{A}_{0n}.
\end{aligned}
$$

But notice that for any $t \geq 0$

$$
\int_0^t \frac{1}{\hat{g}_n(u)}d\mathbb{G}_n(u) = \frac{1}{(k-1)!}\widehat{H}_n^{(k-1)}(t).
$$

It follows that

$$
\begin{aligned}
&\int_{x_0}^{x_0+tn^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_1}\frac{1}{\hat{g}_n(v)}d\mathbb{G}_n(v)dv_1\dots dv_{k-1}\\
&\quad = \frac{1}{(k-1)!}\int_{x_0}^{x_0+n^{-r_k}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_1}\left(\widehat{H}_n^{(k-1)}(v_1)-\widehat{H}_n^{(k-1)}(x_0)\right)dv_1\dots dv_{k-1}\\
&\quad = \frac{1}{(k-1)!}\left(\widehat{H}_n(x_0+tn^{-r_k})-\sum_{j=0}^{k-1}\frac{t^j n^{-jr_k}}{j!}\widehat{H}_n^{(j)}(x_0)\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\widehat{H}_n^{loc}(t)-\widehat{\mathbb{Y}}_n^{loc}(t)\\
&\quad = n^{2k/(2k+1)}g_0(x_0)\Big\{-\frac{\widehat{H}_n(x_0+tn^{-1/(2k+1)})}{(k-1)!}+\frac{t^k}{k!}n^{-k/(2k+1)}\\
&\qquad\qquad + \sum_{j=0}^{k-1}\frac{t^j n^{-j/(2k+1)}}{(k-1)!j!}\widehat{H}_n^{(j)}(x_0)\Big\}\\
&\qquad + \widehat{A}_{(k-1)n}t^{k-1}+\cdots+\widehat{A}_{0n}\\
&\quad = n^{2k/(2k+1)}\frac{g_0(x_0)}{k!}\Big\{-k\widehat{H}_n(x_0+tn^{-1/(2k+1)})+t^k n^{-k/(2k+1)}\\
&\qquad + \sum_{j=0}^{k-1}\frac{t^j n^{-j/(2k+1)}}{j!}k\left(\widehat{H}_n^{(j)}(x_0)-\frac{1}{k}\frac{k!}{(k-j)!}x_0^{k-j}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{j=0}^{k-1} \frac{k!}{j!(k-j)!} t^j n^{-j/(2k+1)} x_0^{k-j} \Bigg\} + \widehat{A}_{(k-1)n} t^{k-1} + \cdots + \widehat{A}_{0n} \\
= \;& n^{2k/(2k+1)} \frac{g_0(x_0)}{k!} \Bigg\{ - k \widehat{H}_n(x_0 + tn^{-1/(2k+1)}) + (x_0 + tn^{-1/(2k+1)})^k \Bigg\}
\end{aligned}
$$

by replacing the coefficients $\widehat{A}_{jn}$, $0 \le j \le k-1$ by their definitions. It follows that

$$
\begin{aligned}
&\widehat{H}_n^{loc}(t) - \widehat{\mathbb{Y}}_n^{loc}(t) \\
= \;& n^{2k/(2k+1)} \frac{g_0(x_0)}{(k-1)!} \left(\frac{1}{k} (x_0 + tn^{-1/(2k+1)})^k - \widehat{H}_n(x_0 + tn^{-1/(2k+1)}) \right) \\
\ge \;& 0.
\end{aligned}
$$

■

Proof of Lemma 3.4. Groeneboom, Jongbloed, and Wellner (2001b) chose the "canonical process" to be

$$Y(t) = \int_0^t W(y) dy + t^4,$$

so that with $X(t) = Y'(t) = W(t) + 4t^3$ we have

$$dX(t) = 12t^2 dt + dW(t) \equiv f_0(t) dt + dW(t) \tag{6.33}$$

where $f_0(t) = 12t^2$ is convex. Here we make a different choice, namely $f_0(t) = (-1)^k t^k$ (so that $f_0(t) = t^2$ in the case $k = 2$). Thus we will rescale the limiting process $Y_{a,\sigma}$ so that we obtain the "canonical process"

$$Y_k(t) = \int_0^t \int_0^{u_{k-1}} \cdots \int_0^{u_2} W(u_1) du_1 du_2 \ldots du_{k-1} + (-1)^k \frac{k!}{(2k)!} t^{2k}, \tag{6.34}$$

for $t \ge 0$. Let $\sigma = \sqrt{g_0(x_0)}$ and $a = (-1)^k g_0^{(k)}(x_0)/k!$. Then

$$
\begin{aligned}
Y_{a,\sigma}(t) \;=\;& \sqrt{g_0(x_0)} \int_0^t \int_0^{u_{k-1}} \cdots \int_0^{u_2} W(u_1) du_1 \cdots du_{k-1} \\
&+ \frac{(-1)^k}{k!} g_0^{(k)}(x_0)(-1)^k \frac{k!}{(2k)!} t^{2k} .
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
Y_{a,\sigma}(t) &= a(-1)^k \frac{k!}{(2k)!} t^{2k} + \sigma \int_0^t \int_0^{u_{k-1}} \cdots \int_0^{u_2} W(u_1) du_1 \cdots du_{k-1} \\
&\stackrel{d}{=} a(-1)^k \frac{k!}{(2k)!} t^{2k} + \alpha^{-1/2}\sigma \int_0^t \int_0^{u_{k-1}} \cdots \int_0^{u_2} W(\alpha u_1) du_1 \cdots du_{k-1} \\
&\stackrel{d}{=} a(-1)^k \frac{k!}{(2k)!} t^{2k} + \alpha^{-1/2}\sigma \int_0^t \int_0^{u_{k-1}} \cdots \int_0^{\alpha u_2} \frac{1}{\alpha} W(u_1) du_1 \cdots du_{k-1} \\
&\vdots \\
&\stackrel{d}{=} \frac{a(-1)^k k!}{(2k)!} t^{2k} + \frac{\sigma}{\sqrt{\alpha}} \int_0^{\alpha t} \int_0^{u_{k-1}} \cdots \int_0^{u_2} W(u_1) \frac{1}{\alpha^{k-1}} du_1 \cdots du_{k-1} \\
&\stackrel{d}{=} \frac{a(-1)^k k!}{(2k)!} t^{2k} + \alpha^{1/2-k}\sigma \int_0^{\alpha t} \int_0^{u_{k-1}} \cdots \int_0^{u_2} W(u_1) du_1 \cdots du_{k-1}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
s_1 Y_{a,\sigma}(s_2 t) \stackrel{d}{=}\ & a(-1)^k \frac{k!}{(2k)!} s_1 (s_2 t)^{2k} \\
& + s_1 \alpha^{1/2-k} \sigma \int_0^{s_2 \alpha t} \int_0^{u_{k-1}} \cdots \int_0^{u_2} W(u_1) du_1 \ldots du_{k-1},
\end{aligned}
$$

and the process on the right side of the last display equals Y_k as defined in (6.34) if $a s_1 s_2^{2k} = 1$, $s_1 \alpha^{1/2-k}\sigma = 1$, and $s_2 \alpha = 1$. Solving this system of equations yields $\alpha = (a/\sigma)^{2/(2k+1)}$, and therefore s_1 and s_2 are given by (3.2) and (3.3) respectively. Thus, $Y_{a,\sigma}(t) \stackrel{d}{=} Y_k(t/s_2)/s_1$ ■

Proof of Lemma 3.5. We will give the proof only for the LSE as the arguments are very similar for the MLE. Also, Let $j \in \{0, \ldots, k-1\}$ and denote $\tilde{\Delta}_n(x) = \tilde{H}_n(x) - \mathbb{Y}_n(x)$ for all $x \geq 0$. Here, we show first Lemma 3.5 for $j = k-1$ and use and an induction argument for $j < k-1$.

