Nonparametric estimation of a k-monotone density: A new asymptotic distribution theory.

Fadoua Balabdaoui

Nonparametric Estimation of a k-monotone Density: A New Asymptotic Distribution Theory

Fadoua Balabdaoui

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

2004

Program Authorized to Offer Degree: Statistics

University of Washington

Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Fadoua Balabdaoui

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Chair of Supervisory Committee:

Jon A. Wellner

Reading Committee:

Jon A. Wellner

Tilmann Gneiting

Piet Groeneboom

Date:

University of Washington

Abstract

Nonparametric Estimation of a k-monotone Density:
A New Asymptotic Distribution Theory

by Fadoua Balabdaoui

Chair of Supervisory Committee:

Professor Jon A. Wellner
Department of Statistics

In this dissertation, we consider the problem of nonparametric estimation of a k-monotone density on $(0, \infty)$ for a fixed integer $k \geq 1$ via the methods of Maximum Likelihood (ML) and Least Squares (LS).

In the introduction, we present the original question that motivated us to look into this problem and also put other existing results in our general framework. In Chapter 2, we study the MLE and LSE of a k-monotone density g_0 based on n i.i.d. observations. Here, our study of the estimation problem is local in the sense that we only study the estimator and its derivatives at a fixed point $x_0 > 0$. Under some specific working assumptions, asymptotic minimax lower bounds for estimating $g_0^{(j)}(x_0)$, $j = 0, \cdots, k-1$ are derived. These bounds show that the rates of convergence of any estimator of $g_0^{(j)}(x_0)$ can be *at most* $n^{-(k-j)/(2k+1)}$. Furthermore, under the same working assumptions we prove that this rate is achieved by the j-th derivative of either the MLE or LSE if a certain conjecture concerning the error in a particular Hermite interpolation problem holds.

To make the asymptotic distribution theory complete, the limiting distribution needs to be determined. This distribution depends on a very special stochastic process H_k which is almost surely uniquely defined on $\mathbb{R}$. Chapter 3 is essentially devoted to an effort to prove the existence of such a process and to establish conditions characterizing it. It turns out

that we can establish the existence and uniqueness of the process H_k if the same conjecture mentioned above with the finite sample problem holds. If Y_k is the $(k-1)$-fold integral of two-sided Brownian motion + $(k!/(2k)!)\, t^{2k}$, then H_k is a random spline of degree $2k-1$ that stays above Y_k if k is even and below it if k is odd. By applying a change of scale, our results include the special cases of estimation of monotone densities ($k=1$), and monotone and convex densities ($k=2$) for which an asymptotic distribution theory is available.

Iterative spline algorithms developed to calculate the estimators and approximate the process H_k on finite intervals are described in Chapter 4. These algorithms exploit both the spline structure of the estimators and the process H_k as well as their characterizations and are based on iterative addition and deletion of the knot points.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First of all, I wish to express my deepest gratitude to my supervisor Professor Jon A. Wellner who, when I asked him the first time whether it is possible to work with him, did not hesitate to accept. I would like to take this opportunity to thank him for being always available and for encouraging me to give my best.

I would like to thank Professors Piet Groeneboom and Eric cator for many stimulating discussions about my research during my visit to Delft University of Technology. Many thanks are also due to Professors Tilmann Gneiting and Peter Guttorp for their great support and encouragement, to Professors Marina Meila and Peter Hoff for being available whenever I needed their help. I also thank Professor Paul Cho, my GSR, for serving in my committee.

Special thanks are due to Professor Geurt Jongbloed, Free University, and Karim Filali, University of Washington, for their valuable help with the computational aspect of this work. I am also very much indebted to Professors Nira Dyn, Tel-Aviv University, and Carl de Boor, University of Wisconsin-Madison, for their inestimable contribution to the progress of this research.

I am grateful to our Program Coordinator Kristin Sprague for her immediate help with administrative matters whenever I needed it. I also thank my friends and colleagues for always keeping my spirits high.

Finally, I would like to thank my parents for their continuous moral support. I owe special thanks to my husband for his great love and constant encouragement.

DEDICATION

To Mom, Dad, Dirk and Nisrine

Chapter 1

INTRODUCTION

Our interest in nonparametric estimation of a k-monotone density was first motivated by Jewell (1982); Jewell considered the nonparametric Maximum Likelihood estimator of a scale mixture of Exponentials g,

$$g(x) = \int_0^\infty t \exp(-tx) dF(t), \qquad x > 0$$

where F is some distribution function concentrated on $(0, \infty)$. Such a scale mixture of Exponentials is a possible model for lifetime distributions when the population that is at risk of failure or deterioration is nonhomogenous and when one is not willing to assume the number of its components to be known. See Jewell (1982) for a survey of the application of the model in different fields.

Suppose that $X_1, \cdots, X_n$ are n independent observations from a common scale mixture of Exponentials g. Jewell (1982) established that the Maximum Likelihood estimator (MLE), of the mixing distribution F, $\hat{F}_n$ say, exists and is discrete with at most n support points. This implies that the MLE of the true mixed density g, $\hat{g}_n$ say, is a finite mixture of Exponentials with at most n components. This result also follows from the work of Lindsay (1983a), Lindsay (1983b), and Lindsay (1995) on nonparametric maximum likelihood in a very general mixture model setting. Jewell (1982) was also able to establish uniqueness and strong consistency of the MLE and used an EM algorithm to compute it. As in other mixture models, there are two main estimation problems of interest when considering a scale mixture of Exponentials: the *direct* and *inverse* problems. In the first one, the goal is to estimate the mixed density g *directly* from the observed data, whereas in the second one the focus is on the underlying mixing distribution F. To our knowledge, the exact rate of convergence of the MLE is still unknown in both problems and thus the asymptotic distribution theory is yet to be developed. In the inverse problem and under additional assumptions on the mixing distribution, asymptotic lower bounds on the rate of convergence of a consistent estimator were derived. For example, Millar (1989) assumed that the mixing distribution F belongs to the class $G_{m,M}$ of all mixing distributions defined on some subset $A \subset \mathbb{R}$ and have a density f that is m-differentiable and such that $\sup_{x \in A} |f^{(j)}(x)| < M, j = 0, \cdots, m$. Using characteristic function techniques, Millar (1989) could establish that

$$(\log n)^{-m} \text{ and } (\log n)^{-(m+1)}$$

are uniform asymptotic lower bounds on the rate of estimation of the mixing density f and the distribution function F at a fixed point x_0 respectively. See Millar (1989) for more details about the definition of uniformity.

Although we want to consider the class of all mixing distributions, this result can be used at least heuristically to derive bounds in more general settings. For $m = 0$, where we impose the minimal smoothness constraints on the mixing distribution F, the asymptotic lower bound for estimating $F(x_0)$ specializes to $1/\log n$. The logarithmic order of these lower bounds show how slow the rate of convergence can be in this kind of nonparametric setting. The estimation problem is far from being regular and therefore one should expect the rate of convergence to be slower than $\sqrt{n}$. In mixture models with smoother kernels, this rate of convergence is expected to be slower. The scale mixture of Exponentials is one example of a "smooth mixture". Another good example is location mixtures of Gaussians. This model is very often used to take measurement error into account. Formally, if X is some random variable with an unknown distribution function F, one gets to observe only $Y = X + Z$, where $Z \sim \mathcal{N}(0, \sigma_0^2)$ and $\sigma_0 > 0$ is supposed to be known. The density of X is given by the convolution of ϕ, the normal density and the distribution function F. Several authors were interested in the inverse problem which is also known as the Gaussian deconvolution problem. The work of Stefanski and Carroll (1990), Carroll and Hall (1988) , and Fan (1991) suggest that the rate of convergence of a consistent estimator of the underlying distribution F, if achieved, would be of the order of $1/\sqrt{\log n}$. Note that this rate is even slower than the expected $\log n$ in the case of scale of mixture of Exponentials.

In the direct problem where the focus is on the mixed density, the sieve MLE was studied by Ghosal and Van der Vaart (2001). By considering a particular class of mixing distributions, the authors could show that $\log n/\sqrt{n}$ is an upper bound for its rate of convergence. This bound is much faster when compared to the one obtained in the inverse problem. But this is not surprising if we associate the difficulty of estimation to the "size" of the class to which the distribution function or the density belongs. In this particular case, the mixed density belongs to a small class of densities that have to be equal to the convolution of the normal density and some distribution function F. It follows that any element of this class has to be infinitely differentiable. But on the other hand, this same smoothness makes the task of "untangling" the underlying distribution F from the Gaussian noise to be statistically hard.

As for the scale mixture of Exponentials, the exact asymptotic distribution of the MLE in the mixture of Gaussians is still to be derived. Although the two models are very different, one can see that some mathematical connection can be made through the exponential form of their kernels. We have not pursued thoroughly this thought as it is beyond the scope of this thesis, but we believe that getting a better understanding of the asymptotics of the MLE in scale mixture of Exponentials might be helpful in achieving the same thing for mixture of Gaussians.

Part of the difficulty of knowing more about the asymptotic behavior of the MLE in these kind of nonparametric models is primarily due to the implicit nature of the characterizations of the estimators. For the scale mixture of Exponentials, Jewell (1982) established that $\hat{g}_n$ is the MLE of the mixed density if and only if

$$\int_0^\infty \frac{\lambda \exp(-\lambda x)}{\hat{g}_n(x)} d\mathbb{G}_n(x) \begin{cases} \leq 1, & \lambda > 0 \\ = 1, & \text{if } \lambda \text{ is a support point of } \hat{F}_n \end{cases}$$

where $\mathbb{G}_n$ is the empirical distribution function. For the characterization of the MLE in a location mixture of Gaussians, see Groeneboom and Wellner (1992), Proposition 2.3, page 58. However, although there are no standard methods available to make these characterizations easily exploitable to derive the exact asymptotic distribution of the MLE, it seems that more is known about the class of scale mixture of Exponentials itself. Indeed, Jewell (1982) noted that g is a scale mixture of Exponentials if and only if the complement of its distribution function is the Laplace transform of some distribution function F. Jewell (1982) also recalled the fact that the class of scale mixtures of Exponentials can be identified as the class of completely monotone densities (Bernstein's theorem) where by definition, a function f on $(0,\infty)$ is completely monotone if and only if f is infinitely differentiable on $(0,\infty)$ and $(-1)^k f^{(k)} \geq 0$, for $k \in \mathbb{N}$ (see, e.g., Widder (1941), Feller (1971), Williamson (1956), Gneiting (1999)).

Now, if we suppose that the density g is only differentiable up to a finite degree but that its existing derivatives alternate in sign, then g is said to be k-monotone if and only if $(-1)^j g^{(j)}$ is nonnegative, nonincreasing and convex for $j = 0, \cdots, k-2$ if $k \geq 2$ and simply nonnegative and nonincreasing if $k = 1$ (see, e.g., Williamson (1956), Gneiting (1999)). One can see that the class of completely monotone densities is the intersection of all the classes of k-monotone densities, $k \geq 1$ (see e.g. Gneiting (1999)) and a completely monotone density can be viewed then as an "∞-monotone" density.

To prepare the ground for establishing the exact rate of convergence of the MLE for scale mixtures of Exponentials or equivalently for completely monotone densities, it seems natural to work on establishing an asymptotic distribution theory for the MLE for k-monotone densities.

When $k = 1$, the problem specializes to estimating a nonincreasing density g_0 and was first solved by Prakasa Rao (1969) and revisited by Groeneboom (1985). Groeneboom (1985) used a geometric interpretation of the MLE (the Grenander estimator) to reprove that

$$n^{1/3}\left(\hat{g}_n(x_0) - g_0(x_0)\right) \to_d \left(\frac{1}{2} g_0(x_0)|g_0'(x_0)|\right)^{1/3} C'(0),$$

where $x_0 > 0$ is a fixed point such that $g_0'(x_0) < 0$ and g_0' is continuous in a neighborhood of x_0, $\hat{g}_n$ is the Grenander estimator, and C is the greatest convex minorant of two-sided Brownian motion starting at 0 plus t^2, $t \in \mathbb{R}$. For $k = 2$, Groeneboom, Jongbloed, and Wellner (2001b) considered both the MLE and LSE and established that if the true convex density g satisfies $g_0''(x_0) > 0$ and g_0'' is continuous in a neighborhood of x_0, then

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

where $\bar{g}_n$ is the either the MLE or LSE, H is a random cubic spline function such that H'' is convex, H stays above the integrated two-sided Brownian motion plus $t^4, t \in \mathbb{R}$ and touches it exactly at those points where H'' changes its slope (see Groeneboom, Jongbloed, and Wellner (2001a)).

Under the working assumption that the true k-monotone density g_0 is k-times differentiable at x_0 such that $(-1)^k g_0^{(k)}(x_0) > 0$ and $g_0^{(k)}$ is continuous in a neighborhood of x_0, asymptotic mimimax lower bounds for the rates of convergence of estimating $g_0^{(j)}(x_0)$ are derived in Chapter 2 and found to be $n^{-(k-j)/(2k+1)}$ for $j = 0, \cdots, k-1$. This result implies that no estimator of $g_0^{(j)}(x_0)$ can converge at a rate faster than $n^{-(k-j)/(2k+1)}$.

The major result of this research is to prove that the above rates are achievable by both the MLE and LSE and that the joint asymptotic distribution of their j-th derivatives at x_0, $\bar{g}_n^{(j)}(x_0), j = 0, \cdots, k-1$ is given by

$$\begin{pmatrix} n^{\frac{k}{2k+1}}\left(\bar{g}_n(x_0) - g_0(x_0)\right) \\ n^{\frac{k-1}{2k+1}}\left(\bar{g}_n^{(1)}(x_0) - g_0^{(1)}(x_0)\right) \\ \vdots \\ n^{\frac{1}{2k+1}}\left(\bar{g}_n^{(k-1)}(x_0) - g_0^{(k-1)}(x_0)\right) \end{pmatrix} \to_d \begin{pmatrix} c_0(g_0) H_k^{(k)}(0) \\ c_1(g_0) H_k^{(k+1)}(0) \\ \vdots \\ c_{k-1}(g_0) H_k^{(2k-1)}(0) \end{pmatrix} \tag{1.1}$$

where H_k is a process characterized by:

(i) $(-1)^k (H_k(t) - Y_k(t)) \geq 0, \quad t \in \mathbb{R}$.

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

(iii) For any $t \in \mathbb{R}$, $H_k(t) = Y_k(t)$ if and only if $H_k^{(2k-2)}$ changes slope at t; equivalently,

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

Y_k is the $(k-1)$-fold integral of two-sided Brownian motion $+(k!/(2k)!)t^{2k}$, $t \in \mathbb{R}$; i.e.,

$$Y_k(t) = \begin{cases} \int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_2} W(t_1)dt_2 \cdots dt_{k-1} + (k!/(2k)!)t^{2k}, & t \geq 0 \\ \int_t^0 \int_{t_{k-1}}^0 \cdots \int_{t_2}^0 -W(t_1)dt_2 \cdots dt_{k-1} + (k!/(2k)!)t^{2k}, & t < 0, \end{cases}$$

and finally the constants $c_j(g_0), j = 0, \cdots, k-1$ are given by

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

The existence of the process H_k is the other major outcome of this work and is established in Chapter 3. By applying a change of scale, the greatest convex minorant of two-sided Brownian motion $+t^2, t \in \mathbb{R}$ and the "invelope" H can be viewed as the two first elements of the sequence $(H_k)_{k \geq 1}$.

In general, the process H_k is a random spline of degree $2k-1$ that stays above Y_k when k is even and below it when k is odd. Furthermore, this spline is of a very particular shape since its $(2k-2)$-th derivative has to be convex. At the points of strict increase of the process $H_k^{(2k-1)}$ (note that the existence of this derivative follows from the convexity assumption), the processes H_k and Y_k have to touch each other. To be more accurate, it is still conjectured that $H_k^{(2k-1)}$ is a jump process. Although the numerical results strongly supports this conjecture, the possibility that $H_k^{(2k-1)}$ is a Cantor type function has not been yet excluded even for the particular case $k = 2$ (Groeneboom, Jongbloed and Wellner (2001A)). The proof of existence and almost surely uniqueness of the process H_k is inspired from the work of Groeneboom, Jongbloed, and Wellner (2001a). In our setting, the process H_k is connected with the Gaussian problem

$$dX_k(t) = t^k dt + dW(t), \qquad t \in \mathbb{R}$$

which can be viewed as an estimation problem with t^k being the "true" function . To "estimate" t^k, we define for a fixed $c > 0$ a Least Squares problem over the class of k-convex functions g on $[-c, c]$; i.e., $g^{(k-2)}$ exists and convex. The process H_k can be then obtained by taking the limit (in an appropriate sense) of the k-fold integral of the solution of the LS problem as $c \to \infty$.

We find that there is a nice parallelism between the problems of estimating the true k-monotone density g_0 and the k-convex function t^k via the Least Squares method. The two problems have many aspects in common and this is one important feature that makes the Least Squares method very appealing. On the computational side, this parallelism helps in reducing the problems of calculating the LSE and approximating the process H_k on finite intervals to one basic algorithm. Described in Chapter 4 in more details, the iterative $(2k-1)$-th spline algorithm is based on iterative addition and deletion of the knot points of the k-fold integral of the LSE and those of the process H_k, which are both splines of degree $2k-1$. As for the MLE, although the same principle applies, a different version of the algorithm is needed to suit the nonlinear form of its characterization.

Chapter 2

ASYMPTOTICS OF THE MAXIMUM LIKELIHOOD AND LEAST SQUARES ESTIMATORS

2.1 Introduction

Let $X_1, \cdots, X_n$ be n independent observations from a common k-monotone density g_0. We consider two estimators corresponding to different estimation procedures: the Maximum Likelihood (ML) and Least Squares (LS) estimators. Both estimators were considered by Groeneboom, Jongbloed, and Wellner (2001b) in the special case of estimating a monotone and convey density. We first establish a mixture representation for k-monotone functions which proves to be very useful in showing existence of both estimators. This result is to some extent similar to Bernstein's theorem for completely monotone functions (see, e.g., Widder (1941), Feller (1971)). Whereas existence of the MLE follows easily from the work of Lindsay (1983a), Lindsay (1983b), and Lindsay (1995)) on nonparametric Maximum Likelihood estimators in a very general mixture model setting, establishing existence of the LSE is a much more difficult task. Beside a compactness argument, the proof of existence in the particular case $k = 2$ uses the fact that the LSE is a piecewise linear function (see Groeneboom, Jongbloed, and Wellner (2001b)) but a different reasoning is needed when $k > 2$. In the general case, the MLE and LSE belong to a special subclass of k-monotone functions: they are k-monotone splines of degree $k-1$. For the MLE, this particular form follows immediately from Theorem 22 of Lindsay (1995). As for the LSE, the proof relies, in the special case $k = 2$, on the simple fact that given any decreasing and convex function g and a finite number of fixed points on its graph, there exists a piecewise decreasing and convex function $\tilde{g}$, passing through the points and staying below g. For more details on this proof, see Groeneboom, Jongbloed, and Wellner (2001b). For $k > 2$, such a property is hard to generalize for any number of points (see Balabdaoui (2004)) and hence there is a need for a different argument to show that the LSE is a spline.

Characterizations of the MLE and LSE are established in Section 2. These characterizations appear to be natural extensions of those obtained in the case $k = 2$ by Groeneboom, Jongbloed, and Wellner (2001b). Beside that they give necessary and sufficient conditions for a k-monotone function to be the solution of the corresponding optimization problem, they are very useful in proving strong consistency of the estimators and their derivatives. In Section 3, we show that for $j = 0, \cdots, k-1$, the j-th derivative of either the MLE or LSE is strongly consistent and that this consistency is uniform on intervals of the form $[c, \infty)$, $c > 0$ for $0 \leq j \leq k-2$.

In a step towards an asymptotic distribution theory, asymptotic minimax lower bounds for the rate of convergence of estimating $g_0^{(j)}(x_0)$, $j = 0, \cdots, k-1$ are derived in Section 4. Here, we are interested in local estimation at a fixed point $x_0 > 0$. We assume that the true density g_0 is k-times differentiable at x_0, the derivative $g_0^{(k)}$ is continuous in a small neighborhood of x_0 and $(-1)^k g_0^{(k)}(x_0) > 0$. Under this working assumptions, the asymptotic lower bound for estimating $g_0^{(j)}(x_0)$ is found to be $n^{-(k-j)/(2k+1)}$, $j = 0, \cdots, k-1$. This result extends the lower bounds obtained in estimation of a decreasing density and that of a decreasing and convex density and its first derivative at a fixed point (see Groeneboom, Jongbloed, and Wellner (2001b)). The result implies that no estimator of $g_0^{(j)}(x_0)$ can converge (in the sense of minimax risk) at rate faster than $n^{-(k-j)/(2k+1)}$. Although these asymptotic bounds cannot be a substitute for the exact rates of convergence, they give a good idea about what one should expect these rates to be.

Under the same working hypotheses, we prove in Section 6 that $n^{-(k-j)/(2k+1)}$ is achieved by the j-th derivative of the MLE and LSE, $j = 0, \cdots, k-1$. The assumption that $(-1)^k g_0^{(k)}(x_0) > 0$ along with consistency of the $(k-1)$-th derivative "force" the number of knot points of the estimators, that are in a small neighborhood of x_0, to diverge to infinity almost surely as $n \to \infty$. This fact is very important for proving the rate achievement. More precisely, the major argument that goes into the proof is the fact that the distance between two successive knots (or jump points of the $(k-1)$-th derivative of the estimators) in a small neighborhood of x_0 is $O_p(n^{-1/(2k+1)})$. The entire Section 5 is devoted to this problem that we refer to as the "gap problem".

In the last section, we derive the joint asymptotic distribution of the derivatives of the MLE and LSE. The limiting distributions depend on a stochastic process H_k whose existence and characterization are established in Chapter 3. In addition, these distributions involve constants that depend on $g_0(x_0)$ and $g_0^{(k)}(x_0)$. An asymptotic distribution is also derived for the associated mixing distribution using an explicit inversion formula established in Section 2.

2.2 *The Maximum Likelihood and Least Squares estimators of a k-monotone density*

2.2.1 Mixture representation of a k-monotone density

Williamson (1956) gave a very useful characterization of a k-monotone function on $(0, \infty)$ by establishing the following theorem:

Theorem 2.2.1 *(Williamson, 1956) A function g is k-monotone on $(0, \infty)$ if and only if there exists a nondecreasing function γ bounded at 0 such that*

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

where $y_+ = y 1_{(0,\infty)}(y)$.

The next theorem gives an inversion formula for the measure γ:

Theorem 2.2.2 *(Williamson, 1956) If g is of the form (2.1) with $\gamma(0) = 0$, then at a continuity point $t > 0$, γ is given by*

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

Proof of Theorems 2.2.1 and 2.2.2: See Williamson (1956). ∎

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

Lemma 2.2.1 *A function g is an integrable k-monotone function if and only if it is of the form*

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

where F is nondecreasing and bounded on $(0, \infty)$.

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

Lemma 2.2.2 *If F in (2.2) satisfies $\lim_{t\to\infty} F(t) = \int_0^\infty g(x)dx$, then at a continuity point $t > 0$, F is given by*

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

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

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

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

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

Now, using integration by parts, we can write

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

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

The characterization in (2.2) is more relevant for us since we are dealing with k-monotone densities. It is easy to see that if g is a density, and F is chosen to be right-continuous and to satisfy the condition of Lemma 2.2.2, then F is a distribution function. For $k = 1$ ($k = 2$), note that the characterization matches with the well known fact that a density is nondecreasing (nondecreasing and convex) on $(0, \infty)$ if and only if it is a mixture of uniform densities (triangular densities). More generally, the characterization establishes a one-to-one correspondance between the class of k-monotone densities and the class of scale mixture of Beta's with parameters 1 and k. From the inversion formula in (2.3), one can see that a natural estimator for the mixing distribution F is obtained by plugging in an estimator for the density g and it becomes obvious that the rate of estimating F is controlled by that of estimating the highest derivative $g^{(k-1)}$. When k increases the densities become much smoother and therefore, the inverse problem of estimating the mixing distribution F becomes harder.

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

2.2.2 The Maximum Likelihood estimator of a k-monotone density

Let

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

be the "adjusted" log-likelihood function defined on $\mathcal{M}_k \cap L_1(\lambda)$, where λ is Lebesgue measure on $\mathbb{R}$. Using the integral representation established in the previous subsection, ψ_n can also be rewritten as

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

where F is bounded and nondecreasing.

Lemma 2.2.3 *The functional ψ_n admits a maximizer $\hat{g}_n$ in the class $\mathcal{D}_k$. Moreover, the density $\hat{g}_n$ is of the form*

$$\hat{g}(x) = w_1 \frac{k(a_1 - x)_+^{k-1}}{a_1^k} + \cdots + w_m \frac{k(a_m - x)_+^{k-1}}{a_m^k},$$

where $w_1, \cdots, w_m$ and $a_1, \cdots, a_m$ are respectively the weights and the support points of the maximizing mixing distribution $\hat{F}_n$.

Proof. First, we prove that there exists a density $\hat{g}_n$ that maximizes the "usual" log-likelihood $l_n = \int_0^\infty \log g(x) d\mathbb{G}_n(x)$ over the class $\mathcal{D}_k$. For g in $\mathcal{D}_k$, let F be the distribution function such that

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

The unicomponent likelihood curve Γ as defined by Lindsay (1995)) is then

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

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

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

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

since $\log(c) \leq c - 1$. Thus ψ_n is maximized over $\mathcal{M}_k \cap L_1(\lambda)$ by $\hat{g}_n \in \mathcal{D}_k$. ■

The following lemma gives a necessary and sufficient condition for a point t to be in the support of the maximizing distribution function $\hat{F}_n$.

Lemma 2.2.4 *Let $X_1, \cdots, X_n$ be i.i.d. random variables from the true density g_0, and let $\hat{F}_n$ and $\hat{g}_n$ be the MLE of the mixing and mixed distribution respectively. Then, for all $t > 0$,*

$$\frac{1}{n}\sum_{j=1}^{n}\frac{k(t-X_j)_+^{k-1}/t^k}{\hat{g}_n(X_j)} \leq 1, \tag{2.4}$$

with equality if and only if $t \in supp(\hat{F}_n) = \{a_1, \cdots, a_m\}$.

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

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

it follows that for all $t > 0$

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

This yields

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

or

$$\frac{1}{n}\sum_{j=1}^{n}\frac{k(t-X_j)_+^{k-1}/t^k}{\hat{g}_n(X_j)} \leq 1. \tag{2.5}$$

Now, let M_n be the set defined by

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

We will prove now that $M_n = \text{supp}(\hat{F}_n)$. We write $P_{\hat{F}_n}$ for the probability measure associated with $\hat{F}_n$. Integrating the left hand side of (2.5) with respect to $\hat{F}_n$, we have

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

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

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

and so

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

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

Remark 2.2.1 *The above characterization can be also given in the following form: The k-monotone density $\hat{g}_n$ is the MLE if and only if*

$$\int_0^\infty \frac{(t-x)_+^{k-1}}{\hat{g}_n(x)} d\mathbb{G}_n(x) \begin{cases} \leq \frac{t^k}{k}, & \textit{for all } t \geq 0 \\ = \frac{t^k}{k}, & \textit{if and only if } t \textit{ is a support point of } \hat{F}_n. \end{cases}$$

This form generalizes the characterization of the MLE of a nonincreasing and convex density ($k = 2$) obtained by Groeneboom, Jongbloed, and Wellner (2001b).

Remark 2.2.2 *The main reason for using the "adjusted" log-likelihood is to obtain a "nice" characterization for the MLE since the maximization is performed over the cone of all integrable k-monotone functions (not necessarily densities).*

For $k = 2$, Groeneboom, Jongbloed, and Wellner (2001b) proved that there exists at most one change of slope of the MLE between two successive observations and used this fact to show that the estimator is unique. For $k > 2$, proving uniqueness seems to be harder. However, we were able to do it for the special case $k = 3$. In the following, we give a proof of this result.

Lemma 2.2.5 *Let $k = 3$. The MLE $\hat{g}_n$ of a 3-monotone density is unique.*

Proof. We start by establishing the fact that the MLE has at most one knot between two successive observations. For that, we take $k > 2$ to be arbitrary and define the function $\hat{H}_n$ by

$$\hat{H}_n(t) = \frac{1}{n} \sum_{j=1}^{n} \frac{k(t-X_j)_+^{k-1}}{t^k \hat{g}_n(X_j)}, \quad t > 0.$$

By strict concavity of the log-likelihood, the vector $(\hat{g}_n(X_{(1)}), \cdots, \hat{g}_n(X_{(n)}))$ is unique. As the support points $a_1, \cdots, a_m$ are the solutions of the equation $\hat{H}_n(t) = 1$, it follows that they are uniquely determined. On the other hand, from the characterization of the MLE in (2.4), $\hat{H}_n(t) \leq 1$ if and only if $t \in \{a_1, \cdots, a_m\}, m \leq n$ the set of knots or equivalently the set of jump points of $\hat{g}_n^{(k-1)}$. This implies that the derivative

$$\hat{H}'_n(t) = \frac{1}{n} \sum_{j=1}^{n} \frac{k(t - X_j)_+^{k-2}(-t + kX_j)}{t^{k+1}\hat{g}_n(X_j)}, \quad t > 0$$

is equal to 0 at a_r for $r = 1, \cdots, m$. The derivative $\hat{H}'_n$ can be rewritten as

$$\hat{H}'_n \quad = \quad \frac{1}{n} \sum_{j=1}^{n} \frac{k(t - X_{(j)})_+^{k-2}(-t + kX_{(j)})}{t^{k+1}\hat{g}_n(X_{(j)})} = \frac{1}{n}\frac{1}{t^{k+2}} Q_n(t)$$

where

$$Q_n(t) = \sum_{j=1}^{n} \lambda_j (t - X_{(j)})_+^{k-2}(-t + kX_{(j)})$$

with

$$\lambda_j = \frac{k}{\hat{g}_n(X_{(j)})}.$$

Note that the first support point a_1 has to be strictly larger than $X_{(1)}$. Indeed, $a_1 \leq X_{(1)}$ implies that $\hat{H}_n(a_1) = 0$ and this is impossible since $\hat{H}_n(a_1) = 1$.

Now let $k = 3$. In the following, we are going to show that $a_r > X_{(r)}$ for all $r \in \{1, \cdots, m\}$. The assertion is true for $r = 1$. If $m = 1$, there is nothing else to be proved. Now we assume that $m > 1$ and that the claim is true for all $1 < r \leq m - 1$. Suppose that it is not true for $r + 1$. This implies that

$$X_{(r)} < a_r < a_{r+1} \leq X_{(r+1)}.$$

Since $\hat{H}_n$ takes the value 1 at both points a_r and a_{r+1}, it follows by the mean value theorem that the derivative $\hat{H}'_n$ has another zero between a_r and a_{r+1}. Therefore, Q_n has three different zeros in $[X_{(r)}, X_{(r+1)})$. But note that on this interval, Q_n is given by

$$Q_n(t) = \sum_{j=1}^{r} \lambda_j (t - X_{(j)})(-t + kX_{(j)})$$

and therefore, Q_n is a polynomial of degree 2. The latter implies that $Q_n \equiv 0$ on $[X_{(r)}, X_{(r+1)})$, which is impossible. We conclude that

$$a_r \geq X_{(r)} \tag{2.6}$$

for all $r \in \{1, \cdots, m\}$.

Now, let $p_1, \cdots, p_m$ be the masses corresponding to the support points $a_1, \cdots, a_m$. For $j = 1, \cdots, n$, we have

$$\hat{g}_n(X_{(j)}) = \sum_{r=1}^{m} p_r \frac{k(a_r - X_{(j)})_+^2}{a_r^3}. \tag{2.7}$$

Suppose that $\{q_1, \cdots, q_m\}$ is another set of masses that satisfy the same system in (2.7). If we denote $\beta_r = p_r - q_r$, then we have for all $j \in \{1, \cdots, n\}$

$$\sum_{r=1}^{m} \beta_r (a_r - X_{(j)})_+^2 = 0. \tag{2.8}$$

To prove that $\beta_r = 0$ for $r = 1, \cdots, m$, we need to prove first that $a_m > X_{(n)}$ (this is true for all $k > 2$). We have

$$\begin{aligned} 1 &= \int_0^\infty \hat{g}_n(x)dx \\ &= p_1 \frac{a_1^k}{a_1^k} + \cdots + p_m \frac{a_m^k}{a_m^k} \\ &= \frac{p_1}{a_1^k} \int_0^{a_1} \frac{k(a_1 - x)^{k-1}}{\hat{g}_n(x)} d\mathbb{G}_n(x) + \cdots + \frac{p_m}{a_m^k} \int_0^{a_m} \frac{k(a_m - x)^{k-1}}{\hat{g}_n(x)} d\mathbb{G}_n(x) \end{aligned}$$

where in the last equality, we used Lemma 2.2.4. But using the chain rule, we can rewrite the right side of this equality as

$$\begin{aligned} &\frac{p_1}{a_1^k} \int_0^{a_1} \frac{k(a_1 - x)^{k-1}}{\hat{g}_n(x)} d\mathbb{G}_n(x) + \cdots + \frac{p_m}{a_m^k} \int_0^{a_m} \frac{k(a_m - x)^{k-1}}{\hat{g}_n(x)} d\mathbb{G}_n(x) \\ &= \int_0^{a_1} \left(p_1 \frac{k(a_1 - x)^{k-1}}{a_1^k} + \cdots + p_m \frac{k(a_m - x)^{k-1}}{a_m^k} \right) \frac{1}{\hat{g}_n(x)} d\mathbb{G}_n(x) \\ &\quad + \int_{a_1}^{a_2} \left(p_2 \frac{k(a_2 - x)^{k-1}}{a_2^k} + \cdots + p_m \frac{k(a_m - x)^{k-1}}{a_m^k} \right) \frac{1}{\hat{g}_n(x)} d\mathbb{G}_n(x) \\ &\quad \vdots \\ &\quad + \int_{a_{m-1}}^{a_m} p_m \frac{k(a_m - x)^{k-1}}{a_m^k} \frac{1}{\hat{g}_n(x)} d\mathbb{G}_n(x) \end{aligned}$$

$$
\begin{aligned}
&= \int_0^{a_1} \frac{\hat{g}_n(x)}{\hat{g}_n(x)} d\mathbb{G}_n(x) + \int_{a_1}^{a_2} \frac{\hat{g}_n(x)}{\hat{g}_n(x)} d\mathbb{G}_n(x) + \cdots + \int_{a_{m-1}}^{a_m} \frac{\hat{g}_n(x)}{\hat{g}_n(x)} d\mathbb{G}_n(x) \\
&= \mathbb{G}_n(a_m).
\end{aligned}
$$

It follows that $\mathbb{G}(a_m) = 1$ and hence $a_m \geq X_{(n)}$. But $a_m \neq X_{(n)}$ because otherwise $\hat{g}_n(X_{(n)}) = 0$ and $l_n = -\infty$. Therefore, $a_m > X_{(n)}$. However, a_m is the only support point that is bigger than $X_{(n)}$. In fact, if there exists another support point $a_j, j < m$ such that $X_{(n)} \leq a_j < a_m$, then the nontrivial polynomial Q_n of degree 2 would have three different zeros in $[X_{(n)}, \infty)$ (here, we assume that $m \geq 2$). By plugging $j = n$ in (2.8), we obtain that $\beta_m = 0$ and therefore

$$
\beta_1 (a_1 - X_{(j)})_+^2 + \cdots + \beta_{m-1} (a_{m-1} - X_{(j)})_+^2 = 0 \tag{2.9}
$$

for all $1 \leq j \leq n-1$. Now, let $j_0 = \max\{1 \leq j \leq n-1 : X_{(j)} \leq a_{m-1} \leq X_{(j+1)}\}$. By the same reasoning as before, a_{m-1} is the only support point in $[X_{(j_0)}, X_{(j_0+1)})$. By plugging $j = j_0$ in (2.9), we obtain that $\beta_{m-1} = 0$. Using induction,we show that $\beta_r = 0$ for $1 \leq r \leq m-2$ and uniqueness of the masses follows. ∎

2.2.3 The Least Squares estimator of a k-monotone density

The least squares criterion is

$$
Q_n(g) = \frac{1}{2} \int_0^\infty g^2(x) dx - \int g(x) d\mathbb{G}_n(x) \, . \tag{2.10}
$$

We want to minimize this over $g \in \mathcal{D}_k \cap L_2(\lambda)$, the subset of square integrable $k-$monotone functions. Instead we will actually solve the somewhat easier optimization problem of minimizing $Q_n(g)$ over $\mathcal{M}_k \cap L_2(\lambda)$ and show that even though the resulting estimator does not necessarily have total mass one it consistently estimates $g_0 \in \mathcal{D}_k$. Using arguments similar to those in the proof of Theorem 1 in Williamson (1956), one can show that $g \in \mathcal{M}_k$ if and only if

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

for a positive measure μ on $(0, \infty)$. Thus we can rewrite the criterion in terms of the corresponding measures μ: note that

$$
\int_0^\infty g^2(x) dx \quad = \quad \int_0^\infty \int_0^\infty (t-x)_+^{k-1} d\mu(t) \int_0^\infty (t'-x)_+^{k-1} d\mu(t') dx
$$

$$= \int_0^\infty \int_0^\infty r_k(t,t')d\mu(t)d\mu(t')$$

where

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

and

$$\begin{aligned}\int_0^\infty g(x)d\mathbb{G}_n(x) &= \int_0^\infty \int_0^\infty (t-x)_+^{k-1}d\mu(t)d\mathbb{G}_n(x)\\ &= \int_0^\infty \frac{1}{n}\sum_{i=1}^n (t-X_i)_+^{k-1}d\mu(t) \equiv \int_0^\infty s_{n,k}(t)d\mu(t)\,.\end{aligned}$$

Hence it follows that, with $g = g_\mu$

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

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

Proposition 2.2.1 *The problem of minimizing $\Phi(\mu)$ over all non-negative measures μ has a unique solution $\tilde{\mu}$.*

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

We identity X of Zeidler's theorem with the space $\mathcal{X}$ of nonnegative measures on $[0,\infty)$, and we show that we can take M of Zeidler's theorem to be

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

for some constant $D < \infty$.