Consider k successive jump points, $\tau_1, \ldots, \tau_k$, of $\tilde{g}_n^{(k-1)}$ where τ_1 is the first jump after x_0. By the mean value theorem, there exist $\tau_1^{(1)} \in (\tau_1, \tau_2)$, $\tau_2^{(1)} \in (\tau_2, \tau_3)$, ..., $\tau_{k-1}^{(1)} \in (\tau_{k-1}, \tau_k)$ such that $\tilde{\Delta}_n'(\tau_i^{(1)}) = 0$ for $1 \leq i \leq k-1$. Also, by the same theorem there exist $\tau_1^{(2)} \in (\tau_1^{(1)}, \tau_2^{(1)}), \ldots, \tau_{k-2}^{(2)} \in (\tau_{k-2}^{(1)}, \tau_{k-1}^{(1)})$ such that $\tilde{\Delta}_n''(\tau_i^{(2)}) = 0$ for $1 \leq i \leq k-2$. It is easy to see that we can carry on this reasoning up to the $(k-1)$-st level of differentiation and so there exists $\tau^{(k-1)}$ such that

$$\tilde{\Delta}_n^{(k-1)}(\tau^{(k-1)}) = 0.$$

Denote $\tau = \tau^{(k-1)}$. We can write

$$\tilde{\Delta}_n^{(k-1)}(x_0) = \tilde{\Delta}_n^{(k-1)}(x_0) - \tilde{\Delta}_n^{(k-1)}(\tau).$$

But

$$\tilde{\Delta}_n^{(k-1)}(x) = \int_0^x d(\tilde{G}_n(t) - \mathbb{G}_n(t)), \quad \text{for } x \geq 0,$$

implies that

$$\begin{aligned}
|\tilde{\Delta}_n^{(k-1)}(x_0)| &= \left| \int_{x_0}^{\tau} d(\tilde{G}_n(t) - \mathbb{G}_n(t)) \right| \\
&\leq \left| \int_{x_0}^{\tau} d(\tilde{G}_n(t) - G_0(t)) \right| + \left| \int_{x_0}^{\tau} d(\mathbb{G}_n(t) - G_0(t)) \right| \\
&= \left| \int_{x_0}^{\tau} (\tilde{g}_n(t) - g_0(t)) dt \right| + \left| \int_{x_0}^{\tau} d(\mathbb{G}_n(t) - G_0(t)) \right| \\
&\leq \int_{x_0}^{\tau} |\tilde{g}_n(t) - g_0(t)|\, dt + \left| \int_{x_0}^{\tau} d(\mathbb{G}_n(t) - G_0(t)) \right|.
\end{aligned}$$

Fix $0 < \epsilon < 1$. By Lemma 3.1 and Proposition 3.1, we can find $M > 0$ and $c > 0$ such that with probability greater than $1 - \epsilon$

$$x_0 \leq \tau \leq x_0 + M n^{-1/(2k+1)}$$

and

$$\left| \tilde{g}_n(t) - g_0(x_0) - g_0'(x_0)(t - x_0) - \cdots - \frac{g_0^{(k-1)}(x_0)}{(k-1)!}(t - x_0)^{k-1} \right| \leq c n^{-k/(2k+1)}$$

for $x_0 - M n^{-1/(2k+1)} \leq t \leq x_0 + M n^{-1/(2k+1)}$. On the other hand, using Taylor expansion, we can find $d > 0$ that

$$\begin{aligned}
\left| g_0(t) - g_0(x) + g_0'(x_0)(t - x_0) - \cdots - \frac{g_0^{(k-1)}(x_0)}{(k-1)!}(t - x_0)^{k-1} \right| &\leq d\,(t - x_0)^k \\
&\leq c' n^{-k/(2k+1)}
\end{aligned}$$

for $x_0 - Mn^{-1/(2k+1)} \le t \le x_0 + Mn^{-1/(2k+1)}$ and where $c' = dM^k$. It follows that

$$\begin{aligned}\int_{x_0}^{\tau} |\tilde{g}_n(t) - g_0(t)|\, dt &\le (c+c')n^{-k/(2k+1)} \int_{x_0}^{\tau} dt \\ &= (c+c')n^{-k/(2k+1)} \times (\tau - x_0) \\ &\le (c+c')Mn^{-(k+1)/(2k+1)}.\end{aligned}$$

Now,

$$\int_{x_0}^{\tau} d(\mathbb{G}_n(t) - G_0(t)) = \int_0^{\infty} 1_{[x_0,\tau]}(t) d(\mathbb{G}_n(t) - G_0(t)).$$

Consider the empirical process

$$U_n(y,z) = \int_0^{\infty} 1_{[y,z]}(t) d(\mathbb{G}_n(t) - G_0(t))$$

for $0 < y \le z$ and the class of functions

$$\mathcal{F}_{y,R} = \left\{ f_{y,z} : f_{y,z}(t) = 1_{[y,z]}(t), y \le z \le y + R \right\}$$

for a fixed $y > 0$ and $R > 0$. By application of Lemma 7.1 with $d = 1$ and $\alpha = k$, it follows that for each $\epsilon > 0$ there exist $\delta > 0$ and $R > 0$ such that

$$|U_n(y,z)| \le \epsilon (z-y)^{k+1} + O_p(n^{-(k+1)/(2k+1)})$$

for all $|y - x_0| \le \delta$, $z \in [y, y+R]$. Thus we conclude that

$$\begin{aligned}\left| \int_{x_0}^{\tau} d(\mathbb{G}_n(t) - G_0(t)) \right| &= o_p\left((\tau - x_0)^{k+1} \right) + O_p(n^{-(k+1)/(2k+1)}) \\ &= O_p((n^{-(k+1)/(2k+1)})\end{aligned}$$

and the result follows for $j = k-1$.

Now, let $j = k-2$. We have,

$$\tilde{\Delta}_n^{(k-2)}(x_0) = \int_0^{x_0} (x_0 - t) d(\tilde{G}_n(t) - \mathbb{G}_n(t)).$$

Let τ be a zero of $\tilde{\Delta}_n^{(k-2)}$ (we can find such a zero the same way we did for $\tilde{\Delta}_n^{(k-1)}$). We can write

$$\begin{aligned}\tilde{\Delta}_n^{(k-2)}(x_0) &= \tilde{\Delta}_n^{(k-2)}(x_0) - \tilde{\Delta}_n^{(k-2)}(\tau) \\ &= \int_0^{x_0}(x_0-t)d(\tilde{G}_n(t)-\mathbb{G}_n(t)) - \int_0^{\tau}(\tau-t)d(\tilde{G}_n(t)-\mathbb{G}_n(t)) \\ &= -\int_{x_0}^{\tau}(x_0-t)d(\tilde{G}_n(t)-\mathbb{G}_n(t)) - (\tau-x_0)\int_0^{\tau} d(\tilde{G}_n(t)-\mathbb{G}_n(t)) \\ &= -\int_{x_0}^{\tau}(x_0-t)d(\tilde{G}_n(t)-\mathbb{G}_n(t)) - (\tau-x_0)\tilde{\Delta}_n^{(k-1)}(\tau).\end{aligned}$$

Let $M>0$ be such that $x_0 \le \tau \le x_0 + Mn^{-1/(2k+1)}$. By the previous result, there exists $c>0$ such that

$$\left|(\tau-x_0)\tilde{\Delta}_n^{(k-1)}(\tau)\right| \le cn^{-(k+2)/(2k+1)}$$

with large probability. Now

$$\begin{aligned}\left|\int_{x_0}^{\tau}(x_0-t)d(\tilde{G}_n(t)-\mathbb{G}_n(t))\right| &\le \int_{x_0}^{\tau}(t-x_0)|\tilde{g}_n(t)-g_0(t)|dt \\ &\qquad + \left|\int_{x_0}^{\tau}(t-x_0)d(\mathbb{G}_n(t)-G_0(t))\right|.\end{aligned}$$

We can find $d>0$ such that

$$\left|\tilde{g}_n(t)-g_0(x_0)-g_0'(x_0)(t-x_0)-\cdots-\frac{g_0^{(k-1)}(x_0)}{(k-1)!}(t-x_0)^{k-1}\right| \le dn^{-k/(2k+1)}$$

and

$$\left|g_0(t)-g_0(x_0)-g_0'(x_0)(t-x_0)-\cdots-\frac{g_0^{(k-1)}(x_0)}{(k-1)!}(t-x_0)^{k-1}\right| \le dn^{-k/(2k+1)}$$

for all $t \in [x_0 - Mn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$ with large probability. It follows that

$$\begin{aligned}\int_{x_0}^{\tau}(t-x_0)|\tilde{g}_n(t)-g_0(t)|dt &\le 2d\, n^{-k/(2k+1)}\int_{x_0}^{\tau}(t-x_0)dt \\ &= d\, n^{-k/(2k+1)}(\tau-x_0)^2 \\ &\le 4dM^2\, n^{-(k+2)/(2k+1)}\end{aligned}$$

with large probability. Finally, via empirical processes arguments and Lemma 7.1 with $d = 2$, it follows that

$$\left| \int_{x_0}^{\tau} (t - x_0)(\mathbb{G}_n(t) - G_0(t)) \right| = O_p(n^{-(k+2)/(2k+1)})$$

and the result follows for $j = k - 2$. The same result holds if we replace x_0 by any $x \in [x_0 - Mn^{-1/(2k+1)}, n^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$, for some $M > 0$; i.e., we can find $K > 0$ independent of x such that

$$\left| \tilde{\Delta}_n^{(k-2)}(x) \right| \le K n^{-(k+2)/(2k+1)}$$

with large probability. Now let $0 \le j \le k - 3$ and fix $\epsilon > 0$. Suppose that for all $j' > j$ and $M > 0$, there exists $c > 0$ such that for all $z \in [x_0 - Mn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$,

$$(k - 1 - j')! |\tilde{\Delta}_n^{(j')}(z)| \le c n^{-(2k-j')/(2k+1)}$$

with probability greater than $1 - \epsilon$. We can write

$$\begin{aligned}
&(k-1-j)! \tilde{\Delta}_n^{(j)}(y) \\
&= \int_0^y (y-t)^{k-1-j} d(\tilde{G}_n(t) - \mathbb{G}_n(t)) \\
&= \int_0^y ((y-x) + (x-t))^{k-1-j} \, d(\tilde{G}_n(t) - \mathbb{G}_n(t)) \\
&= \sum_{l=0}^{k-1-j} \binom{k-1-j}{l} (y-x)^l \int_0^y (x-t)^{k-1-j-l} d(\tilde{G}_n(t) - \mathbb{G}_n(t)) \\
&= \sum_{l=1}^{k-1-j} \binom{k-1-j}{l} (y-x)^l \int_0^y (x-t)^{k-1-j-l} d(\tilde{G}_n(t) - \mathbb{G}_n(t)) \\
&\quad + \int_0^y (x-t)^{k-1-j} d(\tilde{G}_n(t) - \mathbb{G}_n(t)) \\
&= \sum_{l=1}^{k-1-j} \binom{k-1-j}{l} (y-x)^l \tilde{\Delta}_n^{(j+l)}(y) \\
&\qquad + \tilde{\Delta}_n^{(j)}(x) + \int_x^y (x-t)^{k-1-j} d(\tilde{G}_n(t) - \mathbb{G}_n(t)) \, .
\end{aligned}$$

Take x to be a zero of $\tilde{\Delta}_n^{(j)}$ (such a zero can be constructed using the mean value theorem as we did for $j = k-2$ and $j = k-1$). Thus there exists $M > 0$ such that $x_0 - Mn^{-1/(2k+1)} \le x \le x_0 + Mn^{-1/(2k+1)}$. Now by applying the induction hypothesis, there exists $c > 0$ such that for all $y \in [x_0 - Mn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$, we have

$$\begin{aligned}\left|(k-1-j)!\tilde{\Delta}_n^{(j)}(y)\right| \;\; &\le \;\; c \sum_{l=1}^{k-1-j} \binom{k-1-j}{l} |y-x|^l n^{-(2k-(j+l))/(2k+1)} \\ &\quad + \left| \int_x^y (x-t)^{k-1-j} d(\tilde{G}_n(t) - \mathbb{G}_n(t)) \right|.\end{aligned}$$

But,

$$\begin{aligned}&\sum_{l=1}^{k-1-j} \binom{k-1-j}{l} |y-x|^l n^{-(2k-(j+l))/(2k+1)} \\ &\qquad \le \left(\sum_{l=1}^{k-1-j} \binom{k-1-j}{l} (2M)^l \right) n^{-(2k-j)/(2k+1)}\end{aligned}$$

and

$$\left| \int_x^y (x-t)^{k-1-j} d(\tilde{G}_n(t) - \mathbb{G}_n(t)) \right| = O_p(n^{-(2k-j)/(2k+1)})$$

by using empirical processes arguments. Therefore, the result holds for j and hence for all $j = 0, \ldots, k-1$. ∎

Proof of Lemma 3.6. The arguments are very similar to those used in Groeneboom, Jongbloed and Wellner (Groeneboom, Jongbloed, and Wellner (2001b)). We prove the lemma for $\tilde{H}_n^l$ as the arguments are similar for $\widehat{H}_n^l$. Let $c > 0$. On $[-c, c]$, define the vector-valued stochastic process

$$\begin{aligned}Z_n(t) \;\; = \;\; &\Big(\tilde{H}_n^l(t), \ldots, (\tilde{H}_n^l)^{(2k-2)}(t), \mathbb{Y}_n^l(t), \\ &\quad \ldots, (\mathbb{Y}_n^l)^{(k-2)}(t), (\tilde{H}_n^l)^{(2k-1)}(t), (\mathbb{Y}_n^l)^{(k-1)}(t) \Big).\end{aligned}$$

This stochastic process belongs to the space

$$E_k[-c, c] = (C[-c, c])^{3k-2} \times (D[-c, c])^2$$

where $C[-c,c]$ and $D[-c,c]$ are respectively the space of continuous and right-continuous functions on $[-c,c]$. We endow the space $E_k[-c,c]$ with the product topology induced by the uniform topology on $C[-c,c]$ and the Skorohod topology on $D[-c,c]$. By Proposition 3.1 and Lemma 3.5, we know that $(\tilde{H}_n^l)^{(j)}$ is tight in $C[-c,c]$ for $j = 0,\ldots,2k-2$. It follows from the same lemma together with the monotonicity of $(\tilde{H}_n^l)^{(2k-1)}$ that the latter is tight in $D[-c,c]$. On the other hand, since the processes $\left(\mathbb{Y}_n^l,\ldots,(\mathbb{Y}_n^l)^{(k-2)}\right)$ and $(\mathbb{Y}_n^l)^{(k-1)}$ converge weakly, they are tight in $(C[-c,c])^{k-1}$ and $D[-c,c]$ respectively. Now, for a fixed $\epsilon > 0$, there exists a $M > 0$ such that with probability greater than $1-\epsilon$, the process Z_n belongs to $E_{k,M}[-c,c]$ where $E_{k,M} = (C_M[-c,c])^{3k-2} \times (D_M[-c,c])^2$, and $C_M[-c,c]$ and $D_M[-c,c]$ are respectively the subset of functions in $C[-c,c]$ and the subset of monotone functions in $D[-c,c]$ that are bounded by M. Since the subspace $E_{k,M}[-c,c]$ is compact, we can extract from any arbitrary sequence $\{Z_{n'}\}$ a further subsequence $\{Z_{n''}\}$ that is weakly converging to some process

$$Z_0 = \left(H_0,\ldots,H_0^{(2k-2)},Y_0,\ldots,Y_0^{(k-2)},H_0^{(2k-1)},Y_0^{(k-1)}\right) \tag{6.35}$$

in $E_k[-c,c]$ and where $Y_0 = Y_k$. Now, consider the functions ϕ_1 and ϕ_2 : $E_k[-c,c] \mapsto \mathbb{R}$ defined by

$$\phi_1(z_1,\ldots,z_{3k}) = \inf_{t\in[-c,c]} (z_1(t) - z_{2k}(t)) \wedge 0$$

and

$$\phi_2(z_1,\ldots,z_{3k}) = \int_{-c}^{c} (z_1(t) - z_{2k}(t))dz_{3k-1}(t).$$

It is easy to check that the functions ϕ_1 and ϕ_2 are both continuous. By the continuous mapping theorem, it follows that $\phi_1(Z_0) = \phi_2(Z_0) = 0$ since $\phi_1(Z_{n''}) = \phi_2(Z_{n''}) = 0$ and therefore,

$$H_0(t) \geq Y_k(t),$$

for all $t \in [-c,c]$ and

$$\int_{-c}^{c} (H_0(t) - Y_k(t))dH_0^{(2k-1)}(t) = 0.$$

It is easy to see check that $(-1)^k H_0^{(2k-2)}$ is convex. Since $c > 0$ is arbitrary, we see that H_0 satisfies conditions (i) and (iii) of Theorem 3.1. Furthermore, outside the interval $[-c, c]$ we can take $\tilde{H}_n^l$ and $\mathbb{Y}_n^l$ to be identically 0. With this choice, the condition (iv) of Theorem 3.1 is satisfied. By uniqueness of the process H_k, it follows that $H_0 = H_k$. Since the limit is the same for any subsequence $\{Z_{n_l}\}$, we conclude that the sequence $\{Z_n\}$ converges weakly to

$$Z_k = \left(H_k, \ldots, H_k^{(2k-2)}, Y_k, \ldots, Y_k^{(k-2)}, H_k^{(2k-1)}, Y_k^{(k-1)}\right)$$

and in particular $Z_n(0) \to_d Z_k(0)$ and $(\tilde{H}_n^l)^{(j)}(0) \to_d H_k^{(j)}(0)$ for $j = 0, \ldots, 2k-1$. ∎