First, we can, without loss, restrict the minimization to the space of non-negative measures on $[X_{(1)},\infty)$ where $X_{(1)} > 0$ is the first order statistic of the data. To see this, note that we can decompose any measure μ as $\mu = \mu_1 + \mu_2$ where μ_1 is concentrated on $[0, X_{(1)})$

and μ_2 is concentrated on $[X_{(1)}, \infty)$. Since the second term of Φ is zero for μ_1, the contribution of the μ_1 component to $\Phi(\mu)$ is always non-negative, so we make $\inf \Phi(\mu)$ no larger by restricting to measures on $[X_{(1)}, \infty)$.

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

For $s, t \ge t_0 > 0$ we have

$$r_k(s,t) \ge \frac{(1-e^{-v_0})t_0}{2k} s^{k-1} t^{k-1} \tag{2.11}$$

where $v_0 \approx 1.59$. To prove (2.11) we will use the inequality

$$(1 - v/k)^{k-1} \ge e^{-v}, \qquad 0 \le v \le v_0, \quad k \ge 2\,. \tag{2.12}$$

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

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

since

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

But we also have

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

for $s, t \geq t_0$, so we conclude that (2.11) holds.

From the inequality (2.11) we conclude that for measures μ concentrated on $[X_{(1)}, \infty)$ we have

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

On the other hand,

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

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

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

This lower bound is strictly positive if

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

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

$$m_{k-1} \leq 1/A\,. \tag{2.13}$$

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

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

if

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

and hence we can restrict to measures μ with

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

for every $M \geq 1$. But this implies that μ satisfies

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

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

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

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

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

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

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

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

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

In the following, we give a characterization of the least squares estimator.

Proposition 2.2.2 *Define $\mathbb{Y}_n$ and $\tilde{H}_n$ respectively by*

$$\mathbb{Y}_n(x) = \int_0^x \int_0^{t_{k-1}} \cdots \int_0^{t_2} \mathbb{G}_n(t_1)dt_1dt_2\cdots dt_{k-1}, \qquad x \geq 0,$$

and

$$\tilde{H}_n(x) = \int_0^x \int_0^{t_k} \cdots \int_0^{t_2} \tilde{g}_n(t_1)dt_1dt_2\cdots dt_k, \qquad x \geq 0.$$

Then $\tilde{g}_n$ is the LS estimator over $\mathcal{M}_k \cap L_2(\lambda)$ if and only if the following conditions are satisfied for $\tilde{g}_n$ and $\tilde{H}_n$:

$$\begin{cases} \tilde{H}_n(x) \geq \mathbb{Y}_n(x), \qquad \textit{for} \;\; x \geq 0, \\ \textit{and} \\ \int_0^\infty \left(\tilde{H}_n(x) - \mathbb{Y}_n(x)\right) d\tilde{g}_n^{(k-1)}(x). \end{cases} \tag{2.14}$$

Remark 2.2.3 *Note that $\mathbb{Y}_n$ and $\tilde{H}_n$ can be written in the more compact form*

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

and

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

Proof. Let $\tilde{g}_n \in \mathcal{M}_k \cap L_2(\lambda)$ satisfy (2.14), and let g be an arbitrary function in $\mathcal{M}_k \cap L_2(\lambda)$. Then

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

Now, using integration by parts

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

and

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

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

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

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

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

But condition (2.14) implies that

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

Therefore,

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

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

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

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

We have:

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

Using integration by parts, we obtain

$$0 \le \lim_{\epsilon \to 0} \frac{Q_n(\tilde{g}_n + \epsilon g_x) - Q_n(\tilde{g}_n)}{\epsilon} = \tilde{H}_n(x) - \mathbb{Y}_n(x) \, .$$

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

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

which holds if and only if the equality in (2.14) holds. ∎

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

Lemma 2.2.6 *Let $[a,b] \subseteq (0,\infty)$ and let g be a nonnegative and nonincreasing function on $[a,b]$. For any polynomial P_{k-1} of degree $\le k-1$ on $[a,b]$, if the function*

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

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

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

Remark 2.2.4 *In the previous lemma, the assumption that Δ has infinitely many zeros can be weakened. Indeed, we obtain the same conclusion if we assume that Δ has $k+1$ distinct zeros in $[a,b]$.*

Now, we will use the characterization of the LSE $\tilde{g}_n$ together with the previous lemma to show that it is a finite mixture of $Beta(1,k)$'s. We know from Proposition 2.14 that $\tilde{g}_n$ is the LSE if and only if

$$\tilde{H}_n(t) \geq \mathbb{Y}_n(t), \quad \text{for } t > 0, \tag{2.15}$$

and

$$\int_0^\infty \left(\tilde{H}_n(t) - \mathbb{Y}_n(t)\right) d\tilde{g}_n^{(k-1)}(t) = 0 \tag{2.16}$$

where

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

and

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

The condition in (2.16) implies that $\tilde{H}_n$ and $\mathbb{Y}_n$ have to be equal at any point of increase of the monotone function $(-1)^{k-1}\tilde{g}_n^{(k-1)}$. Therefore, the set of points of increase of $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ is included in the set of zeros of the function $\tilde{\Delta}_n = \tilde{H}_n - \mathbb{Y}_n$. Now, note that $\mathbb{Y}_n$ can be given by the explicit expression:

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

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

Proposition 2.2.3 *There exists* $m \in \mathbb{N}\backslash\{0\}$, $\tilde{a}_1, \cdots, \tilde{a}_m$ *and* $\tilde{w}_1, \cdots, \tilde{w}_m$ *such that for all* $x > 0$, *the LSE* $\tilde{g}_n$ *is given by*

$$\tilde{g}_n(x) = \tilde{w}_1 \frac{k(\tilde{a}_1 - x)_+^{k-1}}{\tilde{a}_1^k} + \cdots + \tilde{w}_m \frac{k(\tilde{a}_m - x)_+^{k-1}}{\tilde{a}_m^k}. \tag{2.17}$$

Proof. We need to consider two cases:

(i) The number of zeros of $\tilde{\Delta}_n = \tilde{H}_n - \mathbb{Y}_n$ is finite. This implies by (2.16) that the number of points of increase of $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ is also finite. Therefore, $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ is discrete with finitely many jumps and hence $\tilde{g}_n$ is of the form given in (2.17).

(ii) Now, suppose that $\tilde{\Delta}_n$ has infinitely many zeros. Let j be the smallest integer in $\{0, \cdots, n-1\}$ such that $[X_{(j)}, X_{(j+1)}]$ contains infinitely many zeros of $\tilde{\Delta}_n$ (with $X_{(0)} = 0$ and $X_{(n+1)} = \infty$). By Lemma 2.2.6, if t_j is the smallest zero of $\tilde{g}_n$ in $[X_{(j)}, X_{(j+1)}]$, then $\tilde{g}_n \equiv 0$ on $[t_j, X_{(j+1)}]$ and $\tilde{g}_n > 0$ on $[X_{(j)}, t_j)$ if $t_j > X_{(j)}$. Note that from the proof of Proposition 2.2.1, we know that the minimizing measure $\tilde{\mu}_n$ does not put any mass on $(0, X_{(1)}]$, and hence the integer j has to be strictly greater than 0.

Now, by definition of j, $\tilde{\Delta}_n$ has finitely many zeros to the left of $X_{(j)}$, which implies that $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ has finitely many points of increase in $(0, X_{(j)})$. We also know that $\tilde{g}_n \equiv 0$ on $[t_j, \infty)$. Thus we only need to show that the number of points of increase of $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ in $[X_{(j)}, t_j)$ is finite, when $t_j > X_{(j)}$. This can be argued as follows: Consider z_j to be the smallest zero of $\tilde{\Delta}_n$ in $[X_{(j)}, X_{(j+1)})$. If $z_j \geq t_j$, then we cannot possibly have any point of increase of $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ in $[X_{(j)}, t_j)$ because it would imply that we have a zero of $\tilde{\Delta}_n$ that is strictly smaller than z_j. If $z_j < t_j$, then for the same reason, $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ has no point of increase in $[X_{(j)}, z_j)$. Finally, $(-1)^{k-1}\tilde{g}_n^{(k-1)}$ cannot have infinitely many points of increase in $[z_j, t_j)$ because that would imply that $\tilde{\Delta}_n$ has infinitely zeros in (z_j, t_j), and hence by Lemma 2.2.6, we can find $t'_j \in (z_j, t_j)$ such that $\tilde{g}_n \equiv 0$ on $[t'_j, t_j]$. But this impossible since $\tilde{g}_n > 0$ on $[X_{(j)}, t_j)$. ∎

2.3 Consistency of the estimators

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

$c > 0$.

2.3.1 The Maximum Likelihood estimator

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

Lemma 2.3.1 *If g is a k-monotone density function then*

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

for all $x > 0$.

Proof. We have

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

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

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

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

Proposition 2.3.1 *Let g_0 be a k-monotone density on $(0, \infty)$ and fix $c > 0$. Then*

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

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

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

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

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

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

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

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

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

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

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

and this is equivalent to

$$\int_0^\infty \frac{g_n(x)}{\hat{g}_n(x)} d\mathbb{G}_n(x) \leq 1. \tag{2.1}$$

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

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

where

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

The previous convergence is uniform on intervals of the form $[d, \infty)$, $d > 0$. This follows since $\hat{g}_l$ and $\hat{g}$ are monotone and $\hat{g}$ is continuous.

Much of the following is along the lines of JONGBLOED (1995), pages 117-119, and GROENEBOOM, JONGBLOED, AND WELLNER (2001B), pages 1674-1675. We are going to show that $\hat{g}$ and the true density g_0 have to be the same. For $0 < \alpha < 1$ define $\eta_\alpha = G_0^{-1}(1-\alpha)$. Fix ϵ so small that $\epsilon < \eta_\epsilon$. By (2.1) there is a number $D_\epsilon > 0$ such that $\hat{g}_l(1/\epsilon) \geq D_\epsilon$ for sufficiently large l. To see this, note that (2.1) implies that

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

and hence

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

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

Hence, by continuity of g_l and the bound in Lemma 3.4

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

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

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

In fact,

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

while

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

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

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

uniformly on $[\epsilon, \eta_\epsilon]$. For sufficiently large l, we have using (2.1)

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

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

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

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

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

which is equivalent to

$$\int_0^\infty \frac{g_0^2(x)}{\hat{g}(x)} dx \le 1. \tag{2.2}$$

Define $\tau = \int_0^\infty \hat{g}(x)dx$. Then $\hat{h} = \tau^{-1}\hat{g}$ is a k-monotone density. By (2.2), we have that

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

Now consider the function

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

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

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

It is easy to see that the integrand is minimized pointwise by taking $g(x) = g_0(x)$. Hence $\inf_{\mathcal{C}_d} K(g) \ge 1$. In particular, $K(\hat{h}) \ge 1$ which implies that $\tau = 1$. Now, if $g \ne g_0$ at a point x, it follows that $g \ne g_0$ on an interval of positive length. Hence, $g_0 \ne g \Rightarrow K(g) > 1$. We conclude that we have necessarily $\hat{h} = \hat{g} = g_0$.

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

Corollary 2.3.1 *Let $c > 0$. For $j = 1, \cdots, k-2$,*

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

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

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

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

2.3.2 The Least Squares estimator

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

Proposition 2.3.2 *Fix $c > 0$ and suppose that the true k-monotone density g_0 satisfies $\int_0^\infty x^{-1/2} dG_0(x) < \infty$. Then*

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

Proof. The main difficulty here is that we don't know whether the LSE $\tilde{g}_n$ is a genuine density; i.e. $\tilde{g}_n \in \mathcal{M}_k$ but not necessarily $\tilde{g}_n \in \mathcal{D}_k$. But if only one knew that $\tilde{g}_n$ stays bounded in some sense with high probability, the proof of consistency will be much like the one used for $k = 2$; i.e., consistency of the LSE of a convex and decreasing density (see Groeneboom, Jongbloed, and Wellner (2001b)). The proof for $k = 2$ is based on the very important fact that the LSE is a density, which helps in showing that $\tilde{g}_n$ at the last jump point $\tau_n \in [0, \delta]$ of $\tilde{g}_n'$ for a fixed $\delta > 0$ is uniformly bounded. The proof would have been similar if we only knew that

$$\int_0^\infty \tilde{g}_n(x) dx = O_p(1)\,.$$

Here we will first show that $\int_0^\infty \tilde{g}_n^2 d\lambda = O(1)$ almost surely. From the last display in the proof of Proposition 2.2.2

$$\int_0^\infty \tilde{g}_n^2(x)dx = \int_0^\infty \tilde{g}_n(x)d\mathbb{G}_n(x)$$

and hence

$$\sqrt{\int_0^\infty \tilde{g}_n^2(x)dx} = \int_0^\infty \tilde{u}_n(x)d\mathbb{G}_n(x), \tag{2.3}$$

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

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

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

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

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

since g is decreasing. Therefore

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

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

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

and hence by (2.3) the integral $\int_0^\infty \tilde{g}_n^2 d\lambda$ is bounded (almost surely) by some constant M_k.

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

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

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

$$\begin{aligned}
\tilde{g}_n(\tau_n) &\leq \tilde{g}_n(\delta/2) \leq (2/\delta)(\delta/2)\tilde{g}_n(\delta/2) \leq (2/\delta)\int_0^{\delta/2} \tilde{g}_n(x)dx \\
&\leq (2/\delta)\sqrt{\delta/2}\sqrt{\int_0^{\delta/2} \tilde{g}_n^2(x)dx} \leq \sqrt{2/\delta}\sqrt{\int_0^{\infty} \tilde{g}_n^2(x)dx} \\
&= \sqrt{2M_k/\delta}. \qquad (2.4)
\end{aligned}$$

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

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

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

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

and hence

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

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

$$\tilde{g}_n(\tau_n) \leq 2k\sqrt{M_k/\delta} = C_k. \qquad (2.5)$$

Now, since $\tilde{g}_n(\delta) \le \tilde{g}_n(\tau_n)$, the sequence $\tilde{g}_n(x)$ is uniformly bounded almost surely for all $x \ge \delta$. Using a Cantor diagonalization argument, we can find a subsequence $\{n_l\}$ so that, for each $x \ge \delta$, $g_{n_l}(x) \to \tilde{g}(x)$, as $l \to \infty$. By Fatou's lemma, we have

$$\int_\delta^\infty (\tilde{g}(x) - g_0(x))^2 dx \le \liminf_{l\to\infty} \int_\delta^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2 dx. \tag{2.6}$$

On the other hand, the function $\tilde{g}_{n_l} + \epsilon g_0$ is a square integrable k-monotone function for all $\epsilon > 0$. Therefore, from the characterization of $\tilde{g}_{n_l}$ it follows that

$$\int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x)) d(\tilde{G}_{n_l}(x) - \mathbb{G}_{n_l}(x)) \le 0\,.$$

Thus we can write

$$\begin{aligned}
&\int_\delta^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2 dx \\
&\le \int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x))^2 dx \\
&= \int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x)) d(\tilde{G}_{n_l}(x) - G_0(x)) \\
&= \int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x)) d(\tilde{G}_{n_l}(x) - \mathbb{G}_{n_l}(x)) + \int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x)) d(\mathbb{G}_{n_l}(x) - G_0(x)) \\
&\le \int_0^\infty (\tilde{g}_{n_l}(x) - g_0(x)) d(\mathbb{G}_{n_l}(x) - G_0(x)) \to_{a.s.} 0,
\end{aligned} \tag{2.7}$$

as $l \to \infty$. The last convergence is justified as follows: since $\int_0^\infty \tilde{g}_{n_l}^2 d\lambda$ is bounded almost surely, we can find a constant $C > 0$ such that $\tilde{g}_{n_l} - g_0$ admits $G(x) = C/\sqrt{x}, x > 0$, as an envelope. Since $G \in L_1(G_0)$ by hypothesis and since the class of functions $\{(g - g_0)1_{[G \le M]} : g \in \mathcal{M}_k \cap L_2(\lambda)\}$ is a Glivenko-Cantelli class for every $M > 0$ (each element is a difference of two bounded monotone functions) (2.7) holds. From (2.6), we conclude that

$$\int_\delta^\infty (\tilde{g}(x) - g_0(x))^2 dx \le 0\,,$$

and therefore, $\tilde{g} \equiv g_0$ on $(0, \infty)$ since $\delta > 0$ can be chosen arbitrarily small. We have proved that there exists Ω_0 with $P(\Omega_0) = 1$ and such that for each $\omega \in \Omega_0$ and any given subsequence $\tilde{g}_{n_k}(\cdot, \omega)$, we can extract a further subsequence $\tilde{g}_{n_l}(\cdot, \omega)$ that converges to g_0 on $(0, \infty)$. It follows that $\tilde{g}_n$ converges to g_0 on $(0, \infty)$, and this convergence is uniform on intervals of the form $[c, \infty)$, $c > 0$ by the monotonicity and continuity of g_0. ∎

Corollary 2.3.2 *Let $c > 0$. Under the assumption of Proposition 2.3.2, we have for $j = 1, \cdots, k-2$,*

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

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

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

Proof. See the proof of Corollary 2.3.1. ■

2.4 Asymptotic minimax lower bounds

In this section we derive asymptotic minimax lower bounds for the behavior of *any estimator* of a $k-$monotone density g and its first $k-1$ derivatives at a point x_0 for which the $k-$th derivative exists and is non-zero. The proof will rely upon the basic Lemma 4.1 of GROENEBOOM (1996); see also JONGBLOED (2000). This basic method seems to go back to DONOHO AND LIU (1987) and DONOHO AND LIU (1991)). As before, let $\mathcal{D}_k$ denote the class of $k-$monotone densities on $[0,\infty)$. Here is the notation we will need. Consider estimation of the $j-$th derivative of $g \in \mathcal{D}_k$ at x_0 for $j \in \{0, 1, \ldots, k-1\}$. If $\hat{T}_n$ is an arbitrary estimator of the real-valued functional T of g, then the (L_1-)minimax risk based on a sample $X_1, \ldots, X_n$ of size n from g which is known to be in a suitable subset $\mathcal{D}_{k,n}$ of $\mathcal{D}_k$ is defined by

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

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

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

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

Lemma 2.4.1 *Assume that there exists some subset* $\{g_\epsilon : \epsilon > 0\}$ *of densities in* $\mathcal{D}_{k,n}$ *such that, as* $\epsilon \downarrow 0$,

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

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

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

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

Here is the main result of this section:

Proposition 2.4.1 *Let* $g_0 \in \mathcal{D}_k$ *and* x_0 *be a fixed point in* $(0, \infty)$ *such that* g_0 *is* k *times differentiable at* x_0 $(k \ge 2)$. *An asymptotic lower bound for the local minimax risk of any estimator* $\hat{T}_{n,j}$ *for estimating the functional* $T_j g_0 = g_0^{(j)}(x_0)$, *is given by:*

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

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

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

where

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

and

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

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

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

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

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

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

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

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

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

Then, we can write $\tilde{g}_\mu$ as

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

Then for $0 \le j \le k$

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

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

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

But for μ small enough, the sign of $(-1)^j g_\mu^{(j)}$ will be that of $(-1)^j g_0^{(j)}(x_0)$. For $j = k$,

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

Assume that $r^{(k)}(0) \neq 0$. Set

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

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

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

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

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

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

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

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

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

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

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

If m is even, we obtain

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

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

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

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

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

and hence

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

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

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

When j is odd, the first term vanishes and

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

We denote

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

and $r^{(k)}(0) = C_k$, which specializes to

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

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

$$C_k = \begin{cases} 2 \times (-1)^{k/2}(k+1)(k-1)!\binom{k}{k/2-1}, & \text{if } k \text{ is even} \\ (-1)^{(k-1)/2}k!\binom{k+1}{(k-1)/2}, & \text{if } k \text{ is odd.} \end{cases} \tag{2.1}$$

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

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

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

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

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

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

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

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

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

for any integers n and p, using the convention

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

when $p=0$. We have,

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

since

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

and hence

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

$$
\begin{aligned}
&= 2^{4k+6}\frac{(2(k+1))!(2k+3)!}{(4k+6)!}\frac{\binom{2k+3}{2k+3}}{\binom{4k+7}{4k+6}} + \frac{2^{4k+5}\left((2(k+1))!\right)^2}{(4k+5)!} \\
&= 2^{4k+5}\frac{\left((2(k+1))!\right)^2}{(4k+6)!}\left(\frac{2(2k+3)}{4k+7} + (4k+6)\right) \\
&= 2^{4k+5}\frac{\left((2(k+1))!\right)^2}{(4k+7)!}\left((4k+6) + (4k+6)(4k+7)\right) \\
&= 2^{4k+5}\frac{\left((2(k+1))!\right)^2}{(4k+7)!}(4k+6)(4k+8) \\
&= 2^{4(k+2)}(2k+3)(k+2)\frac{\left((2(k+1))!\right)^2}{(4k+7)!}. \qquad (2.2)
\end{aligned}
$$

Combining and (2.1) and (2.2), we find that $\lambda_{k,2}$ is given by

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

and

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

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

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

so that $\mu = (\epsilon/b_k)^{1/(2k+1)}(1+o(1))$, then for $0 \le j \le k-1$, the modulus of continuity, m_j, of the functional T_j satisfies

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

The result is that

$$m_j(\epsilon) \ge (r_{k,j}\epsilon)^{\frac{k-j}{2k+1}}(1+o(1)),$$

where

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

and hence

$$\sup_{\tau>0} \lim_{n\to\infty} \inf n^{\frac{k-j}{2k+1}} MMR_1(n, T_j, \mathcal{D}_{k,n,\tau}) \geq \frac{1}{4}\left(4\frac{k-j}{2k+1}e^{-1}\right)^{\frac{k-j}{2k+1}} (r_{k,j})^{\frac{k-j}{2k+1}}, \tag{2.3}$$

which can be rewritten as

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

for $j = 0, \cdots, k-1$. ■

Remark 2.4.1 *It might seem that a more natural choice for a perturbation would have been*

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

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

2.5 The gap problem

2.5.1 Introduction

Recall that it was assumed that g_0 is k-times continuously differentiable at x_0 and that $(-1)^k g_0^{(k)}(x_0) > 0$. This hypothesis together with strong consistency of the $(k-1)$-st derivative of the MLE and LSE imply that the number of jump points of this derivative, in a small neighborhood of x_0, has to diverge 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: we would like to determine an upper bound for this rate of convergence.

Using the characterizations of the MLE and LSE and the "mid-point property" (that we will describe later), Groeneboom, Jongbloed, and Wellner (2001b) could prove that for $k = 2$, this gap is of the order $n^{-1/5}$. For $k = 1$, the same property can be used to see that the gap in this case is of the order $n^{-1/3}$. As a function of k, it is natural to think that the order of the gap takes the general form $n^{-1/(2k+1)}$. In the problem of nonparametric regression via splines, Mammen and van de Geer conjectured the same form for the knot points of the regression spline but did not suggest any method to prove the conjecture (see Mammen and van de Geer (1997), page 400).

In the following subsection, we describe the difficulty of establishing this result for $k > 2$. In the general case, the problem exhibits a high level of complexity and the situation becomes fundamentally different from the one encountered in the case $k = 2$. In fact, the arguments used in this special case cannot be applied in our general case but rather, one should think of a general way of arguing the result and in which the proof for $k = 2$ would only be recognized as a very special case.

2.5.2 Fundamental differences

Let τ_n^- and τ_n^+ be the last and first jump points of the $(k-1)$-sh derivative of either the MLE or LSE, 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 $k = 2$. Here, we restrict ourselves only to the LSE since it is a simpler case to deal with than the MLE. Recall that for $k = 2$ the characterization of the LSE, $\tilde{g}_n$, is given by

$$\tilde{H}_n(x) \begin{cases} \geq \mathbb{Y}_n(x), & x \geq 0 \\ = \mathbb{Y}_n(x), & \text{if and only if } x \text{ is a jump point of } \tilde{g}_n' \end{cases} \tag{2.1}$$

where

$$\tilde{H}_n(x) = \int_0^x (x-t)\tilde{g}_n(t)dt, \quad \text{and} \quad \mathbb{Y}_n(x) = \int_0^x (x-t)d\mathbb{G}_n(t),$$

and $\mathbb{G}_n$ is the empirical distribution function. For ease of notation, we omit writing the subscript n on the jump points, but their dependence on n should be kept in mind. On the interval $[\tau^-, \tau^+)$, the function $\tilde{g}_n'$ is constant since they are no more jump points in this interval. This implies that $\tilde{H}_n$ is polynomial of degree 3 on $[\tau^-, \tau^+)$. But, from the characterization in (2.1), it follows that

$$\tilde{H}_n(\tau^-) = \mathbb{Y}_n(\tau^-), \qquad \tilde{H}_n'(\tau^-) = \mathbb{Y}_n'(\tau^-)$$

and

$$\tilde{H}_n(\tau^+) = \mathbb{Y}_n(\tau^+), \qquad \tilde{H}_n'(\tau^+) = \mathbb{Y}_n'(\tau^+).$$

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

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

Groeneboom, Jongbloed and Wellner refer to this as the "mid-point property". By applying the first condition (the inequality condition) in (2.1), it follows that

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

The inequality in the last display can be rewritten as

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

where G_0 and Y_0 are the true counterparts of $\mathbb{G}_n$ and $\mathbb{Y}_n$ respectively, and $\mathbb{E}_n$ a random error. Using techniques from empirical processes, Groeneboom, Jongbloed, and Wellner (2001b) could prove that

$$|\mathbb{E}_n| = O_p(n^{-4/5}) + o_p((\tau^+ - \tau^-)^4). \tag{2.2}$$

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

$$\begin{aligned} &\frac{Y_0(\tau^-) + Y_0(\tau^+)}{2} - \frac{(G_0(\tau^+) - G_0(\tau^-))\,(\tau^+ - \tau^-)}{8} \\ &= -Cg_0''(x_0)(\tau^+ - \tau^-)^4 + \ o_p((\tau^+ - \tau^-)^4). \end{aligned} \tag{2.3}$$

Combining the results in (2.2) and (2.3), it follows that

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

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 $[\tau^-, \tau^+]$ and therefore it can be evaluated at any point between τ^- and τ^+. Second of all, it can 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 $\bar{\tau}$, a property that helps establishing the formula derived in (2.3).

When $k > 2$, we have established in Proposition 2.2.2 that $\tilde{g}_n$ is the LSE if and only if

$$\tilde{H}_n(x) \begin{cases} \geq \mathbb{Y}_n(x), & x \geq 0 \\ = \mathbb{Y}_n(x), & \text{if and only if } x \text{ is a jump point of } \tilde{g}_n^{(k-1)} \end{cases}$$

where

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

and

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

If τ is an arbitrary jump point of $\tilde{g}_n^{(k-1)}$, then the equalities

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

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

$$\tilde{H}_n^{(j)}(\tau^-) = \mathbb{Y}_n^{(j)}(\tau^-), \qquad \tilde{H}_n^{(j)}(\tau^+) = \mathbb{Y}_n^{(j)}(\tau^+) \tag{2.4}$$

for $j = 0, \cdots, k-1$. For example, in the case of $k = 3$, the polynomial $\tilde{H}_n$ of degree 5 would be identically equal to the polynomial $\tilde{P}_n$ given by

$$\tilde{P}_n(t) = \frac{\alpha_0}{5!}(\tau^+ - t)^5 + \frac{\alpha_1}{4!}(\tau^+ - t)^4(t - \tau^-) + \cdots + \frac{\alpha_{2k-1}}{5!}(t - \tau^-)^5$$

for $t \in [\tau^-, \tau^+]$, where

$$\begin{aligned}
\alpha_0 &= 5!\frac{\mathbb{Y}_n(\tau^-)}{(\tau^+ - \tau^-)^5} \\
\alpha_1 &= 5!\frac{\mathbb{Y}_n(\tau^-)}{(\tau^+ - \tau^-)^5} + 4!\frac{\mathbb{Y}_n'(\tau^-)}{(\tau^+ - \tau^-)^4} \\
\alpha_2 &= 5!\frac{\mathbb{Y}_n(\tau^-)}{(\tau^+ - \tau^-)^5} + 2\cdot 4!\frac{\mathbb{Y}_n'(\tau^-)}{(\tau^+ - \tau^-)^4} + 3!\frac{\mathbb{Y}_n''(\tau^-)}{(\tau^+ - \tau^-)^3}
\end{aligned}$$

and

$$\begin{aligned}
\alpha_3 &= 5!\frac{\mathbb{Y}_n(\tau^+)}{(\tau^+ - \tau^-)^5} \\
\alpha_4 &= 5!\frac{\mathbb{Y}_n(\tau^+)}{(\tau^+ - \tau^-)^5} - 4!\frac{\mathbb{Y}_n'(\tau^+)}{(\tau^+ - \tau^-)^4} \\
\alpha_5 &= 5!\frac{\mathbb{Y}_n(\tau^+)}{(\tau^+ - \tau^-)^5} - 2\cdot 4!\frac{\mathbb{Y}_n'(\tau^+)}{(\tau^+ - \tau^-)^4} + 3!\frac{\mathbb{Y}_n''(\tau^+)}{(\tau^+ - \tau^-)^3}.
\end{aligned}$$

For $n = 6$ and $n = 10$, we simulated n i.i.d. random variables from a standard Exponential and in each case, the LSE was calculated using the iterative $(2k-1)$-th spline algorithm (see Chapter 4). The plots in Figures 2.1, 2.2 show clearly that $\tilde{H}_n$ and $\tilde{P}_n$ are two different polynomials. A similar conclusion is reached with $n = 50$ and $k = 4$ (see Figure 2.3).

Two jump points are clearly not sufficient to determine the polynomial $\tilde{H}_n$. However, if we consider $p > 2$ jump points $\tau_0 < \cdots < \tau_{p-1}$ (all located e.g. after x_0), $\tilde{H}_n$ is a spline of degree $2k-1$ that is $(2k-2)$-times differentiable at its knot points $\tau_0, \cdots, \tau_{p-1}$. In the next subsection, we prove that if $p = 2k-2$, the spline $\tilde{H}_n$ is completely determined on $[\tau_0, \tau_{2k-3}]$ by the conditions

$$\tilde{H}_n(\tau_i) = \mathbb{Y}(\tau_i), \text{ and } \tilde{H}_n'(\tau_i) = \mathbb{Y}'(\tau_i) \tag{2.5}$$

for $i = 0, \cdots, 2k-3$. 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.

2.5.3 A Hermite interpolation problem

In the next lemma, we prove that given $\tau_0 < \cdots < \tau_{2k-3}$, $2k-2$ jump points of $\tilde{g}_n^{(k-1)}$, $\tilde{H}_n$ is the unique solution of the Hermite problem given by (2.5). But before that, we need the following lemma which gives a definition of B-splines.

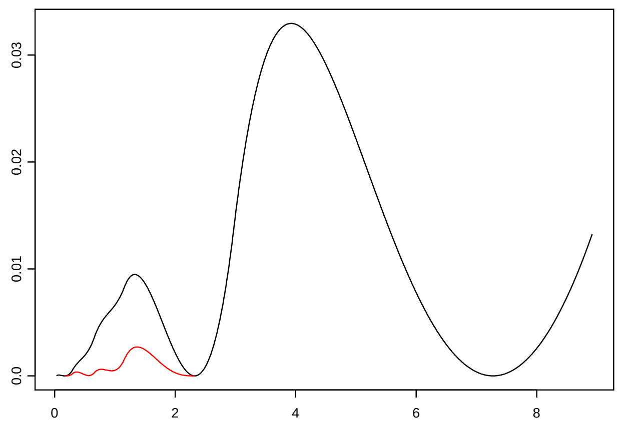

Figure 2.1: Plots of $\tilde{H}_n - \mathbb{Y}_n$ in black and $\tilde{P}_n - \mathbb{Y}_n$ on $[\tau^-, \tau^+]$ in red, where $k = 3$, $n = 6$, $\tau^- = 0.169$ and $\tau^+ = 2.319$.

Lemma 2.5.1 *Let $m \geq 1$ be an integer and $x_1 < \cdots < x_{m+1}$ be arbitrary $(m+1)$ points in $\mathbb{R}$. There exists a unique vector $(a_1, \cdots, a_{m+1}) \in \mathbb{R}^{m+1}$ such that the spline*

$$B(t) = \sum_{i=1}^{m+1} a_i (t - x_i)_+^{m-1}, \qquad t \in \mathbb{R}$$

satisfies

$$B(t) = 0, \qquad \text{if } t \leq x_1 \quad \text{or} \quad t \geq x_{m+1} \tag{2.6}$$

$$B_k(t) > 0, \qquad \text{if } t \in (x_1, x_{m+1}) \tag{2.7}$$

$$\int_{x_1}^{x_{m+1}} B(t) dt = 1. \tag{2.8}$$

B is called the B-spline of degree $m - 1$ with support $[x_1, x_{m+1}]$. Furthermore,

$$B(t) = [x_1, \cdots, x_{m+1}](-1)^m m(t - \cdot)_+^{m-1}, \qquad t \in \mathbb{R}; \tag{2.9}$$

thus $B(t)$ is the divided difference of order m of the function $x \mapsto (-1)^m m(t-x)_+^{m-1}$, $x \in \mathbb{R}$ with respect to the knots $x_1, \ldots, x_{m+1}$.

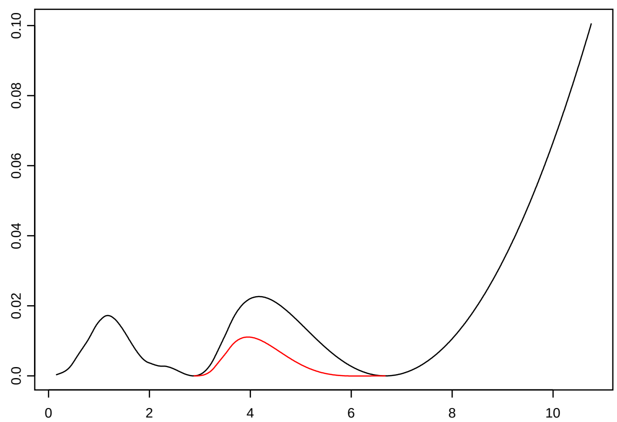

Figure 2.2: Plots of $\tilde{H}_n - \mathbb{Y}_n$ in black and $\tilde{P}_n - \mathbb{Y}_n$ on $[\tau^-, \tau^+]$ in red, where $k = 3$, $n = 10$, $\tau^- = 2.880$ and $\tau^+ = 6.680$.

Proof. See e.g. NÜRNBERGER (1989), Theorems 2.2 and 2.9, pages 96 and 99. ■

Remark 2.5.1 *Note that for any a and b in $\mathbb{R}$, we have*

$$(b-a)^{m-1} = (b-a)_+^{m-1} + (-1)^{m-1}(a-b)_+^{m-1}.$$

On the other hand, we can write

$$\begin{aligned}\sum_{i=1}^{m+1} a_i (t-x_i)^{m-1} &= \sum_{i=1}^{m+1} a_i \sum_{l=0}^{m-1} \binom{m-1}{l} x_i^l \, t^{m-1-l} \\ &= \sum_{l=0}^{m-1} \binom{m-1}{l} \left(\sum_{i=1}^{m+1} a_i x_i^l \right) t^{m-1-l} = 0, \quad \textit{for } t \in \mathbb{R},\end{aligned}$$

where the last equality follows from the identities in (2.4) of Theorem 2.2 in NÜRNBERGER (1989)*. Therefore, B can also be given by*

$$B(t) = (-1)^m \sum_{i=1}^{m+1} a_i (x_i - t)_+^{m-1} \qquad t \in \mathbb{R},$$

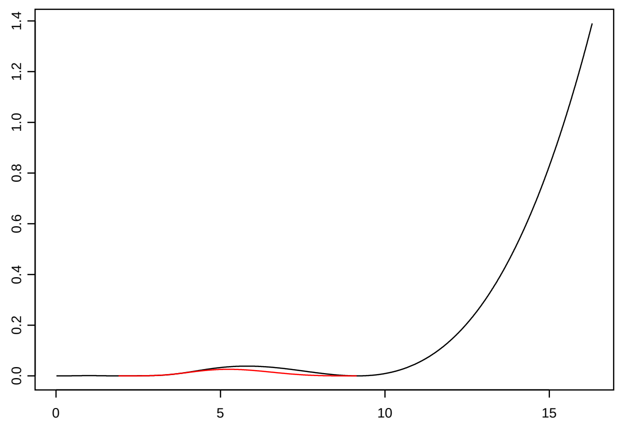

Figure 2.3: Plots of $\tilde{H}_n - \mathbb{Y}_n$ in black and $\tilde{P}_n - \mathbb{Y}_n$ on $[\tau^-, \tau^+]$ in red, where $k = 4$, $n = 50$, $\tau^- = 1.901$ and $\tau^+ = 9.141$.

or equivalently

$$B(t) = [x_1, \cdots, x_{m+1}]m(\cdot - t)_+^{m-1}. \tag{2.10}$$

The latter form will be used in the rest of this chapter.

Lemma 2.5.2 *Let $k \geq 2$. Given any $2k-2$ successive jump points of $\tilde{H}_n^{(2k-1)}$, $\tau_0 < \cdots < \tau_{2k-3}$, the $(2k-1)$-th spline $\tilde{H}_n$ is uniquely determined on $[\tau_0, \tau_{2k-3}]$ by the values of the empirical process $\mathbb{Y}_n$ and of its derivative $\mathbb{Y}_n'$ at $\tau_0, \cdots, \tau_{2k-3}$. Furthermore, for any arbitrary points $\tau_{-(2k-1)} < \cdots < \tau_{-1}$ to the left of τ_0 and $\tau_{2k-2} < \cdots < \tau_{4k-4}$ to the right of τ_{2k-3}, there exist coefficients $\alpha_{-(2k-1)}, \cdots, \alpha_{2k-4}$ depending on $\mathbb{Y}_n(\tau_i)$ and $\mathbb{Y}_n'(\tau_i)$, $i = 0, \cdots, 2k-3$, such that the spline $\tilde{H}_n$ can be written as*

$$\tilde{H}_n(t) = \sum_{i=-(2k-1)}^{2k-4} \alpha_i B_i(t), \tag{2.11}$$

for all $t \in [\tau_0, \tau_{2k-3}]$ where, for $i = -(2k-1), \cdots, 2k-4$, B_i is the B-spline of degree $2k-1$ corresponding to the set of knots $\{\tau_i, \cdots, \tau_{i+2k}\}$.

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 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).

Now, let $p = 2k - 2$ and consider successive $2k - 2$ jump points $\tau_0 < \cdots < \tau_{2k-3}$. We denote $\tau_0 = x_0 = a$, $\tau_{2k-3} = x_{2k-3} = b$ and $\tau_1 = x_1, \cdots, \tau_{2k-4} = x_{2k-4}$. Also, for $i = 1, \cdots, 4k - 4$, consider the points t_i such that $t_1 = t_2 = x_0$, $t_3 = t_4 = x_1, \ldots,$ $t_{4k-5} = t_{4k-4} = x_{2k-3}$. Using this notation, we see that the $(2k-1)-th$ spline $\tilde{H}_n$ satisfies

$$\tilde{H}_n(t_i) = \mathbb{Y}_n(t_i) \quad \text{and} \quad \tilde{H}_n'(t_i) = \mathbb{Y}_n'(t_i) \tag{2.12}$$

for all $i = 1, \cdots, 4k - 4$. Furthermore, we can check that for all $i = 1, \cdots, 2k - 4$, we have

$$t_i < x_i < t_{i+2k}.$$

Indeed, for a given $i = 1, \cdots, 2k - 4$, we know that $x_i = t_{2i+1} = t_{2i+2}$ and it is easy to see that

$$t_i < t_{2i+1} = t_{2i+2} < t_{i+2k}.$$

Therefore, by Theorem 3.7 in Nürnberger (1989), page 109, the *Hermite interpolation problem* defined in (2.12) has a unique solution in $S_{2k-1}(x_1, \cdots, x_{2k-4})$, the space of splines of degree $2k-1$ that are $(2k-2)$-times continuously differentiable at the knots $x_1, \cdots, x_{2k-4}$ (or, see DeVore and Lorentz (1993), Theorem 9.2, page 162). Notice that in Nürnberger's notation (see Nürnberger (1989)), the parameters $p - 2$ and $2k - 1$ play the role of k and m respectively. Also, note that the integer $p = 2k-2$ was chosen here so that the number of equations $(2p)$ and the dimension of the space $S_{2k-1}(x_1, \cdots, x_p)$ $(\dim(S_{2k-1}(x_1, \cdots, x_p)) = p - 2 + 2k)$ are equal. It follows that we can find $\alpha_{-(2k-1)}, \cdots, \alpha_{2k-4}$ such that

$$\tilde{H}_n(t) = \sum_{i=-(2k-1)}^{2k-4} \alpha_i B_i(t)$$

for all $t \in [a,b] \equiv [\tau_0, \tau_{2k-3}]$, where $\underline{\alpha}^t = (\alpha_{-(2k-1)}, \cdots, \alpha_{2k-4})^t$ is the unique solution of the linear system

$$M\alpha \equiv \begin{pmatrix} B_{-(2k-1)}(\tau_0) & \cdots & B_{2k-4}(\tau_0) \\ (B_{-(2k-1)})'(\tau_0) & \cdots & (B_{2k-4})'(\tau_0) \\ \vdots & \vdots & \vdots \\ B_{-(2k-1)}(\tau_{2k-3}) & \cdots & B_{2k-4}(\tau_{2k-3}) \\ (B_{-(2k-1)})'(\tau_{2k-3}) & \cdots & (B_{2k-4})'(\tau_{2k-3}) \end{pmatrix} \underline{\alpha} = \begin{pmatrix} \mathbb{Y}_n(\tau_0) \\ \mathbb{Y}'_n(\tau_0) \\ \vdots \\ \mathbb{Y}_n(\tau_{2k-3}) \\ \mathbb{Y}'_n(\tau_{2k-3}) \end{pmatrix} \tag{2.13}$$

and $B_i, i = -(2k-1), \cdots, 2k-4$, are $(4k-4)$ linearly independent B-splines of degree $2k-1$ and knots $\tau_i, \cdots < \tau_{i+2k}$. ■

In the following lemma, we prove a preparatory result that will be used later for deriving the stochastic order of the distance between the jump points.

Lemma 2.5.3 *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 $y^{2k}/(2k)!$ at the points $\tau_0, \cdots, \tau_{2k-3}$, then*

$$-g_0^{(k)}(\bar{\tau})e_k(\bar{\tau}) \leq \mathbb{E}_n + \mathbb{R}_n$$

where $\mathbb{E}_n$ defined in (2.15) is a random error and $\mathbb{R}_n$ defined in (2.17) is a remainder that both depend on the knots $\tau_0, \cdots, \tau_{2k-3}$ and the point $\bar{\tau}$.

Proof. In this proof, we use the explicit B-splines representation of $\tilde{H}_n$ that was introduced in the previous lemma. Let $A = (a_{ij})_{ij}$ and $B = (b_{ij})_{ij}$ be the $(4k-4) \times (k-1)$ sub-matrices obtained by extracting the odd and even columns of the inverse of the matrix M given in (2.13). We can write,

$$\tilde{H}_n(t) = \sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} (a_{ij}\mathbb{Y}_n(\tau_j) + b_{ij}\mathbb{Y}'_n(\tau_j)) \right) B_i(t)$$

for all $t \in [\tau_0, \tau_{2k-3}]$. Fix $t = \bar{\tau} \in \cup_{i=0}^{2k-4}(\tau_i, \tau_{i+1})$. From the inequality condition in the characterization of the LSE , it follows that

$$\sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} (a_{ij}\mathbb{Y}_n(\tau_j) + b_{ij}\mathbb{Y}'_n(\tau_j)) \right) B_i(\bar{\tau}) \geq \mathbb{Y}_n(\bar{\tau})$$

or equivalently

$$\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3}(a_{ij}Y_0(\tau_j)+b_{ij}Y_0'(\tau_j))\right)B_i(\bar\tau)-Y_0(\bar\tau)\geq -\mathbb{E}_n \tag{2.14}$$

where Y_0 is the k-fold integral of the true density g_0 and $\mathbb{E}_n$ is given by

$$\mathbb{E}_n=\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3}(a_{ij}(\mathbb{Y}_n-Y_0)(\tau_j)+b_{ij}(\mathbb{Y}_n'-Y_0')(\tau_j))\right)B_i(\bar\tau)+Y_0(\bar\tau)-\mathbb{Y}_n(\bar\tau). \tag{2.15}$$

Based on the working assumptions, the function Y_0 is $(2k)$-times continuously differentiable in a small neighborhood of x_0. Using Taylor expansion of $Y_0(\tau_j)$ and $Y_0'(\tau_j)$ around $\bar\tau$ up to the orders $2k$ and $2k-1$ respectively, the inequality in (2.14) can be rewritten as

$$\begin{aligned}
&\left(\sum_{i=-(2k-1)}^{2k-4}\left\{\sum_{j=0}^{2k-3}a_{ij}\right\}B_i(\bar\tau)-1\right)Y_0(\bar\tau)\\
&+\left(\sum_{i=-(2k-1)}^{2k-4}\left\{\sum_{j=0}^{2k-3}a_{ij}(\tau_j-\bar\tau)+b_{ij}\right\}B_i(\bar\tau)\right)Y_0'(\bar\tau)\\
&\vdots\\
&+\left(\sum_{i=-(2k-1)}^{2k-4}\left\{\sum_{j=0}^{2k-3}a_{ij}\frac{(\tau_j-\bar\tau)^{2k}}{(2k)!}+b_{ij}\frac{(\tau_j-\bar\tau)^{2k-1}}{(2k-1)!}\right\}B_i(\bar\tau)\right)Y_0^{(2k)}(\bar\tau)\\
&+\mathbb{R}_n\\
&\geq -\mathbb{E}_n
\end{aligned} \tag{2.16}$$

where $\mathbb{R}_n$ is the remainder of the Taylor expansion and can be given in the integral form

$$\begin{aligned}
\mathbb{R}_n \;=\; &\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3}a_{ij}\int_{\bar\tau}^{\tau_j}\frac{(\tau_j-t)^{2k-1}}{(2k)!}(g_0^{(k)}(t)-g_0^{(k)}(x_0))dt\right. \\
&\left.+\,b_{ij}\int_{\bar\tau}^{\tau_j}\frac{(\tau_j-t)^{2k-2}}{(2k-2)!}(g_0^{(k)}(t)-g_0^{(k)}(x_0))dt\right)B_i(\bar\tau).
\end{aligned} \tag{2.17}$$

The remainder $\mathbb{R}_n$ can be viewed as the error of Hermite interpolation at the point $\bar\tau$ where

$$x\mapsto\int_{\bar\tau}^{x}\frac{(x-t)^{2k-1}}{(2k-1)!}(g_0^{(k)}(t)-g_0^{(k)}(x_0))dt$$

is the function being interpolated. The order of $\mathbb{R}_n$ will be determined in a coming subsection. Now, note that

$$\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3}a_{ij}\right)B_i(\bar\tau)-1 \;=\; 0 \tag{2.18}$$

$$\begin{aligned}
\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3} a_{ij}(\tau_j-\bar\tau)+b_{ij}\right)B_i(\bar\tau) &= 0\\
&\vdots\\
\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3} a_{ij}\frac{(\tau_j-\bar\tau)^{2k-1}}{(2k-1)!}+b_{ij}\frac{(\tau_j-\bar\tau)^{2k-2}}{(2k-2)!}\right)B_i(\bar\tau) &= 0.
\end{aligned}$$

Indeed, since the space of splines of degree $2k-1$ and with simple knots $\tau_0,\cdots,\tau_{2k-3}$ includes all the polynomials of degree $\le 2k-1$, the solution of the Hermite problem when the interpolated function is a polynomial of degree $\le 2k-1$ is the polynomial itself. Therefore, if we consider $P_0(t)=1, P_1(t)=t-\bar\tau,\cdots,P_{2k-1}(t)=(t-\bar\tau)^{2k-1}/(2k-1)!$, the previous terms are identically zero since they are exactly equal to $P_j(\bar\tau)=0$, $j=0,\cdots,2k-1$.

Now

$$\sum_{i=-(2k-1)}^{2k-4}\sum_{j=0}^{2k-3}\left(a_{ij}\frac{(\tau_j-\bar\tau)^{2k}}{(2k)!}+b_{ij}\frac{(\tau_j-\bar\tau)^{2k-1}}{(2k-1)!}\right)B_i(\bar\tau)$$

can be recognized as the Hermite interpolation error at the point $\bar\tau$ when $(y-\bar\tau)^{2k}/(2k)!$ is the function being interpolated at the knots $\tau_0,\cdots,\tau_{2k-3}$. But this error is equal to $e_k(\bar\tau)$. Indeed, using the binomial identity, we can write

$$\begin{aligned}
&\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3} a_{ij}\frac{(\tau_j-\bar\tau)^{2k}}{(2k)!}+b_{ij}\frac{(\tau_j-\bar\tau)^{2k-1}}{(2k-1)!}\right)B_i(\bar\tau)\\
&= \sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3} a_{ij}\frac{(\tau_j)^{2k}}{(2k)!}+b_{ij}\frac{(\tau_j)^{2k-1}}{(2k-1)!}\right)B_i(\bar\tau)\\
&\quad+\sum_{r=1}^{2k-1}\left(\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3} a_{ij}\binom{2k}{r}\frac{(\tau_j)^{2k-r}}{(2k)!}+b_{ij}\binom{2k-1}{r}\frac{(\tau_j)^{2k-1-r}}{(2k-1)!}\right)B_i(\bar\tau)\right)(-1)^r\bar\tau^r\\
&\quad+\left(\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3} a_{ij}\right)B_i(\bar\tau)\right)\frac{\bar\tau^{2k}}{(2k)!}.
\end{aligned}$$

Using the identity

$$\binom{2k-1}{r}=\frac{2k-r}{2k}\binom{2k}{r}$$

for all $r\in\{0,\cdots,2k\}$, it follows that

$$\sum_{i=-(2k-1)}^{2k-4}\left(\sum_{j=0}^{2k-3} a_{ij}\binom{2k}{r}\frac{(\tau_j)^{2k-r}}{(2k)!}+b_{ij}\binom{2k-1}{r-1}\frac{(\tau_j)^{2k-1-r}}{(2k-1)!}\right)B_i(\bar\tau)$$

$$
\begin{aligned}
&= \binom{2k}{r} \sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} a_{ij} \frac{(\tau_j)^{2k-r}}{(2k)!} + b_{ij}(2k-r)\frac{(\tau_j)^{2k-1-r}}{(2k)!} \right) B_i(\bar{\tau}) \\
&= \binom{2k}{r} \frac{\bar{\tau}^{2k-r}}{(2k)!}
\end{aligned}
$$

since for all $t \in [\tau_0, \tau_{2k-3}]$ and $1 \leq r \leq 2k-1$

$$
\sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} a_{ij}(\tau_j)^{2k-r} + b_{ij}(2k-r)(\tau_j)^{2k-1-r} \right) B_i(t) = t^{2k-r}.
$$

Therefore,

$$
\begin{aligned}
&\sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} a_{ij} \frac{(\tau_j - \bar{\tau})^{2k}}{(2k)!} + b_{ij} \frac{(\tau_j - \bar{\tau})^{2k-1}}{(2k-1)!} \right) B_i(\bar{\tau}) \\
&= \sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} a_{ij} \frac{(\tau_j)^{2k}}{(2k)!} + b_{ij} \frac{(\tau_j)^{2k-1}}{(2k-1)!} \right) B_i(\bar{\tau}) \\
&\quad + \left(\sum_{r=1}^{2k} (-1)^r \binom{2k}{r} \right) \frac{\bar{\tau}^{2k}}{(2k)!} \\
&= \sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} a_{ij} \frac{(\tau_j)^{2k}}{(2k)!} + b_{ij} \frac{(\tau_j)^{2k-1}}{(2k-1)!} \right) B_i(\bar{\tau}) - \frac{\bar{\tau}^{2k}}{(2k)!} \\
&= e_k(\bar{\tau})
\end{aligned}
$$

since $\sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} a_{ij} \right) B_i(\bar{\tau}) = 1$ and $\sum_{r=0}^{2k} (-1)^r \binom{2k}{r} = 0$. We conclude that the inequality in (2.16) can be rewritten as stated in the lemma. ■

2.5.4 The order of the gap

In this subsection, we give the solution of the gap problem. We restrict here ourselves to the LSE. For the MLE, the proof follows the same steps except that the notation is much more cumbersome. The error $e_k(t)$ defined in the previous lemma 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. As a first step, we will derive an upper bound for the random error $\mathbb{E}_n$. But before that, we need the following lemma:

Lemma 2.5.4 *Let $a = x_0 < x_1 < \cdots < x_{2k-3} = b$ be $2k-2$ arbitrary points and $1 \leq r \leq 2k-1$. Suppose that f that is a function that is r-times differentiable on $[a,b]$ except for a finite number of points. If Hf denotes the unique interpolating spline of degree $2k-1$ that solves the Hermite problem:*

$$Hf(x_j) = f(x_j), \text{ and } (Hf)'(x_j) = f'(x_j)$$

for $j = 0, \cdots, 2k-3$, then there exists a constant $C > 0$ (depending only on k) such that

$$\sup_{t \in [a,b]} |Hf(t) - f(t)| \leq C\omega(f^{(r)}; b-a)\,(b-a)^r$$

where $\omega(f^{(r)}; \cdot)$ is the modulus of continuity of $f^{(r)}$ on $[a,b]$:

$$\omega(h; \delta) = \sup\{|h(t_2) - h(t_1)| : \ t_1, t_2 \in [a,b], \ |t_2 - t_1| \leq \delta\}.$$

The above lemma still needs to be proved. In the case of quasi-interpolation, a similar result is available and was proved by de Boor and Fix (1973); see e.g. Nürnberger (1989), page 189. However, we believe that such a result should also be true for our Hermite interpolation problem. Although the literature seems to be more concerned with the approximation error of other types of interpolating splines, we believe that there is no reason that our spline fails to satisfy a similar property especially that it tries to "recover" better the original function f by interpolating its tangent at the knots as well. Also, it should be mentioned that it is known that, given an interval $[a,b]$, the minimal deviation of a function f from the space of splines $S_m(x_1, \cdots, x_p)$ satisfies

$$d_\infty(f, S_m(x_1, \cdots, x_p)) \leq K\delta^r \omega(f^{(r)}; \delta)$$

if $f^{(r)} \in C[a,b]$ for some $r \in \{0, \cdots, m\}$, where $K > 0$ is a universal constant that depends only on r and $\delta = \max_{0 \leq i \leq p} |x_{i+1} - x_i|$ with $x_0 = a$ and $x_{p+1} = b$ (see e.g. Nürnberger (1989), Theorem 4.27, page 159).

Lemma 2.5.5 *If Lemma 2.5.4 holds, then the random error $\mathbb{E}_n$ satisfies*

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

Proof. Let f be the function given by

$$f(t) = \sum_{i=-(2k-1)}^{2k-3} \left(\sum_{j=0}^{2k-4} (a_{ij} \frac{(\tau_j - t)^{k-1}}{(k-1)!} + b_{ij} \frac{(\tau_j - t)^{k-2}}{(k-2)!}) 1_{[\tau_j, \bar{\tau}]}(t) \right) B_i(\bar{\tau}),$$

where $[\tau_j, \bar{\tau}] \equiv [\bar{\tau}, \tau_j]$ if $\tau_j > \bar{\tau}$. Then, the error $\mathbb{E}_n$ can be rewritten as

$$\mathbb{E}_n = \int_0^\infty f(t) d(\mathbb{G}_n(t) - G_0(t)). \tag{2.19}$$

Indeed, we found in the previous subsection that $\mathbb{E}_n$ is given by

$$\mathbb{E}_n = \sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} (a_{ij} (\mathbb{Y}_n - Y_0)(\tau_j) + b_{ij} (\mathbb{Y}_n' - Y_0')(\tau_j)) \right) B_i(\bar{\tau}) + Y_0(\bar{\tau}) - \mathbb{Y}_n(\bar{\tau}).$$

Let us denote $\mathbb{D}_n = \mathbb{Y}_n - Y_0$. The error $\mathbb{E}_n$ can be rewritten as

$$\mathbb{E}_n = \sum_{i=-(2k-1)}^{2k-4} (\sum_{j=0}^{2k-3} (a_{ij} \mathbb{D}_n(\tau_j) + b_{ij} \mathbb{D}_n'(\tau_j)) B_i(\bar{\tau}) - \mathbb{D}_n(\bar{\tau}).$$

Now for arbitrary x and y, we can write

$$\mathbb{D}_n(y) = \mathbb{D}_n(x) + (y - x)\mathbb{D}_n'(x) + \cdots + \int_x^y \frac{(y-t)^{k-1}}{(k-1)!} d(\mathbb{G}_n(t) - G_0(t))$$

and similarly

$$\mathbb{D}_n'(y) = \mathbb{D}_n'(x) + (y - x)\mathbb{D}_n''(x) + \cdots + \int_x^y \frac{(y-t)^{k-2}}{(k-2)!} d(\mathbb{G}_n(t) - G_0(t)).$$

Taking $x = \bar{\tau}$ and $y = \tau_j$ for $j = 0, \cdots, 2k-3$ and using the identities in (2.18) up to the order $(k-2)$, it follows that

$$\begin{aligned}
\mathbb{E}_n &= \sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} \int_{\bar{\tau}}^{\tau_j} (a_{ij} \frac{(\tau_j - t)^{k-1}}{(k-1)!} + b_{ij} \frac{(\tau_j - t)^{k-2}}{(k-2)!}) d(\mathbb{G}_n(t) - G_0(t)) \right) B_i(\bar{\tau}) \\
&= \sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} \int_0^\infty (a_{ij} \frac{(\tau_j - t)^{k-1}}{(k-1)!} + b_{ij} \frac{(\tau_j - t)^{k-2}}{(k-2)!}) 1_{[\bar{\tau}, \tau_j]}(t) d(\mathbb{G}_n(t) - G_0(t)) \right) \\
&\quad B_i(\bar{\tau}) \\
&= \int_0^\infty \Bigg[\sum_{i=-(2k-1)}^{2k-4} \left(\sum_{j=0}^{2k-3} (a_{ij} \frac{(\tau_j - t)^{k-1}}{(k-1)!} + b_{ij} \frac{(\tau_j - t)^{k-2}}{(k-2)!}) 1_{[\bar{\tau}, \tau_j]}(t) \right) \\
&\quad B_i(\bar{\tau}) \Bigg] d(\mathbb{G}_n(t) - G_0(t))
\end{aligned}$$

which is the form claimed in (2.19).

Even if the function f is formally integrated on $(0,\infty)$, it is clear that we can assume that f is compactly supported on $[\tau_0, \tau_{2k-3}]$. For a fixed $t \in [\tau_0, \tau_{2k-3}]$, there are two possibilities: $t < \bar{\tau}$ or $t \geq \bar{\tau}$. Suppose without loss of generality that $t \geq \bar{\tau}$. Then, $f(t)$ which can be also given by

$$\begin{aligned} f(t) &= \sum_{i=-(2k-1)}^{2k-4} \left\{ \sum_{j=0}^{2k-3} \left(a_{ij} \frac{(\tau_j - t)^{k-1}}{(k-1)!} + b_{ij} \frac{(\tau_j - t)^{k-2}}{(k-2)!} \right) \right\} 1_{[\tau_j \geq t]} B_i(\bar{\tau}) \\ &= \sum_{i=-(2k-1)}^{2k-4} \left\{ \sum_{j=0}^{2k-3} \left(a_{ij} g_t(\tau_j) + b_{ij} g_t'(\tau_j) \right) \right\} B_i(\bar{\tau}) \end{aligned}$$

with

$$g_t(x) = \frac{(x-t)^{k-1}}{(k-1)!} 1_{[x \geq t]},$$

is nothing but the error at the point $\bar{\tau}$ of the Hermite interpolation of g_t at the points $\tau_0, \cdots, \tau_{2k-3}$. Note that g_t is a spline of degree $k-1$ that is $(k-1)$-times differentiable except at its unique knot t. By Lemma 2.5.4, there exists $C > 0$, such that

$$|f(t)| \leq C\omega(g_t^{(k-1)}, \tau_{2k-3} - \tau_0)(\tau_{2k-3} - \tau_0)^{k-1}.$$

But

$$\omega(g_t^{(k-1)}, \tau_{2k-3} - \tau_0) \leq 1.$$

Therefore, it follows that

$$\sup_{t \in [\tau_0, \tau_{2k-3}]} |f(t)| \leq C(\tau_{2k-3} - \tau_0)^{k-1}. \tag{2.20}$$

Now, since the function $f(t)$ depends on the knots $\tau_0, \cdots, \tau_{2k-3}$ and the point $\bar{\tau}$ (which is a fixed point in $\cup_{j=0}^{2k-4}(\tau_j, \tau_{j+1})$, it can be viewed as an element of the class

$$\mathcal{F}_{x,\underline{r}} = \left\{ f_{x,y_1,\cdots,y_{2k-2}} : x \leq y_1 \leq x + r_1, \cdots, y_{2k-3} \leq y_{2k-2} \leq y_{2k-3} + r_{2k-2} \right\}$$

where $x > 0$ and $\underline{r} = (r_1, \cdots, r_{2k-2}) : r_j > 0,\ j = 1, \cdots, 2k-2$ is a fixed $(2k-2)$-vector. To make the link between the members of the class $\mathcal{F}_{x,\underline{r}}$ and the function $f(t)$, the latter can be written as

$$f(t) = f_{\tau_0,\tau_1,\cdots,\bar{\tau},\cdots,\tau_{2k-3}}(t), \quad t \in [\tau_0, \tau_{2k-3}].$$

In this case, $x = \tau_0$, $y_1 = \tau_1$, $y_{2k-2} = \tau_{2k-3}$ and $\{y_1, \cdots, y_{2k-2}\} = \{\tau_1, \cdots, \tau_{2k-3}\} \cup \{\bar{\tau}\}$.

Let Q be an arbitrary measure on $(0, \infty)$. The collection $\mathcal{F}_{x,\underline{r}}$ admits a finite covering number with respect to $L_2(Q)$. In fact, any element $f_{x,y_1,\cdots,y_{2k-2}} \in \mathcal{F}_{x,\underline{r}}$ is $(k-2)$-times differentiable on $[x, y_{2k-2}]$. Therefore, for every $\epsilon > 0$, the collection $\mathcal{F}_{x,\underline{r}}$ admits a finite bracketing number that is bounded by $(K/\epsilon)^{1/(k-2)}$, for some $0 < K < \infty$. More specifically, there exists a constant $K > 0$ depending only on k and $R = r_1 + \cdots + r_{2k-2}$ (an upper bound for the length of the interval $[x, y_{2k-2}]$) such that

$$\log N_{[]}(\epsilon, \mathcal{F}_{x,\underline{r}}, L_2(Q)) \le K \left(\frac{1}{\epsilon}\right)^{\frac{1}{k-2}} \tag{2.21}$$

(see e.g. van der Vaart and Wellner (1996), Corollary 2.7.2, page 157). It follows that

$$\int_0^1 \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{F}_{x,\underline{r}}, L_2(G_0))}\, d\epsilon < \infty.$$

On the other hand, using Lemma 2.5.4, we have

$$|f_{x,y_1,\cdots,y_{2k-2}}(t)| \quad \le \quad C(y_{2k-2} - x)^{k-1} 1_{[x,y_{2k-2}]}(t)$$

(compare with the bound in 2.20) and hence the function $F_{x,R}$ given by

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

is an envelope for the class $\mathcal{F}_{x,\underline{r}}$. On the other hand, if x belongs to a small neighborhood $[x_0 - \delta, x_0 + \delta]$ for some small $\delta > 0$, then we can find some constant $M > 0$ depending only on δ, R and $g_0(x_0)$ such that $0 < \sup_{t \in [x_0-\delta, x_0+\delta+R]} g_0(t) < M$. Therefore,

$$EF_{x,R}^2(X_1) = C^2 R^{2(k-1)} \int_x^{x+R} g_0(x)dx \le C^2 M R^{2k-1}.$$

By Theorem 2.14.2 in van der Vaart and Wellner (1996), page 240, it follows that

$$E\left\{\left(\sup_{f_{x,y_1,\cdots,y_{2k-2}} \in \mathcal{F}_{x,\underline{r}}} \left|(\mathbb{G}_n - G_0)(f_{x,y_1,\cdots,y_{2k-2}})\right|\right)^2\right\} \le \frac{K'}{n} EF_{x,R}^2(X_1) = O(n^{-1}R^{2k-1}) \tag{2.22}$$

for some constant $K' > 0$ depending only on x_0, δ and R.

We denote

$$(\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{2k-2}}) = (\mathbb{G}_n - G_0)(f_{x,y_1,\cdots,y_{2k-2}})$$

where $f_{x,y_1,\cdots,y_{2k-2}}$ is an element in $\mathcal{F}_{x,\underline{r}}$ and define M_n as

$$\begin{aligned} M_n &= \inf\Big\{D > 0 : \Big|(\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{2k-3},y})\Big| \le \epsilon(y-x)^{2k} \\ &\qquad + n^{-2k/(2k+1)}D, \ \text{ for all } \ y \in [x, x+R]\Big\}. \end{aligned}$$

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

$$\begin{aligned} &P(M_n > m) \\ &\le \sum_{1\le j\le j_n} P\Big\{(j-1)n^{-1/(2k+1)} \le y - x \le jn^{-1/(2k+1)}, \\ &\qquad \Big|(\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{2k-3},y})\Big| > \epsilon(y-x)^{2k} + n^{-2k/(2k+1)}m\Big\} \\ &= \sum_{1\le j\le j_n} P\Big\{(j-1)n^{-1/(2k+1)} \le y - x \le jn^{-1/(2k+1)}, \\ &\qquad \Big|(\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{2k-3},y})\Big| > \epsilon(y-x)^{2k} + n^{-2k/(2k+1)}m\Big\} \\ &\le \sum_{1\le j\le j_n} P\Big\{(j-1)n^{-1/(2k+1)} \le y - x \le jn^{-1/(2k+1)}, \\ &\qquad n^{2k/(2k+1)}\Big|(\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{2k-3},y})\Big| > \epsilon(j-1)^{2k} + m\Big\} \\ &\le \sum_{1\le j\le j_n} n^{4k/(2k+1)} \frac{E\left\{\left(\sup_{y:0\le y-x<jn^{-1/(2k+1)}} \left|(\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{2k-3},y})\right|\right)^2\right\}}{\left(\epsilon(j-1)^{2k}+m\right)^2} \\ &= \sum_{1\le j\le j_n} n^{4k/(2k+1)} \frac{E\left\{\left(\sup_{f_{x,y_1,\cdots,y_{2k-3},y}\in\mathcal{F}_{x,jn^{-1/(2k+1)}}} \left|(\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{2k-3},y})\right|\right)^2\right\}}{\left(\epsilon(j-1)^{2k}+m\right)^2} \\ &\le C \sum_{1\le j\le j_n} n^{4k/(2k+1)} n^{-1} n^{-(2k-1)/(2k+1)} \frac{j^{2k-1}}{\left(\epsilon(j-1)^{2k}+m\right)^2} \\ &= C \sum_{1\le j\le j_n} \frac{j^{2k-1}}{\left(\epsilon(j-1)^{2k}+m\right)^2} \end{aligned}$$

$$\leq \; C\sum_{j=1}^{\infty} \frac{j^{2k-1}}{\left(\epsilon(j-1)^{2k}+m\right)^2} \; \searrow 0 \;\; \text{as } m \nearrow \infty$$

where $C > 0$ is a constant that is independent of $x \in [x_0-\delta, x_0+\delta]$. Therefore, $M_n = O_p(1)$ and hence it follows that

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

which holds for *all* $f_{x,y_1,\cdots,y_{2k-3},y} \in \mathcal{F}_{x,\underline{r}}$ and x in some small neighborhood $[x_0-\delta, x_0+\delta]$ of x_0. It follows that

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

■

To show that $\tau_{2k-3} - \tau_0 = O_p(n^{-1/(2k+1)})$, we need the following result:

Lemma 2.5.6 *The error $e_k(t)$ has no other zeros than $\tau_0, \cdots, \tau_{2k-3}$ in $[\tau_0, \tau_{2k-3}]$.*

Proof. The result follows from Proposition 1 of Michelli (1972) and de Boor (2004).

■

Recall that $e_k(t)$ is a monospline of degree $2k$ with $2k-2$ simple knots $\tau_0, \cdots, \tau_{2k-3}$. Furthermore, by construction, these knots are also double zeros; i.e. $e_k(\tau_j) = e_k'(\tau_j) = 0$ for $j = 0, \cdots, 2k-3$. Now, we state two preparatory lemmas that will help determine the sign of the error $e_k(t)$ at any point $t \in \cup_{j=0}^{2k-4}(\tau_j, \tau_{j+1})$.

Lemma 2.5.7 *Let $k \geq 2$ be an integer. The monospline M_k of degree $2k$ with simple knots $\xi_0 = -k+3/2, \xi_1 = -k+5/2, \cdots, \xi_{2k-4} = k+1/2, \xi_{2k-3} = k-3/2$ and such that $M_k(\xi_j) = M_k'(\xi_j) = 0$ for $j = 0, \cdots, 2k-3$ has a constant sign: +1 (-1) if k is odd (even).*

Proof. Let $\mathcal{B}_{2k}$ be the Bernoulli monospline of degree $2k$. The function $\mathcal{B}_{2k}(t-1/2) - \mathcal{B}_{2k}(0)$ is equal to the error of the Hermite interpolation of $t^{2k}/(2k)!$ at the equispaced knots $\xi_0, \cdots, \xi_{2k-3}$. By uniqueness, it follows that

$$M_k(t) = \mathcal{B}_{2k}(t-1/2) - \mathcal{B}_{2k}(0)$$

for all $t \in [-k+3/2, k-3/2]$. The Bernoulli monospline $\mathcal{B}_{2k}$ is the 1-periodic extension of the Bernoulli polynomial p_{2k} of degree $2k$ which takes extreme values at 0 when considered as a function on $[0,1]$. It follows that M_k is of one sign on $[-k+3/2, k-3/2]$. Furthermore, $p_{2k}(1/2) < p_{2k}(0)$ if k is even and $p_{2k}(1/2) > p_{2k}(0)$. Therefore, M_k is nonpositive if k is even and nonnegative if k is odd. ∎

Lemma 2.5.8 *If $t \in \cup_{j=0}^{2k-4}(\tau_j, \tau_{j+1})$, then*

$$(-1)^{k-1} e_k(t) > 0;$$

i.e., $e_k(t)$ is nonnegative (nonpositive) if k is odd (even).

Proof. Let $\bar{\tau}$ be a fixed point in $\cup_{j=0}^{2k-4}(\tau_j, \tau_{j+1})$. We can assume without loss of generality that $\bar{\tau} \in (\tau_0, \tau_1)$. There exists $\lambda \in (0,1)$ such that $\bar{\tau} = \lambda\tau_0 + (1-\lambda)\tau_1$. Consider now the function

$$(\tau_0, \cdots, \tau_{2k-3}) \mapsto \frac{e_k(\bar{\tau}) + |e_k(\bar{\tau})|}{2e_k(\bar{\tau})}.$$

Note that it is possible to divide by $e_k(\bar{\tau})$ since $e_k(\bar{\tau}) \neq 0$ as $\bar{\tau}$ is different from the knots. It is easy to see that the function is continuous in $\tau_0, \cdots, \tau_{2k-3}$. 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 2.5.7 that the constant is 1 (0) if k is odd (even). It follows that $(-1)^{k-1} e_k(\bar{\tau}) > 0$. ∎

We can finally state the main result of this section:

Lemma 2.5.9 *Let $k \geq 2$. If $g_0 \in \mathcal{D}_k$ satisfies $g_0^{(k)}(x_0) \neq 0$ and Lemma 2.5.4 holds, then $\tau_{2k-3} - \tau_0 = O_p(n^{-1/(2k+1)})$.*

Proof. Let $j_0 \in \{0, \cdots, 2k-4\}$ be such that $[\tau_{j_0}, \tau_{j_0+1}]$ be the largest knot interval; i.e., $\tau_{j_0+1} - \tau_{j_0} = \max_{0 \leq j \leq 2k-4}(\tau_{j+1} - \tau_j)$. Let $a = \tau_0$, $b = \tau_{2k-3}$.

By Lemma 2.5.4, there exists a constant $C > 0$ depending only on k such that

$$|\mathbb{R}_n| \leq C \sup_{t \in [\tau_0, \tau_{2k-3}]} |g_0^{(k)}(t) - g_0^{(k)}(x_0)| \, (b-a)^{2k}$$

using the fact that if f is $\in C^{2k}[a,b]$, then

$$\omega(f^{(2k-1)}, b-a) \leq \sup_{t \in [a,b]} |f^{(2k)}(t)|\,(b-a).$$

Therefore, it follows that

$$|\mathbb{R}_n| \leq C \sup_{t \in [\tau_0, \tau_{2k-3}]} |g_0^{(k)}(t) - g_0^{(k)}(x_0)| (\tau_{2k-3} - \tau_0)^{2k} = o_p((\tau_{2k-3} - \tau_0)^{2k}).$$

Using the result of Lemma 2.5.3 and since the bounds on $\mathbb{R}_n$ and $\mathbb{E}_n$ (see Lemma 2.5.5) are independent of the choice of $\bar{\tau}$ in $\cup_{j=0}^{2k-4} (\tau_j, \tau_{j+1})$, it follows that

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

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

$$t \mapsto t^{2k} - \left(\frac{\tau_{j_0+1} - \tau_{j_0}}{2} \right)^{2k} \frac{1}{2^{2k-1}} T_{2k} \left(\frac{2t - (\tau_{j_0} + \tau_{j_0+1})}{\tau_{j_0+1} - \tau_{j_0}} \right), \tag{2.23}$$

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

$$\begin{aligned} (-1)^{k-1} e_k(\bar{\tau}) &\geq \left\| \frac{T_{2k}}{2^{4k-1}(2k)!} \right\|_\infty (\tau_{j_0+1} - \tau_{j_0})^{2k} \\ &= \frac{1}{2^{4k-1}(2k)!} (\tau_{j_0+1} - \tau_{j_0})^{2k} \end{aligned} \tag{2.24}$$

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

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

It follows that

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

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

2.6 Rates of convergence of the estimators

Now, we are going to use the result of the previous section to derive the rates of convergence of $\bar{g}_n^{(j)}$, $j = 0, \cdots, k-1$ at a fixed point $x_0 > 0$.

Consider the event $J_n = J_n^{(1)} \cap J_n^{(2)}$ where $J_n^{(i)}$, $i = 1, 2$, are defined by

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

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 2.6.1 *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 the hypothesis of Proposition 2.3.2 holds. Then, if the conjectured Lemma 2.5.4 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 $\epsilon > 0$. We will consider first the LSE and we will start with $j = 0$. Fix $\epsilon > 0$. 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 2.5.9, 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 $\tau_{k+1} - \tau_1 \ge k n^{-1/(2k+1)}$. Fix $c > 0$ and consider the event

$$\inf_{t \in [\tau_1, \tau_{k+1}]} |\tilde{g}_n(t) - g_0(t)| > c n^{-k/(2k+1)}. \tag{2.25}$$

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

$$\left| \int_{\tau_1}^{\tau_{k+1}} \left(\tilde{g}_n(t) - g_0(t)\right) g(t)dt \right| \geq cn^{-k/(2k+1)} \int_{\tau_n^-}^{\tau_n^+} g(t)dt. \tag{2.26}$$

Now, let B be the B-spline of degree $k-1$ and with support $[x_1, x_{k+1}]$. Recall from (2.10) in Section 5 that B can be given by

$$B(t) = [\tau_1, \cdots, \tau_{k+1}]k \ (\cdot - t)_+^{k-1}$$

where $[x_1, \cdots, x_m]g$ denotes the divided difference of degree m with respect to the points $x_1, \cdots, x_m$. After some algebra, we find that B can be given by

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

for all $t \in [\tau_1, \tau_{k+1}]$.

Let $|\eta| > 0$ and consider the perturbation function

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

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 $0 \leq j \leq k-2$, which guarantees that the perturbed function $\tilde{g}_{\eta,n}$ belongs to $C^{k-2}(0, \infty)$. For $0 \leq j \leq k-3$, 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 piecewise linear and hence not strictly convex on $(0, \infty)$. Since p is a spline of degree $k-1$, the function $(-1)^{k-2}\tilde{g}_{\eta,n}^{(k-2)}$ is also piecewise linear and one can check that it is nonincreasing and convex for very small values of η. 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 (2.26), we obtain

$$\begin{aligned}\left|\int_{\tau_1}^{\tau_{k+1}} p(t) d(\mathbb{G}_n(t) - G_0(t))\right| &\geq cn^{-k/(2k+1)} \int_{\tau_1}^{\tau_{k+1}} p(t)dt \\ &= cn^{-k/(2k+1)} \prod_{1\leq i<j\leq k+1} (\tau_j - \tau_i) \qquad (2.27) \\ &\geq cn^{-k/(2k+1)} \left(n^{-1/(2k+1)}\right)^{k(k+1)/2} \qquad (2.28) \\ &= cn^{-(3+k)k/(2(2k+1))}\end{aligned}$$

where in (2.27), we used the fact that B-splines integrate to 1, whereas in (2.28) we used the facts that there are $k(k+1)/2$ terms in the product $\prod_{1\leq i<j\leq k+1}(\tau_j - \tau_i)$ and that $\tau_j - \tau_i \geq n^{-1/(2k+1)}, 1 \leq i < j \leq k+1$.