Proof of Theorem 3.2. We start with the direct estimation problems. For the LSE, we have for $j = 0, \ldots, k-1$

$$(\tilde{H}_n^l)^{(j)}(0) \quad = \quad s_1 s_2^j (\tilde{H}_n^{loc})^{(k)}(0) = n^{(k-j)/(2k+1)} c_j(g_0)(\tilde{g}_n^{(j)}(x_0) - g_0^{(j)}(x_0)),$$

whereas in the Maximum Likelihood case, we have

$$\begin{aligned}(\widehat{H}_n^l)^{(j)}(0) \quad &= \quad n^{(k-j)/(2k+1)} c_j(g_0) g_0(x_0) \left(\frac{(\hat{g}_n(x_0) - g_0(x_0))}{\hat{g}_n(x_0)} \right) \\ &\quad + \sum_{i=0}^{j-1} \binom{j}{i} \left(\hat{g}_n^{(i)}(x_0) - g_0^{(i)}(x_0)\right) \left(\frac{1}{\hat{g}_n(x)}\right)_{x=x_0}^{(j-i)}\end{aligned}$$

using the convention $\binom{j}{i} = 0$ if $i < 0$.

Weak convergence follows immediately from Lemma 3.6. Note that, for the MLE, the factor $g_0(x_0)/\hat{g}_n(x_0)$ converges in probability to 1. Moreover, for $j = 1, \ldots, k-1$ it can be shown, using Proposition 3.1 and uniform consistency of $\hat{g}_n^{(j)}$ in the neighborhood of x_0 (see Balabdaoui and Wellner (2004a), Propositions 3.2 and 3.3), that the second terms in the above expressions converge to 0 in probability, and hence

$$|(\widehat{H}_n^l)^{(k+j)}(0) - n^{(k-j)/(2k+1)} c_{k-j}(g_0)(\hat{g}_n^{(j)}(x_0) - g_0^{(j)}(x_0)| \to_p 0$$

for $j = 0, \ldots, k-1$.

For the inverse problem, the claim follows from Lemma 3.6 and the inverse formula given in (2.9) (see Section 2). ■

7. Appendix 2 - Proofs from Empirical processes theory

The following proposition is a slight generalization of Lemma 4.1 of Kim and Pollard (1990), page 201.

Lemma 7.1 *Let $\mathcal{F}$ be a collection of functions defined on $[x_0 - \delta, x_0 + \delta]$, with $\delta > 0$ small and let $\alpha > 0$. Suppose that for a fixed $x \in [x_0 - \delta, x_0 + \delta]$ and $R > 0$ such that $[x, x + R] \subseteq [x_0 - \delta, x_0 + \delta]$, the collection*

$$\mathcal{F}_{x,R} = \left\{ f_{x,y} \equiv f 1_{[x,y]}, \quad f \in \mathcal{F}, \quad x \le y \le x + R \right\}$$

admits an envelope $F_{x,R}$ such that

$$EF_{x,R}^2(X_1) \le K R^{2d-1}, \qquad R \le R_0,$$

for some $d \ge 1/2$ where $K > 0$ depends only on x_0, δ, and α. Moreover, suppose that

$$\sup_Q \int_0^1 \sqrt{\log N(\eta \|F_{x,R}\|_{Q,2}, \mathcal{F}_{x,R}, L_2(Q))} d\eta < \infty. \tag{7.36}$$

Then, for each $\epsilon > 0$ there exist random variables $M_n = M_n(\alpha)$ of order $O_p(1)$ such that

$$|(\mathbb{G}_n - G_0)(f_{x,y})| \ \le \ \epsilon |y - x|^{\alpha+d} + n^{-\frac{\alpha+d}{2\alpha+1}} M_n \qquad \text{for} \ \ |y - x| \le R_0. \tag{7.37}$$

Proof. By van der Vaart and Wellner (1996), theorem 2.14.1, page 239, it follows that

$$E \left\{ \sup_{f_{x,y} \in \mathcal{F}_{x,R}} |(\mathbb{G}_n - G_0)(f_{x,R})| \right\}^2 \ \le \ \frac{K}{n} EF_{x,R}^2(X_1) = O\left(n^{-1} R^{2d-1}\right) \tag{7.38}$$

for some constant $K > 0$ depending only on x_0, δ, and the entropy integral in (7.36). For any $f_{x,y} \in \mathcal{F}_{x,R}$, we write

$$(\mathbb{P}_n - P_0)(f_{x,y}) = (\mathbb{G}_n - G_0)(f_{x,y})$$

and define M_n by

$$\begin{aligned} M_n &= \inf\Big\{D > 0 : \Big|(\mathbb{P}_n - P_0)(f_{x,y})\Big| \le \epsilon(y-x)^{\alpha+d} + n^{-(\alpha+d)/(2\alpha+1)} D, \\ &\qquad \text{for all } \ f_{x,y} \in \mathcal{F}_{x,R}\Big\} \end{aligned}$$

and $M_n = \infty$ if no $D > 0$ satisfies the required inequality. For $1 \le j \le \lfloor R n^{1/(2\alpha+1)} \rfloor = j_n$, we have

$$\begin{aligned} & P(M_n > m) \\ &\le P\Big(\Big|(\mathbb{P}_n - P_0)(f_{x,y})\Big| > \epsilon(y-x)^{\alpha+d} \\ &\qquad + n^{-(\alpha+d)/(2\alpha+1)} m \ \text{ for some } f_{x,y} \in \mathcal{F}_{x,R}\Big) \\ &\le \sum_{1 \le j \le j_n} P\Big\{ n^{(\alpha+d)/(2\alpha+1)} \Big|(\mathbb{P}_n - P_0)(f_{x,y})\Big| > \epsilon(j-1)^{\alpha+d} + m \\ &\qquad \text{for some } f_{x,y} \in \mathcal{F}_{x,R}, (j-1)n^{-1/(2\alpha+1)} \le y - x \le j n^{-1/(2\alpha+1)} \Big\} \\ &\le \sum_{1 \le j \le j_n} n^{2(\alpha+d)/(2\alpha+1)} \frac{E\Big\{\sup_{y: 0 \le y-x < j n^{-1/(2\alpha+1)}} \Big|(\mathbb{P}_n - P_0)(f_{x,y-x})\Big|\Big\}^2}{(\epsilon(j-1)^{\alpha+d} + m)^2} \\ &= \sum_{1 \le j \le j_n} n^{2(\alpha+d)/(2\alpha+1)} \frac{E\Big\{\sup_{f_{x,y-x} \in \mathcal{F}_{x, j n^{-1/(2\alpha+1)}}} \Big|(\mathbb{P}_n - P_0)(f_{x,y-x})\Big|\Big\}^2}{(\epsilon(j-1)^{\alpha+d} + m)^2} \\ &\le C \sum_{1 \le j \le j_n} n^{2(\alpha+d)/(2\alpha+1)} n^{-1} n^{-(2d-1)/(2\alpha+1)} \frac{j^{2d-1}}{(\epsilon(j-1)^{\alpha+d} + m)^2} \\ &= C \sum_{1 \le j \le j_n} \frac{j^{2d-1}}{(\epsilon(j-1)^{\alpha+d} + m)^2} \le C \sum_{j=1}^{\infty} \frac{j^{2d-1}}{(\epsilon(j-1)^{\alpha+d} + m)^2} \searrow 0 \end{aligned}$$

as $m \nearrow \infty$ where $C > 0$ is a constant that depends only on x_0, δ, and α. Therefore it follows that (7.37) holds. ∎

In the following, we present VC-subgraph proofs for Lemma 4.4 and Proposition 6.1.

Proposition 7.1 *For $k \geq 2$ the classes of functions $\mathcal{F}^{(1)}_{y_0,R}$ and $\mathcal{F}^{(2)}_{y_0,R}$ given in (4.11) and (6.23) are VC-subgraph classes of functions.*

Proof. Consider the collection $\mathcal{F}^{(1)}_{y_0,R}$. We want to show that the class of subgraphs

$$\mathcal{D} = \{\{(t,c) \in \mathbb{R}^+ \times \mathbb{R} : c < f_t(x)\} : \\ x \in [\tau_0, \tau_{2k-3}],\ x_0 - \delta \leq y_0 < y_1 < \cdots < y_{2k-3} \leq y_0 + R\}$$

is a VC class of sets in $\mathbb{R}^+ \times \mathbb{R}$. If we show this, then the class of functions (4.11) is VC-subgraph. Alternatively, from van der Vaart and Wellner (1996), problem 11, page 152, it suffices to show that the "between graphs"

$$\mathcal{D}_1 = \big\{\{(t,c) \in \mathbb{R}^+ \times \mathbb{R} : 0 \leq c \leq f_t(x) \text{ or } f_t(x) \leq c \leq 0\} : \\ x \in [y_0, y_{2k-3}],\ x_0 - \delta \leq y_0 < y_1 < \cdots < y_{2k-3} \leq y_0 + R\}$$

is a VC class of sets. Let

$$\begin{aligned} \mathcal{D}_{1,j} &= \Big\{\{(t,c) \in \mathbb{R}^+ \times \mathbb{R} : 0 \leq c \leq f_t(x) 1_{[y_{j-1},y_j]}(t) \\ &\quad \text{or } f_t(x) 1_{[y_{j-1},y_j]}(t) \leq c \leq 0\} : x \in [\tau_0, \tau_{2k-3}], \\ &\quad x_0 - \delta \leq y_0 < y_1 < \cdots < y_{2k-3} \leq y_0 + R\Big\} \end{aligned}$$

for $j = 1, \ldots, 2k-3$. Since $t \mapsto f_t(x) 1_{[y_{j-1},y_j]}(t)$ is a polynomial of degree at most $k-1$ for each $j = 1, \ldots, k$, the classes $\mathcal{D}_{1,j}$ are all VC classes. Also note that

$$\mathcal{D}_1 \subset \mathcal{D}_{1,1} \sqcup \ldots \sqcup \mathcal{D}_{1,2k-3} \equiv \mathcal{D}_{\sqcup k}.$$

By Dudley (1999), theorem 2.5.3, page 153, $\mathcal{D}_{\sqcup k}$ is a VC class (or see van der Vaart and Wellner (1996), Lemma 2.6.17, part (iii), page 147). Hence $\mathcal{D}_1$ is a VC class and $\mathcal{F}^{(1)}$ is a VC - subgraph class.

The proof for the class $\mathcal{F}^{(2)}_{y_0,R}$ is similar.

Proposition 7.2 *The collection of functions*

$$\begin{aligned}\mathcal{F}_{y_0,R,\gamma} &= \Big\{ f_{y_0,y_1,\cdots,y_{2k-3},\lambda,h} : y_0 \le y_1 \le \cdots \le y_{2k-4} \le y_{2k-3} \le y_0 + R, \\ &\qquad \lambda \in [0,1], \text{ and } h \in C_\gamma^{k-2}[x_0-\delta, x_0+\delta] \Big\}\end{aligned}$$

defined below in (8.44) is a VC-class. Furthermore, we have

$$\|\mathbb{P}_n - P_0\|_{\mathcal{F}_{y_0,R,\gamma}} = o_p(y_{2k-3} - y_0)^{2k} + O_p(n^{-2k/(2k+1)})$$

where

$$\|\mathbb{P}_n - P_0\|_{\mathcal{F}_{y_0,R,\gamma}} = \sup_{f_{y_0,y_1,\cdots,y_{2k-3},\lambda,h} \in \mathcal{F}_{y_0,R,\gamma}} |(\mathbb{P}_n - P_0)(f_{y_0,y_1,\cdots,y_{2k-3},\lambda,h})|.$$

Proof. Fix $\eta > 0$, and let Q be a probability measure on $(0,\infty)$. Now using the same arguments as in the proof of Proposition 7.1, the collection $\mathcal{F}_{y_0,R}$ is VC, and we can find $D_1 = D_1(\delta, k) < \infty$ such that

$$\log N(\eta \|F_{y_0,R}\|_{Q,2}, \mathcal{F}_{y_0,R}, L_2(Q)) \le D_1 \log \frac{1}{\eta}.$$

On the other hand, by Theorem 2.7.1 of van der Vaart and Wellner (1996), page 155, there exists $D_2 = D_2(\delta, k) < \infty$

$$\log N(\eta\gamma, C_\gamma^{k-2}[x_0-\delta, x_0+\delta], \|\cdot\|_\infty) \le D_2 \left(\frac{1}{\eta}\right)^{\frac{1}{k-2}},$$

where the constant D_2 depends on k, and δ, but not on x_0. Note that $\mathcal{F} \equiv \mathcal{F}_{y_0,R}$ has bounded envelope function $F \equiv F_{y_0,R}$. Thus if $\{f_j\}$ is an $\eta\|F\|_{y_0,R}-$net with respect to $L_2(Q)$ for $\mathcal{F} \equiv \mathcal{F}_{y_0,R}$ and $\{g_{j'}\}$ is an $\eta\gamma-$net with respect to $\|\cdot\|_\infty$ for $\mathcal{G} \equiv C_\gamma^{k-2}[x_0-\delta, x_0+\delta]$, then $\{f_j \cdot g_{j'}\}$ is a $2\eta\gamma\|F\|_{Q,2}-$net for $\mathcal{F}\cdot\mathcal{G}$ with respect to $L_2(Q)$: for f, g with $\|f - f_j\|_{Q,2} \le \epsilon\|F\|_{Q,2}$ and $\|g - g_{j'}\|_\infty \le \epsilon\gamma$,

$$\begin{aligned}\|f\cdot g - f_j g_{j'}\|_{Q,2} &\le \gamma\|f - f_j\|_{Q,2} + \|F\|_{Q,2}\|g - g_{j'}\|_\infty \\ &\qquad \text{by the triangle inequality followed by use} \\ &\qquad \text{of the sup norm} \\ &\le \gamma\eta\|F\|_{Q,2} + \|F\|_{Q,2}\eta\gamma = 2\eta\gamma\|F\|_{Q,2}.\end{aligned}$$

It follows that

$$
\begin{aligned}
&N(2\eta\gamma\|F\|_{Q,2}, \mathcal{F}_{y_0,R,\gamma}, L_2(Q)) \\
&\qquad \le N(\eta\|F\|_{Q,2}, \mathcal{F}_{y_0,R}, L_2(Q)) \cdot N(\eta\gamma, C_\gamma^{k-2}[x_0-\delta, x_0+\delta], L_2(Q)).
\end{aligned}
\tag{7.39}
$$