Let $0 < x < y_1 < \cdots < y_{k-1} < y$ be $(k+1)$ points in $(0, \infty)$ and consider the function $f_{x,y_1,\cdots,y_{k-1},y}$ defined by

$$f_{x,y_1,\cdots,y_{k-1},y_k}(t) = (-1)^k k \prod_{0\leq i<j\leq k} (y_j - y_i) \left(\frac{(y_0 - t)_+^{k-1}}{\prod_{j\neq 0}(y_j - y_0)} + \cdots + \frac{(y_{k-1} - t)_+^{k-1}}{\prod_{j\neq k-1}(y_j - y_{k-1})}\right)$$

where $y_0 = x$. Let $\underline{r} = (r_1, \cdots, r_k)$, $r_i > 0$ for $i = 1, \cdots, k$, be a fixed k-vector and consider the collection of functions

$$\mathcal{F}_{x,\underline{r}} = \left\{f_{x,y_1,\cdots,y_{k-1},y_k} \;:\; x < y_1 \leq x + r_1, \cdots, y_{k-1} < y_k \leq y_{k-1} + r_k\right\}.$$

For a fixed $x > 0$ and $\underline{r}$, the collection $\mathcal{F}_{x,\underline{r}}$ has a finite covering number with respect to $L_2(Q)$ where Q is an arbitrary probability measure. In fact, denote

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

and consider the collections of functions

$$\mathcal{F}_{x,R_j} = \left\{t \mapsto \alpha_j (y_j - t)_+^{k-1} 1_{[x,y_k]}(t), x \leq y_j \leq x + R_j, x \leq y_k \leq x + R\right\}$$

where $R_j = r_1 + \cdots + r_j$ for $j = 1, \cdots, k$ and $R = R_k$. By Lemmas 2.6.16 and 2.6.18 in van der Vaart and Wellner (1996), the collections $\mathcal{F}_{x,R_j}, j = 1, \cdots, k-1$ are VC-subgraph classes. Furthermore, the function

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

is a common envelope for these classes. To see that, notice that for $j = 0, \cdots, k$, the product $\prod_{j' \neq j}(y_{j'} - y_j)$ contains k terms and hence α_j is a product of $k(k+1)/2 - k = k(k-1)/2$ that are at most R distant from one another. It follows that

$$\alpha_j \leq kR^{k(k-1)/2}, \quad \text{for } j = 0, \cdots, k.$$

For an arbitrary probability measure Q, we have

$$\|F_{x,R}\|^2_{Q,2} = k^2 R^{k(k-1)} \int_x^{x+R} (t-x)^{2k-2} dQ(t) \leq k^2 R^{k(k+1)-2}$$

which is independent of Q. By Theorem 2.6.7 in van der Vaart and Wellner (1996), there exist a universal constant $K > 0$, two constants $D_j > 0$ and $V_j > 0$ that depend only on x, R_j and R such that the $\epsilon\|F_{x,R}\|^2_{Q,2}$-covering number of $\mathcal{F}_{x,R_j}$ with respect to $L_2(Q)$ is given by

$$N\left(\epsilon\|F_{x,R}\|^2_{Q,2}, \mathcal{F}_{x,R_j}, L_2(Q)\right) \leq KD_j \left(\frac{1}{\epsilon}\right)^{V_j}.$$

It follows that the collection $\mathcal{F}_{x,\underline{r}}$ admits a finite ϵ-covering number with respect to $L_2(Q)$. Furthermore, it is easy to see that the function $k \times F_{x,R}$ is an envelope for this collection. Therefore, there exist a universal constant $K > 0$, $D > 0$ and $V > 0$ depending only on x and $\underline{R}_j$, $j = 1, \cdots, k$ such that

$$N\left(\epsilon\|F_{x,R}\|^2_{Q,2}, \mathcal{F}_{x,\underline{r}}, L_2(Q)\right) \leq KD \left(\frac{1}{\epsilon}\right)^{V}$$

and therfore

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

On the other hand, if x 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 $[x, x+R]$ for all $x \in [x_0 - \delta, x_0 + \delta]$. It follows that

$$\begin{aligned} EF_{x,R}^2(X_1) &\leq k^2 R^{k(k-1)} \int_x^{x+R} (t-x)^{2k-2} g_0(x)dx \\ &\leq \frac{k^2 C}{2k-1} R^{k(k-1)} R^{2k-1} = \frac{k^2 C}{2k-1} R^{k(k+1)-1}. \end{aligned}$$

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

$$\begin{aligned} &E\left\{ \left(\sup_{f_{x,y_1,\cdots,y_k} \in \mathcal{F}_{x,\underline{r}}} \left| (\mathbb{G}_n - G_0)(f_{x,y_1,\cdots,y_k}) \right| \right)^2 \right\} \\ &\quad \leq \frac{K'}{n} EF_{x,R}^2(X_1) = O(n^{-1} R^{k(k+1)+1}), \end{aligned} \tag{2.29}$$

for some constant K' depending only on x_0, δ and R.

We denote

$$(\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{k-1},y}) \quad = \quad (\mathbb{G}_n - G_0)(f_{x,y_1,\cdots,y_{k-1},y})$$

where $f_{x,y_1,\cdots,y_{k-1},y} \in \mathcal{F}_{x,R}$ and define M_n as

$$\begin{aligned} M_n \quad = \quad \inf \Bigg\{ D > 0 : & \left| (\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{k-1},y}) \right| \leq \epsilon (y-x)^{(3k+1)k/2} \\ & + n^{-(3+k)k/(2(2k+1))} D, \quad \text{for all} \quad y \in [x, x+R] \Bigg\}; \end{aligned}$$

note that M_n is possibly equal to infinity if no $D > 0$ satisfies the required inequality. Let $n > N$. For $1 \leq j \leq \lfloor R n^{1/(2k+1)} \rfloor = j_n$, we have

$$\begin{aligned} &P(M_n > m) \\ &\leq \sum_{1 \leq j \leq j_n} P\Bigg\{ \exists\, y : \ (j-1) n^{-1/(2k+1)} \leq y - x \leq j n^{-1/(2k+1)}, \\ &\qquad \left| (\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{k-1},y}) \right| > \epsilon (y-x)^{(3+k)k/2} + \ n^{-(3+k)k/(2(2k+1))} \ m \Bigg\} \\ &\leq \sum_{1 \leq j \leq j_n} P\Bigg\{ \exists\, y : \ 0 \leq y - x \leq j n^{-1/(2k+1)}, \\ &\qquad n^{(3+k)k/(2(2k+1))} \left| (\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{k-1},y}) \right| > \epsilon (j-1)^{(3+k)k/2} + \ m \Bigg\} \\ &\leq \sum_{1 \leq j \leq j_n} n^{(3+k)k/(2k+1)} \frac{E\left\{ \left(\sup_{y: 0 \leq y - x < j n^{-1/(2k+1)}} \left| (\mathbb{P}_n - P_0)(f_{x,y_1,\cdots,y_{k-1},y}) \right| \right)^2 \right\}}{\left(\epsilon (j-1)^{(3+k)k/2} + m \right)^2} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{1\le j\le j_n} n^{(3+k)k/(2k+1)} \frac{E\left\{\left(\sup_{f_{x,y_1,\cdots,y_{k-1},y}\in\mathcal{F}_{x,jn^{-1/(2k+1)}}}\left|(\mathbb{P}_n-P_0)(f_{x,y_1,\cdots,y_{k-1},y})\right|\right)^2\right\}}{\left(\epsilon(j-1)^{(3+k)k/2}+m\right)^2}\\
&\le C\sum_{1\le j\le j_n} n^{(3+k)k/(2k+1)} n^{-1} n^{-(k(k+1)-1)/(2k+1)} \frac{j^{k(k+1)-1}}{\left(\epsilon(j-1)^{(3+k)k/2}+m\right)^2}\\
&= C\sum_{1\le j\le j_n} \frac{j^{k(k+1)-1}}{\left(\epsilon(j-1)^{(3+k)k/2}+m\right)^2}\\
&\le C\sum_{j=1}^{\infty} \frac{j^{k(k+1)-1}}{\left(\epsilon(j-1)^{(3+k)k/2}+m\right)^2}, \ \searrow 0 \ \text{ as } m\to\infty,
\end{aligned}
$$

where $C>0$ is a constant independent of $x\in[x_0-\delta,x_0+\delta]$. Therefore, $M_n=O_p(1)$ and hence

$$
\left|(\mathbb{P}_n-P_0)(f_{x,y_1,\cdots,y_{k-1},y})\right| \le \epsilon(y-x)^{(3+k)k/2}+O_p\left(n^{-(3+k)k/(2(2k+1))}\right)
$$

uniformly in x,y. 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 (2.25) is arbitrarily small. This proves the result for $j=0$.

Now let $1\le j\le k-1$. 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 $i = 0, \cdots, k+j-2$, 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

$$\inf_{t \in [\tau_1, \tau_{k+1+j}]} \left| \tilde{g}_n^{(j)}(t) - g_0^{(j)}(t) \right| > c\, n^{-(k-j)/(2k+1)} \tag{2.30}$$

then

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

Using similar empirical process arguments as in the proof for $j = 0$ it can be shown 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. For the MLE, the result can be proved similarly by using the same perturbation functions and also consistency of the MLE. ■

Proposition 2.6.2 *Let $x_0 > 0$ and g_0 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 the conjectured Lemma 2.5.4 holds, then for each $M > 0$ we have,*

$$\sup_{|t|\le M} \left| \bar{g}_n^{(k-1)}(x_0 + n^{-1/(2k+1)}t) - g_0^{(k-1)}(x_0) \right| = O_p(n^{-1/(2k+1)}) \tag{2.31}$$

and

$$\sup_{|t|\le M} \left| \bar{g}_n^{(j)}(x_0 + n^{-1/(2k+1)}t) - \sum_{i=j}^{k-1} \frac{n^{-(i-j)/(2k+1)} g_0^{(i)}(x_0)}{(i-j)!} t^{i-j} \right| = O_p(n^{-(k-j)/(2k+1)}) \tag{2.32}$$

for $j = 0, \cdots, k-2$.

Proof. To prove (2.32), 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 (2.31). Let $M > 0$ and $0 < \epsilon < 1$. We consider two sequences of $(k+1)$ jump points $\tau_{1,1}, \cdots, \tau_{k+1,1}$ and $\tau_{1,2}, \cdots, \tau_{k+1,2}$ as described in the previous theorem, 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 $\tau_{1,-1} \cdots, \tau_{k+1,-1}$ and $\tau_{1,-2}, \cdots, \tau_{k+1,-2}$ 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 $1 - \epsilon$

$$\begin{aligned}
(-1)^{k-2}\bar{g}_n^{(k-1)}(t-) &\le (-1)^{k-2}\bar{g}_n^{(k-1)}(t+) \\
&\le \frac{(-1)^{k-2}\bar{g}_n^{(k-2)}(\xi_2) - (-1)^{k-2}\bar{g}_n^{(k-2)}(\xi_1)}{\xi_2 - \xi_1} \\
&\le \frac{(-1)^{k-2}g_0^{(k-2)}(\xi_2) - (-1)^{k-2}g_0^{(k-2)}(\xi_1) + 2cn^{-2/(2k+1)}}{\xi_2 - \xi_1} \\
&\le (-1)^{k-2}g_0^{(k-1)}(\xi_2) + 2cn^{-1/(2k+1)}
\end{aligned}$$

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 (2.31).

Using similar arguments in the proof of Lemma 4.4 in Groeneboom, Jongbloed, and Wellner (2001b), we can show (2.32) for $j = k-2$ which specializes to

$$\sup_{|t|\leq M}\left|\bar{g}_n^{(k-2)}(x_0 + n^{-1/(2k+1)}t) - g_0^{(k-2)}(x_0) - n^{-1/(2k+1)}tg_0^{(k-1)}(x_0)\right| = O_p(n^{-2/(2k+1)})$$

for all $M > 0$. Indeed, since the jump points $\tau_{j,i}, j = 1, \cdots, k+1, i = -2, -1, 1, 2$ 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 $t \in [x_0 - Mn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$

$$\begin{aligned}
&(-1)^{k-2}\bar{g}_n^{(k-2)}(t) \\
&\geq (-1)^{k-2}\bar{g}_n^{(k-2)}(\xi_1) + (-1)^{k-2}\bar{g}_n^{(k-1)}(\xi_1)(t-\xi_1) \\
&\geq (-1)^{k-2}g_0^{(k-2)}(\xi_1) - cn^{-2/(2k+1)} + ((-1)^{k-2}g_0^{(k-1)}(x_0) + c'n^{-1/(2k+1)})(t-\xi_1) \\
&\geq (-1)^{k-2}g_0^{(k-2)}(x_0) + (\xi_1 - x_0)(-1)^{k-2}g_0^{(k-1)}(x_0) + (t-\xi_1)(-1)^{k-2}g_0^{(k-1)}(x_0) \\
&\quad -cn^{-2/(2k+1)} - c'n^{-1/(2k+1)}(\xi_1 - t) \qquad (2.33) \\
&\geq (-1)^{k-2}g_0^{(k-2)}(x_0) + (t-x_0)(-1)^{k-2}g_0^{(k-1)}(x_0) - (c+2Kc')n^{-2/(2k+1)}.
\end{aligned}$$

where in (2.33), 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

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

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 (2.32) holds for $j = k-2$.

Now, suppose that (2.32) 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 (2.32) 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 there 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 $1-\epsilon$,

$$
\begin{aligned}
-c' n^{-(k-j')/(2k+1)} &\leq \bar{g}_n^{(j')}(x_0 + n^{-1/(2k+1)} t) - \sum_{i=j'}^{k-1} \frac{n^{-(i-j')/(2k+1)} g_0^{(i)}(x_0)}{(i-j')!} t^{i-j'} \\
&\leq c' n^{-(k-j')/(2k+1)} \qquad (2.34)
\end{aligned}
$$

for all $|t| \leq M$ and $j' > j-1$.

Using convexity of $\bar{g}_n^{(j-1)}$ “from below”, we have for all $|t - x_0| \leq M n^{-1/(2k+1)}$ with probability greater than $1 - 2\epsilon$,

$$
\begin{aligned}
&\bar{g}_n^{(j-1)}(t) \\
&\geq \bar{g}_n^{(j-1)}(\xi_1) + \bar{g}_n^{(j)}(\xi_1)(t-\xi_1) + \cdots + \frac{1}{(k-j)!} \bar{g}_n^{(k-1)}(\xi_1)(t-\xi_1)^{k-j} \\
&\geq g_0^{(j-1)}(\xi_1) - c n^{-(k-j+1)/(2k+1)} + \left(\sum_{i=j}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j)!} (\xi_1 - x_0)^{i-j} (t - \xi_1) \right) \\
&\quad + \left(\sum_{i=j+1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j-1)!} (\xi_1 - x_0)^{i-j-1} \right) \frac{(t-\xi_1)^2}{2!} + \cdots + g_0^{(k-1)}(x_0) \frac{(t-\xi_1)^{k-j}}{(k-j)!} \\
&\quad + c' n^{-(k-j)/(2k+1)} (t - \xi_1) - c' n^{-(k-j-1)/(2k+1)} \frac{(t-\xi_1)^2}{2!} \\
&\quad + \cdots - c' n^{-1/(2k+1)} \frac{(t-\xi_1)^{k-j}}{(k-j)!}. \qquad (2.35)
\end{aligned}
$$

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

$$
|t - \xi_1| \leq K n^{-1/(2k+1)},
$$

the right side of (2.35) can be bounded below by

$$
\sum_{i=j-1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j+1)!} (\xi_1 - x_0)^{i-j+1} + \sum_{i=j}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j)!} (\xi_1 - x_0)^{i-j} (t - \xi_1)
$$

$$
\begin{aligned}
& + \sum_{i=j+1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j-1)!}(\xi_1 - x_0)^{i-j-1}\frac{(t-\xi_1)^2}{2!} + \cdots + g_0^{(k-1)}(x_0)\frac{(t-\xi_1)^{k-j}}{(k-j)!} \\
& - \left(c + c'\sum_{p=1}^{k-j}\frac{K^p}{p!}\right) n^{-(k-j+1)/(2k+1)} + \frac{g_0^{(k)}(\nu)}{(k-j+1)!}(\xi_1 - x_0)^{k-j+1} \\
= \; & g_0^{(j-1)}(x_0) + g_0^{(j)}(x_0)(t-x_0) \\
& + \frac{g_0^{(j+1)}(x_0)}{2!}\left((\xi_1 - x_0)^2 + 2(\xi_1 - x_0)(t-\xi_1) + (t-\xi_1)^2\right) \\
& + \cdots + \frac{g_0^{(k-1)}(x_0)}{(k-j)!}\sum_{p=0}^{k-j}\frac{(k-j)!}{(k-j-p)!p!}(\xi_1 - x_0)^{k-j-p}(t-\xi_1)^p \\
& - \left(c + c'\sum_{p=1}^{k-j}\frac{K^p}{p!}\right) n^{-(k-j+1)/(2k+1)} + \frac{g_0^{(k)}(\nu)}{(k-j+1)!}(\xi_1 - x_0)^{k-j+1} \\
= \; & g_0^{(j-1)}(x_0) + g_0^{(j)}(x_0)(t-x_0) + \cdots + \frac{g^{(k-1)}(x_0)}{(k-j)!}(t-x_0)^{k-j} \\
& - \left(c + c'\sum_{p=1}^{k-j}\frac{K^p}{p!}\right) n^{-(k-j+1)/(2k+1)} - \frac{D_1 K^{k-j+1}}{(k-j+1)!} n^{-(k-j+1)/(2k+1)}
\end{aligned}
$$

since $0 \le \xi_1 - x_0 \le Kn^{-1/(2k+1)}$.

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 that

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

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

$$
\begin{aligned}
& \bar{g}_n^{(j-1)}(t) \\
& \quad \le \; \bar{g}_n^{(j-1)}(\xi_{-1}) + \bar{g}_n^{(j)}(\xi_{-1})(t-\xi_{-1}) + \bar{g}_n^{(j+1)}(\xi_{-1})\frac{(t-\xi_{-1})^2}{2!} \\
& \qquad + \cdots + \frac{g_0^{(k-2)}(\xi_1) - g_0^{(k-2)}(\xi_{-1}) + 2cn^{-2/(2k+1)}}{\xi_1 - \xi_{-1}}\frac{(t-\xi_{-1})^{k-j}}{(k-j)!}
\end{aligned}
$$

$$
\begin{aligned}
&\leq g_0^{(j-1)}(\xi_{-1}) + cn^{-(k-j+1)/(2k+1)} \\
&\quad + \left(\sum_{i=j}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j)!}(\xi_{-1}-x_0)^{i-j} + c' n^{-(k-j)/(2k+1)} \right)(t-\xi_{-1}) \\
&\quad + \left(\sum_{i=j+1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j-1)!}(\xi_{-1}-x_0)^{i-j-1} + c' n^{-(k-j-1)/(2k+1)} \right) \frac{(t-\xi_{-1})^2}{2!} \\
&\quad + \cdots + \left(g_0^{(k-1)}(\xi_1) + \frac{c}{K} n^{-1/(2k+1)} \right) \frac{(t-\xi_{-1})^{k-j}}{(k-j)!} \\
&\leq \sum_{i=j-1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j+1)!}(\xi_{-1}-x_0)^{i-j+1} + \frac{g^{(k)}(\nu)}{k!}(\xi_{-1}-x_0)^{k-j+1} \\
&\quad + \left(\sum_{i=j}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j)!}(\xi_{-1}-x_0)^{i-j} \right)(t-\xi_{-1}) \\
&\quad + \cdots + \left(\sum_{i=j+1}^{k-1} \frac{g_0^{(i)}(x_0)}{(i-j-1)!}(\xi_{-1}-x_0)^{i-j-1} \right) \frac{(t-\xi_{-1})^2}{2!} \\
&\quad + \left(g_0^{(k-1)}(x_0) + cn^{-1/(2k+1)} \right) \frac{(t-\xi_{-1})^{k-j}}{(k-j)!} \\
&\quad + \left(c(1+K^{k-j}) + c' \sum_{p=1}^{k-j} \frac{K^p}{p!} + \frac{D_1 K^{k-j+1}}{k!} \right) n^{-(k-j+1)/(2k+1)} \\
&= g_0^{(j-1)}(x_0) + g_0^{(j)}(x_0)(t-x_0) + \cdots + g_0^{(k-j)}(x_0)\frac{(t-x_0)^{k-j}}{(k-j)!} + K' n^{-(k-j+1)/(2k+1)}
\end{aligned}
$$

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 (2.32) holds for $j-1$. ■

2.7 Asymptotic distribution

Recall that the characterization of the LSE $\tilde{g}_n$ involved the processes $\mathbb{Y}_n$ and $\tilde{H}_n$ defined by

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

and

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

Since we are interested in estimating the true density or its l-th derivative ($l \leq k-1$)

at a point $x_0 > 0$, we need to define a local version of these processes. 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

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

where

$$\begin{aligned}\tilde{A}_{(k-1)n} &= \frac{n^{(k+1)/(2k+1)}}{(k-1)!} \left(\tilde{H}_n^{(k-1)}(x_0) - \mathbb{Y}_n^{(k-1)}(x_0) \right) = \frac{n^{(k+1)/(2k+1)}}{(k-1)!} \left(\tilde{G}_n(x_0) - \mathbb{G}_n(x_0) \right) \\ \tilde{A}_{(k-2)n} &= \frac{n^{(k+2)/(2k+1)}}{(k-2)!} \left(\tilde{H}_n^{(k-2)}(x_0) - \mathbb{Y}_n^{(k-2)}(x_0) \right) \\ &\vdots \\ \tilde{A}_{1n} &= n^{(2k-1)/(2k+1)} \left(\tilde{H}_n'(x_0) - \mathbb{Y}_n'(x_0) \right) \\ \tilde{A}_{0n} &= n^{2k/(2k+1)} \left(\tilde{H}_n(x_0) - \mathbb{Y}_n(x_0) \right),\end{aligned}$$

and $\tilde{G}_n(x) = \int_0^x \tilde{g}_n(y)dy$.

Example 2.7.1 $k = 3$

$$\begin{aligned}\mathbb{Y}_n^{loc}(t) &= n^{6/7} \int_{x_0}^{x_0+tn^{-1/7}} \int_{x_0}^{w} \Big\{ \mathbb{G}_n(v) - \mathbb{G}_n(x_0) \\ &\qquad - \int_{x_0}^{v} \left(g_0(x_0) + (u-x_0) g_0'(x_0) + \frac{1}{2}(u-x_0)^2 g_0''(x_0) \right) du \Big\} dv dw,\end{aligned}$$

and

$$\begin{aligned}\tilde{H}_n^{loc}(t) &= n^{6/7}\int_{x_0}^{x_0+tn^{-1/7}}\int_{x_0}^{w}\int_{x_0}^{v}\Big\{\tilde{g}_n(u)-g_0(x_0)-(u-x_0)g_0'(u)\\ &\qquad -\frac{1}{2}(u-x_0)^2g_0''(x_0)\Big\}dudvdw+\tilde{A}_{2n}t^2+\tilde{A}_{1n}t+\tilde{A}_{0n}\end{aligned}$$

where

$$\begin{aligned}\tilde{A}_{2n} &= \frac{n^{4/7}}{2}\Big(\tilde{G}_n(x_0)-\mathbb{G}_n(x_0)\Big),\\ \tilde{A}_{1n} &= n^{5/7}\Big(\tilde{H}_n'(x_0)-\mathbb{Y}_n'(x_0)\Big),\end{aligned}$$

and

$$\tilde{A}_{0n} = n^{6/7}\Big(\tilde{H}_n(x_0)-\mathbb{Y}_n(x_0)\Big).$$

In the following lemma, we will give the asymptotic distribution of the local process $\mathbb{Y}_n^{loc}$ 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-differentiable at x_0 and continuous in an open neighborhood around x_0.

Lemma 2.7.1 *Let x_0 be a point where g_0 is k-differentiable and $g_0^{(k)}$ is continuous at x_0. Then*

$$\mathbb{Y}_n^{loc}(t)\Rightarrow\begin{cases}\sqrt{g_0(x_0)}\int_0^t\int_0^{s_{k-1}}\cdots\int_0^{s_2}W(s_1)ds_1\cdots ds_{k-1}+\frac{1}{2k!}t^{2k}g_0^{(k)}(x_0),\ t\geq 0\\ \sqrt{g_0(x_0)}\int_t^0\int_{s_{k-1}}^0\cdots\int_{s_2}^0 W(s_1)ds_1\cdots ds_{k-1}+\frac{1}{2k!}t^{2k}g_0^{(k)}(x_0),\ t<0\end{cases}$$

in $D[-K,K]$ for every $K>0$ and where W is standard Brownian motion starting at 0.

Proof. Fix $K>0$. We will prove the lemma for $t\geq 0$ and similar arguments can be used for $t\in[-K,0)$. We have

$$\begin{aligned}\mathbb{Y}_n^{loc}(t) &= n^{2k/(2k+1)}\int_{x_0}^{x_0+tn^{-1/(2k+1)}}\int_{x_0}^{v_{k-1}}\cdots\int_{x_0}^{v_2}\Big\{\mathbb{G}_n(v_1)-\mathbb{G}_n(x_0)\\ &\quad -\int_{x_0}^{v_1}\Big(g_0(x_0)+(u-x_0)g_0'(x_0)+\cdots+\frac{1}{(k-1)!}(u-x_0)^{k-1}g_0^{(k-1)}(x_0)\Big)du\Big\}\\ &\qquad dv_1dv_2\cdots dv_{k-1}\\ &= A_n+B_n,\end{aligned}$$

where

$$A_n = n^{2k/(2k+1)} \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \left\{ \mathbb{G}_n(v_1) - \mathbb{G}_n(x_0) - (G_0(v_1) - G_0(x_0)) \right\} dv_1 dv_2 \cdots dv_{k-1},$$

and

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

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 $\xi_1, \cdots, \xi_n$ are i.i.d. $U(0,1)$ random variables, we have

$$\begin{aligned} A_n & \stackrel{d}{=} n^{2k/(2k+1)-1/2} \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Big(\mathbb{U}_n(G_0(v_1)) - \mathbb{U}_n(G_0(x_0)) \Big) \Big) \\ & \qquad dv_1 dv_2 \cdots dv_{k-1} \\ & = n^{\frac{2k-1}{2(2k+1)}} \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Big(\mathbb{U}_n(G_0(v_1)) - \mathbb{U}_n(G_0(x_0)) \Big) \Big) \\ & \qquad dv_1 dv_2 \cdots dv_{k-1}, \end{aligned}$$

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

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

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

$$
\begin{aligned}
B_{n2} &= n^{\frac{2k}{2k+1}} \frac{1}{(k+1)!} g_0^{(k)}(x_0) \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_3} \frac{1}{k+2}(v_2 - x_0)^{k+2} dv_2 \cdots dv_{k-1} \\
&= n^{\frac{2k}{2k+1}} \frac{1}{(k+3)!} g_0^{(k)}(x_0) \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_4} (v_3 - x_0)^{k+3} dv_4 \cdots dv_{k-1} \\
&\vdots \\
&= n^{\frac{2k}{2k+1}} \frac{1}{(2k-1)!} g_0^{(k)}(x_0) \int_{x_0}^{x_0+tn^{-1/(2k+1)}} (v_{k-1} - x_0)^{2k-1} dv_{k-1} \\
&= n^{\frac{2k}{2k+1}} g_0^{(k)}(x_0) \frac{1}{(2k)!} \left(\frac{t}{n^{1/2k+1}} \right)^{2k} \\
&= \frac{1}{(2k)!} g_0^{(k)}(x_0) t^{2k}.
\end{aligned}
$$

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

$$
B_n \to \frac{1}{(2k)!} g_0^{(k)}(x_0) t^{2k}, \tag{2.1}
$$

as $n \to \infty$ uniformly in $0 \leq t \leq K$.

Using the identity

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

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

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

But,

$$
\begin{aligned}
A_{n1} &\leq 2n^{\frac{2k-1}{2(2k+1)}} \|\mathbb{U}_n - \mathbb{U}\|_\infty \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} dv_1 \cdots dv_{k-1} \\
&= 2n^{\frac{2k-1}{2(2k+1)}} \|\mathbb{U}_n - \mathbb{U}\|_\infty \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_3} (v_2 - x_0) dv_2 \cdots dv_{k-1} \\
&= 2n^{\frac{2k-1}{2(2k+1)}} \|\mathbb{U}_n - \mathbb{U}\|_\infty \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_4} \frac{1}{2}(v_3 - x_0)^2 dv_3 \\
&\vdots \\
&= 2n^{\frac{2k-1}{2(2k+1)}} \|\mathbb{U}_n - \mathbb{U}\|_\infty \frac{1}{(k-2)!} \int_{x_0}^{x_0+tn^{-1/(2k+1)}} (v_{k-1} - x_0)^{k-2} dv_{k-1} \\
&= 2n^{\frac{2k-1}{2(2k+1)}} \|\mathbb{U}_n - \mathbb{U}\|_\infty \frac{1}{(k-1)!} \left(\frac{t}{n^{1/(2k+1)}} \right)^{k-1} \\
&= 2t^{k-1} n^{\frac{1/2}{2k+1}} O\left(\frac{\log(n)^2}{n^{1/2}} \right) \\
&= 2t^{k-1} O\left(\frac{\log(n)^2}{n^{k/(2k+1)}} \right) \qquad (2.2)
\end{aligned}
$$

since $\|\mathbb{U}_n - \mathbb{U}\|_\infty = O\left(n^{-1/2} (\log(n))^2\right)$ 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

$$
\begin{aligned}
A_{n3} &\leq |W(1)| g_0(x_0) n^{\frac{2k-1}{2(2k+1)}} \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} (v_1 - x_0) dv_1 \cdots dv_{k-1} \\
&= |W(1)| g_0(x_0) n^{\frac{2k-1}{2(2k+1)}} \int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_3} \frac{1}{2}(v_1 - x_0)^2 dv_2 \\
&\vdots \\
&= |W(1)| g_0(x_0) n^{\frac{2k-1}{2(2k+1)}} \frac{1}{(k-1)!} \int_{0}^{x_0+tn^{-1/(2k+1)}} (v_{k-1} - x_0)^{k-1} dv_{k-1} \\
&= |W(1)| g_0(x_0) n^{\frac{2k-1}{2(2k+1)}} \frac{1}{k!} \left(\frac{t}{n^{1/(2k+1)}} \right)^{k} \\
&= |W(1)| g_0(x_0) t^k n^{-\frac{1}{2(2k+1)}} \to_p 0, \qquad (2.3)
\end{aligned}
$$

as $n \to \infty$ uniformly in $0 \leq t \leq K$.

Finally, using the change of variables $s_j = n^{1/(2k+1)}(v_j - x_0)$ for $j = 1, \ldots, k-1$, we have

$$\begin{aligned}
A_{n2} &= n^{\frac{2k-1}{2(2k+1)}} \int_{x_0}^{x_0+tn^{\frac{-1}{2k+1}}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Big(W(G_0(v_1)) - W(G_0(x_0)) \Big) dv_1 \cdots dv_{k-1} \\
&= n^{\frac{2k-1}{2(2k+1)}} n^{-\frac{(k-1)}{(2k+1)}} \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} \Big(W(G_0(n^{\frac{-1}{2k+1}} s_1 + x_0)) - W(G_0(x_0)) \Big) \\
&\qquad\qquad ds_1 \cdots ds_{k-1} \\
&\stackrel{d}{=} n^{\frac{1}{2(2k+1)}} \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{v_2} W\Big(G_0(n^{\frac{-1}{2k+1}} s_1 + x_0) - G_0(x_0) \Big) ds_1 \cdots ds_{k-1} \\
&\stackrel{d}{=} \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W\Big(n^{\frac{1}{2k+1}} (G_0(n^{\frac{-1}{2k+1}} s_1 + x_0) - G_0(x_0)) \Big) ds_1 \cdots ds_{k-1} \\
&\to \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1 g_0(x_0)) ds_1 \cdots ds_{k-1} \quad \text{as } n \to \infty \\
&\stackrel{d}{=} \sqrt{g_0(x_0)} \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1) ds_1 \cdots ds_{k-1}. \qquad (2.4)
\end{aligned}$$

Therefore, combining (2.1), (2.2), (2.3) and (2.4) yields

$$\mathbb{Y}_n^{loc}(t) \Rightarrow \sqrt{g_0(x_0)} \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1) ds_1 \cdots ds_{k-1} + \frac{1}{(2k)!} t^{2k} g_0^{(k)}(x_0)$$

for $0 \leq t \leq K$. A similar argument for $-K \leq t < 0$ yields the conclusion. ■

We will now rescale this limiting process to obtain a "canonical" version. In the case of $k = 2$, Groeneboom, Jongbloed and Wellner (Groeneboom, Jongbloed, and Wellner (2001b)) chose the "canonical process" to be

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

and one can establish a link between estimating a non-decreasing convex density and the following Gaussian problem:

$$dX(t) = f_0(t)dt + dW(t) \qquad (2.5)$$

where f_0 is convex. Integrating (2.5) twice and choosing $f_0(t) = 12t^2$, we have

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

Similarly, one can establish a link between the k-monotone density estimation problem and the Gaussian problem:

$$dX(t) = f_0(t)dt + dW(t)$$

where $(-1)^k f_0$ has a convex $(k-2)$-th derivative. If we choose $f_0(t) = t^k$ and integrate the previous stochastic differential equation $k-1$ times, we get

$$\begin{aligned}
X(t) &= \frac{1}{k+1}t^{k+1} + W(t) \\
X_1(t) &= \int_0^t X(s)ds = \frac{1}{(k+1)(k+2)}t^{k+2} + \int_0^t W(s)ds \\
X_2(t) &= \int_0^t \int_0^{s_2} X(s_1)ds_1 ds_2 = \frac{k!}{(k+3)!}t^{k+3} + \int_0^t \int_0^{s_2} W(s_1)ds_1 ds_2 \\
&\vdots \\
X_{k-1}(t) &= \frac{k!}{(2k)!}t^{2k} + \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1)ds_1 ds_2 \cdots ds_{k-1} \stackrel{d}{=} Y_k(t).
\end{aligned}$$

Here we will rescale the limiting process so that we obtain the "canonical process"

$$Y_k(t) = \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1)ds_1 ds_2 \cdots ds_{k-1} + (-1)^k \frac{k!}{(2k)!}t^{2k}, \quad t \geq 0.$$

Let us denote by σ and a, the multiplicative term $\sqrt{g_0(x_0)}$ and $(-1)^k g_0^{(k)}(x_0)/k!$, the leading coefficient of the drift term in the limiting process

$$Y_{a,\sigma}(t) = \sqrt{g_0(x_0)} \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1)ds_1 \cdots ds_{k-1} + \frac{(-1)^k}{k!} g_0^{(k)}(x_0)(-1)^k \frac{k!}{(2k)!}t^{2k}$$

respectively. In the following, we are going to find constants r_1 and r_2 such that

$$r_1 Y_{a,\sigma}(r_2 t) \stackrel{d}{=} Y_k(t).$$

We have,

$$\begin{aligned}
Y_{a,\sigma}(t) &= a(-1)^k \frac{k!}{(2k)!}t^{2k} + \sigma \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1)ds_1 \cdots ds_{k-1} \\
&\stackrel{d}{=} a(-1)^k \frac{k!}{(2k)!}t^{2k} + \alpha^{-1/2}\sigma \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(\alpha s_1)ds_1 \cdots ds_{k-1} \\
&\stackrel{d}{=} a(-1)^k \frac{k!}{(2k)!}t^{2k} + \alpha^{-1/2}\sigma \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{\alpha s_2} \frac{1}{\alpha} W(s_1)ds_1 \cdots ds_{k-1} \\
&\stackrel{d}{=} a(-1)^k \frac{k!}{(2k)!}t^{2k} + \alpha^{-1/2}\sigma \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{\alpha s_3} \int_0^{s_2} \frac{1}{\alpha^2} W(s_1)ds_1 \cdots ds_{k-1} \\
&\vdots \\
&\stackrel{d}{=} a(-1)^k \frac{k!}{(2k)!}t^{2k} + \alpha^{-1/2}\sigma \int_0^{\alpha t} \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1) \frac{1}{\alpha^{k-1}} ds_1 \cdots ds_{k-1} \\
&\stackrel{d}{=} a(-1)^k \frac{k!}{(2k)!}t^{2k} + \alpha^{1/2-k}\sigma \int_0^{\alpha t} \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1)ds_1 \cdots ds_{k-1}.
\end{aligned}$$

Therefore,

$$r_1 Y_{a,\sigma}(r_2 t) \stackrel{d}{=} a(-1)^k \frac{k!}{(2k)!} r_1 (r_2 t)^{2k} + r_1 \alpha^{1/2-k} \sigma \int_0^{r_2 \alpha t} \int_0^{s_{k-1}} \cdots \int_0^{s_2} W(s_1) ds_1 \cdots ds_{k-1},$$

and

$$\begin{cases} a r_1 r_2^{2k} = 1, \\ r_1 \alpha^{1/2-k} \sigma = 1, \\ r_2 \alpha = 1. \end{cases}$$

Solving the previous system of equations yields

$$\alpha = \left(\frac{a}{\sigma}\right)^{2/(2k+1)}$$

and therefore

$$r_1 = \frac{1}{\sqrt{g_0(x_0)}} \left(\frac{(-1)^k g_0^{(k)}(x_0)}{k! \sqrt{g_0(x_0)}} \right)^{(2k-1)/(2k+1)} \quad \text{and} \tag{2.6}$$

$$r_2 = \left(\frac{\sqrt{g_0(x_0)}}{\frac{(-1)^k g_0^{(k)}(x_0)}{k!}} \right)^{2/(2k+1)}. \tag{2.7}$$

Thus,

$$\begin{aligned} Y_{a,\sigma}(t) &\stackrel{d}{=} \frac{1}{r_1} Y_k\left(\frac{t}{r_2}\right) \\ &= \sqrt{g_0(x_0)} \left(\frac{k! \sqrt{g_0(x_0)}}{(-1)^k g_0^{(k)}(x_0)} \right)^{(2k-1)/(2k+1)} Y_k\left(\left(\frac{k! \sqrt{g_0(x_0)}}{(-1)^k g_0^{(k)}(x_0)} \right)^{-2/(2k+1)} t \right). \end{aligned}$$

Note that (2.6) specializes to A.9 in Groeneboom, Jongbloed, and Wellner (2001a), page 1651 when $k = 2$.