By (7.39) and dominance of the second entropy bound as $\eta \searrow 0$, we conclude that

$$
\log N(\eta\gamma\|F\|_{Q,2}, \mathcal{F}_{y_0,R,\gamma}, L_2(Q)) \le K\left(\frac{1}{\eta}\right)^{\frac{1}{k-2}}.
$$

where K depends on k and δ (but not on R or Q). On the other hand, it follows from the error boundedness Conjecture (1.7) (also see Balabdaoui and Wellner (2005)) that $\mathcal{F}_{x,R,\gamma}$ admits the function

$$
F_{y_0,R,\gamma}(t) = C\gamma R^{k-1} 1_{[y_0,y_0+R]}(t)
$$

as an envelope, where $C > 0$ depends only on k. Now we can find a constant $D > 0$ depending only on η and g_0 and such that $0 < \sup_{t\in[x_0-\delta,\ x_0+\delta]} g_0(t) \le D$. We can write

$$
EF^2_{x,R,\gamma}(X_1) = C^2\gamma^2 R^{2(k-1)} \int_{y_0}^{y_0+R} g_0(t)dt \le C^2 D\gamma^2 R^{2k-1},
$$

and hence by van der Vaart and Wellner (1996), Theorem 2.14.2, page 240, there exists a constant K' depending only on x_0 and δ such that

$$
\begin{aligned}
&E\left\{\Big(\sup_{f_{y_0,y_1,\ldots,y_{2k-3},\lambda,h}\in\mathcal{F}_{y_0,R,\gamma}} |(\mathbb{G}_n - G_0)(f_{y_0,y_1,\ldots,y_{2k-3},\lambda,h})|\Big)^2\right\} \\
&\qquad \le \frac{K'}{n} EF^2_{y_0,R,\gamma}(X_1) = K'' n^{-1}\gamma^2 R^{2k-1}.
\end{aligned}
$$

Now, define

$$
\begin{aligned}
M_n &= \inf\Big\{ m > 0 : \big|(\mathbb{P}_n - P_0)(f_{y_0,y_1,\ldots,y_{2k-3},\lambda,h}\big| \le \epsilon(y_{2k-3}-y_0)^{2k} \\
&\qquad\qquad + n^{-2k/(2k+1)}m, \quad \text{for all } f_{y_0,y_1,\ldots,y_{2k-3},\lambda,h} \in \mathcal{F}_{y_0,R,\gamma}\Big\}
\end{aligned}
$$

and $M_n = \infty$ if no $m > 0$ satisfies the required inequality. Using arguments very similar to those of the proof of Lemma 7.1, we can show that $M_n = O_p(1)$, which proves our second claim. ■

8. Appendix 3 - Proofs for Subsections 4.2 and 4.3

To prove Lemma 4.3, we need the following lemma:

Lemma 8.1 *Let $k \geq 2$ be an integer. The monospline M_k with simple knots $\xi_0 = -k + 3/2$, $\xi_1 = -k + 5/2$, ... , $\xi_{2k-4} = k - 5/2$, $\xi_{2k-3} = k - 3/2$ and such that $M_k(\xi_j) = M_k'(\xi_j) = 0$ for $j = 0, \ldots, 2k - 3$ satisfies $(-1)^k M_k \geq 0$ on $[-k + 3/2, k + 3/2] \equiv [\xi_0, \xi_{2k-3}]$.*

Proof. Consider the function D_{2k} defined on $[-k + 3/2, k - 3/2]$ by

$$D_{2k}(t) = \mathcal{B}_{2k}(t - \xi_j) - B_{2k}, \quad \text{on } [\xi_j, \xi_{j+1}] \equiv [\xi_j, \xi_j + 1]$$

for $j = 0, \ldots, 2k - 3$, where $\mathcal{B}_{2k}$ is the normalized Bernoulli polynomial of degree $2k$ (defined on $[0, 1]$) and $B_{2k} = \mathcal{B}_{2k}(0)$. By the known properties of Bernoulli polynomials (see e.g. Bojanov, Hakopian and Sahakian (1993), pages 117-124), we have $D_{2k}^{(l)}(\xi_j-) = D_{2k}^{(l)}(\xi_j+)$ for $l = 0, \ldots, 2k - 2$. Hence, D_{2k} is a monospline of degree $2k$. Furthermore, since $D_{2k}(\xi_j) = D_{2k}'(\xi_j) = 0$, it follows that $M_k = D_{2k}$ on $[-k + 3/2, k - 3/2]$. Now, the sign of M_k is the same as the sign of $\mathcal{B}_{2k} - B_{2k}$ on $[0, 1]$. But the latter is determined by the sign of $\mathcal{B}_{2k}(1/2) - B_{2k}$ as 0 and 1 are the only zeros of $\mathcal{B}_{2k} - B_{2k}$ on $[0, 1]$. Using the formula

$$\mathcal{B}_{2k}(1/2) = -(1 - 2^{1-2k})B_{2k}$$

(see e.g. Abramowitz and Stegun (1972), formula 23.1.21, page 805) and the fact that $B_{2k} > 0$ (< 0) when k is odd (even), it follows that $M_k \leq 0$ (≥ 0) when k is odd (even), i.e. $(-1)^k M_k \geq 0$. ■

Proof of Lemma 4.3. The first part of the claim follows from Proposition 1 of Michelli (1972); see also de Boor (2004). For the second part, let ξ be a fixed point in $\cup_{j=0}^{2k-4}(\tau_j, \tau_{j+1})$. We can assume without loss of generality that $\xi \in (\tau_0, \tau_1)$. There exists $\lambda \in (0,1)$ such that $\xi = \lambda\tau_0 + (1-\lambda)\tau_1$. Consider now the function

$$(\tau_0, \ldots, \tau_{2k-3}) \mapsto \frac{e_k(\xi) + |e_k(\xi)|}{2e_k(\xi)}.$$

Note that it is possible to divide by $e_k(\xi)$ since a zero of e_k must be a knot. It is easy to see that the function defined above is continuous in $\tau_0, \ldots, \tau_{2k-3}$. Furthermore, it can only take two possible values, 0 or 1, and therefore has to be constant. But, when the knots are equally distant, we know from Lemma 8.1 that the constant is 0 (1) if k is odd (even). It follows that $(-1)^k e_k(\xi) > 0$. ■

To prove Lemma 4.7, we need to establish the following result.

Lemma 8.2 *For any $\epsilon > 0$, there exists $K > 0$ (depending on k) such that for $j = 1, \ldots, 2k-3$ the event*

$$0 < (-1)^{k-1}(\hat{g}_n^{(k-1)}(\tau_j) - \hat{g}_n^{(k-1)}(\tau_{j-1})) < K\ (\tau_j - \tau_{j-1})$$

occurs with probability greater than $1 - \epsilon$.

Proof. A picture is sufficient to prove the lemma, but more formally we have for $x \in [x_0 - \delta, x_0 + \delta]$ for small $\delta > 0$

$$\begin{aligned} &\frac{\hat{g}_n^{(k-2)}(x-h) - \hat{g}_n^{(k-2)}(x)}{-h} \\ &\leq \ \hat{g}_n^{(k-1)}(x-) \leq \hat{g}_n^{(k-1)}(x+) \leq \frac{\hat{g}_n^{(k-2)}(x+h) - \hat{g}_n^{(k-2)}(x)}{h} \qquad (8.40) \end{aligned}$$

(we assume here that k is even). We denote by $\Delta\hat{g}_n^{(k-1)}(x)$ the height of the jump of $\hat{g}_n^{(k-1)}$ at the point x; i.e., $\Delta\hat{g}_n^{(k-1)}(x) = \hat{g}_n^{(k-1)}(x+) - \hat{g}_n^{(k-1)}(x-)$, and by Δx the value of the corresponding gap (if $x = \tau_j$, then $\Delta x = \tau_j - \tau_{j-1}$). The inequality in (8.40) implies that for all $0 < h < \Delta x$, we have

$$\begin{aligned} 0 \quad \leq \quad & \limsup_{n\to\infty} \frac{\Delta \hat{g}_n^{(k-1)}(x)}{\Delta x} \\ & \leq \frac{1}{h}\left\{ \frac{g_0^{(k-2)}(x+h) - g_0^{(k-2)}(x)}{h} - \frac{g_0^{(k-2)}(x-h) - g_0^{(k-2)}(x)}{-h} \right\}. \end{aligned}$$