Let us now have a closer look at the difference of the two local processes $\mathbb{Y}_n^{loc}$ and $\tilde{H}_n^{loc}$. The asymptotic behavior of this difference, as we will show later, will have a crucial role in establishing the asymptotic theory of the LSE.

We have,

$$\begin{aligned} &\tilde{H}_n^{loc}(t) - \mathbb{Y}_n^{loc}(t) \\ &= n^{\frac{2k}{2k+1}} \int_{x_0}^{x_0 + t n^{-\frac{1}{2k+1}}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_2} \Big\{ \Big((\tilde{G}_n(v_1) - \tilde{G}_n(x_0)) - (\mathbb{G}_n(v_1) - \mathbb{G}_n(x_0)) \Big) \end{aligned}$$

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

by the first Fenchel condition satisfied by the LSE.

A natural thing to do is to rescale the processes $\mathbb{Y}_n^{loc}(t)$ and $\tilde{H}_n^{loc}(t)$ so that the rescaled

$\mathbb{Y}_n^{loc}(t)$ converges to the process Y_k we defined already. Since the scaling of $\mathbb{Y}_n^{loc}(t)$ will be exactly the same as the one we used for Y_k, we define $\tilde{H}_n^l$ as

$$\tilde{H}_n^l(t) = r_1 \tilde{H}_n^{loc}(r_2 t)$$

where

$$r_1 = \frac{1}{\sqrt{g_0(x_0)}} \left(\frac{(-1)^k g_0^{(k)}(x_0)}{\sqrt{g_0(x_0)} k!} \right)^{(2k-1)/(2k+1)}, \quad r_2 = \left(\frac{(-1)^k g_0^{(k)}(x_0)}{\sqrt{g_0(x_0)} k!} \right)^{-2/(2k+1)}.$$

Now, we can write

$$\begin{aligned}
(\tilde{H}_n^l)^{(k)}(0) &= r_1 r_2^k (\tilde{H}_n^{loc})^{(k)}(0) = n^{k/(2k+1)} c_k(g_0)(\tilde{g}_n(x_0) - g_0(x_0)) \\
(\tilde{H}_n^l)^{(k+1)}(0) &= r_1 r_2^{k+1} (\tilde{H}_n^{loc})^{(k+1)}(0) = n^{(k-1)/(2k+1)} c_{k-1}(g_0)(\tilde{g}_n'(x_0) - g_0'(x_0)) \\
(\tilde{H}_n^l)^{(k+2)}(0) &= r_1 r_2^{k+2} (\tilde{H}_n^{loc})^{(k+2)}(0) = n^{(k-2)/(2k+1)} c_{k-2}(g_0)(\tilde{g}_n''(x_0) - g_0''(x_0)) \\
&\vdots \\
(\tilde{H}_n^l)^{(2k-1)}(0) &= r_1 r_2^{2k-1} (\tilde{H}_n^{loc})^{(2k-1)}(0) = n^{1/(2k+1)} c_1(g_0)(\tilde{g}_n^{(k-1)}(x_0) - g_0^{(k-1)}(x_0)).
\end{aligned}$$

Now, let us consider the MLE $\hat{g}_n$. Recall that the characterization of this estimator involves the process $\widehat{H}_n$ given by

$$\widehat{H}_n(t) = \int_0^t \frac{(t-u)^{k-1}}{\hat{g}_n(u)} d\mathbb{G}_n(t), \quad \text{for all} \quad t \geq 0$$

and that

$$\widehat{H}_n(t) \begin{cases} \leq \frac{t^k}{k}, & t \geq 0 \\ = \frac{t^k}{k}, & \text{when } \ t \ \text{ is a jump point of } \ \hat{g}_n^{(k-1)} \end{cases}$$

is a necessary and sufficient condition for $\hat{g}_n$ to be the MLE. Note that $\widehat{H}_n$ and $\hat{H}_n$ defined in Lemma 2.2.5 in Section 2 are different: $\widehat{H}_n = (t^k/k)\hat{H}_n$.

We define the local processes $\widehat{Y}_n^{loc}$ and $\widehat{H}_n^{loc}$ as

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

and

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

where for $0 \le j \le k-1$

$$\widehat{A}_{jn} = -\frac{n^{(2k-j)/(2k+1)}}{(k-1)!j!} g_0(x_0) \left(\widehat{H}_n^{(j)}(x_0) - \frac{(k-1)!}{(k-j)!} x_0^{k-j} \right).$$

With this particular choice of $\widehat{A}_{jn}$, $0 \le j \le k-1$, we have

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

But notice that for any $t \ge 0$

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

It follows that

$$\begin{aligned}
&\int_{x_0}^{x_0+tn^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_1} \frac{1}{\hat{g}_n(v)} d\mathbb{G}_n(v) dv_1 \cdots dv_{k-1} \\
&= \frac{1}{(k-1)!} \int_{x_0}^{x_0+n^{-1/(2k+1)}} \int_{x_0}^{v_{k-1}} \cdots \int_{x_0}^{v_1} \left(\widehat{H}_n^{(k-1)}(v_1) - \widehat{H}_n^{(k-1)}(x_0) \right) dv_1 \cdots dv_{k-1} \\
&= \frac{1}{(k-1)!} \left(\widehat{H}_n(x_0 + tn^{-1/(2k+1)}) - \sum_{j=0}^{k-1} \frac{t^j n^{-j/(2k+1)}}{j!} \widehat{H}_n^{(j)}(x_0) \right).
\end{aligned}$$

Therefore,

$$\widehat{H}_n^{loc}(t) - \widehat{Y}_n^{loc}(t)$$

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

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

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

As for the LSE, we define $\widehat{Y}_n^l$ and $\widehat{H}_n^l$ by

$$\widehat{Y}_n^l(t) = r_1 \widehat{Y}_n^{loc}(r_2 t)$$

and

$$\widehat{H}_n^l(t) = r_1 \widehat{H}_n^{loc}(r_2 t).$$

Lemma 2.7.2 *Let $K > 0$. Then*

$$\widehat{Y}_n \Rightarrow Y_k$$

in $D[-K,K]$.

Proof. We apply the same arguments in the proof of Lemma 2.7.1 in the case of the LSE.

∎

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

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

and

$$\widehat{A}_{jn} = -\frac{n^{(2k-j)/(2k+1)}}{(k-1)!j!} g_0(x_0) \left(\widehat{H}_n^{(j)}(x_0) - \frac{(k-1)!}{(k-j)!} x_0^{k-j} \right).$$

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

Lemma 2.7.3 *For all $j \in \{0, \ldots, k-1\}$, let $\bar{A}_{jn}$ denote either $\tilde{A}_{jn}$ or $\widehat{A}_{jn}$. If the conjectured Lemma 2.5.4 holds, then*

$$\bar{A}_{jn} = O_p(1). \tag{2.8}$$

Proof. We will show the lemma only for the LSE as the arguments are very similar for the MLE. Let $j \in \{0, \ldots, k-1\}$ and denote $\tilde{\Delta}_n(x) = \tilde{H}_n(x) - \mathbb{Y}_n(x)$ for all $x \geq 0$. We will start by proving (2.8) for $j = k-1$ and $k-2$ and then use induction for $2 \leq j \leq k-3$. Proving (2.8) for $j = k-1$ would have been sufficient but we wanted to show it for $j = k-2$ to give a better idea about how the proof works.

Now consider k successive jump points, $\tau_1, \cdots, \tau_k$, of $\tilde{g}_n^{(k-1)}$ where τ_1 is the first jump after x_0. By the mean value theorem, there exist $\tau_1^{(1)} \in (\tau_1, \tau_2)$, $\tau_2^{(1)} \in (\tau_2, \tau_3)$, $\ldots$, $\tau_{k-1}^{(1)} \in (\tau_{k-1}, \tau_k)$ 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 since

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

we can write,

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

Fix $0 < \epsilon < 1$. By Lemma 2.5.9 and Proposition 2.6.2, we can find $M > 0$ and $c > 0$ such that with probability greater than $1 - \epsilon$

$$x_0 \le \tau \le x_0 + Mn^{-1/(2k+1)}$$

and

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

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

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

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

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

To finish off the proof, we only need to check that

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

But this can be shown using similar arguments to those in the proof of Proposition 2.6.1. Indeed,

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

is an empirical process indexed by the point $\tau \in [x_0, x_0 + Mn^{-1/(2k+1)}]$.

Consider now the empirical process

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

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

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

for a fixed $y > 0$ and $R > 0$. One can prove that there exist, $\delta > 0$ and $R > 0$ such that

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

for all $|y - x_0| \leq \delta$, $z \in [y, y+R]$ and for all $\epsilon > 0$. It follows 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 $j = k-1$. Note that we obtain the same result if we replace x_0 by any x in an neighborhood of x_0 of the form $]x_0 - Mn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$, for some constant $K > 0$; i.e., we can find $K > 0$ indenpendent of x such that

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

with large probability.

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 as we did for $\tilde{\Delta}_n^{(k-1)}$). We can write

$$\tilde{\Delta}_n^{(k-2)}(x_0) \quad = \quad \tilde{\Delta}_n^{(k-2)}(x_0) - \tilde{\Delta}_n^{(k-2)}(\tau)$$

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

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

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

with large probability.

Now,

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

We can find $d > 0$ such that

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

and

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

for all $t \in [x_0 - M n^{-1/(2k+1)}, x_0 + M n^{-1/(2k+1)}]$ 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, using again empirical processes arguments, we can show 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 $x \in [x_0 - M n^{-1/(2k+1)}, n^{-1/(2k+1)}, x_0 + M n^{-1/(2k+1)}]$, for some $M > 0$; i.e., we can find

$K > 0$ indenpendent of x such that

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

with large probability.

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

$$(k-1-j')!|\tilde{\Delta}_n^{(j')}(z)| \leq cn^{-(2k-j')/(2k+1)}.$$

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

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

Take x to be a zero of $\tilde{\Delta}_n^{(j)}$ (such 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 $x_0 - Mn^{-1/(2k+1)} \leq x \leq x_0 + Mn^{-1/(2k+1)}$. Now by applying the induction hypothesis, there exists $c > 0$ such that we have for all $y \in [x_0 - Mn^{-1/(2k+1)}, x_0 + Mn^{-1/(2k+1)}]$, we have

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

But,

$$\sum_{l=1}^{k-1-j} \binom{k-1-j}{l} |y-x|^l n^{-(2k-(j+l))/(2k+1)} \leq \left(\sum_{l=1}^{k-1-j} \binom{k-1-j}{l} (2M)^l \right) n^{-(2k-j)/(2k+1)}$$

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 $j = 0, \cdots, k-1$. ∎

Theorem 2.7.1 *For all $k \geq 1$, let Y_k denote the same stochastic process defined before; i.e.,*

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

There exists an almost surely uniquely defined stochastic process H_k characterized by the three following conditions:

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

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

(ii) $(-1)^k H_k$ is $2k$-convex; i.e. $(-1)^k H_k^{(2k-2)}$ exists and 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 $j = 0, \cdots, (k-2)/2$; 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 $j = 0, \cdots, (k-3)/2$.

Proof. Existence of the processes H_k follows from Corollary 3.2.1 in Chapter 3. ∎

Lemma 2.7.4 *Let $0 \leq j \leq 2k-1$ and $c > 0$. Let $\bar{H}_n^l$ denote either $\widehat{H}_n^l$ or $\tilde{H}_n^l$. Then*

$$(\bar{H}_n^l)^{(j)} \Rightarrow H_k^{(j)}$$

in $D[-c, c]$ for $j = 0, \cdots, 2k-1$ and where H_k is the stochastic process defined in Theorem 2.7.1.

Proof. The arguments are very similar to the ones used in Groeneboom, Jongbloed and Wellner (Groeneboom, Jongbloed, and Wellner (2001b)). We show 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

$$Z_n(t) = \left(\tilde{H}_n^l(t), \cdots, (\tilde{H}_n^l)^{(2k-2)}(t), \mathbb{Y}_n^l(t), \cdots, (\mathbb{Y}_n^l)^{(k-2)}(t), (\tilde{H}_n^l)^{(2k-1)}(t), (\mathbb{Y}_n^l)^{(k-1)}(t)\right).$$

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 Lemma 2.7.3, we know that $(\tilde{H}_n^l)^{(j)}$ is tight in $C[-c, c]$ for $j = 0, \cdots, 2k-2$. 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 $\left(\mathbb{Y}_n^l, \cdots, (\mathbb{Y}_n^l)^{(k-2)}\right)$ 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 an $M > 0$ 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

$$Z_0 = \left(H_0, \cdots, H_0^{(2k-1)}, Y_0, \cdots, Y_0^{(k-2)}, H_0^{(2k-1)}, Y_0^{(k-1)}\right) \tag{2.9}$$

in $E_k[-c, c]$ and where $Y_0 = Y_k$.

Now, consider the functions ϕ_1 and $\phi_2 : E_k[-c, c] \mapsto \mathbb{R}$ defined by

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

and

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

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 2.7.1. 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 2.7.1 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

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

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 $j = 0, \cdots, 2k-1$. ■

Now we are able to state the main result of this chapter:

Theorem 2.7.2 *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 the conjectured Lemma 2.5.4, then*

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

and

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

where

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

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

Proof. For the direct problems, we apply Lemma 2.7.4 at $t = 0$ together with the fact that for $j = 0, \cdots, k-1$,

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

and

$$(\widehat{H}_n^l)^{k+j}(0) - c_j(g_0) n^{(k-j)/(2k+1)} (\hat{g}_n(x_0) - g_0(x_0)) \to_p 0 \quad \text{as } n \to \infty$$

which follow from the respective definitions of $\tilde{H}_n^l$ and $\widehat{H}_n^l$, and also strong consistency of the MLE (for $\widehat{H}_n^l$). For the inverse problem, the claim follows from Lemma 2.7.4 and the inverse formula in (2.3). ∎

Chapter 3

LIMITING PROCESSES: INVELOPES AND ENVELOPES

3.1 Introduction

In the previous chapter, it is claimed that the limiting distribution of the MLE and LSE and their derivatives involves a particular stochastic process H_k. This chapter is completely devoted to proving the existence of such a process. If W is two-sided Brownian motion starting at 0 and k is an integer greater or equal to 1, we define Y_k as the $(k-1)$ fold integral of $W + (k!/(2k)!)t^{2k}$. The process H_k is characterized by: (i) H_k stays above (below) Y_k if k is even (odd), (ii) H_k is $2k$-convex; i.e., $H_k^{(2k-2)}$ exists and convex and (iii) H_k touches Y_k if $H_k^{(2k-2)}$ changes its slope, (iv) $\lim_{|t|\to\infty}(H_k^{(2j)}(t) - Y_k^{(2j)}(t)) = 0$ for $j = 0, \cdots, (k-2)/2$, if k is even, and $\lim_{t\to\infty}(H_k(t) - Y_k(t)) = 0$, $\lim_{|t|\to\infty}(H_k^{(2j+1)}(t) - Y_k^{(2j+1)}(t)) = 0$ for $j = 0, \cdots, (k-3)/2$ if k is odd. In the particular cases $k = 1$ and 2, it takes only a change of scale to see that the processes H_1 and H_2 are very closely related to the greatest convex minorant of $W + t^2$ (Groeneboom (1985), Groeneboom (1989)) and to the "invelope" of the first integral of $W + t^4$ (Groeneboom, Jongbloed, and Wellner (2001a)) respectively. To have more intuition about the process H_k, one might think first about the drift $(k!/(2k)!)t^{2k}$ as the k-fold integral of the "canonical" function t^k. We can then define the following Gaussian problem:

$$dX_k(t) = t^k dt + dW(t), \qquad t \in \mathbb{R}.$$

It is an estimation problem that goes in parallel with the original one where the k-monotone density g_0 is replaced by the k-convex function t^k and $dX_k(t)$ plays the role of the observed data $X_1, \cdots, X_n$. Note that the process Y_k is nothing but the k-fold integral of dX_k. How could we "estimate" t^k? As in the original problem of estimation of a k-monotone density, one can define a Least Squares problem whose solution would be the "closest" k-convex function in the L_2-norm to the function t^k plus Gaussian noise, on a finite interval $[-c, c]$. By construction, the process H_k is the limit (in an appropriate sense) of the k-fold integral of the LS solution, $H_{c,k}$ say, as $c \to \infty$.

As it was mentioned in the introdution, the process H_k is a random spline of degree $2k-1$ whose knots are exactly the points where it touches Y_k. This fact is certainly true for $k = 1$ (Groeneboom (1989)). However, it is still conjectured for $k \geq 2$. In the particular case $k = 2$, Groeneboom, Jongbloed, and Wellner (2001a) could only prove that the points of touch between H_k and Y_k form a set a Lebesgue measure 0 and conjectured that they are isolated.

The proof of existence and uniqueness of the process H_k relies heavily on showing the following fact: For any point $t \in (-c, c)$, if τ_c^- (τ_c^+) is the last (first) point of touch between $H_{c,k}$ and Y_k before (after) t, then $\tau_c^+ - \tau_c^- = O_p(1)$ as $c \to \infty$. This problem is very similar to the problem of determining the stochastic order of the distance between two knot points of the MLE or LSE, when these knots are in a small neighborhood of x_0. Our results show that the above "fact" is indeed true if the conjectured Lemma 2.5.4 holds.

3.2 The Main Result

Suppose that $k \geq 1$ and let W be a two-sided Brownian motion starting from 0 at 0. Define the Gaussian processes $\{Y_k(t) : t \in \mathbb{R}\}$ by

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

and set $X_k(t) \equiv Y_k^{(k-1)}(t) = W(t) + (k+1)^{-1} t^{k+1}$ for $t \in \mathbb{R}$. Thus

$$dX_k(t) = t^k dt + dW(t) \equiv f_{k,0}(t) dt + dW(t)$$

where $f_{k,0}$ is monotone for $k = 1$, convex for $k = 2$, and, for $k \geq 3$ the $(k-2)$-th derivative $f_{k,0}^{(k-2)}(t) = (k!/2) t^2$ is convex. Thus we can consider "estimation" of the function $f_{k,0}$ in Gaussian noise $dW(t)$ subject to the constraint of convexity of $f^{(k-2)}$ (or monotonicity of f in the case $k = 1$).

Here is our main result.

Theorem 3.2.1 *If the conjectured Lemma 2.5.4 holds, then for all $k \geq 1$, there exists an almost surely uniquely defined stochastic process H_k characterized by the four following conditions:*

(i) $(-1)^k(H_k(t) - Y_k(t)) \geq 0, \quad t \in \mathbb{R}.$

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

(iii) *For any $t \in \mathbb{R}$, $H_k(t) = Y_k(t)$ if and only if $H_k^{(2k-2)}$ changes slope at t; equivalently,*

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

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

Note that H_k is below Y_k for k odd (and hence is an "envelope"), while H_k lies above Y_k for k even (and hence is an "invelope", a term that was coined by Groeneboom, Jongbloed, and Wellner (2001a) to describe the situation in the case $k = 2$). One can view $H_k^{(k)} \equiv f_k$ as an "estimator" of $f_{k,0}$, and $H_k^{(k+j)}$ as estimators of $f_{k,0}^{(j)}$, $j = 1, \ldots, k-1$.

Note that in Chapter 2, Section 7, the drift term in the limiting process is equal to $(-1)^k\, (k!/(2k)!)\ t^{2k}$ and hence a slightly different version of Theorem 3.2.1 is needed:

Corollary 3.2.1 *Let $k \geq 1$ and suppose that Lemma 2.5.4 holds. If Z_k is the $(k-1)$-fold integral of two-sided Brownian motion $+\ (-1)^k\, (k!/(2k)!)\ t^{2k}$, then there exists an almost surely uniquely defined stochastic process G_k characterized by the four following conditions:*

(i) $G_k(t) \geq Z_k(t) \geq 0, \quad t \in \mathbb{R}.$

(ii) $(-1)^k G_k$ *is $2k$-convex.*

(iii) *For any* $t \in \mathbb{R}$, $G_k(t) = Z_k(t)$ *if and only if* $G_k^{(2k-2)}$ *changes slope at* t*; equivalently,*

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

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

Proof. Since for all $k \geq 1$, $(-1)^k W \stackrel{d}{=} W$, it follows that $(-1)^k Z_k \stackrel{d}{=} Y_k$, or $Z_k \stackrel{d}{=} (-1)^k Y_k$. From Theorem 3.2.1, it follows that the process $G_k =_{a.s.} (-1)^k H_k$ is almost surely uniquely defined by the conditions (i)-(iv) of Corollary 3.2.1. ■

Our proof of Theorem 3.2.1 proceeds along the general lines of the proof for the case $k = 2$ in Groeneboom, Jongbloed, and Wellner (2001a). We first establish the existence and give characterizations of processes $H_{c,k}$ on $[-c, c]$, we then show that these processes are tight and converge to the limit process H_k as $c \to \infty$. But there are a number of new difficulties and complications. For example, we have not yet found analogues of the "mid-point relations" given in Lemma 2.4 and Corollary 2.2 of Groeneboom, Jongbloed, and Wellner (2001a). Those arguments are replaced by new more general results involving perturbations by B-splines. Several of our key results for the general case involve the theory of splines as given in Nürnberger (1989) and DeVore and Lorentz (1993). Some of the arguments sketched in Groeneboom, Jongbloed, and Wellner (2001a) are given in more detail (and greater generality) here. Throughout the remainder of this Chapter we assume that the conjectured Lemma 2.5.4 holds. The tightness claims in this Chapter are all dependent of the validity of Lemma 2.5.4.

This chapter is organized as follows: In section 3 we establish existence and give characterizations of processes $H_{c,k}$ on compact intervals $[-c, c]$ as solutions of certain minimization problems that can be viewed in terms of "estimation" of the "canonical" $k-$convex function t^k and its derivatives in Gaussian white noise $dW(t)$. These problems are slightly different for k even and k odd due to the different boundary conditions involved, and hence are treated separately for even and odd k's. In section 4 we establish tightness of the processes

$H_{c,k}$ and derivatives $H_{c,k}^{(j)}$ for $j \in \{1, \ldots, 2k-1\}$ as $c \to \infty$. These arguments rely on the crucial fact that two successive changes of slope τ_c^+ and τ_c^- of $H_{c,k}^{(2k-2)}$ to the right and left of a fixed point t satisfy $\tau_c^+ - t = O_p(1)$ and $t - \tau_c^- = O_p(1)$ as $c \to \infty$. In section 5 we combine the results from sections 3 and 4 to complete the proof of Theorem 3.2.1.

3.3 The processes $H_{c,k}$ on $[-c, c]$

To prepare for the proof of Theorem 3.2.1, we first consider the problem of minimizing the criterion function

$$\Phi_c(f) = \frac{1}{2}\int_{-c}^{c} f^2(t)dt - \int_{-c}^{c} f(t)dX_k(t) \tag{3.1}$$

over the class of k-convex functions on $[-c, c]$ and which satisfy two different sets of boundary conditions depending on the parity of k. We will start by considering the case k even, $k > 2$.

3.3.1 Existence and Characterization of $H_{c,k}$ for k even

Throughout this subsection k is assumed to be an even integer, $k > 2$ (since the case $k = 2$ is covered by Groeneboom, Jongbloed, and Wellner (2001a)). Let $c > 0$ and $\underline{m}_1$ and $\underline{m}_2 \in \mathbb{R}^l$, where $k = 2l$. Consider the problem of minimizing Φ_c over $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$ the class of k-convex functions satisfying

$$(f^{(k-2)}(-c), \cdots, f^{(2)}(-c), f(-c)) = \underline{m}_1 \quad \text{and} \quad (f^{(k-2)}(c), \cdots, f^{(2)}(c), , f(c)) = \underline{m}_2.$$

Proposition 3.3.1 *The functional Φ_c admits a unique minimizer in $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$.*

We preface the proof of the proposition by the following lemma:

Lemma 3.3.1 *Let g be a convex function defined on $[0, 1]$ such that $g(0) = k_1$ and $g(1) = k_2$ where k_1 and k_2 are arbitrary real constants. If there exists $t_0 \in (0, 1)$ such that $g(t_0) < -M$, then $g(t) < -M/2$ on the interval $[t_L, t_U]$ where*

$$t_L = \frac{k_1 + M/2}{k_1 + M} t_0, \quad t_U = \frac{(k_2 + M/2)t_0 + M/2}{k_2 + M}.$$

Proof. Since g is convex, it is below the chord joining the points $(0, k_1)$ and $(t_0, -M)$ and the chord joining the points $(t_0, -M)$ and $(1, k_2)$. We can easily verify that these chords intercept the horizontal line $y = -M/2$ at the points $(t_L, -M/2)$ and $(t_U, -M/2)$ where t_L and t_U are the ones defined in the lemma. ■

Proof of Proposition 3.3.1 We first prove that we can restrict ourselves to the class of functions

$$\mathcal{C}_{k,\underline{m}_1,\underline{m}_2,M} = \left\{ f \in \mathcal{C}_{k,\underline{m}_1,\underline{m}_2}, f^{(k-2)} > -M \right\}$$

for some $M > 0$. Without loss of generality, we assume that $f^{(k-2)}(-c) \geq f^{(k-2)}(c)$; i.e., $m_{1,1} \geq m_{1,2}$. Now, by integrating $f^{(k-2)}$ twice ($k \geq 4$), we have

$$f^{(k-4)}(x) = \int_{-c}^{x} (x-s) f^{(k-2)}(s) ds + \alpha_1 (x+c) + \alpha_0, \tag{3.2}$$

where

$$\alpha_0 = f^{(k-4)}(-c) = m_{1,2}$$

and

$$\begin{aligned} \alpha_1 &= \left(f^{(k-4)}(c) - f^{(k-4)}(-c) - \int_{-c}^{c} (c-s) f^{(k-2)}(s) ds \right) / (2c) \\ &= \left(m_{2,2} - m_{1,2} - \int_{-c}^{c} (c-s) f^{(k-2)}(s) ds \right) / (2c). \end{aligned}$$

Using the change of variable $x = (2t-1)c$, $t \in [0,1]$, and denoting

$$d_{k-2}(t) = f^{(k-2)}\left((2t-1)c\right) - m_{1,1}$$

we can write, for all $t \in [0,1]$

$$\begin{aligned} & f^{(k-4)}\left((2t-1)c\right) \\ &= (2c)^2 \left(\int_0^t (t-s) d_{k-2}(s) ds - t \int_0^1 (1-s) d_{k-2}(s) ds \right) \\ &+ (2c)^2 m_{1,1} \left(\int_0^t (t-s) ds - t \int_0^1 (1-s) ds \right) + (m_{2,2} - m_{1,2}) t + m_{1,2} \\ &= (2c)^2 \left((t-1) \int_0^t s\, d_{k-2}(s) ds - t \int_t^1 (1-s) d_{k-2}(s) ds \right) \\ &+ (2c)^2 m_{1,1} \left(\frac{t^2 - t}{2} \right) + (m_{2,2} - m_{1,2}) t + m_{1,2}. \end{aligned}$$

If there exists $x_0 \in [-c, c]$ such that $-3M/2 + m_{1,1} < f^{(k-2)}(x_0) < -M + m_{1,1}$ for $M > 0$ large, then $-3M/2 < d_{k-2}(t_0) < -M$ where $x_0 = (2t_0 - 1)c$. Let t_L and t_U be the same numbers defined in Lemma 3.3.1.

Now, since $d_{k-2} \leq 0$ on $[0,1]$ (recall that it was assumed that $f^{(k-2)}(-c) > f^{(k-2)}(c)$), we have for all $0 \leq t \leq 1$

$$f^{(k-4)}\left((2t-1)c\right) \geq (2c)^2 m_{1,1}\left(\frac{t^2-t}{2}\right) + (m_{2,2} - m_{1,2})t + m_{1,2}$$

and in particular, if $t \in [t_L, t_U]$, we have

$$\begin{aligned} f^{(k-4)}\left((2t-1)c\right) &\geq (2c)^2(1-t)\int_{t_L}^{t} s\,(-d_{k-2})(s)ds \\ &\quad + (2c)^2 m_{1,1}\left(\frac{t^2-t}{2}\right) + (m_{2,2}-m_{1,2})t + m_{1,2} \\ &\geq \frac{M(2c)^2}{2}(1-t)\int_{t_L}^{t} s\;ds + (2c)^2 m_{1,1}\left(\frac{t^2-t}{2}\right) \\ &\quad + (m_{2,2}-m_{1,2})t + m_{1,2} \\ &= \frac{M(2c)^2}{4}(1-t)(t^2-t_L^2) + (2c)^2 m_{1,1}\left(\frac{t^2-t}{2}\right) \\ &\quad + (m_{2,2}-m_{1,2})t + m_{1,2}. \end{aligned} \tag{3.3}$$

Hence, if $k = 4$, this implies that $\int_{t_L}^{t_U} f^2\left((2t-1)c\right)dt$ is of the order of M^2. In fact, if M is chosen to be large enough so that the term in (3.3) is positive for all $t \in [t_L, t_U]$, it is easy to establish that, using the fact that $1 - t \geq 1 - t_U$ and $t + t_L \geq 2t_L$

$$\int_{t_L}^{t_U} f^2\left((2t-1)c\right)dt \geq \alpha_2 M^2 + \alpha_1 M$$

where

$$\alpha_2 = c^4(1-t_U)^2(2t_L)^2(t_U-t_L)^3/3,$$

and

$$\begin{aligned} \alpha_1 &= \frac{1}{2}\Bigg(\frac{m_{1,1}(2c)^2}{2}\int_{t_L}^{t_U}(1-t)(t^2-t_L^2)(t^2-t)dt \\ &\quad + (m_{2,2}-m_{1,2})\int_{t_L}^{t_U} t(1-t)(t^2-t_L^2)dt + m_{1,2}\int_{t_L}^{t_U}(1-t)(t^2-t_L^2)dt\Bigg). \end{aligned}$$

But α_2 does not vanish as $M \to \infty$ since $t_L \to t_0/2$, $t_U \to (t_0+1)/2$ and $t_U - t_L \to 1/2$. Therefore, for $k = 4$, if there exists x_0 such that $f^{(2)}(x_0) < -M$, then we can find real constants $c_2 > 0$, c_1 and c_0 such that

$$\begin{aligned} \Phi_c(f) &= \frac{1}{2}\int_{-c}^{c} f^2(t)dt - \int_{-c}^{c} f(t)dX_4(t) \\ &\geq c\int_{t_L}^{t_U} f^2\left((2t-1)c\right)dt - \int_{-c}^{c} f(t)dX_4(t) \qquad (3.4) \\ &\geq c_2M^2 + c_1M + c_0, \end{aligned}$$

since the second term in (3.4) is of the order of M. Indeed, using integration by parts, we can write

$$\int_{-c}^{c} f(t)dX_4(t) = X_4(c)f(c) - X_4(-c)f(-c) - \int_{-c}^{c} f'(t)X_4(t)dt$$

where for all $t \in (-c, c)$

$$f'(t) = \int_{-c}^{t} f^{(2)}(s)ds + \left(m_{2,2} - m_{1,2} - \int_{-c}^{c}(c-s)f^{(2)}(s)ds\right)/(2c).$$

Hence,

$$\begin{aligned} |f'(t)| &\leq \frac{3M}{2}\int_{-c}^{t} ds + \left(|m_{2,2} - m_{1,2}| + \frac{3M}{2}\int_{-c}^{c}(c-s)ds\right)/(2c) \\ &\leq 6M\,c + \frac{|m_{2,2} - m_{1,2}|}{2c} \end{aligned}$$

and

$$\left|\int_{-c}^{c} f(t)dX_4(t)\right| \leq (12Mc + |m_{2,2} - m_{1,1}| + |m_{1,2}| + |m_{2,2}|)\sup_{[-c,c]}|X_4(t)|.$$

This implies that the functions in $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$ have to be bounded in order to be possible candidates for the minimization problem.

Suppose now that $k > 4$. In order to reach the same conclusion, we are going to show that in this case too, there exist constants $c_2 > 0$, c_1, and c_0 such that

$$\frac{1}{2}\int_{-c}^{c} f^2(t)dt - \int_{-c}^{c} f(t)dX_k(t) \geq c_2M^2 + c_1M + c_0.$$

For this purpose we use induction. Suppose that for $2 \leq j < k/2$, there exists a polynomial $P_{1,j}$ whose coefficients depend only on c and the first j components of $\underline{m}_1$ and $\underline{m}_2$ such that we have for all $t \in [0,1]$

$$(-1)^j f^{(k-2j)}\left((2t-1)c\right) \geq P_{1,j}(t),$$

and suppose that there exists a polynomial Q_j depending only on t_L and c such that $Q_j > 0$ on (t_L, t_U) and lastly $P_{2,j}$ a polynomial whose coefficients depend on t_L, c and the first j components of $\underline{m}_1$ and $\underline{m}_2$ such that for all $t \in [t_L, t_U]$, we have

$$(-1)^j f^{(k-2j)}\left((2t-1)c\right) \geq MQ_j(t) + P_{2,j}(t).$$

By integrating $f^{(k-2j)}$ twice, we have

$$f^{(k-2j-2)}(x) = \int_{-c}^{x} (x-s) f^{(k-2j)}(s)ds + \alpha_{1,j}(x+c) + \alpha_{0,j},$$

where

$$\alpha_{0,j} = f^{(k-2j-2)}(-c) = m_{1,j+1}$$

and

$$\begin{aligned}
\alpha_{1,j} &= \left(f^{(k-2j-2)}(c) - f^{(k-2j-2)}(-c) - \int_{-c}^{c}(c-s)f^{(k-2j-2)}(s)ds\right)/(2c) \\
&= \left(m_{2,j+1} - m_{1,j+1} - \int_{-c}^{c}(c-s)f^{(k-2j-2)}(s)ds\right)/(2c).
\end{aligned}$$

For $2 \leq j < k/2$, we denote

$$d_{k-2j}(t) = f^{(k-2j)}\left((2c-1)t\right), \quad \text{for } t \in [0,1].$$

By the same change of variable we used before, we can write for all $t \in [0,1]$

$$\begin{aligned}
&(-1)^j f^{(k-2j-2)}(c(2t-1)) \\
&= (2c)^2\left(\int_0^t (t-s)(-1)^j d_{k-2j}(s)ds - t\int_0^1 (1-s)(-1)^j d_{k-2j}(s)ds\right) \\
&+ (m_{2,j+1} - m_{1,j+1})t + m_{1,j+1} \\
&= (2c)^2\left((t-1)\int_0^t s(-1)^j d_{k-2j}(s)ds - t\int_t^1 (1-s)(-1)^j d_{k-2j}(s)ds\right) \\
&+ (m_{2,j+1} - m_{1,j+1})t + m_{1,j+1}.
\end{aligned}$$

Hence, by using the induction hypothesis, we have for all $t \in [0,1]$

$$\begin{aligned}(-1)^j f^{(k-2j-2)}\left((2t-1)c\right) \quad &\leq \quad (2c)^2 \left((t-1)\int_0^t sP_{1,j}(s)ds - t\int_t^1 (1-s)P_{1,j}(s)ds\right) \\ &\quad + (m_{2,j+1} - m_{1,j+1})t + m_{1,j+1}\end{aligned}$$

which is equivalent to

$$\begin{aligned}(-1)^{j+1} f^{(k-2j-2)}\left((2t-1)c\right) \quad &\geq \quad (2c)^2 \left((1-t)\int_0^t sP_{1,j}(s)ds + t\int_t^1 (1-s)P_{1,j}(s)ds\right) \\ &\quad - (m_{2,j+1} - m_{1,j+1})t - m_{1,j+1} = P_{1,j+1}(t),\end{aligned}$$

and if $t \in [t_L, t_U]$

$$\begin{aligned}&(-1)^j f^{(k-2j-2)}\left((2t-1)c\right) \\ &\leq \quad (2c)^2\Bigg((t-1)\int_0^{t_L} sP_{1,j}(s)ds + (t-1)\int_{t_L}^t s(MQ_j(s) + P_{2,j}(s))ds \\ &\qquad - t\int_t^1 (1-s)P_{1,j}(s)ds\Bigg) + (m_{2,j+1} - m_{1,j+1})t + m_{1,j+1}.\end{aligned}$$

This can be rewritten

$$\begin{aligned}(-1)^{j+1} f^{(k-2j-2)}\left((2t-1)c\right) \quad &\geq \quad (2c)^2\Bigg(M(1-t)\int_{t_L}^t sQ_j(s)ds + (1-t)\int_0^{t_L} sP_{1,j}(s)ds \\ &\quad + (1-t)\int_{t_L}^t P_{2,j}(s)ds + t\int_t^1 (1-s)P_{1,j}(s)ds\Bigg) \\ &\quad - (m_{2,j+1} - m_{1,j+1})t - m_{1,j+1} \\ &= \quad MQ_{j+1}(t) + P_{2,j+1}(t),\end{aligned}$$

where $P_{1,j+1}$, $P_{1,j+1}$ and Q_{j+1} satisfy the same properties assumed in the induction hypothesis. Therefore, there exist two polynomials P and Q such that for all $t \in [t_L, t_U]$,

$$(-1)^{k/2} f\left((2t-1)c\right) \geq MQ(t) + P(t)$$

and $Q > 0$ on (t_L, t_U). Thus, for M chosen large enough

$$\Phi_c(f) \geq M^2 \int_{t_L}^{t_U} Q^2(t)dt + O_p(M)$$

since it can be shown using induction and similar arguments as for the case $k = 4$ that

$$\left|\int_{-c}^{c} f(t)dX_k(t)\right| = O_p(M).$$

We conclude that there exists some $M > 0$ such that we can restrict ourselves to the space $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2,M}$ while searching for the minimizer of Φ_c.

Let us endow the space $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2,M}$ with the distance

$$d(g,h) = \|g^{(k-2)} - h^{(k-2)}\|_\infty = \sup_{t\in[-c,c]} |g^{(k-2)}(t) - h^{(k-2)}(t)|.$$

d is indeed a distance since $d(g,h) = 0$ if an only if $g^{(k-2)}$ and $h^{(k-2)}$ are equal on $[-c,c]$ and hence $g = h$ using the boundary conditions; i.e., $g^{(k-2p)}(\pm c) = h^{(k-2p)}(\pm c)$, for $2 \le p \le k/2$.