On the other hand, we know from our working assumptions that $g_0^{(k-2)}$ is twice continuously differentiable in the neighborhood of x_0. Therefore, using Taylor expansion, we have

$$g_0^{(k-2)}(x+h) = g_0^{(k-2)}(x) + h g_0^{(k-1)}(x) + \frac{1}{2} h^2 g_0^{(k)}(x) + o(h^2)$$

and

$$g_0^{(k-2)}(x-h) = g_0^{(k-2)}(x) - h g_0^{(k-1)}(x) + \frac{1}{2} h^2 g_0^{(k)}(x) + o(h^2)$$

and hence

$$\begin{aligned} \frac{1}{h}\left\{ \frac{g_0^{(k-2)}(x+h) - g_0^{(k-2)}(x)}{h} - \frac{g_0^{(k-2)}(x-h) - g_0^{(k-2)}(x)}{-h} \right\} \quad &= \quad g_0^{(k)}(x) + o(1) \\ &\leq \quad K \end{aligned}$$

where K can be taken e.g. to be equal to $2\sup_{t\in[x_0-\delta, x_0+\delta]} |g_0^{(k)}(t)|$. It follows that for n large enough and for all $j \in 1, \ldots, 2k-3$,

$$0 < \hat{g}_n^{(k-1)}(\tau_j) - \hat{g}_n^{(k-1)}(\tau_{j-1}) \leq K\ (\tau_j - \tau_{j-1})$$

with large probability. ■

Proof of Lemma 4.7. We start by showing that

$$\sup_{\bar{\tau}\in[\tau_0, \tau_{2k-3}]} |\mathcal{E}_2(\bar{\tau})| = o_p((\tau_{2k-3} - \tau_0)^{2k}) + O_p(n^{-(2k)/(2k+1)}). \tag{8.41}$$

We have

$$|\mathcal{E}_2(\bar{\tau})|$$

$$= \left| \int_0^\infty \left\{ (\bar\tau - t)^{k-1} 1_{[\tau_0,\bar\tau]}(t) \left(\frac{1}{\hat g_n(t)} - \frac{1}{g_0(\tau_0)} \right) \right.\right.$$
$$\left.\left. - \mathcal{H}_k \left[(\cdot - t)^{k-1} 1_{[\tau_0,\cdot]}(t) \left(\frac{1}{\hat g_n(t)} - \frac{1}{g_0(\tau_0)} \right) \right] (\bar\tau) \right\} d(\mathbb{G}_n(t) - G_0(t)) \right|.$$

In other words, one can view $\mathcal{E}_2(\tau)$ in the following way: For a fixed $t \in [\tau_0, \tau_{2k-3}]$, we compute the value at the point $\bar\tau$ of the Hermite interpolation error for interpolating the function

$$x \mapsto (x-t)^{k-1} 1_{[\tau_0,x]}(t) \left(\frac{1}{\hat g_n(t)} - \frac{1}{g_0(\tau_0)} \right)$$

or, since $1_{[\tau_0,x]}(t) = 1_{[x \ge t]}$ since $t \ge \tau_0$,

$$\begin{aligned} x &\mapsto (x-t)^{k-1} 1_{[x \ge\ t]} \left(\frac{1}{\hat g_n(t)} - \frac{1}{g_0(\tau_0)} \right), \\ &= (x-t)_+^{k-1} \left(\frac{1}{\hat g_n(t)} - \frac{1}{g_0(\tau_0)} \right). \end{aligned} \tag{8.42}$$

This yields a function of t, which is then integrated with respect to $(\mathbb{G}_n - G_0)$. Let us denote by

$$f_{\tau_0,\ldots,\tau_{2k-3},\bar\lambda,\hat h_n}(t)$$

the function that assigns to each $t \in [\tau_0, \tau_{2k-4}]$ the Hermite interpolation error at the point $\bar\tau$ for interpolating the function defined in (8.42), where $\bar\lambda = (\bar\tau - \tau_0)/(\tau_{2k-3} - \tau_0)$ and $\hat h_n = 1/\hat g_n - 1/g_0(\tau_0)$.

Let $\epsilon > 0$, and $\delta > 0$ such that $[\tau_0, \tau_{2k-4}] \subset [x_0 - \delta, x_0 + \delta]$. Now, uniform consistency of the derivative $\hat g_n^{(j)}, j = 0, \cdots, k-2$ implies uniform boundedness of $\hat h_n^{(j)}$. Hence, for $\gamma > 0$ large enough there exists a $N = N_\gamma \in \mathbb{N}$ such that for $n > N$, the probability of the event

$$J_{x_0,\delta,\gamma} \equiv \left\{ \omega : \max_{0 \le j \le k-2} \sup_{t \in [x_0-\delta, x_0+\delta]} \left| \left(\frac{1}{\hat g_n(\omega,t)} - \frac{1}{g_0(\tau_0)} \right)^{(j)} \right| \le \gamma \right\} \tag{8.43}$$

is greater than $1 - \epsilon$. Note that for $j = 0$, γ can be taken arbitrarily small which will yield a stronger result. However, this is not true for $j > 0$ since the term $-1/g_0(\tau_0)$ disappears. In what follows, we consider the case where the previous event occurs.

Now fix $y_0 \in [x_0 - \delta, x_0 + \delta - R]$. Consider the collection

$$\begin{aligned} \mathcal{F}_{y_0,R,\gamma} \quad = \quad & \Big\{ f_{y_0,y_1,\cdots,y_{2k-3},\lambda,h} : y_0 \le y_1 \le \cdots \le y_{2k-4} \le y_{2k-3} \le y_0 + R, \\ & \lambda \in [0,1], \text{and } h \in C_\gamma^{k-2}[x_0 - \delta, x_0 + \delta] \Big\} \end{aligned} \tag{8.44}$$

where $C_\gamma^{k-2}[x_0 - \delta, x_0 + \delta]$ is the set of functions on $[x_0 - \delta, x_0 + \delta]$ whose j-th derivative, $j = 0, \cdots, k-2$, is uniformly bounded by γ. Explicitly, a function in the previous collection can be written as

$$\begin{aligned} & f_{y_0,y_1,\cdots,y_{2k-3},\lambda,h}(t) \\ & = \Big\{ (\lambda y_0 + (1-\lambda) y_{2k-3} - t)_+^{k-1} - [\mathcal{H}_k(\cdot - t)_+^{k-1}](\lambda y_0 + (1-\lambda) y_{2k-3}) \Big\}\, h(t). \end{aligned}$$

We recall here that $h \in C_\gamma^{k-2}[x_0 - \delta, x_0 + \delta]$, where $C_\gamma^{k-2}[x_0 - \delta, x_0 + \delta]$ is defined above. If we denote by $\mathcal{F}_{y_0,R}$, the collection of functions appearing in the first term on the right side of the previous display, we have

$$\mathcal{F}_{y_0,R,\gamma} \subset \mathcal{F}_{y_0,R} \cdot C_\gamma^{k-2}[x_0 - \delta, x_0 + \delta].$$

By Proposition 7.2, we have

$$\sup_{\bar{\lambda} \in [0,1]} \left| (\mathbb{P}_n - P_0)(f_{\tau_0,\tau_1,\ldots,\tau_{2k-3},\bar{\lambda},\hat{h}_n}) \right| = o_p((\tau_{2k-3} - \tau_0)^{2k}) + O_p(n^{-2k/(2k+1)})$$

or equivalently

$$\sup_{\bar{\tau} \in [\tau_0, \tau_{2k-3}]} |\mathcal{E}_2(\bar{\tau})| = o_p((\tau_{2k-3} - \tau_0)^{2k}) + O_p(n^{-2k/(2k+1)}).$$

Finally, we focus on the error term, $\mathcal{E}_1$. Recall that the corresponding interpolated function is given by

$$\check{f}_n(\bar{\tau}) = -\int_{\tau_0}^{\bar{\tau}} (\bar{\tau} - t)^{k-1} \frac{1}{\hat{g}_n(t)} (\hat{g}_n(t) - g_0(\tau_0))(\hat{g}_n(t) - g_0(t)) dt$$

for all $\bar{\tau} \in [\tau_0, \tau_{2k-3}]$. Note that the function is $(2k-1)$-times differentiable on $[\tau_0, \tau_{2k-3}]$, and we have

$$\begin{aligned}
&\|\mathcal{H}_k \check{f}_n - \check{f}_n\|_\infty \\
&\le \frac{1}{(2k-1)!} \int_{\tau_0}^{\tau_{2k-3}} \|\mathcal{H}_k[(t-\cdot)_+^{2k-1}] - (t-\cdot)_+^{2k-1}\|_\infty \, |d\check{f}_n^{(2k-1)}(t)| \\
&\le d_k (\tau_{2k-3} - \tau_0)^{2k-1} \int_{\tau_0}^{\tau_{2k-3}} |d\check{f}_n^{(2k-1)}(t)|,
\end{aligned}$$