Consider a sequence $(f_n)_n$ in $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2,M}$. Denote

$$g_n = f_n^{(k-2)}.$$

Since $(g_n)_n$ is uniformly bounded and convex on the interval $[-c,c]$, there exists a subsequence $(g_k)_k$ of $(g_n)_n$ and a convex function g such that $g(-c) = m_{1,1}$, $g(c) = m_{2,1}$, $g \ge -M$ and $(g_k)_k$ converges uniformly to g on $[-c,c]$ (e.g. Roberts and Varberg (1973), pages 17 and 20). Define f as the $(k-2)$-fold integral of the limit g that satisfies $f^{(k-4)}(-c) = m_{1,2}, \cdots, f(-c) = m_{1,k-2}$ and $f^{(k-4)}(c) = m_{2,2}, \cdots, f(c) = m_{2,k-2}$. Then, f belongs to $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2,M}$ and

$$d(f_k, f) \to 0, \quad \text{as } k \to \infty.$$

Thus, the space $\left(\mathcal{C}_{k,\underline{m}_1,\underline{m}_2,M}, d\right)$ is compact. It remains to show now that Φ_c is continuous with respect to d and that the minimizer is unique. Fix a small $\epsilon > 0$ and consider f and g two elements in $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2,M}$.

$$\begin{aligned} |\Phi_c(g) - \Phi_c(f)| &= \left| \frac{1}{2}\int_{-c}^{c} \left(g^2(t) - f^2(t)\right) dt - \int_{-c}^{c} (g(t) - f(t))\, dX_k(t) \right| \\ &\le \frac{1}{2}\left| \int_{-c}^{c} \left(g^2(t) - f^2(t)\right) dt \right| + \left| \int_{-c}^{c} (g(t) - f(t))\, dX_k(t) \right|. \end{aligned}$$

Suppose that $k = 4$. By using the expression obtained in (3.2), we can write

$$g(t) - f(t) = \int_{-c}^{t} (t-s)\left(g^{(2)}(s) - f^{(2)}(s)\right) ds + \alpha_1 (t+c), \qquad t \in [-c,c]$$

where

$$\alpha_1 = -\int_{-c}^{c} (c-s)\left(g^{(2)}(s) - f^{(2)}(s)\right) ds/(2c)$$

since $f(\pm c) = g(\pm c)$ and $f^{(2)}(\pm c) = g^{(2)}(\pm c)$. Therefore, for all $t \in [-c, c]$, we have

$$\begin{aligned}
|g(t) - f(t)| &\leq \left(\int_{-c}^{t} (t-s)ds\right) d(f,g) + \left(\frac{\int_{-c}^{c}(c-s)ds}{2c}\right)(t+c)d(f,g) \\
&= \left(\frac{(t+c)^2}{2} + \frac{(2c)^2}{2}\frac{(t+c)}{2c}\right) d(f,g) \\
&\leq \left(\frac{(2c)^2}{2} + \frac{(2c)^2}{2}\right) d(f,g) \\
&= (2c)^2 d(f,g).
\end{aligned}$$

Also, we obtain using the same expression

$$\begin{aligned}
|f(t)| &\leq \left(\int_{-c}^{t} (t-s)ds + \int_{-c}^{c} (c-s)ds\right) \max\left(|m_{1,1}|, |m_{2,1}|, M\right) + |m_{1,2}| + |m_{2,2}| \\
&\leq 4\,c^2 \max\left(|m_{1,1}|, |m_{2,1}|, M\right) + |m_{1,2}| + |m_{2,2}|
\end{aligned}$$

for all $t \in [-c, c]$ and the same inequality holds for g. By denoting

$$K_0 = 4\,c^2 \max\left(|m_{1,1}|, |m_{2,1}|, M\right) + |m_{1,2}| + |m_{2,2}|,$$

it follows that

$$\begin{aligned}
\frac{1}{2}\left|\int_{-c}^{c} \left(g^2(t) - f^2(t)\right) dt\right| &\leq \frac{1}{2}\int_{-c}^{c} |g(t) + f(t)| \cdot |g(t) - f(t)| dt \\
&\leq K_0 \int_{-c}^{c} |g(t) - f(t)| dt \\
&\leq (2c) K_0 \sup_{t \in [-c,c]} |g(t) - f(t)| \\
&\leq (2c)^3 K_0 \;\; d(f,g).
\end{aligned} \tag{3.5}$$

Now, using integration by parts and again the fact that $f(\pm c) = g(\pm c)$, we can write

$$\int_{-c}^{c} \left(g(t) - f(t)\right) dX_k(t) = -\int_{-c}^{c} \left(g'(t) - f'(t)\right) X_k(t) dt \tag{3.6}$$

But,

$$\left(g'(t) - f'(t)\right) - \left(g'(-c) - f'(-c)\right) = \int_{-c}^{t} \left(g^{(2)}(s) - f^{(2)}(s)\right) ds \tag{3.7}$$

for all $t \in [-c, c]$. On the other hand, we obtain using integration by parts

$$-\int_{-c}^{c} (c-s)\left(g^{(2)}(s) - f^{(2)}(s)\right) ds/(2c) = g'(-c) - f'(-c). \tag{3.8}$$

By the triangle inequality, we obtain

$$\begin{aligned}
|g'(t)-f'(t)| &\leq |g'(-c)-f'(-c)| + \int_{-c}^{t} |g^{(2)}(s)-f^{(2)}(s)|ds \\
&\leq \int_{-c}^{c} (c-s)|g^{(2)}(s)-f^{(2)}(s)|ds/(2c) + \int_{-c}^{t} |g^{(2)}(s)-f^{(2)}(s)|ds \\
&\leq \frac{2c}{2}\ d(f,g) + (t+c)d(f,g) \\
&\leq \left(\frac{2c}{2} + 2c\right) d(f,g) \\
&= (3c)\ d(f,g). \qquad (3.9)
\end{aligned}$$

Combining (3.5) and (3.9), it follows that

$$|\Phi_c(g) - \Phi_c(f)| \leq \left((2c)^3 K_0 + (3c) \int_{-c}^{c} |X_k(t)|dt \right)\ d(f,g).$$

Now, let $k > 4$ be an even integer. We have

$$g^{(k-4)}(t) - f^{(k-4)}(t) = \int_{-c}^{t} (t-s) \left(g^{(k-2)}(s) - f^{(k-2)}(s) \right) ds + \alpha_1 (t+c), \qquad t \in [-c,c]$$

where

$$\alpha_1 = -\int_{-c}^{c} (c-s) \left(g^{(k-2)}(s) - f^{(k-2)}(s) \right) ds/(2c)$$

we obtain, applying the same techniques used for $k = 4$, that

$$\left| g^{(k-4)}(t) - f^{(k-4)}(t) \right| \leq (2c)^2\ d(f,g), \qquad t \in [-c,c].$$

By induction and using the fact that for $j = 3, \cdots, k/2$

$$g^{(k-2j)}(t) - f^{(k-2j)}(t) = \int_{-c}^{t} (t-s) \left(g^{(k-2j+2)}(s) - f^{(k-2j+2)}(s) \right) ds + \alpha_{1,j}(t+c),$$

for $t \in [-c,c]$ where

$$\alpha_{1,j} = -\int_{-c}^{c} (c-s) \left(g^{(k-2j+2)}(s) - f^{(k-2j+2)}(s) \right) ds/(2c),$$

it follows that

$$\sup_{t\in[-c,c]} |g^{(k-2j)}(t) - f^{(k-2j)}(t)| \leq (2c)^{2j-2}\ d(f,g),$$

and in particular

$$\sup_{t\in[-c,c]} |g(t) - f(t)| \leq (2c)^{k-2} \ d(f,g).$$

Now, notice that the identities in (3.6), (3.7), (3.8), and the inequality in (3.9) continue to hold. It follows that there exist constants $K_{k-2j} > 0, j = 2, \cdots, k/2$ such that for all $t \in [-c, c]$

$$|f^{(k-2j)}(t)|, |g^{(k-2j)}(t)| \leq K_{k-2j}$$

where for $j = 3, \cdots, k/2$

$$K_{k-2j} \leq 4\, c^2 K_{k-2j+2} + |m_{2,j} - m_{1,j}| + |m_{1,j}|.$$

On the other hand, we have

$$\begin{aligned}
|g'(t) - f'(t)| &\leq |g'(-c) - f'(-c)| + \int_{-c}^{t} |g^{(2)}(s) - f^{(2)}(s)|ds \\
&\leq \int_{-c}^{c} (c-s)|g^{(2)}(s) - f^{(2)}(s)|ds/(2c) + \int_{-c}^{t} |g^{(2)}(s) - f^{(2)}(s)|ds \\
&\leq \frac{2c}{2}(2c)^{k-4} \ d(f,g) + (t+c)(2c)^{k-4} \ d(f,g) \\
&\leq \left(\frac{(2c)^{k-3}}{2} + (2c)^{k-3} \right) d(f,g) \\
&= \frac{3}{2}(2c)^{k-3} \ d(f,g)
\end{aligned}$$

and hence

$$|\Phi_c(g) - \Phi_c(f)| \leq \left((2c)^{k-1} K_0 + (3/2)(2c)^{k-3} \int_{-c}^{c} |X_k(t)|dt \right) \ d(f,g).$$

We conclude that the functional Φ_c admits a minimizer in the class $\mathcal{C}_{\underline{m}_1,\underline{m}_2,M}$ and hence in $\mathcal{C}_{\underline{m}_1,\underline{m}_2}$. This minimizer is unique by the strict convexity of Φ_c. ■

The next proposition gives a characterization of the minimizer.

Proposition 3.3.2 *The function $f_{c,k} \in \mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$ is the minimizer of Φ_c if and only if*

$$H_{c,k}(t) \geq Y_k(t), \qquad t \in [-c, c], \tag{3.10}$$

and

$$\int_{-c}^{c} (H_{c,k}(t) - Y_k(t))\, df_{c,k}^{(k-1)}(t) = 0, \tag{3.11}$$

where $H_{c,k}$ *is the k-fold integral of* $f_{c,k}$ *satisfying*

$$H_{c,k}(-c) = Y_k(-c), H_{c,k}^{(2)}(-c) = Y_k^{(2)}(-c), \cdots, H_{c,k}^{(k-2)}(-c) = Y_k^{(k-2)}(-c),$$

and

$$H_{c,k}(c) = Y_k(c), H_{c,k}^{(2)}(c) = Y_k^{(2)}(c), \cdots, H_{c,k}^{(k-2)}(c) = Y_k^{(k-2)}(c).$$

Our proof of Proposition 3.3.2 will use the following lemma.

Lemma 3.3.2 *Let* $t_0 \in [-c, c]$*. The probability that there exists a polynomial* P *of degree* k *such that*

$$P(t_0) = Y_k(t_0),\ \ P'(t_0) = Y_k'(t_0),\ \ \cdots,\ \ P^{(k-1)}(t_0) = Y_k^{(k-1)}(t_0) \tag{3.12}$$

and satisfies $P \geq Y_k$ *or* $P \leq Y_k$ *in a small neighborhood of* t_0 *(right (resp. left) neighborhood if* $t_0 = -c$ *(resp.* $t_0 = c$*)) is equal to* 0*.*

Proof. Without loss of generality, we assume that $0 \leq t_0 < c$. As a consequence of Blumenthal's 0-1 law and the Markov property of a Brownian motion, the probability that a straight line intercepting a Brownian motion W at the point $(t_0, W(t_0))$ is above or below W in a neighborhood of t_0 is equal to 0 since W crosses the horizontal line $y = W(t_0)$ infinitely many times in such neighborhood with probability 1 (see e.g. Durrett (1984), (5), page 14). Suppose that there exist $\delta > 0$ and a polynomial P satisfying the condition in (3.12) and $P(t) \geq Y_k(t)$ for all $t \in [t_0, t_0 + \delta]$ (the case $P \leq Y_k$ can be handled similarly). Denote $\Delta = P - Y_k$. Using the condition in (3.12) and successive integrations by parts, we can establish for all $t \in \mathbb{R}$ the identity

$$P(t) - Y_k(t) = \int_{t_0}^{t} \frac{(t-s)^{k-2}}{(k-2)!} \Delta^{(k-1)}(s) ds.$$

Moreover, we have for all $t \in [t_0, t_0 + \delta]$

$$\int_{t_0}^{t} \frac{(t-s)^{k-2}}{(k-2)!} \Delta^{(k-1)}(s) ds \geq 0. \tag{3.13}$$

This implies that there exists a subinterval $[t_0+\delta_1, t_0+\delta_2] \subset [t_0, t_0+\delta]$ such that

$$\Delta^{(k-1)}(t) = P^{(k-1)}(t) - Y_k^{(k-1)}(t) \geq 0, \qquad t \in [t_0+\delta_1, t_0+\delta_2] \tag{3.14}$$

since otherwise, the integral in (3.13) would be strictly negative. But a polynomial P of degree k satisfying (3.12) can be written as

$$P(t) = Y_k(t_0) + Y_k'(t_0)(t-t_0) + \cdots + Y_k^{(k-1)}(t_0)\frac{(t-t_0)^{k-1}}{(k-1)!} + P^{(k)}(t_0)\frac{(t-t_0)^k}{k!},$$

and therefore, it follows from the inequality in (3.14) that

$$Y_k^{(k-1)}(t_0) + P^{(k)}(t_0)(t-t_0) \geq Y_k^{(k-1)}(t), \qquad t \in [t_0+\delta_1, t_0+\delta_2]\,,$$

or equivalently

$$W(t_0) + \frac{1}{k+1}t_0^{k+1} + P^{(k)}(t_0)(t-t_0) \geq W(t) + \frac{1}{k+1}t^{k+1}, \qquad t \in [t_0+\delta_1, t_0+\delta_2].$$

The latter event occurs with probability 0 since the law of the process $\{W(t)+\frac{t^{k+1}}{k+1} : t \in [0,c])$ is equivalent to the law of the Brownian motion process $\{W(t) :\ t \in [0,c]\}$, and the result follows. ∎

Proof of Proposition 3.3.2. Let $f_{c,k}$ be a function in $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$ satisfying (3.10) and (3.11). To avoid conflicting notations, we replace $f_{c,k}$ by f. For an arbitrary function g in $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$, we have

$$g^2 - f^2 = (g-f)^2 + 2f(g-f) \geq 2f(g-f), \tag{3.15}$$

and therefore

$$\Phi_c(g) - \Phi_c(f) \geq \int_{-c}^{c} f(t)\,(g(t)-f(t))\,dt - \int_{-c}^{c} (g(t)-f(t))\,dX_k(t)\,.$$

Using the fact that $H_{c,k}^{(j)}$ is the $(k-j)$-fold integral of f for $j = 1, \cdots, k$,

$$g^{(2i)}(\pm c) = f^{(2i)}(\pm c), \quad \text{for } \ i = 0, \cdots, (k-2)/2$$

and

$$H_{c,k}^{(2j)}(\pm c) = Y_k^{(2j)}(\pm c), \qquad \text{for } \ j = 0, \cdots, (k-2)/2\,,$$

we obtain, using successive integrations by parts,

$$
\begin{aligned}
&\int_{-c}^{c} f(t)\,(g(t)-f(t))\,dt - \int_{-c}^{c} (g(t)-f(t))\,dX_k(t) \\
&= \Big[\Big(H_{c,k}^{(k-1)}(t) - Y_k^{(k-1)}(t)\Big)\,(g(t)-f(t)) \Big]_{-c}^{c} \\
&\quad - \int_{-c}^{c} \Big(H_{c,k}^{(k-1)}(t) - Y_k^{(k-1)}(t)\Big)\,\big(g'(t)-f'(t)\big)\,dt \\
&= -\int_{-c}^{c} \Big(H_{c,k}^{(k-1)}(t) - Y_k^{(k-1)}(t)\Big)\,\big(g'(t)-f'(t)\big)\,dt \\
&= -\Big[\Big(H_{c,k}^{(k-2)}(t) - Y_k^{(k-2)}(t)\Big)\,(g'(t)-f'(t)) \Big]_{-c}^{c} \\
&\quad + \int_{-c}^{c} \Big(H_{c}^{(k-2)}(t) - Y_k^{(k-2)}(t)\Big)\,\big(f''(t)-f_c''(t)\big)\,dt \\
&= \int_{-c}^{c} \Big(H_{c,k}^{(k-2)}(t) - Y_k^{(k-2)}(t)\Big)\,\big(g''(t)-f''(t)\big)\,dt \\
&\ \vdots \\
&= \int_{-c}^{c} (H_{c,k}(t) - Y_k(t))\,\Big(dg^{(k-1)}(t) - df^{(k-1)}(t)\Big)
\end{aligned}
$$

which yields, using the condition in (3.11),

$$
\begin{aligned}
&\int_{-c}^{c} f(t)\,(g(t)-f(t))\,dt - \int_{-c}^{c} (g(t)-f(t))\,dX_k(t) \\
&= \int_{-c}^{c} (H_{c,k}(t) - Y_k(t))\,dg^{(k-1)}(t).
\end{aligned}
$$

Using condition (3.10) and the fact that $g^{(k-1)}$ is nondecreasing, we conclude that

$$\Phi_c(g) \geq \Phi_c(f).$$

Since g was arbitrary, f is the minimizer. In the previous proof, we used implicitly the fact that $f^{(k-1)}$ and $g^{(k-1)}$ exist at $-c$ and c. Hence, we need to check that such an assumption can be made. First, notice that with probability 1, there exists $j \in \{1, \cdots, k-1\}$ such that $H_{c,k}^{(j)}(c) \neq Y_k^{(j)}(c)$. If such a j does not exist, it will follow that there exists a polynomial P of degree k such that

$$P^{(i)}(c) = Y_k^{(i)}(c), \quad \text{for } i = 0, \cdots, k-1$$

and $P(t) \geq Y_k(t)$, for t in a left neighborhood of c. Indeed, using Taylor expansion of $H_{c,k}$ at the point c, we have for some small $\delta > 0$ and $u \in [c-\delta, c)$

$$H_{c,k}(u)$$

$$
\begin{aligned}
&= H_{c,k}(c) + H'_{c,k}(c)(u-c) + \cdots + \frac{H^{(k-1)}_{c,k}(c)}{(k-1)!}(u-c)^{k-1} + \frac{H^{(k)}_{c,k}(c)}{k!}(u-c)^k \\
&\qquad + o((u-c)^k) \\
&= Y_k(c) + Y'_k(c)(u-c) + \cdots + \frac{Y^{(k-1)}_{k}(c)}{(k-1)!}(u-c)^{k-1} + \frac{H^{(k)}_{c,k}(c)}{k!}(u-c)^k \\
&\qquad + o((u-c)^k) \\
&\geq Y_k(u).
\end{aligned}
$$

Hence, there exists $\delta_0 > 0$ such that the polynomial P given by

$$
P(u) = Y_k(c) + Y'_k(c)(u-c) + \cdots + \frac{Y^{(k-1)}_{k}(c)}{(k-1)!}(u-c)^{k-1} + \frac{H^{(k)}_{c,k}(c)+1}{k!}(u-c)^k
$$

satisfies $P \geq Y_k$ on $[c-\delta_0, c)$. But by Lemma 3.3.2, we know that the probability of the latter event is equal to 0.

Consider j_0 the smallest integer in $\{1, \cdots, k-1\}$ such that $H^{(j_0)}_{c,k}(c) \neq Y^{(j_0)}_k(c)$. Notice first that j_0 has to be odd. Besides, since $H_{c,k} \geq Y_k$, $H^{(j_0)}_{c,k}(c) \neq Y^{(j_0)}_k(c)$ implies $H^{(j_0)}_{c,k}(c) < Y^{(j_0)}_k(c)$, and by continuity there exists a left neighborhood $[c-\delta, c)$ of c such that $H^{(j_0)}_{c,k}(t) < Y^{(j_0)}_k(t)$ for all $t \in [c-\delta, c)$. Hence, if we suppose that $g^{(k-1)}(t) \to \infty$ as $t \uparrow c$, where $g \in \mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$ then

$$
\int_{c-\delta}^{u} g^{(k-1)}(t)\left(H^{(j_0)}_{c,k}(t) - Y^{(j_0)}_k(t)\right)dt \to -\infty \quad \text{as} \quad u \uparrow c.
$$

Now, if $j_0 = k-1$ we have

$$
\begin{aligned}
&\int_{c-\delta}^{c} g^{(k-1)}(t)\left(H^{(k-1)}_{c,k}(t) - Y^{(k-1)}_k(t)\right)dt \\
&= \left[g^{(k-2)}(t)\left(H^{(k-1)}_{c,k}(t) - Y^{(k-1)}_k(t)\right)\right]_{c-\delta}^{c} - \int_{c-\delta}^{c} g^{(k-2)}(t)f(t)dt + \int_{c-\delta}^{c} g^{(k-2)}(t)dX_k(t)
\end{aligned}
$$

and hence

$$
\begin{aligned}
\lim_{u\uparrow c}\int_{c-\delta}^{c} g^{(k-1)}(t)(H_{c,k}(t) - Y_k(t))dt &= g^{(k-2)}(c)(H^{(k-1)}_{c,k}(c) - X_k(c)) \\
&\quad - g^{(k-2)}(c-\delta)(H^{(k-1)}_{c,k}(c-\delta) - X_k(c-\delta)) \\
&\quad - \int_{c-\delta}^{c} g^{(k-2)}(t)f(t)dt + \int_{c-\delta}^{c} g^{(k-2)}(t)dX_k(t) \\
&> -\infty.
\end{aligned}
$$

Therefore, when $t \uparrow c$, $g^{(k-1)}(t)$ converges to a finite limit and we can assume that $g^{(k-1)}(c)$ is finite. Using a similar arguments, we can show that $\lim_{t \downarrow -c} g^{(k-1)}(t) > -\infty$. The same conclusion is reached when $j_0 < k-1$.

Now, suppose that f minimizes Φ_c over $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$. Fix a small $\epsilon > 0$ and let $t \in (-c, c)$. We define the function $f_{t,\epsilon}$ on $[-c, c]$ by

$$\begin{aligned} f_{t,\epsilon}(u) &= f(u) + \epsilon \Bigg(\frac{(u-t)_+^{k-1}}{(k-1)!} + \alpha_{k-1} \frac{(u+c)^{k-1}}{(k-1)!} \\ &\qquad + \alpha_{k-3} \frac{(u+c)^{k-3}}{(k-3)!} + \cdots + \alpha_1 (u+c) \Bigg) \\ &= f(u) + \epsilon p_t(u) \end{aligned}$$

satisfying

$$p_t^{(2i)}(\pm c) = 0, \quad \text{for} \quad i = 0, \cdots, (k-2)/2. \tag{3.16}$$

For this choice of a perturbation function, we have for all $u \in [-c, c]$

$$f_{t,\epsilon}^{(k-2)}(u) = f^{(k-2)}(u) + \epsilon \left((u-t)_+ + \alpha_{k-1}(u+c) \right).$$

Thus, for any $\epsilon > 0$, $f_{t,\epsilon}^{(k-2)}$ is the sum of two convex functions and so it is convex. The condition (3.16) ensures that $f_{t,\epsilon}$ remains in the class $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$ and the parameters α_j, $j = 1, 3, \cdots, k-1$ are uniquely determined:

$$\begin{aligned} \alpha_{k-1} &= -\frac{(c-t)}{2c} \\ \alpha_{k-3} &= -\alpha_{k-1} \frac{(2c)^3}{3!} - \frac{(c-t)^3}{3!} \\ &\vdots \\ \alpha_1 &= -\alpha_{k-1} \frac{(2c)^{k-1}}{(k-1)!} - \cdots - \alpha_3 \frac{(2c)^3}{3!} - \frac{(c-t)^{k-1}}{(k-1)!}. \end{aligned}$$

Since f is the minimizer of Φ_c, we have

$$\lim_{\epsilon \searrow 0} \frac{\Phi_c(f_{\epsilon,t}) - \Phi_c(f)}{\epsilon} \geq 0.$$

On the other hand,

$$\lim_{\epsilon \searrow 0} \frac{\Phi_c(f_{\epsilon,t}) - \Phi_c(f)}{\epsilon}$$

$$
\begin{aligned}
&= \int_{-c}^{c} f(u)p_t(u)du - \int_{-c}^{c} p_t(u)dX_k(u) \\
&= \left[\left(H_{c,k}^{(k-1)}(u) - Y_k^{(k-1)}(u)\right) p_t(u)\right]_{-c}^{c} - \int_{-c}^{c} \left(H_{c,k}^{(k-1)}(u) - Y_k^{(k-1)}(u)\right) p_t'(u)du \\
&= -\left[\left(H_{c,k}^{(k-2)}(u) - Y_k^{(k-2)}(u)\right) p_t'(u)\right]_{-c}^{c} + \int_{-c}^{c} \left(H_{c,k}^{(k-2)}(u) - Y_k^{(k-2)}(u)\right) p_t^{(2)}(u)du \\
&= \int_{-c}^{c} \left(H_{c,k}^{(k-2)}(u) - Y_k^{(k-2)}(u)\right) p_t^{(2)}(u)du \\
&\vdots \\
&= \int_{-c}^{c} \left(H_{c,k}(u) - Y_k(u)\right) dp_t^{(k-1)}(u)du \\
&= H_{c,k}(t) - Y_k(t)\,,
\end{aligned}
$$

and therefore the condition in (3.10) is satisfied.

Similarly, consider the function f_ϵ defined as

$$
\begin{aligned}
f_\epsilon(u) &= f(u) + \epsilon\Bigg(f(u) + \beta_{k-1}\frac{(u+c)^{k-1}}{(k-1)!} + \beta_{k-2}\frac{(u+c)^{k-2}}{(k-2)!} \\
&\qquad + \cdots + \beta_1(u+c) + \beta_0\Bigg). \\
&= f(u) + \epsilon h(u)
\end{aligned}
$$

Notice first that,

$$
f_\epsilon^{(k-2)}(u) = (1+\epsilon)f^{(k-2)}(u) + \epsilon\beta_{k-1}(u+c)
$$

which is convex for $|\epsilon| > 0$ sufficiently small. In order to have f_ϵ in the class $\mathcal{C}_{\epsilon,\underline{m}_1,\underline{m}_2}$, we choose $\beta_{k-1}, \beta_{k-2}, \cdots, \beta_0$ such that

$$
h^{(2i)}(\pm c) = 0, \quad \text{for} \quad i = 0, \cdots, (k-2)/2.
$$

It is easy to check that the latter conditions determine $\beta_{k-1}, \cdots, \beta_0$ uniquely. Thus, we have

$$
\begin{aligned}
0 = \lim_{\epsilon \to 0} \frac{\Phi_c(f_\epsilon) - \Phi_c(f)}{\epsilon} &= \int_{-c}^{c} f(u)h(u)du - \int_{-c}^{c} h(u)dX_k \\
&= \int_{-c}^{c} \left(H_{c,k}^{(k-1)}(u) - Y_k^{(k-1)}(u)\right) h'(u)du \\
&\vdots
\end{aligned}
$$

$$= \int_{-c}^{c} (H_{c,k}(u) - Y_k(u))\, dh^{(k-1)}(u)$$
$$= \int_{-c}^{c} (H_{c,k}(u) - Y_k(u))\, df^{(k-1)}(u)$$

and hence condition (3.11) is satisfied. ■

3.3.2 Existence and Characterization of $H_{c,k}$ for k odd

In the previous section, we proved that the minimization problem for $k = 2$ studied in Groeneboom, Jongbloed, and Wellner (2001a) can be generalized naturally for any even $k > 2$. For k odd, the problem remains to be formalized. For the particular case $k = 1$, it is very well known that the stochastic process involved in the limiting distribution of the MLE of a monotone density at a fixed point x_0 (under some regularity conditions) is determined by the slope at 0 of the greatest convex minorant of the process $(W(t) + t^2, t \in \mathbb{R})$. In this case, a "switching" relationship was exploited as a fundamental tool to derive the asymptotic distribution of the MLE. It is based on the observation that if $\hat{g}_n$ is the MLE (the Grenander estimator); i.e., the left derivative of the greatest concave majorant of the empirical distribution $\mathbb{G}_n$ based on an i.i.d. sample from the true monotone density, then for a fixed $a > 0$

$$\left[\sup\left\{s \geq 0 : \mathbb{G}_n(s) - as \;\text{ is maximal}\right\}\right] = \left[\hat{g}_n(t) \leq a\right]$$

(see Groeneboom (1985)). A similar relationship is currently unknown when $k > 1$. The difficulty is apparent already for $k = 2$ and hence there was a need to formalize the problem differently.

As we did for even integers $k \geq 2$, we need to pose an appropriate minimization problem for odd integers $k > 1$. Wellner (2003) revisited the case $k = 1$ and established a necessary and sufficient condition for a function in the class of monotone functions g such that $\|g\|_{\infty,[-c,c]} \leq K$ to be the minimizer of the functional

$$\Psi_c(g) = \frac{1}{2}\int_{-c}^{c} g^2(t)dt - \int_{-c}^{c} g(t)d(W(t) + t^2)$$

(see Theorem 3.1 in Wellner (2003)). However, the characterization involves two Lagrange parameters which makes the resulting optimizer hard to study. Wellner (2003) pointed out that when $K = K_c \to \infty$, the Lagrange parameters will vanish as $c \to \infty$. Here we define the minimization problem differently. Let $k > 1$ be an odd integer, $c > 0$, $m_0 \in \mathbb{R}$ and $\underline{m}_1$ and $\underline{m}_2 \in \mathbb{R}^l$ where $k = 2l + 1$. Consider the problem of minimizing the same criterion function Φ_c introduced in (3.1) over the class $\mathcal{C}_{k,m_0,\underline{m}_1,\underline{m}_2}$ of k-convex functions satisfying

$$(f^{(k-2)}(-c), \cdots, f^{(1)}(-c)) = \underline{m}_1 \quad \text{and} \quad (f^{(k-2)}(c), \cdots, f^{(1)}(c)) = \underline{m}_2,$$

and $f(c) = m_0$.

Proposition 3.3.3 *Φ_c defined in (3.1) admits a unique minimizer in the class $\mathcal{C}_{k,m_0,\underline{m}_1,\underline{m}_2}$.*

Proof. The proof is very similar to the one we used for k even. ■

The following proposition gives a characterization for the minimizer. Although the techniques are similar to those developed for k even, we prefer to give a detailed proof in order to show clearly the differences between the cases k even and k odd.

Proposition 3.3.4 *The function $f_{c,k} \in \mathcal{C}_{k,m_0,\underline{m}_1,\underline{m}_2}$ is the minimizer of Φ_c if and only if*

$$H_{c,k}(t) \le Y_k(t), \qquad t \in [-c, c] \tag{3.17}$$

and

$$\int_{-c}^{c} (H_{c,k}(t) - Y_k(t))\, df_{c,k}^{(k-1)}(t) = 0, \tag{3.18}$$

where $H_{c,k}$ is the k-fold integral of $f_{c,k}$ satisfying

$$H_{c,k}(-c) = Y_k(-c), H_{c,k}^{(2)}(-c) = Y_k^{(2)}(-c), \cdots, H_{c,k}^{(k-3)}(-c) = Y_k^{(k-3)}(-c),$$

$$H_{c,k}(c) = Y_k(c), H_{c,k}^{(2)}(c) = Y_k^{(2)}(c), \cdots, H_{c,k}^{(k-3)}(c) = Y_k^{(k-3)}(c),$$

and

$$H_{c,k}^{(k-1)}(-c) = Y^{(k-1)}(-c).$$

Proof. To avoid conflicting notations, we replace $f_{c,k}$ by f. Let f be a function in $\mathcal{C}_{k,m_0,\underline{m}_1,\underline{m}_2}$ satisfying (3.17) and (3.18). Using the inequality in (3.15), we have for an arbitrary function g in $\mathcal{C}_{k,m_0,\underline{m}_1,\underline{m}_2}$

$$\Phi_c(g) - \Phi_c(f) \geq \int_{-c}^{c} f(t)\,(g(t) - f(t))\,dt - \int_{-c}^{c} (g(t) - f(t))\,dX_k(t).$$

Using the fact that $H_{c,k}^{(j)}$ is the $(k-j)$-fold integral of f for $j = 1, \cdots, k$ and the fact that

$$g(c) = f(c), \qquad H_{c,k}^{(k-1)}(-c) = Y_k^{(k-1)}(-c)\,,$$

$$g^{(2i+1)}(\pm c) = f^{(2i+1)}(\pm c), \quad \text{for } \ i = 0, \cdots, (k-3)/2\,,$$

and

$$H_{c,k}^{(2j)}(\pm c) = Y_k^{(2j)}(\pm c), \quad \text{for } \ j = 0, \cdots, (k-3)/2\,,$$

we obtain by successive integrations by parts

$$\begin{aligned}
&\int_{-c}^{c} f(t)\,(g(t) - f(t))\,dt - \int_{-c}^{c} (g(t) - f(t))\,dX_k(t) \\
&= \left[\left(H_{c,k}^{(k-1)}(t) - Y_k^{(k-1)}(t)\right)(g(t) - f(t))\right]_{-c}^{c} \\
&\quad - \int_{-c}^{c} \left(H_{c,k}^{(k-1)}(t) - Y_k^{(k-1)}(t)\right)\left(g'(t) - f'(t)\right)dt \\
&= -\int_{-c}^{c} \left(H_{c,k}^{(k-1)}(t) - Y_k^{(k-1)}(t)\right)\left(g'(t) - f'(t)\right)dt \\
&= -\left[\left(H_{c,k}^{(k-2)}(t) - Y_k^{(k-2)}(t)\right)\left(g'(t) - f'(t)\right)\right]_{-c}^{c} \\
&\quad + \int_{-c}^{c} \left(H_{c,k}^{(k-2)}(t) - Y_k^{(k-2)}(t)\right)\left(g''(t) - f''(t)\right)dt \\
&= \int_{-c}^{c} \left(H_{c,k}^{(k-2)}(t) - Y_k^{(k-2)}(t)\right)\left(g''(t) - f''(t)\right)dt \\
&\ \vdots \\
&= -\int_{-c}^{c} \left(H_{c,k}(t) - Y_k(t)\right)\left(dg^{(k-1)}(t) - df^{(k-1)}(t)\right).
\end{aligned}$$

This yields, using the condition in (3.18),

$$\begin{aligned}
&\int_{-c}^{c} f(t)\,(g(t) - f(t))\,dt - \int_{-c}^{c} (g(t) - f(t))\,dX_k(t) \\
&= -\int_{-c}^{c} \left(H_{c,k}(t) - Y_k(t)\right) dg^{(k-1)}(t)\,.
\end{aligned}$$

Now, using condition (3.17) and the fact that $g^{(k-1)}$ is nondecreasing, we conclude that

$$\Phi_c(g) \geq \Phi_c(f)$$

and that f is the minimizer of Φ_c.

Conversely, suppose that f minimizes Φ_c over the class $\mathcal{C}_{k,m_0,\underline{m}_1,\underline{m}_2}$. Fix a small $\epsilon > 0$ and let $t \in (-c, c)$. We define the function $f_{t,\epsilon}$ on $[-c, c]$ by

$$\begin{aligned} f_{t,\epsilon}(u) &= f(u) + \epsilon\left(\frac{(u-t)_+^{k-1}}{(k-1)!} + \alpha_{k-1}\frac{(u+c)^{k-1}}{(k-1)!} + \alpha_{k-3}\frac{(u+c)^{k-3}}{(k-3)!}\right. \\ &\qquad \left. + \cdots + \alpha_2\frac{(u+c)^2}{2!} + \alpha_0\right) \\ &= f(u) + \epsilon p_t(u) \end{aligned}$$

satisfying

$$p_t^{(2i+1)}(\pm c) = 0, \qquad \text{for } i = 0, \cdots, (k-3)/2 \tag{3.19}$$

and

$$p_t(c) = 0. \tag{3.20}$$

For this choice of a perturbation function, we have for all $u \in [-c, c]$

$$f_{t,\epsilon}^{(k-2)}(u) = f^{(k-2)}(u) + \epsilon((u-t)_+ + \alpha_{k-1}(u+c)).$$

Thus, $f_{t,\epsilon}$ is convex for any $\epsilon > 0$ as a sum of two convex functions. The conditions (3.19) and (3.20) ensures that $f_{t,\epsilon}$ remains in the class $\mathcal{C}_{k,m_0,\underline{m}_1,\underline{m}_2}$ and the parameters $\alpha_{k-1}, \alpha_{k-3}, \cdots, \alpha_0$ are uniquely determined:

$$\begin{aligned} \alpha_{k-1} &= -\frac{(c-t)}{2c} \\ \alpha_{k-3} &= -\frac{1}{2c}\left(\alpha_{k-1}\frac{(2c)^3}{3!} + \frac{(c-t)^3}{3!}\right) \\ &\vdots \\ \alpha_2 &= -\frac{1}{2c}\left(\alpha_{k-1}\frac{(2c)^{k-2}}{(k-2)!} + \cdots + \alpha_4\frac{(2c)^3}{3!} + \frac{(2c)^{k-2}}{(k-2)!}\right) \\ \alpha_0 &= -\left(\alpha_{k-1}\frac{(2c)^{k-1}}{(k-1)!} + \cdots + \alpha_2\frac{(2c)^2}{2!} + \frac{(c-t)^{k-1}}{(k-1)!}\right). \end{aligned}$$

Since f is the minimizer of Φ_c, we have

$$\lim_{\epsilon \searrow 0} \frac{\Phi_c(f_\epsilon) - \Phi_c(f)}{\epsilon} \geq 0.$$

But

$$\begin{aligned}
&\lim_{\epsilon \searrow 0} \frac{\Phi_c(f_\epsilon) - \Phi_c(f)}{\epsilon} \\
&= \int_{-c}^{c} f(u) p_t(u) du - \int_{-c}^{c} p_t(u) dX_k(u) \\
&= \left[\left(H_{c,k}^{(k-1)}(u) - Y_k^{(k-1)}(u)\right) p_t(u)\right]_{-c}^{c} - \int_{-c}^{c} \left(H_{c,k}^{(k-1)}(u) - Y_k^{(k-1)}(u)\right) p_t'(u) du \\
&= -\left[\left(H_{c,k}^{(k-2)}(u) - Y_k^{(k-2)}(u)\right) p_t'(u)\right]_{-c}^{c} + \int_{-c}^{c} \left(H_{c,k}^{(k-2)}(u) - Y_k^{(k-2)}(u)\right) p_t^{(2)}(u) du \\
&\vdots \\
&= -\int_{-c}^{c} \left(H_{c,k}(u) - Y_k(u)\right) dp_t^{(k-1)}(u) \\
&= -\left(H_{c,k}(t) - Y_k(t)\right),
\end{aligned}$$

and therefore the condition in (3.17) is satisfied. Similarly, consider the function f_ϵ defined as

$$\begin{aligned}
f_\epsilon(u) &= f(u) + \epsilon \left(f(u) + \beta_{k-1} \frac{(u+c)^{k-1}}{(k-1)!} + \beta_{k-2} \frac{(u+c)^{k-2}}{(k-2)!} + \cdots + \beta_1 (u+c) + \beta_0 \right) \\
&= f(u) + \epsilon h(u).
\end{aligned}$$

Notice first that,

$$f_\epsilon^{(k-2)}(u) = (1+\epsilon) f^{(k-2)}(u) + \epsilon \beta_{k-1}(u+c)$$

which is convex for $|\epsilon|$ small enough. In order to have f_ϵ in the class $\mathcal{C}_{m_0, \underline{m}_1, \underline{m}_2}$, we choose the coefficients $\beta_{k-1}, \beta_{k-2}, \cdots, \beta_0$ such that

$$h^{(2i+1)}(\pm c) = 0, \quad \text{for } i = 0, \cdots, (k-3)/2,$$

and $h(c) = 0$. It is easy to check that the previous equations admit a unique solution. Thus, we have

$$0 = \lim_{\epsilon \to 0} \frac{\Phi_c(f_\epsilon) - \Phi_c(f)}{\epsilon} = \int_{-c}^{c} f(u) h(u) du - \int_{-c}^{c} h(u) dX_k(u)$$

$$
\begin{aligned}
&= \int_{-c}^{c} \left(H_{c,k}^{(k-1)}(u) - Y_k^{(k-1)}(u) \right) h'(u) du \\
&\vdots \\
&= -\int_{-c}^{c} \left(H_{c,k}(u) - Y_k(u) \right) dh^{(k-1)}(u) \\
&= -\int_{-c}^{c} \left(H_{c,k}(u) - Y_k(u) \right) df^{(k-1)}(u),
\end{aligned}
$$

and hence condition (3.18) is satisfied. ■

3.4 *The tightness problem*

3.4.1 Existence of points of touch

Although the characterizations given in Propositions 3.3.2 and 3.3.4, indicate that $f_{c,k}^{(k-2)}$ is piecewise linear and the k-fold integral of $f_{c,k}$ touches Y_k whenever $f_{c,k}^{(k-2)}$ changes its slope, they do not provide us with any information about the number of the jump points of $f_{c,k}^{(k-1)}$. It is possible, at least in principle, that $f_{c,k}^{(k-1)}$ does not have any jump point, in which case $f_{c,k}^{(k-2)}$ is a straight line. However, if we take

$$
\underline{m}_1 = \underline{m}_2 = \left(\frac{k!}{2!} c^2, \frac{k!}{4!} c^4, \cdots, c^k \right)
$$

when k is even, and

$$
m_0 = c^k, \qquad \underline{m}_1 = \underline{m}_2 = \left(\frac{k!}{2!} c^2, \frac{k!}{4!} c^4, \cdots, \frac{k!}{(k-1)!} c^{k-1} \right)
$$

when k is odd, then with an increasing probability, $H_{c,k}$ and Y_k have to touch each other in $(-c, c)$ as $c \to \infty$. The next proposition establishes this basic fact.