for some $d_k > 0$. On the other hand, we have

$$\begin{aligned}
\check{f}_n^{(2k-1)}(t) &= \left[\left(\hat{g}_n(t) - g_0(t) \right) \left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)} \right) \right]^{(k-1)} \\
&= \sum_{j=0}^{k-1} \binom{k-1}{j} \left(\hat{g}_n^{(j)}(t) - g_0^{(j)}(t) \right) \left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)} \right)^{(k-1-j)},
\end{aligned}$$

and hence

$$\begin{aligned}
&d\check{f}_n^{(2k-1)}(t) \\
&= (\hat{g}_n(t) - g_0(t))\, d\left[\left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)} \right)^{(k-1)} \right] \\
&\quad + d\left(\hat{g}_n^{(k-1)}(t) - g_0^{(k-1)}(t) \right) \left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)} \right) \\
&\quad + \sum_{j=1}^{k-2} \binom{k-1}{j} \left(\hat{g}_n^{(j+1)}(t) - g_0^{(j+1)}(t) \right) \left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)} \right)^{(k-1-j)} dt \\
&\quad + \sum_{j=1}^{k-2} \binom{k-1}{j} \left(\hat{g}_n^{(j)}(t) - g_0^{(j)}(t) \right) \left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)} \right)^{(k-j)} dt \\
&= dh_1(t) + dh_2(t) + dh_3(t) + dh_4(t).
\end{aligned}$$

The last two functions, dh_3 and dh_4 are easier to handle, since we can see that uniform consistency of the derivatives of the MLE implies that

$$\sup_{t \in [\tau_0, \tau_{2k-3}]} |h_3'(t)| = \sup_{t \in [\tau_0, \tau_{2k-3}]} |h_4'(t)| = o_p(1),$$

and hence

$$(\tau_{2k-3} - \tau_0)^{2k-1} \int_{\tau_0}^{\tau_{2k-3}} |h_3'(t) + h_4'(t)| dt = o_p((\tau_{2k-3} - \tau_0)^{2k}).$$

Now we need go back to study h_1 and h_2 and the corresponding error terms, and we start with h_2. Recall that

$$\begin{aligned}\int_{\tau_0}^{\tau_{2k-3}} |dh_2(t)| &= \int_{\tau_0}^{\tau_{2k-3}} \left| d\left(\hat{g}_n^{(k-1)}(t) - g_0^{(k-1)}(t)\right) \left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)}\right)\right| \\ &\le \int_{\tau_0}^{\tau_{2k-3}} \left| d\hat{g}_n^{(k-1)}(t) \left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)}\right)\right| \\ &\quad + \int_{\tau_0}^{\tau_{2k-3}} \left| dg_0^{(k-1)}(t) \left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)}\right)\right|.\end{aligned}$$

The second term is $o_p((\tau_{2k-3} - \tau_0))$ since $dg_0^{(k-1)}(t) = g_0^{(k)}(t)dt$ (we apply the same argument used for dh_3 and dh_4). As for the first term, we have

$$\begin{aligned}&\int_{\tau_0}^{\tau_{2k-3}} \left| d\hat{g}_n^{(k-1)}(t) \left(\frac{\hat{g}_n(t) - g_0(\tau_0)}{\hat{g}_n(t)}\right)\right| \\ &= \sum_{j=1}^{2k-3} |\hat{g}_n^{(k-1)}(\tau_j) - \hat{g}_n^{(k-1)}(\tau_{j-1})| \left|\frac{\hat{g}_n(\tau_j) - g_0(\tau_0)}{\hat{g}_n(\tau_j)}\right| \\ &\le D \sum_{j=1}^{2k-3} (\tau_j - \tau_{j-1}) \left|\frac{\hat{g}_n(\tau_j) - g_0(\tau_0)}{\hat{g}_n(\tau_j)}\right|, \quad \text{by Lemma 8.2} \\ &\le D(\tau_{2k-3} - \tau_0) \sum_{j=1}^{2k-3} \left|\frac{\hat{g}_n(\tau_j) - g_0(\tau_0)}{\hat{g}_n(\tau_j)}\right| \\ &= o_p((\tau_{2k-3} - \tau_0))\end{aligned}$$

by uniform consistency of the MLE and continuity of g_0 which imply that $\hat{g}_n(\tau_j) - g_0(\tau_0) = o_p(1)$ for $j = 1, \ldots, 2k-3$. Similar arguments can be used for h_1. We conclude that the associated error term is of the order

$$o_p((\tau_{2k-3} - \tau_0)^{2k}),$$

or using our notation above

$$\sup_{\bar{\tau} \in [\tau_0, \tau_{2k-3}]} |\mathcal{E}_1(\bar{\tau})| = o_p((\tau_{2k-3} - \tau_0)^{2k}).$$

■

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

References

Abramowitz, M. and Stegun I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publishers, New York.

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

Balabdaoui, F. and Wellner, J. A. (2004a-c). Estimation of a k-monotone density, parts 1 - 3. Technical Reports 459 - 461, Department of Statistics, University of Washington. Available at: *http://www.stat.washington.edu/www/research/reports/2004/*.

Balabdaoui, F. and Wellner, J. A. (2005). Conjectures of error boundedness in a new Hermite interpolation problem via splines of odd-degree splines. Technical Report 480, Department of Statistics, University of Washington. Available at: *http://www.stat.washington.edu/www/research/reports/2004/*.

Bojanov, B. D., Hakopian, H. A. and Sahakian, A. A. (1993). *Spline Functions and Multivariate Interpolations*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Davidian, M. and Gallant, A. R. (1992). Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. *J. Pharmacokinetics and Biopharmaceutics* **20**, 529 - 556.

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 Rev.* **16**, 531 - 544.

de Boor, C. (2001) *Practical Guide to Splines*. Springer-Verlag, New York.

de Boor, C. (2004). *http://www.cs.wisc.edu/ deboor/toast/pages09.html.*

de Boor, C. and Fix, G. J. (1973). Spline approximation by quasi-interpolants. *J. Approx. Theory* **8**, 19 - 45.

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

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

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.

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. and Jongbloed, G. (1995). Isotonic estimation and rates of convergence in Wicksell's problem. *Ann. Statist.* **23**, 1518 - 1542.

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.

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.

Karlin, S. and Ziegler, Z. (1966). Chebyshevian spline functions. *SIAM J. Numerical Anal.* **3**, 514 - 543.

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

Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample disttribution function. *Z. Wahrsch. verw. Geb.* **32**, 111 - 131.

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

Mallet, A. (1986). A maximum likelihood estimation method for random coefficient regression models. *Biometrika* **73**, 645 - 656.

Mallet, A., Mentre, F., Steimer, J-L., and Lokiec, F. (1988). Nonparametric maximum likelihood estimation for population pharmacokinetics, with application to cyclosporine. *J. Pharmacokinetics and Biopharmaceutics* **16**, 311-327.

Mammen, E. (1991). Nonparametric regression under qualitaive smoothness assumptions *Ann. Statist.* **19**, 741 - 759.

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **25**, 387 - 413.

Michelli, C. (1972). The fundamental theorem of algebra for monosplines with multiplicities. In *Linear Operators and Approximation*, 419 - 430. Birkhäuser, Basel.

Millar, R. (1989). *Estimation of mixing and mixed distribution.* P.h.D dissertation, University of Washington, Department of Statistics.

Miller, D. R. and Sofer, A. (1986). Least-squares regression under convexity and higher-order difference constraints with application to software reliability. *Advances in order restricted statistical inference (Iowa City, Iowa, 1985), 91–124*, *Lecture Notes in Statist.* **37**, Springer, Berlin.

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

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

Schoenberg, I. J. (1938). Metric spaces and completely monotone functions. *Ann. of Math.* **39**, 811 - 841.

Schoenberg, I. J. (1963). On interpolation by spline functions and its minimal property. In *Proceedings at the Conference held at the Mathematical Research Institute at Oberwolfach, August 4-10, 1963.* Birkhäuser Verlag Basel, 109-128.

Schoenberg, I. J. and Whitney, A. (1953). On Pólya frequency functions III: The positivity of translation determinants with application to the interpolation problem by spline curves, *Trans. Amer. Math. Soc.* **74**, 246-259.

Shadrin, A. Yu. (2001). The L_∞-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.

Shorack, G. and Wellner J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley and Sons, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439 - 454.

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

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