Proposition 3.4.1 *Let $\epsilon > 0$ and consider $\underline{m}_1$, $\underline{m}_2$, and m_0 as specified above according to whether k is even or odd. Then, there exists $c_0 > 0$ such that the probability that $H_{c,k}$ and Y_k have at least one point of touch is greater than $1 - \epsilon$ for $c > c_0$; i.e.,*

$$
P(Y_k(\tau) = H_{c,k}(\tau) \ \ \textit{for some } \tau \in [-c, c]) \to 1, \quad \textit{as} \ \ c \to \infty .
$$

Proof. We start with k even. If $H_{c,k}$ and Y_k do not touch each other at any point in $(-c, c)$, it follows that $H_{c,k}$ is a polynomial of degree $2k-1$ in which case $H_{c,k}$ is fully determined by

$$\begin{aligned} H_{c,k}^{(2i)}(\pm c) &= Y_k^{(2i)}(\pm c), \quad \text{for} \quad i = 0, \cdots, (k-2)/2 \\ H_{c,k}^{(2i)}(\pm c) &= \frac{k!}{(2k-2i)!} c^{2k-2i}, \quad \text{for} \quad i = k/2, \cdots, (2k-2)/2. \end{aligned}$$

If we write the polynomial $H_{c,k}$ as

$$H_{c,k}(t) = \frac{\alpha_{2k-1}}{(2k-1)!} t^{2k-1} + \frac{\alpha_{2k-2}}{(2k-2)!} t^{2k-2} + \cdots + \alpha_1 t + \alpha_0,$$

then $\alpha_{2k-1} = 0$ since $H_{c,k}^{(2k-2)}(-c) = H_{c,k}^{(2k-2)}(c)$. Because of the same symmetry, $\alpha_{2k-3} = \alpha_{2k-5} = \cdots = \alpha_{k+1} = 0$. Furthermore, it is easy to establish after some algebra that the coefficients $\alpha_{2k-2}, \alpha_{2k-4}, \cdots, \alpha_k$ are given by

$$\alpha_{2k-2} = \frac{k!}{2!} c^2,$$

and for $j = 2, \cdots, k/2$.

$$\alpha_{2k-2j} = \frac{k!}{(2j)!} c^{2j} - \left(\frac{\alpha_{2k-2}}{(2j-2)!} c^{2j-2} + \cdots + \frac{\alpha_{2k-2j+2}}{2!} c^2 \right)$$

For $\alpha_{k-1}, \cdots, \alpha_0$, we have different expressions:

$$\alpha_{k-1} = \frac{Y_k^{(k-2)}(c) - Y_k^{(k-2)}(-c)}{2c},$$

$$\alpha_{k-2} = \frac{Y_k^{(k-2)}(-c) + Y_k^{(k-2)}(c)}{2} - \left(\frac{\alpha_{2k-2}}{k!} c^k + \cdots + \frac{\alpha_k}{2!} c^2 \right)$$

which can be viewed as the starting values for α_{k-2j-1} and α_{k-2j-2} given by

$$\alpha_{k-2j-1} = \frac{Y_k^{(k-2j-2)}(c) - Y_k^{(k-2j-2)}(-c)}{2c} - \left(\frac{\alpha_{k-1}}{(2j+1)!} c^{2j} + \cdots + \frac{\alpha_{k-2j+1}}{3!} c^2 \right),$$

and

$$\alpha_{k-2j-2} = \frac{Y_k^{(k-2j-2)}(c) + Y_k^{(k-2j-2)}(-c)}{2} - \left(\frac{\alpha_{2k-2}}{(k+2j)!} c^{k+2j} + \cdots + \frac{\alpha_{k-2j}}{2!} c^2 \right)$$

for $j = 1, \cdots, (k-2)/2$.

Let V_k denote the $(k-1)$-fold integral of two-sided Brownian motion; i.e.,

$$Y_k(t) = V_k(t) + \frac{k!}{(2k)!} t^{2k}, \quad t \in \mathbb{R}.$$

We also introduce a_{2k-2j}, for $j = 1, \cdots, k$ defined by

$$a_{2k-2j} = \alpha_{2k-2j}, \qquad \text{for} \quad j = 1, \cdots, k/2 \tag{3.1}$$

and

$$a_{2k-2j} = \alpha_{2k-2j} - \frac{V_k^{(2k-2j)}(-c) + V_k^{(2k-2j)}(c)}{2}, \qquad \text{for} \quad j = (k+2)/2, \cdots, k. \tag{3.2}$$

The coefficients a_{2k-2j}, for $j = 2, \cdots, k$ are given by the following recursive formula

$$a_{2k-2j} = \frac{k!}{(2j)!} c^{2j} - \left(\frac{a_{2k-2}}{(2j-2)!} c^{2j-2} + \cdots + \frac{a_{2k-2j+2}}{2!} c^2 \right),$$

with

$$a_{2k-2} = \frac{k!}{2!} c^2.$$

Now, using the expressions in (3.1) and (3.2), we can write the value of $H_{c,k}$ at the point 0, $H_{c,k}(0)$, as a function of the derivatives of V_k at the boundary points $-c$ and c and the a_j's:

$$\begin{aligned}
H_{c,k}(0) &= \alpha_0 \\
&= \frac{Y_k(c) + Y_k(-c)}{2} - \left(\frac{\alpha_{2k-2}}{(2k-2)!} c^{2k-2} + \cdots + \frac{\alpha_2}{2!} c^2 \right) \\
&\quad - \left(\frac{a_{2k-2}}{(2k-2)!} + \cdots + \frac{a_k}{k!} \right) \\
&\quad - \left(\frac{V_k^{(2)}(c) + V_k^{(2)}(-c)}{2} + \frac{a_{k-2}}{(k-2)!} \right) c^{k-2} \\
&\quad - \cdots - \left(\frac{V_k^{(k-2)}(c) + V_k^{(k-2)}(-c)}{2} + \frac{a_2}{2!} \right) c^2 \\
&= \frac{V_k(c) + V_k(-c)}{2} - \left(\frac{V_k^{(2)}(c) + V_k^{(2)}(-c)}{2} \right) \frac{c^2}{2!} \\
&\qquad - \cdots - \left(\frac{V_k^{(k-2)}(c) + V_k^{(k-2)}(-c)}{2} \right) \frac{c^{k-2}}{(k-2)!}
\end{aligned}$$

$$
\begin{aligned}
& +\frac{k!}{2!}c^{2k} - \left(\frac{a_{2k-2}}{(2k-2)!}c^{2k-2} + \frac{a_{2k-4}}{(2k-4)!}c^{2k-4} + \cdots + \frac{a_2}{2!}\right) \\
= \quad & \frac{V_k(c) + V_k(-c)}{2} - \left(\frac{V_k^{(2)}(c) + V_k^{(2)}(-c)}{2}\right)\frac{c^2}{2!} \\
& - \cdots - \left(\frac{V_k^{(k-2)}(c) + V_k^{(k-2)}(-c)}{2}\right)\frac{c^{k-2}}{(k-2)!} + a_0.
\end{aligned}
$$

By going back to the definition of a_{2k-2j} for $j = 0, \cdots, k$, we can see that a_{2k-2j} is proportional to c^{2j}. Hence, there exists λ_k such that $a_0 = \lambda_k c^{2k}$. One can verify numerically that λ_k is negative. The plot in Figure 3.1 shows the curve of $\log(-\lambda_k)$ versus $k = 4, \cdots, 170$. The reason for taking the logarithmic transformation is that $|\lambda_k|$ becomes very large for increasing values of k, e.g. for $k = 100$, $\lambda_k = -7.094 \times 10^{118}$.

Table 3.1: Table of λ_k and $\log(-\lambda_k)$ for some values of even integers k.

k	λ_k	$\log(-\lambda_k)$
4	-0.82440	-0.19309
20	-4.42832×10^{10}	24.51387
30	-5.77268×10^{20}	47.80483
48	-2.35131×10^{42}	97.56354
100	-7.09477×10^{118}	273.66439

Now, denote

$$
\begin{aligned}
S_k(c) \quad = \quad & \frac{V_k(c) + V_k(-c)}{2} - \left(\frac{V_k^{(2)}(c) + V_k^{(2)}(-c)}{2}\right)\frac{c^2}{2!} \\
& - \cdots - \left(\frac{V_k^{(k-2)}(c) + V_k^{(k-2)}(-c)}{2}\right)\frac{c^{k-2}}{(k-2)!}.
\end{aligned}
$$

However, we have

$$
S_k(c) = O_p\left(c^{k-1/2}\right) \quad \text{as} \quad c \to \infty.
$$

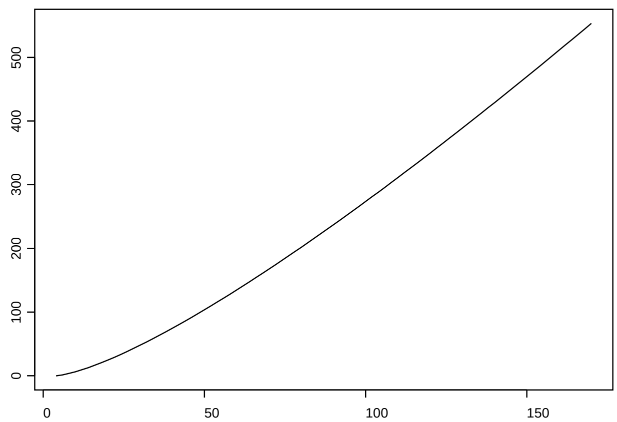

Figure 3.1: The plot of $\log(-\lambda_k)$ versus k for $k = 4, 8, \cdots, 170$.

Indeed, for $0 \le j \le k-2$,

$$V_k^{(j)}(c) \stackrel{d}{=} \int_0^c \frac{(c-t)^{k-1-j}}{(k-1-j)!} dW(t).$$

By using the change of variable $u = ct$ and $W(cu) \stackrel{d}{=} \sqrt{c}W(u)$, we have

$$\begin{aligned} V_k^{(j)}(c) &\stackrel{d}{=} c^{k-j-1} \int_0^1 \frac{(1-u)^{k-1-j}}{(k-1-j)!} dW(cu) \\ &\stackrel{d}{=} c^{k-j-1/2} \int_0^1 \frac{(1-u)^{k-1-j}}{(k-1-j)!} dW(u). \end{aligned}$$

Therefore, $V_k^{(j)}(c) = O_p\left(c^{k-j-1/2}\right)$ as $c \to \infty$. Similarly, $V_k^{(j)}(-c) = O_p\left(c^{k-j-1/2}\right)$ and therefore $S_k(c) = O_p\left(c^{k-1/2}\right)$. But since $\lambda_k < 0$, it follows that

$$\begin{aligned} P(H_{c,k}(0) \ge Y_k(0)) &= P(S_k(c) + \lambda_k c^{2k} \ge 0) \\ &= P(S_k(c) \ge -\lambda_k c^{2k}) \to 0 \quad \text{as} \quad c \to \infty, \end{aligned}$$

that is, with probability converging to 1, $H_{c,k}$ and Y_k have at least one point of touch as $c \to \infty$.

Now, suppose that k is odd. The proof is similar but involves a different "starting polynomial". Let us assume again that $H_{c,k}$ and Y_k do not have any point of touch in $(-c,c)$. Then, $H_{c,k}$ would be a polynomial of degree $2k-1$ which can be fully determined by the boundary conditions

$$H_{c,k}^{(2i)}(\pm c) = \frac{k!}{(2k-2i)!}c^{2k-2i}, \quad \text{for } i = (2k-2)/2, \cdots, (k+1)/2\,, \tag{3.3}$$

$$H_{c,k}^{(k)}(c) \quad = \quad c^k\,, \tag{3.4}$$

$$H_{c,k}^{(k-1)}(-c) \quad = \quad Y_k^{(k-1)}(-c)\,, \tag{3.5}$$

and

$$H_{c,k}^{(2i)}(\pm c) = Y_k^{(2i)}(\pm c), \quad \text{for } i = (k-3)/2, \cdots, 0. \tag{3.6}$$

There exist coefficients $\alpha_{2k-1}, \alpha_{2k-2}, \cdots, \alpha_1, \alpha_0$ such that

$$H_{c,k}(t) = \frac{\alpha_{2k-1}}{(2k-1)!}t^{2k-1} + \frac{\alpha_{2k-2}}{(2k-2)!}t^{2k-2} + \cdots + \alpha_1 t + \alpha_0, \quad t \in [-c,c].$$

The boundary conditions in (3.3) imply that $\alpha_{2k-1} = \alpha_{2k-3} = \cdots = \alpha_{k+2} = 0$. Also, using the same conditions we obtain that

$$\alpha_{2k-2} = \frac{k!}{2!}c^2$$

and for $2 \leq j \leq (k-1)/2$

$$\alpha_{2k-2j} = \frac{k!}{(2j)!}c^{2j} - \left(\frac{\alpha_{2k-2}}{(2j-2)!} + \cdots + \frac{\alpha_{2k-2j+2}}{2!}c^2\right).$$

The "one-sided" conditions (3.4) and (3.5) imply that for $j = 1, \cdots, (k-1)/2$

$$\alpha_k = c^k - \left(\frac{\alpha_{2k-2}}{(k-2)!}c^{k-2} + \cdots + \frac{\alpha_{k+3}}{(k+3)!}c^3 + \alpha_{k+1}c\right)$$

and

$$\alpha_{k-1} = Y_k^{(k-1)}(-c) - \left(\frac{\alpha_{2k-2}}{(k-1)!}c^{k-1} + \cdots + \frac{\alpha_{k+1}}{2!}c^2 - \alpha_k c\right)$$

respectively. Finally, using the boundary conditions in (3.6) we obtain that

$$\alpha_{k-2j} = \frac{Y_k^{(k-2j-1)}(c) - Y_k^{(k-2j-1)}(-c)}{2c} - \left(\frac{\alpha_k}{(2j+1)!} c^{2j} + \cdots + \frac{\alpha_{k-2j+2}}{3!} c^3 \right)$$

and

$$\alpha_{k-2j-1} = \frac{Y_k^{(k-2j-1)}(-c) + Y_k^{(k-2j-1)}(c)}{2} - \left(\frac{\alpha_{2k-2}}{(k+2j-1)!} c^{k+2j-1} + \cdots + \frac{\alpha_{k-2j+1}}{2!} c^2 \right)$$

for $j = 1, \cdots, (k-1)/2$.

Let V_k continue to denote the $(k-1)$-fold integral of two-sided Brownian motion and consider $a_{2k-2}, a_{2k-4}, \cdots, a_{k+1}, a_k, a_{k-1}, \cdots, a_0$ given by

$$a_{2k-2j} = \alpha_{2k-2j}, \qquad \text{for } \ j = 1, \cdots, (k-1)/2$$

$$a_k = c^k - \left(\frac{a_{2k-2}}{(k-2)!} + \cdots + \frac{a_{k+3}}{3!} c^3 + a_{k+1} c \right)$$

$$a_{k-1} = \frac{k!}{(k+1)!} c^{k+1} - \left(\frac{a_{2k-2}}{(k-1)!} c^{k-1} + \cdots + \frac{\alpha_{k+1}}{2!} c^2 - a_k c \right),$$

and

$$a_{k-2j-1} = \frac{k!}{(k+2j+1)!} c^{k+2j+1} - \left(\frac{a_{2k-2}}{(k+2j-1)!} c^{k+2j-1} + \cdots + \frac{a_{k-2j+1}}{2!} c^2 \right)$$

for $j = 1, \cdots, (k-1)/2$. It follows that

$$\begin{aligned}
H_{c,k}(0) &= \alpha_0 \\
&= \frac{Y_k(-c) + Y_k(c)}{2} - \left(\frac{\alpha_{2k-2}}{(2k-2)!} c^{2k-2} + \frac{\alpha_{2k-4}}{(2k-4)!} c^{2k-4} + \cdots + \frac{\alpha_2}{2!} c^2 \right) \\
&= \frac{V_k(-c) + V_k(c)}{2} - \left(\frac{V_k(-c) + V_k(c)}{2} \right) \frac{c^2}{2!} \\
&\qquad - \cdots - \left(\frac{V_k(-c) + V_k(c)}{2} \right) \frac{c^{k-2}}{(k-2)!} + a_0 \\
&= S_k(c) + a_0
\end{aligned}$$

where

$$a_0 = \frac{k!}{(2k)!} c^{2k} - \left(\frac{a_{2k-2}}{(2k-2)!} c^{2k-2} + \cdots + \frac{a_2}{2!} c^2 \right).$$

It is easy to see that the coefficients $a_{2k-2}, a_{2k-4}, \cdots, a_0$ are proportional to $c^2, c^4, \cdots, c^{2k}$ respectively. Therefore, there exists λ_k such that $a_0 = \lambda_k c^{2k}$. We can verify numerically that $\lambda_k > 0$ (see Figure 3.2 and Table 3.2). But since

$$S_k(c) = O_p\left(c^{k-1/2}\right),$$

it follows that

$$\begin{aligned} P(H_{c,k}(0) \le Y_k(0)) &= P(S_k(c) + \lambda_k c^{2k} \le 0) \\ &= P(S_k(c) \le -\lambda_k c^{2k}) \\ &= P(-S_k(c) \ge \lambda_k c^{2k}) \to 0 \quad \text{as } c \to \infty, \end{aligned}$$

which completes the proof. ■

Table 3.2: Table of λ_k and $\log(\lambda_k)$ for some values of odd integers k.

k	λ_k	$\log(\lambda_k)$
3	1.50833	0.41100
19	1.63896×10^{10}	23.51991
29	1.42435×10^{20}	46.40541
57	6.79374×10^{54}	126.25559
99	5.25169×10^{117}	271.06100

Corollary 3.4.1 *Fix $\epsilon > 0$ and let $t \in (-c, c)$. There exists $c_0 > 0$ such that the probability that the process $H_{c,k}$ touches Y_k at two points of touch τ^- and τ^+ before and after the point t is larger than $1 - \epsilon$ for $c > c_0$.*

Proof. We focus on k even as the arguments are very similar for k odd. Consider first $t = 0$. We know by Proposition 3.4.1 that, with very large probability, there exists at least one point of touch (before or after 0) as $c \to \infty$. By symmetry of two-sided Brownian motion originating at 0 and hence by that of the process Y_k, there exist two points of touch

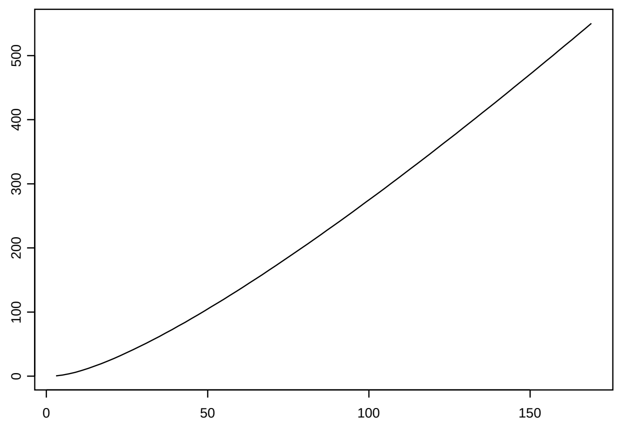

Figure 3.2: Plot of $\log(\lambda_k)$ versus k for $k = 3, 5, \cdots, 169$.

before and after 0 with very large probability as $c \to \infty$. Now, fix $t_0 \neq 0$ and consider the problem of minimizing

$$\begin{aligned} \Phi_{c,t_0}(f) &= \frac{1}{2}\int_{-c+t_0}^{c+t_0} f^2(t)dt - \int_{-c+t_0}^{c+t_0} f(t)dX_k(t) \\ &= \frac{1}{2}\int_{-c+t_0}^{c+t_0} f^2(t)dt - \int_{-c+t_0}^{c+t_0} f(t)(t^k dt + dW(t)) \end{aligned}$$

over the class of k-convex functions satisfying

$$f^{(k-2)}(-c+t_0) = \frac{k!}{2!}(-c+t_0)^2, f^{(k-4)}(-c+t_0) = \frac{k!}{4!}(-c+t_0)^4, \cdots, f(-c+t_0) = (-c+t_0)^k$$

and

$$f^{(k-2)}(c+t_0) = \frac{k!}{2!}(c+t_0)^2, f^{(k-4)}(c+t_0) = \frac{k!}{4!}(c+t_0)^4, \cdots, f(c+t_0) = (c+t_0)^k.$$

Since adding any constant to $-c$ and c is irrelevant to the original minimization problem, all the above results hold and in particular that of existence of two points of touch τ^- and τ^+ before and after 0 with increasing probability as $c \to \infty$.

But using the change of variable $u = t - t_0$, Φ_{c,t_0} can be rewritten as

$$\begin{aligned}
\Phi_{c,t_0}(f) &= \frac{1}{2}\int_{-c}^{c} f^2(u+t_0)du - \int_{-c+t_0}^{c+t_0} f(t)(t^k dt + dW(t)) \\
&= \frac{1}{2}\int_{-c}^{c} f^2(u+t_0)du - \int_{-c}^{c} f(u+t_0)((u+t_0)^k dt + dW(u+t_0)) \\
&\stackrel{d}{=} \frac{1}{2}\int_{-c}^{c} g^2(u)du - \int_{-c}^{c} g(u)((u+t_0)^k dt + dW(u)) \qquad (3.7)
\end{aligned}$$

where in (3.7), we used stationarity of the increments of W and $g(u) = f(u+t_0)$ is k-convex satisfying the above boundary conditions at $-c$ and c. From the latter form of Φ_{c,t_0}, we can see that the "true" k-convex is now $(t+t_0)^k$ defined on $[-c, c]$. However, the "estimation" problem is basically the same expect and hence there exist two points of touch before and after t_0 with increasing probability as $c \to \infty$. ■

3.4.2 Tightness

One very important element in proving the existence of the process H_k is tightness of the process $H_{c,k}$ and its $(2k-1)$ derivatives when $c \to \infty$. The process H_k can be defined as the limit of $H_{c,k}$ as $c \to \infty$ the same way Groeneboom, Jongbloed, and Wellner (2001a) did for the special case $k = 2$. In the latter case, tightness of the process $H_{c,2}$ and its derivatives $H'_{c,k}$, $H^{(2)}_{c,k}$, and $H^{(3)}_{c,k}$ was implied by tightness of the distance between the points of touch of $H_{c,2}$ with respect to Y_2. The authors could prove using martingale arguments, that for a fixed $\epsilon > 0$, there exists $M > 0$ independent of t such that for any fixed $t \in (-c, c)$,

$$\limsup_{c\to\infty} P\left([t-\tau^- > M] \cap [\tau^+ - t > M]\right) \le \epsilon \qquad (3.8)$$

where τ^- and τ^+ are respectively the last point of touch before t and the first point of touch after t.

Before giving any further details about the difficulties of proving such a property when $k > 2$, we explain the difference between the result proven in (3.8) and the one stated in Lemma 3.4.4 and Corollary 3.4.2. By the first result, we only know that not both points of touch τ^- and τ^+ are "out of control" whereas our result implies that they both stay within a bounded distance from the point t with very large probability as $c \to \infty$. Therefore, we are claiming a stronger result than the one proved by Groeneboom, Jongbloed, and Wellner (2001a). Intuitively, tightness has to be a common property of both the points of touch and this can be seen by using symmetry of the process Y_k. Indeed, since the latter has the same law whether the Brownian motion W "runs" from $-c$ to c or vice versa, it is not hard to be convinced that tightness of one point of touch implies tightness of the other. It should be mentioned here that for proving the existence of *two* points of touch before and after any fixed point t, the authors claimed that this follows from arguments that are similar to the ones used to show existence of at least one point of touch. We tried to reproduce such arguments but we found the situation somehow different. In fact, we found that the arguments used in the proof of Lemma 2.1 in Groeneboom, Jongbloed, and Wellner (2001a) cannot be used similarly to prove the existence of two points of touch unless one of these points of touch is "under control". More formally, we need to make sure that the existing point of touch is tight; i.e., there exists some $M > 0$ independent of t such that the distance between t and this point of touch is bounded by M with a large probability as $c \to \infty$. We find that it is simpler to use a symmetry argument as in Corollary 3.4.1 to make the conclusion.

As mentioned before, proving tightness was the most crucial point that led in the end to showing the existence of the process H_2. Groeneboom, Jongbloed, and Wellner (2001a) were able to prove it by using martingale arguments but more importantly the fact that the process $H_{c,2}$, which is a cubic spline, can be explicitly determined on the "excursion" interval $[\tau^-, \tau^+]$. Indeed, in the special case of $k = 2$, the four conditions $H_{c,2}(\tau^-) = Y_2(\tau^-)$, $H_{c,2}(\tau^+) = Y_2(\tau^+)$ and $H'_{c,2}(\tau^-) = Y'_2(\tau^-)$, $H'_{c,2}(\tau^+) = Y_2(\tau^+)$, implied by the fact that $H_{2,c} \geq Y_2$, yield a unique solution. The same conditions hold true for $k > 2$ but are obviously not enough to determine the $(2k-1)$-th spline $H_{c,k}$. To do so, it seems inevitable to consider the whole set of points of touch along with the boundary conditions at $-c$ and c, which is rather infeasible since, in principle, the locations of the other points of touch are unknown. However, we shall see that we only need $2k-2$ points to be able to determine the spline $H_{c,k}$ completely. For $k > 2$, it seems that the Gaussian problem becomes less local as we need more than one excursion interval in order to study the properties of $H_{c,k}$ and its derivatives at a fixed point. Although the special case $k = 2$ gives a lot of insight into the general problem, the arguments by Groeneboom, Jongbloed, and Wellner (2001a) cannot be readapted directly for the general case of $k > 2$. In the proof of Lemma 3.4.4, we skip many technical details as the tightness problem is very similar to the gap problem for the LSE and MLE studied in great detail in Chapter 2. We will also restrict ourselves to k even as the case k odd can be handled similarly.

In order to make use of the techniques developed in Chapter 2 for solving the gap problem, it is very beneficial to first change the minimization problem from its current version to the slightly different one where we minimize,

$$\frac{1}{2}\int_{-c^{\frac{1}{2k+1}}}^{c^{\frac{1}{2k+1}}} g^2(t)dt - \int_{-c^{\frac{1}{2k+1}}}^{c^{\frac{1}{2k+1}}} g(t)(t^k dt + dW(t)) \tag{3.9}$$

over the class of k-convex functions on $[-c^{1/(2k+1)}, c^{1/(2k+1)}]$ satisfying

$$g(c^{\frac{1}{2k+1}}) = c^{\frac{k}{2k+1}}, g''(c^{\frac{1}{2k+1}}) = \frac{k!}{(k-2)!}c^{\frac{k-2}{2k+1}}, \cdots, g^{(k-2)}(c^{\frac{1}{2k+1}}) = \frac{k!}{(2)!}c^{\frac{2}{2k+1}}.$$

Now using the change of variable $t = c^{1/(2k+1)}u$, we can write

$$\begin{aligned}
&\frac{1}{2}\int_{-c^{\frac{1}{2k+1}}}^{c^{\frac{1}{2k+1}}} g^2(t)dt - \int_{-c^{\frac{1}{2k+1}}}^{c^{\frac{1}{2k+1}}} g(t)dX_k(t) \\
&\overset{d}{=} \; c^{\frac{1}{2k+1}}\frac{1}{2}\int_{-1}^{1} g^2(c^{\frac{1}{2k+1}}u)du - \int_{-1}^{1} g(c^{\frac{1}{2k+1}}u)(c^{\frac{k+1}{2k+1}}u^k du + dW(c^{\frac{1}{2k+1}}u)) \\
&\overset{d}{=} \; c^{\frac{1}{2k+1}}\frac{1}{2}\int_{-1}^{1} g^2(c^{\frac{1}{2k+1}}u)du - \int_{-1}^{1} g(c^{\frac{1}{2k+1}}u)\left(c^{\frac{k+1}{2k+1}}u^k du + c^{\frac{1}{2(2k+1)}}dW(u)\right) \\
&\overset{d}{=} \; c^{\frac{1}{2k+1}}\frac{1}{2}\int_{-1}^{1} g^2(c^{\frac{1}{2k+1}}u)du - \int_{-1}^{1} g(c^{\frac{1}{2k+1}}u)\left(c^{\frac{k+1}{2k+1}}u^k du + c^{\frac{1}{2(2k+1)}}\sqrt{c}\,\frac{dW(u)}{\sqrt{c}}\right) \\
&\overset{d}{=} \; c^{\frac{1}{2k+1}}\frac{1}{2}\int_{-1}^{1} g^2(c^{\frac{1}{2k+1}}u)du - \int_{-1}^{1} g(c^{\frac{1}{2k+1}}u)\left(c^{\frac{k+1}{2k+1}}u^k du + c^{\frac{k+1}{2k+1}}\frac{dW(u)}{\sqrt{c}}\right) \\
&\overset{d}{=} \; c^{\frac{1}{2k+1}}\left(\frac{1}{2}\int_{-1}^{1} g^2(c^{\frac{1}{2k+1}}u)du - \int_{-1}^{1} g(c^{\frac{1}{2k+1}}u)c^{\frac{k}{2k+1}}\left(u^k du + \frac{dW(u)}{\sqrt{c}}\right)\right).
\end{aligned}$$

If we set

$$g(c^{\frac{1}{2k+1}}u) = c^{\frac{k}{2k+1}}h(u)$$

then the problem is equivalent to minimizing

$$\left(\frac{1}{2}\int_{-1}^{1} c^{\frac{2k}{2k+1}}h^2(u)du - \int_{-1}^{1} c^{\frac{2k}{2k+1}}h(u)\left(u^k du + \frac{dW(u)}{\sqrt{c}}\right)\right)$$

or simply minimizing

$$\frac{1}{2}\int_{-1}^{1} h^2(u)du - \int_{-1}^{1} h(u)\left(u^k du + \frac{dW(u)}{\sqrt{c}}\right), \tag{3.10}$$

over the class of k-convex function on $[-1,1]$ satisfying

$$h(\pm 1) = 1,\; h''(\pm 1) = \frac{k!}{(k-2)!}, \cdots,\; h^{(k-2)}(\pm 1) = \frac{k!}{2!}. \tag{3.11}$$

With this new criterion function, the situation is very similar to the "finite sample" one. Indeed, as the Gaussian noise vanishes away at a rate of $1/\sqrt{c}$ as $c \to \infty$, one can view $t^k dt + dW(t)/\sqrt{c}$ as a "continuous" analogue to $d\mathbb{G}_n(\mathrm{t})$ ($\mathbb{G}_n$ being the empirical distribution) where the true k-monotone density is replaced by the k-convex function t^k. Existence and characterization of the minimizer of the criterion function in (3.10) follow from arguments that are very similar to the ones used in the original problem. Furthermore, if $\tilde{h}_c$ denotes the minimizer, we claim that the number of jump points of $\tilde{h}_c^{(k-1)}$ that are in the neighborhood of a fixed point t increases to infinity, and the distance between two successive jump points is of the order $c^{-1/(2k+1)}$ as $c \to \infty$. To establish this result, we need the following definition and lemma:

Definition 3.4.1 *Let f be a sufficiently differentiable function on a finite interval $[a,b]$, and $t_1 \le \cdots \le t_m$ be m points in $[a,b]$. The Lagrange interpolating polynomial is the unique polynomial P of degree $m-1$ which passes through $(t_1, f(t_1)), \cdots, (t_m, f(t_m))$. Furthermore, P is given by its Newton form*

$$P(t) = \sum_{j=1}^{m} f(t_j) \prod_{\substack{k=1\\ k\neq j}}^{m} \frac{(t-t_k)}{(t_j - t_k)}$$

or Lagrange form

$$P(t) = f(t_1) + (t-t_1)[t_1,t_2]f + \cdots + (t-t_1)\cdots(t-t_m)[t_1,\cdots,t_m]f$$

where $[x_1,\cdots,x_p]g$ denotes the divided difference of g of order p (see, e.g., de Boor (1978), Nürnberger (1989), DeVore and Lorentz (1993)).

Lemma 3.4.1 *Let g be an m-convex function on a finite interval $[a,b]$; i.e., $g^{(m-2)}$ exists and is convex on (a,b), and let $l_m(g,x,x_1,\cdots,x_m)$ be the Lagrange polynomial of degree $m-1$ interpolating g at the points x_i, $1 \le i \le m$, where $a < x_1 \le x_2 \le \cdots \le x_m < b$. Then*

$$(-1)^{m+i}\left(g(x) - l_m(g,x,x_1,\cdots,x_m)\right) \ge 0, \qquad x \in [x_i, x_{i+1}],\ i = 1,\cdots,m-1.$$

Proof. See, e.g., Ubhaya (1989), (a), page 235 or Kopotun and Shadrin (2003), Lemma 8.3, page 918. ■

The following lemma states consistency of the LS solution. It is very crucial for proving tightness of the distance between successive points of touch of $H_{c,k}$ and Y_k.

Lemma 3.4.2 *For $j \in \{0,\cdots,k-1\}$, we have*

$$\left|\tilde{h}_c^{(j)}(t) - \frac{k!}{(k-j)!}t^{k-j}\right| \to 0, \quad \textit{almost surely as } c \to \infty.$$

Proof. We will prove the result for $t = 0$ as the arguments are similar in the general case. Let us denote

$$\psi_c(h) = \frac{1}{2}\int_{-1}^{1} h^2(t)dt - \int_{-1}^{1} h(t)d\mathbb{H}_c(t)$$

where

$$d\mathbb{H}_c(t) = t^k dt + \frac{dW(t)}{\sqrt{c}}.$$

Since $\tilde{h}_c$ is the minimizer of ψ_c, then

$$\lim_{\epsilon \to 0} \frac{\psi(\tilde{h}_c + \epsilon\tilde{h}_c) - \psi(\tilde{h}_c)}{\epsilon} = 0$$

implying that

$$\int_{-1}^{1} \tilde{h}_c^2(t)dt = \int_{-1}^{1} \tilde{h}_c(t)d\mathbb{H}_c(t). \tag{3.12}$$

Also, for any k-convex function g defined on $(-1,1)$ that satisfies the boundary conditions in (3.11), we have

$$\lim_{\epsilon \searrow 0} \frac{\psi((1-\epsilon)\tilde{h}_c + \epsilon g) - \psi(\tilde{h}_c)}{\epsilon} \ge 0$$

and therefore

$$\int_{-1}^{1}(g(t)-\tilde{h}_c(t))\tilde{h}_c(t)dt - \int_{-1}^{1}(g(t)-\tilde{h}_c(t))d\mathbb{H}_c(t) \geq 0. \tag{3.13}$$

Let us denote $h_0(t) = t^k$, $dH_0(t) = h_0(t)dt$, and $d\tilde{H}_c(t) = \tilde{h}_c(t)dt$. If we take $g = h_0$ in (3.13), it follows that

$$\int_{-1}^{1}(\tilde{h}_c(t)-h_0(t))d(\tilde{H}_c(t)-\mathbb{H}_c(t)) \leq 0. \tag{3.14}$$

Now the equality in (3.12) can be rewritten as

$$\sqrt{\int_{-1}^{1}\tilde{h}_c^2(t)dt} = \int_{-1}^{1}\tilde{u}_c(t)d\mathbb{H}_c(t)$$

where $\tilde{u}_c = \tilde{h}_c/\|\tilde{h}_c\|_2$ is a k-convex function on $[-1,1]$ such that

$$\|\tilde{u}_c\|_2 = 1, \text{ and } \tilde{u}_c^{(2j)}(\pm 1) = \frac{k!}{(k-2j)!\|\tilde{h}_c\|_2} \quad \text{for } j = 0, \cdots, (k-2)/2.$$

We want to show that the function $\lim_{c\to\infty}\tilde{h}_c(t) = h_0(t)$ for all $t \in (-1,1)$. Let us take $c = c(n) = n$. We start by showing that the sequence $(\tilde{h}_n)_n$ is uniformly bounded on $(-1,1)$; i.e., there exists a constant $M > 0$ independent of n such that $\|\tilde{h}_n\|_\infty < M$ for all $n \in \mathbb{N}$. Suppose it is not. This implies that $(\tilde{h}_n^{(k-2)})_n$ is not bounded because if it was, we can find $M > 0$ such that for all $n > 0$,

$$|\tilde{h}_n^{(k-2)}(t)| \leq M,$$

for $t \in (-1,1)$. By integrating $\tilde{h}_n^{(k-2)}$ twice and using the boundary conditions at -1 and 1, it follows that

$$\tilde{h}_n^{(k-4)}(t) = \int_{-1}^{t}(t-s)h_n^{(k-2)}(s)ds - \left(\frac{1}{2}\int_{-1}^{1}(1-s)\tilde{h}_n^{(k-2)}(s)ds\right)(t+1) + \frac{k!}{2!}$$

and therefore

$$\|\tilde{h}_n^{(k-4)}\|_\infty \leq 2M + 2M + \frac{k!}{2!} = 4M + \frac{k!}{2!}.$$

By induction, it follows that $(\tilde{h}_n)_n$ has to be bounded. We conclude that $\tilde{h}_n^{(k-2)}$ is not bounded. Now, using convexity of $\tilde{h}_n^{(k-2)}$ and the same arguments of Proposition 3.3.1, this implies that we can find a subsequence $(\tilde{h}_{n'})_{n'}$ such that $\lim_{n'\to\infty}\|\tilde{h}_{n'}\|_2 = \infty$. Therefore,

$$\lim_{n'\to\infty}\tilde{u}_{n'}^{(2j)}(-1) = \lim_{n'\to\infty}\tilde{u}_{n'}^{(2j)}(1) = 0.$$

for $j \in \{0, \cdots, (k-2)/2\}$.

In the limit, the derivatives of $\tilde{u}_{n'}$ are "pinned down" at ± 1 and this implies that for large n', $\tilde{u}_{n'}^{(2j)}(\pm), j = 0, \cdots, (k-1)/2$ stay close to 0. On the other hand, we know that $\|\tilde{u}_{n'}\|_\infty = 1$. Therefore, the convex function $\tilde{u}_n^{(k-2)}$ has to be uniformly bounded by the same arguments of Proposition 3.3.1. It follows that there exists $M > 0$ such that $\|\tilde{u}_{n'}\|_\infty < M$. By Arzelà-Ascoli's theorem, we can find a subsequence $(\tilde{u}_{n''})_{n''}$ and a function $\tilde{u}$ such that

$$\lim_{n'' \to \infty} \tilde{u}_{n''}(t) = \tilde{u}(t)$$

for all $t \in (-1, 1)$. But since $\int_{-1}^1 |\tilde{u}| dH_0(t) \leq 2M/(k+1) < \infty$, it follows that

$$\lim_{n'' \to \infty} \int_{-1}^1 \tilde{u}_{n''}(t) d\mathbb{H}_{n''}(t) = \int_{-1}^1 \tilde{u}(t) dH_0(t) < \infty. \tag{3.15}$$

But recall that

$$\int_{-1}^1 \tilde{u}_{n''}(t) d\mathbb{H}_{n''}(t) = \|\tilde{h}_{n''}\|_2^2 \to \infty$$

as $n'' \to \infty$. Since this contradicts the result in (3.15), it follows that there exists $M > 0$ such that $\|\tilde{h}_n\|_\infty < M$.

Now, we can find a subsequence $(\tilde{h}_{n_l})_{n_l}$ and a function $\tilde{h}$ such that

$$\lim_{n_l \to \infty} \tilde{h}_{n_l}(t) = \tilde{h}(t)$$

for $t \in (-1, 1)$. By Fatou's lemma, we have

$$\int_{-1}^1 (\tilde{h}(t) - h_0(t))^2 dt \leq \liminf_{n_l \to \infty} \int_{-1}^1 (\tilde{h}_{n_l}(t) - h_0(t))^2 dt.$$

On the other hand, it follows from (3.14) that

$$\int_{-1}^1 (\tilde{h}_{n_l}(t) - h_0(t)) d(\tilde{H}_{n_l}(t) - \mathbb{H}_{n_l}(t)) \leq 0.$$

Thus we can write

$$\begin{aligned} &\int_{-1}^1 (\tilde{h}_{n_l}(t) - h_0(t))^2 dt \\ &= \int_{-1}^1 (\tilde{h}_{n_l}(t) - h_0(t)) d(\tilde{H}_{n_l}(t) - H_0(t)) \\ &= \int_{-1}^1 (\tilde{h}_{n_l}(t) - h_0(t)) d(\tilde{H}_{n_l}(t) - \mathbb{H}_{n_l}(t)) + \int_{-1}^1 (\tilde{h}_{n_l}(t) - h_0(t)) d(\mathbb{H}_{n_l}(t) - H_0(t)) \\ &\leq \int_{-1}^1 (\tilde{h}_{n_l}(t) - h_0(t)) d(\mathbb{H}_{n_l}(t) - H_0(t)) \to_{a.s.} 0, \quad \text{as } n_l \to \infty, \end{aligned}$$

since $\tilde{h}_{n_l} - h_0$ is bounded and $\int_{-1}^{1} h_0(t)dt < \infty$ (which implies that $\tilde{h}_{n_l} - h_0$ has an envelope $\in L_1(H_0)$). We conclude that

$$\int_{-1}^{1} (\tilde{h}(t) - h_0(t))^2 dt \leq 0$$

and therefore $\tilde{h} \equiv h_0$ on $(-1, 1)$. Since the choice $c(n) = n$ is irrelevant for the arguments above, we make the same conclusion with any other increasing sequence c_n such that $c_n \to \infty$. It follows that $\lim_{c\to\infty} \tilde{h}_c(t) = h_0(t)$. What should also be retained from the above arguments is the uniform boundedness of the derivatives of $\tilde{h}_c^{(l)}, l = 1, \cdots, k-2$. This is not guaranteed in general but k-convexity plays together with the fact that $\tilde{h}_c^{(2j}, j = 1, \cdots, (k-2)/2$ have fixed values at -1 and 1 play a crucial role. A proof of this fact follows from using induction and arguments that are similar to the ones used in the proof of Proposition 3.3.1.

Now, fix $t = 0$. We will show that we have also consistency of the derivatives of $\tilde{h}_c$. For that, consider $x_0, x_1, \cdots, x_{k-1} < 1$ to be k points such that $0 = x_0 \leq x_1 \leq \cdots \leq x_{k-1}$. By taking $m = k$ and $i = 2$ in Lemma 3.4.1, we have for all $t \in [x_1, x_2]$

$$\begin{aligned} \tilde{h}_c(t) &\geq \tilde{h}_c(x_0) + (t - x_0)\tilde{h}_c[x_0, x_1] \\ &\quad + \cdots + (t - x_0)(t - x_1) \cdots (t - x_{k-2})\tilde{h}_c[x_0, x_1, \cdots, x_{k-1}]. \end{aligned} \tag{3.16}$$

If we take $x_0 = x_1$, then the inequality in (3.16) can be rewritten as

$$\begin{aligned} \tilde{h}_c(t) &\geq \tilde{h}_c(x_0) + (t - x_0)\tilde{h}'_c(x_0) + (t - x_0)^2\tilde{h}_c[x_0, x_0, x_2] \\ &\quad + \cdots + (t - x_0)^2(t - x_2) \cdots (t - x_{k-2})\tilde{h}_c[x_0, x_0, x_2 \cdots, x_{k-1}] \end{aligned}$$

or equivalently

$$\begin{aligned} \tilde{h}'_c(x_0) &\leq \frac{\tilde{h}_c(t) - \tilde{h}_c(x_0)}{t - x_0} - (t - x_0)\Big(\tilde{h}_c[x_0, x_0, x_2] \\ &\quad + \cdots + (t - x_2) \cdots (t - x_{k-2})\tilde{h}_c[x_0, x_0, x_2 \cdots, x_{k-1}]\Big). \end{aligned}$$

since $t \geq x_0$. Furthermore, since $|\tilde{h}'_c(x_0)|$ is bounded, we can find a sequence $(\tilde{h}_n)_n$ such that the divided differences $\tilde{h}_n[x_0, x_0, x_2], \cdots, \tilde{h}_n[x_0, x_0, x_2, \cdots, x_{k-1}]$ converge to finite limits as $n \to \infty$. For instance, we have

$$\tilde{h}_n[x_0, x_0, x_2] = \frac{1}{x_2 - x_0}\left(\frac{\tilde{h}_n(x_2) - \tilde{h}_n(x_1)}{x_2 - x_0} - \tilde{h}'_n(x_0)\right).$$

If we denote $l(x_0) = \lim_{n\to\infty} \tilde{h}'_n(x_0)$, then

$$\lim_{n\to\infty} \tilde{h}_n[x_0, x_0, x_2] = \frac{1}{x_2 - x_0}\left(\frac{\tilde{h}_0(x_2) - \tilde{h}_0(x_1)}{x_2 - x_0} - l(x_0)\right).$$

The same reasoning can be applied for the remaining divided differences. By letting $n \to \infty$ and then $t \searrow x_0$, it follows that

$$\limsup_{n\to\infty} \tilde{h}'_n(x_0) \le h'_0(x_0); \text{ i.e.,}$$

$$\limsup_{n\to\infty} \tilde{h}'_n(0) \le h'_0(0).$$

Now, we need to exploit the inequality from above and for that consider $x_{-1} \le x_0 \le x_1 \le \cdots \le x_{k-2}$ to be k points, where $x_0 = 0$ and $x_1, \cdots, x_{k-2}$ can be taken to be the same as before. For all $t \in [x_1, x_2]$, we have

$$\begin{aligned} \tilde{h}_c(t) \quad &\le \quad \tilde{h}_c(x_{-1}) + (t - x_{-1})\ \tilde{h}_c[x_{-1}, x_0] \\ &\quad + \cdots + (t - x_{-1})(t - x_0)\cdots(t - x_{k-3})\ \tilde{h}_c[x_{-1}, x_0 \cdots, x_{k-2}]. \end{aligned}$$

In this case, we have $i = 3$ (see Lemma 3.4.1). If we take $x_{-1} = x_0 = x_1$, then for all $t \in [x_0, x_2]$ we have

$$\begin{aligned} \tilde{h}'_c(x_0) \quad &\ge \quad \frac{\tilde{h}_c(t) - \tilde{h}_c(x_0)}{t - x_0} - (t - x_0)\bigg((t - x_0)\ \frac{\tilde{h}''_c(x_0)}{2} \\ &\quad + \ \cdots + (t - x_0)^2 \cdots (t - x_{k-3})\ \tilde{h}_c[x_0, x_0, x_0 \cdots, x_{k-2}]\bigg). \end{aligned}$$

Using the fact that $|h''_c(x_0)|$ is bounded and the same reasoning as before, we obtain that

$$\liminf_{n\to\infty} \tilde{h}'_n(x_0) \ge h'_0(x_0); \text{ i.e.,}$$

$$\liminf_{n\to\infty} \tilde{h}'_n(0) \ge h'_0(0).$$

Combining both inequalities, we can write

$$h'_0(0) \le \liminf_{n\to\infty} \tilde{h}'_n(0) \le \limsup_{n\to\infty} \tilde{h}'_n(0) \le h'_0(0)$$

and hence $\lim_{c\to\infty} \tilde{h}_c'(0) = h_0'(0)$. An induction argument can be used to show that consistency holds true for $\tilde{h}_c^{(j)}(0), j = 2, \cdots, k-2$. As for the last derivative, we apply the well-known chord inequality satisfied by convex functions: For all $h > 0$, we have

$$\frac{\tilde{h}_c^{(k-2)}(0) - \tilde{h}_c^{(k-2)}(-h)}{-h} \leq \tilde{h}_c^{(k-1)}(0-) \leq \tilde{h}_c^{(k-1)}(0+) \leq \frac{\tilde{h}_c^{(k-2)}(h) - \tilde{h}_c^{(k-2)}(0)}{h}.$$

We obtain the result by letting $c \to \infty$ and then $h \searrow 0$. ■

Before we state the main lemma of this section, we give first a characterization for the minimizer $\tilde{h}_c$:

Lemma 3.4.3 *Let Y_c^1 be the process defined on $[-1, 1]$ by*

$$Y_c^1(t) \stackrel{d}{=} \begin{cases} \frac{1}{\sqrt{c}} \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} dW(s) + \frac{k!}{(2k)!} t^{2k}, & \text{if } t \in [0,1] \\ \frac{1}{\sqrt{c}} \int_t^0 \frac{(t-s)^{k-1}}{(k-1)!} dW(s) + \frac{k!}{(2k)!} t^{2k}, & \text{if } t \in [-1,0) \end{cases}$$

and H_c^1 be the k-fold integral of $\tilde{h}_c$ that satisfies the boundary conditions

$$\frac{d^{2j} H_c^1}{dt^{2j}}|_{t=\pm c} = \frac{d^{2j} Y_c^1}{dt^{2j}}|_{t=\pm c},$$

for $j = 0, \cdots, (k-2)/2$. The minimizer $\tilde{h}_c$ is characterized by the conditions:

$$H_c^1(t) \geq Y_c^1(t), \quad \text{for all } t \in [-1, 1]$$

and

$$\int_{-1}^1 \left(H_c^1(t) - Y_c^1(t)\right) d\tilde{h}_c^{(k-1)}(t) = 0.$$

Proof. The arguments are very similar to those used in the proof of Lemma 3.3.2. ■

Lemma 3.4.4 *Let t be a fixed point in $(-1, 1)$ and suppose that the conjectured Lemma 2.5.4 holds. If τ_c^- and τ_c^+ are the last (first) point of touch between of H_c^1 and Y_c^1 before (after) t, then*

$$\tau_c^+ - \tau_c^- = O_p(c^{-1/(2k+1)}).$$

Proof. As the minimization problem was changed so that the setting is very similar to that of the LS problem for estimating a k-monotone density (see Chapter 2), we can apply the result obtained in Lemma 2.5.9. In fact, consistency of $\tilde{h}_c^{(k-1)}$ at the point t and the fact that $h_0(t) = t^k$ is k-times differentiable with $h_0^{(k)}(t) = k! > 0$ force the number of points of change of slope of $\tilde{h}_c^{(k-2)}$ to increase to infinity almost surely as $c \to \infty$. If $\tau_{c,0} < \cdots < \tau_{c,2k-3}$ are $2k-2$ jump points of $\tilde{h}_c^{(k-1}$ that are in a small neighborhood of t, then H_c^1 is a polynomial spline of degree $2k-1$ and simple knots $\tau_{c,0}, \cdots, \tau_{c,2k-3}$. Furthermore, $\tilde{H}_c$ is the unique solution of the following Hermite problem:

$$H_c^1(\tau_j) = Y_c^1(\tau_j), \text{ and } (H_c^1)'(\tau_j) = (Y_c^1)'(\tau_j)$$

for $j = 0, \cdots, 2k-3$. By Lemma 2.5.9, it follows that

$$\tau_{c,2k-3} - \tau_{c,0} = O_p(c^{-1/(2k+1)}).$$

As we are free to choose $\tau_{c,2k-3}$ and $\tau_{c,0}$ to be located to the left and right of t (as long as they are in a small neighborhood of t), it follows that

$$\tau_c^+ - \tau_c^- = O_p(c^{-1/(2k+1)}).$$

■

Corollary 3.4.2 *Let t be a fixed point in $(-c, c)$. If τ_c^- and τ_c^+ now denote the last (first) point of touch between of H_c and Y_c before (after) t, then*

$$\tau_c^+ - \tau_c^- = O_p(1),$$

and hence for any $\epsilon > 0$ there exists $M = M(\epsilon) > 0$ such that

$$\limsup_{c \to \infty} P(\tau_c^+ - t > M \quad or \quad t - \tau^- > M) \le \epsilon.$$

Proof. Recall that

$$g(c^{1/(2k+1)}t) = c^{k/(2k+1)}h(t), \quad \text{for all } t \in [-1, 1]$$

where g and h belong to the k-convex class defined in the original and new minimization problems respectively. Therefore, if t_c^- and t_c^+ are two successive jump points of $\tilde{h}_c^{(k-1)}$ in the neighborhood of some fixed point $t \in (-1, 1)$, then $\tau_c^- = c^{1/(2k+1)} t_c^-$ and $\tau_c^+ = c^{1/(2k+1)} t_c^+$ are successive jump points of $\tilde{g}_c^{(k-1)}$. Therefore,

$$\tau_c^+ - \tau_c^- = c^{1/(2k+1)}(t_c^+ - t_c^-) = O_p(1).$$

■

Remark 3.4.1 *Despite the complexity of the tightness problem for $k > 2$, we can view it in a simple heuristic way. Recall that in the original Gaussian problem defined in (2.5), we want to "estimate" the k-convex function $t \mapsto t^k$. The Least Squares estimate on a finite interval $[-c, c]$ is a spline of degree $k - 1$ whose knots are exactly the points of touch of the process $H_{c,k}$ with respect to Y_k. As $c \to \infty$, we expect that the Least Squares estimator to be close to the estimated function. Since the latter is infinitely differentiable, the knots of the estimator need to stay tight in order to "compensate" the difference of smoothness.*

Lemma 3.4.5 *Let $c > 0$ and $H_{c,k}$ be the k-fold integral of $f_{c,k}$ the minimizer of Φ_c over the class $\mathcal{C}_{k,\underline{m}_1,\underline{m}_2}$ (resp. $\mathcal{C}_{k,m_0,\underline{m}_1,\underline{m}_2}$) with $\underline{m}_1 = \underline{m}_2 = \left((k!/2!)c^2, \cdots, (k!/(k-2)!)c^{k-2}\right)$ (resp. $m_0 = c^k$, $\underline{m}_1 = \underline{m}_2 = \left((k!/2!)c^2, \cdots, (k!/(k-1)!)c^{k-1}\right)$) if k is even (resp. odd) . Then, for a fixed $t \in \mathbb{R}$, the collections $\{f_{c,k}^{(j)}(t) - f_0^{(j)}(t)\}_{c,k \geq |t|}, j = 0, \cdots, k-1$ are tight; here $f_{c,k}^{(k-1)}$ can either be the right or left $(k-1)$-st derivative of f_c.*

Proof. We will prove the lemma for k is even and $t = 0$ (the cases k odd or $t \neq 0$ can be handled similarly). We start with $j = 0$. Fix $\epsilon > 0$ and denote $\Delta = H_c - Y_k$. By Corollary 3.4.2 and for c large enough, there exist $M > 0$ and a point of touch of $\tau_1 \in [M, 3M]$ with probability greater than $1 - \epsilon$. Applying the same reasoning, there exists $M > 0$ (maybe at the cost of increasing M) such that we can find points of touch $\tau_2 \in [4M, 6M]$, $\tau_3 \in [7M, 9M]$, $\cdots$, $\tau_{2^{k-1}} \in \left[\left(3 \cdot 2^{k-1} - 2\right) M, 3 \cdot 2^{k-1} M\right]$ with probability greater than $1 - \epsilon$. Since at any point of touch τ, $\Delta'(\tau) = 0$, then by the mean value theorem, there exist $\tau_1^{(2)} \in (\tau_1, \tau_2)$, $\tau_2^{(2)} \in (\tau_3, \tau_4)$, $\cdots$, $\tau_{2^{k-2}}^{(2)} \in (\tau_{2^{k-1}-1}, \tau_{2^{k-1}})$ such that $\Delta^{(2)}(\tau_j^{(2)}) = 0, j = 1, \cdots, 2^{k-2}$. By applying the mean value theorem successively $k-3$ more times, we can find $\tau_1^{(k-1)} < \tau_2^{(k-1)} \in [M, 3 \cdot 2^{k-1} M]$ such that $\Delta^{(k-1)}(\tau_i^{(k-1)}) = 0, i = 1, 2$ and $\tau_2^{(k-1)} - \tau_1^{(k-1)} \geq M$. Finally, there exists $\tau^{(k)} = \xi_1 \in (\tau_1^{(k-1)}, \tau_2^{(k-1)})$ such that

$$
\begin{aligned}
f_{c,k}^{(k)}(\xi_1) &= H_c^{(k)}(\xi_1) \\
&= \frac{H_{c,k}^{(k-1)}(\tau_2^{(k-1)}) - H_{c,k}^{(k-1)}(\tau_1^{(k-1)})}{\tau_2^{(k-1)} - \tau_1^{(k-1)}} \\
&= \frac{Y_k^{(k-1)}(\tau_2^{(k-1)}) - Y_k^{(k-1)}(\tau_1^{(k-1)})}{\tau_2^{(k-1)} - \tau_1^{(k-1)}} \\
&= \frac{W(\tau_2^{(k-1)}) - W(\tau_1^{(k-1)})}{\tau_2^{(k-1)} - \tau_1^{(k-1)}} + \frac{1}{k+1} \frac{\left(\tau_2^{(k-1)}\right)^{k+1} - \left(\tau_1^{(k-1)}\right)^{k+1}}{\tau_2^{(k-1)} - \tau_1^{(k-1)}}
\end{aligned}
$$

and therefore

$$
\begin{aligned}
|f_{c,k}^{(k)}(\xi_1)| &\leq \frac{\left|W(\tau_2^{(k-1)}) - W(\tau_1^{(k-1)})\right|}{M} + \left(3\, 2^{k-1} M\right)^k \\
&\leq \frac{C}{M} + \left(3 \cdot 2^{k-1} M\right)^k
\end{aligned}
$$

for some constant $C = C(M) > 0$ by tightness of W and stationarity of its increments, and using the fact that $y^{k+1} - x^{k+1} = (y-x)(x^k + x^{k-1}y + \cdots + y^k)$. In general, we can find $k-2$ points $\xi_1 < \cdots < \xi_{k-2}$ to the right of 0 such that $\xi_1 \in [M, 3M]$, the distance between any ξ_i and $\xi_j, i \neq j$ is at least M and $f_{c,k}(\xi_i)$ is tight for $i = 1, \cdots, k-2$. Similarly and this time to the left of 0, we can find two points of touch $\xi_{-2} < \xi_{-1}$ such that $\xi_{-1} \in [-3 \cdot 2^{k-1} M, -M]$, $\xi_{-1} - \xi_{-2} \geq M$ and $f_{c,k}(\xi_{-1})$ and $f_c(\xi_{-2})$ are tight. In total, we have k points that are at least M-distant from each other and we are ready to apply Lemma 3.4.1. Hence, if we take $g = f_{c,k}$, $m = k$, $i = 2$, and $x_1 = \xi_{-2}$, $x_2 = \xi_{-1}$, $x_3 = \xi_1$, $\cdots$, $x_k = \xi_{k-2}$, we have for all $t \in (\xi_{-1}, \xi_1)$

$$
\begin{aligned}
f_{c,k}(t) \;\geq\; & f_{c,k}(\xi_{-2}) + (t - \xi_{-2})\ [\xi_{-2}, \xi_{-1}] f_{c,k} + (t - \xi_{-2})(t - \xi_{-1})\ [\xi_{-2}, \xi_{-1}, \xi_1] f_{c,k} \\
& + \cdots + (t - \xi_{-2})(t - \xi_{-1}) \cdots (t - \xi_{k-3})\ [\xi_{-2}, \xi_{-1} \cdots, \xi_{k-2}] f_{c,k}.
\end{aligned}
$$

In particular, when $t=0$ we have

$$\begin{aligned} f_{c,k}(0) \quad &\geq \quad f_{c,k}(\xi_{-2}) - \xi_{-2}\ [\xi_{-2},\xi_{-1}]f_{c,k} + \xi_{-2}\xi_{-1}\ [\xi_{-2},\xi_{-1},\xi_1]f_{c,k} \\ &\quad + \cdots + (-1)^{k-1}\xi_{-2}\xi_{-1}\cdots\xi_{k-3}\ [\xi_{-2},\xi_{-1},\cdots,\xi_{k-2}]f_{c,k} \end{aligned}$$

which is tight by construction of ξ_i, $i=-2,-1,1,\cdots,k-2$. Now, by adding a point ξ_{k-1} to the right and ξ_{k-2} such that $\xi_{k-1}-\xi_{k-2} \geq M$ and considering the points $\xi_{-1},\xi_1,\cdots,\xi_{k-1}$, we apply Lemma 3.4.1 (with $i=1$) to bound $f_{c,k}(0)$ by above:

$$\begin{aligned} f_{c,k}(0) \quad &\leq \quad f_{c,k}(\xi_{-1}) - \xi_{-1}\ [\xi_{-1},\xi_1]f_{c,k} + \xi_{-1}\xi_1[\xi_{-1},\xi_1,\xi_2]f_{c,k} \\ &\quad + \cdots + \ (-1)^{k-1}\xi_{-1}\xi_1\cdots\xi_{k-2}[\xi_{-1},\xi_1,\cdots,\xi_{k-1}]f_{c,k} \end{aligned}$$

which is again tight.

Now if $j=1,\cdots,k-3$, the argument is entirely similar where $k-j$ is the number of points of touch needed to prove tightness. For $j=k-2$, we can bound $f_{c,k}^{(k-2)}(0)$ from above by considering two points of touch $\xi_{-1}\leq -M$ and $M\leq \xi_1$ and using convexity of $f_{c,k}^{(k-2)}$ (which follows also from Lemma 3.4.1 in the particular case where g is convex). To bound $f_{c,k}^{(k-2)}(0)$ from below, we use a similar argument as in the proof of Proposition 3.3.1. Finally, for $j=k-1$, consider again ξ_{-1} and ξ_1. By convexity of $f_{c,k}^{(k-2)}$, we have

$$\frac{f_{c,k}^{(k-2)}(0) - f_{c,k}^{(k-2)}(\xi_{-1})}{\xi_{-1}} \leq f_{c,k}^{(k-1)}(0-) \leq f_{c,k}^{(k-1)}(0+) \leq \frac{f_{c,k}^{(k-2)}(\xi_1) - f_{c,k}^{(k-2)}(0)}{\xi_1}$$

hence,

$$|f_{c,k}^{(k-1)}(0)| \leq \max\left\{\left|\frac{f_{c,k}^{(k-2)}(0) - f_{c,k}^{(k-2)}(\xi_{-1})}{\xi_{-1}}\right|, \left|\frac{f_{c,k}^{(k-2)}(\xi_1) - f_{c,k}^{(k-2)}(0)}{\xi_1}\right|\right\}$$

which is bounded with large probability by tightness of $f_{c,k}^{(k-2)}(t), t\in(-c,c)$ and construction of ξ_{-1} and ξ_1. ∎

3.5 Proof of Theorem 3.2.1

We use similar arguments as in the proof of Theorem 2.1 in Groeneboom, Jongbloed, and Wellner (2001a) and for convenience, we adopt their notation. We assume here that

k is even since the arguments are very similar for k odd. For $m > 0$ fixed, consider the semi-norm

$$\|H\|_m = \sup_{t \in [-m,m]} \{|H(t)| + |H'(t)| + \cdots + |H^{(2k-2)}(t)|\}$$

on the space of $(2k-2)-$continuously differentiable functions defined on $\mathbb{R}$. By Lemma 3.4.5, we know if we take $c(n) = n$ that the collection $\{f_{n,k}^{(k-2)}(t) - f_0^{(k-2)}(t)\}_{n>M}$ is tight for any fixed $t \in [-M, M]$, in particular for $t = 0$. Furthermore, by the same lemma, we know that the collections $\{f_{n,k}^{(k-1)}(t-)\}$ and $\{f_{n,k}^{(k-1)}(t+)\}$ are also tight for $t \in [-M, M]$. By monotonicity of $f_{n,k}^{(k-1)}$, it follows that the sequence $\left(f_{n,k}^{(k-2)}\right)$ has uniformly bounded derivatives on $[-M, M]$. Therefore, by Arzelà-Ascoli, the sequence $\left(f_{n,k}^{(k-2)}|[-M, M]\right)$ has a subsequence $\left(f_{n_l,k}^{(k-2)}|[-M, M]\right) \equiv \left(H_{n_l,k}^{(2k-2)}|[-M, M]\right)$ converging in the supremum metric on $C[-M, M]$ to a bounded convex function on $[-M, M]$. By the same theorem, we can find a further subsequence $\left(H_{n_p,k}^{(2k-3)}|[-M, M]\right)$ converging in the same metric to a bounded function on $[-M, M]$. Applying Arzelà-Ascoli $(2k-3)$ times, we can find a further subsequence $\left(H_{n_q,k}|[-M, M]\right)$ that converges in the supremum metric on $C[-M, M]$.

Now, fix m in $\mathbb{N}$ and let $n > m$. For any sequence $(H_{n,k})$, we can find a subsequence $(H_{n_j,k})$ so that $(H_{n_j,k}|[-m, m])$ converges in the metric $\|H\|_m$ to a limit $H_k^{(m)}$ that is $(2k)-$convex on $[-m, m]$; i.e., its $(2k-2)$-th derivative, $f_k^{(m)}$, is convex on $[-m, m]$. Finally, by a diagonal argument, we can extract from any sequence $(H_{n,k})$ a subsequence $(H_{n_j,k})$ converging to a limit H_k in the topology induced by the semi-norms $\|H\|_m, m \in \mathbb{N}$. The limit H_k is clearly $2k$-convex. Besides, it preserves by construction the properties (3.10) and (3.11) in the characterization of $H_{n,k} \equiv H_{c(n),k}$. On the other hand, since $H_{n,k}^{(j)}(\pm c) = Y_k^{(j}(\pm c)$ for $j = 0, 2, \cdots, k$, it follows that $\lim_{|t|\to\infty} H_k^{(j)}(t) - Y_k^{(j)}(t) = 0$ for $j = 0, 2, \cdots, k$. Thus H_k satisfies the conditions (i)-(iv) of Theorem 3.2.1. It remains only to show that this process is unique. ■

To prove uniqueness of H_k, we need the following lemma:

Lemma 3.5.1 *Let G_k be a $2k$-convex function on $\mathbb{R}$ that satisfies*

$$\lim_{|t|\to\infty} (G_k^{(k-2)}(t) - Y_k^{(k-2)}(t)) = 0$$

if k is even, and

$$\lim_{|t|\to\infty}(G_k^{(k-3)}(t) - Y_k^{(k-3)}(t)) = 0$$

if k is odd. Let $g_k = G_k^{(k)}$ and fix $\epsilon > 0$. Then,

(i) For any fixed $M_2 \geq M_1 > 0$, and a and b such that $|a| < |b|$ are large enough and $M_2 \geq |b| - |a| \geq M_1$, we can find a positive constant $K = K(\epsilon, M_1, M_2)$ such that

$$P(\|G_k^{(j)} - Y_k^{(j)}\|_{[a,b]} > K) \leq \epsilon$$

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

(ii) For any fixed $M_2 \geq M_1 > 0$, and a and b such that $|a| < |b|$ are large enough and $M_2 \geq |b| - |a| \geq M_1$, we can find a positive constant $K = K(\epsilon, M_1, M_2)$ such that

$$P(\|g_k^{(j)} - f_{0,k}^{(j)}\|_{[a,b]} > K) \leq \epsilon$$

for $j = 0, \cdots, k-1$, where $f_{0,k}(t) = t^k$.

Proof. We develop the arguments only in the case of k even (k odd can be handled similarly). We start by proving (ii) and for that we fix $\delta > 0$. Without loss of generality, we can take $M_1 = M_2 = M$. Since $\lim_{t\to\infty}(G_k^{(k-2)}(t) - Y_k^{(k-2)}(t)) = 0$, then there exists $A > 0$ such that

$$|G_k^{(k-2)}(t) - Y_k^{(k-2)}(t)| < \delta$$

for all $t > A$. Let $t_0 > A$ and $t_1 = t_0 + M$, and $t_2 = t_0 + 2M$, where M is some positive constant. By the mean value theorem, there exists $\xi \in (t_0, t_1)$ such that

$$G_k^{(k-1)}(\xi) - Y_k^{(k-1)}(\xi) = \frac{(G_k^{(k-2)}(t_1) - Y_k^{(k-2)}(t_1)) - (G_k^{(k-2)}(t_0) - Y_k^{(k-2)}(t_0))}{t_1 - t_0} \tag{3.17}$$

and hence

$$\left|G_k^{(k-1)}(\xi) - Y_k^{(k-1)}(\xi)\right| \leq \frac{2\delta}{M}.$$

From now on, we take $\delta = 1$. For all $t \in [t_1, t_2]$, we can write

$$
\begin{aligned}
G_k^{(k-2)}(t) - Y_k^{(k-2)}(t) &= G_k^{(k-2)}(t_1) - Y_k^{(k-2)}(t_1) + \int_{t_1}^{t} (G^{(k-1)}(s) - Y_k^{(k-1)}(s))ds \\
&= G_k^{(k-2)}(t_1) - Y_k^{(k-2)}(t_1) + \int_{t_1}^{t}\int_{\xi}^{s} d(G^{(k-1)}(u) - Y_k^{(k-1)}(u))ds \\
&\quad +(t-t_1)(G_k^{(k-1)}(\xi) - Y_k^{(k-1)}(\xi)) \\
&= G_k^{(k-2)}(t_1) - Y_k^{(k-2)}(t_1) + \int_{t_1}^{t}\int_{\xi}^{s} (g_k(u) - f_{0,k}(u))duds \\
&\quad + \int_{t_1}^{t}\int_{\xi}^{s} dW(u)ds + (t-t_1)(G_k^{(k-1)}(\xi) - Y_k^{(k-1)}(\xi))
\end{aligned} \tag{3.18}
$$

and hence

$$
\inf_{t\in[t_0,t_2]} |g_k(t) - f_{0,k}(t)| < \frac{8(6 + M\ C/2)}{M^2} \tag{3.19}
$$

where $C = C(M, \epsilon)$ such that

$$
P(|W(t)| < C,\ t \in [0, 2M]) > 1 - \epsilon.
$$

Indeed, from (3.18), we have for all $t \in [(t_1 + t_2)/2, t_2]$

$$
\begin{aligned}
&\Big|\int_{t_1}^{t}\int_{\xi}^{s} (g_k(u) - f_{0,k}(u))duds\Big| \\
&\le \Big|G_k^{(k-2)}(t) - Y_k^{(k-2)}(t)\Big| + \Big|G_k^{(k-2)}(t_1) - Y_k^{(k-2)}(t_1)\Big| \\
&\quad + \int_{t_1}^{t} |W(s) - W(\xi)|dsdu + (t-t_1)\Big|G_k^{(k-1)}(\xi) - Y_k^{(k-1)}(\xi)\Big| \\
&\le 2 + (t-t_1)\ C + 2M\frac{2}{M}, \quad \text{using stationarity of the increments of } W \\
&= 6 + M\ C/2
\end{aligned} \tag{3.20}
$$

with probability greater than $1 - \epsilon$. Now, since

$$
\begin{aligned}
\inf_{y\in[t_0,t_2]} |g_k(y) - f_{0,k}(y)| &\le \Big|\int_{t_1}^{t}\int_{\xi}^{s} (g_k(u) - f_{0,k}(u))duds\Big| \Big/ \int_{t_1}^{t}\int_{\xi}^{s} duds \\
&\le \Big|\int_{t_1}^{t}\int_{\xi}^{s} (g_k(u) - f_{0,k}(u))duds\Big| \Big/ \int_{t_1}^{t}\int_{t_1}^{s} duds, \quad \text{since } \xi \le t_1 \\
&= 2\Big|\int_{t_1}^{t}\int_{\xi}^{s} (g_k(u) - f_{0,k}(u))duds\Big| /(t-t_1)^2
\end{aligned}
$$

$$\leq \frac{8}{M^2}\left|\int_{t_1}^{t}\int_{\xi}^{s}(g_k(u)-f_{0,k}(u))duds\right|, \quad \text{since } t-t_1 \geq M/2, \tag{3.21}$$

the inequality in (3.19) follows by combining (3.20) and (3.21).

Now, consider two other points to the left of t_2, $t_3 = t_0 + 3M$ and $t_4 = t_0 + 4M$. By using similar arguments, we can find $\xi_0 \in [t_0, t_2]$ and $\xi_1 \in (t_2, t_3)$ such that

$$\left|g_0(\xi_0) - f_{0,k}(\xi_0)\right| = \inf_{u\in[t_0,t_2]}\left|g_0(u) - f_{0,k}(u)\right|$$

and

$$G_k^{(k-1)}(\xi_1) - Y_k^{(k-1)}(\xi_1) = \frac{(G_k^{(k-2)}(t_3) - Y_k^{(k-2)}(t_3)) - (G_k^{(k-2)}(t_2) - Y_k^{(k-2)}(t_2))}{t_3 - t_2}.$$

For $t \in [(t_3+t_4)/2, t_4]$, we can write

$$\begin{aligned} G_k^{(k-2)}(t) - Y_k^{(k-2)}(t) &= G_k^{(k-2)}(t_3) - Y_k^{(k-2)}(t_3) + \int_{t_3}^{t}\int_{\xi_1}^{s}\int_{\xi_0}^{u}(g_k'(y) - f_{0,k}'(y))dyduds \\ &\quad +(g_k'(\xi_0) - f_{0,k}'(\xi_0))\int_{t_3}^{t}\int_{\xi_1}^{s}duds + \int_{t_3}^{t}\int_{\xi_1}^{s}dW(u)ds \\ &\quad +(t-t_3)(G_k^{(k-1)}(\xi_1) - Y_k^{(k-1)}(\xi_1)). \end{aligned}$$

As argued above, we can find a constant $D > 0$ depending on M and ϵ such that

$$\inf_{u\in[t_0,t_4]}\left|g_k'(u) - f_{0,k}'(u)\right| < D$$

with probability greater than $1-\epsilon$. By induction, we can show that there exist an integer $p_k > 0$ and a constant $D_k > 0$ depending on M and ϵ such that

$$\inf_{u\in[t_0,t_{p_k}]}\left|g_k^{(k-2)}(u) - f_{0,k}^{(k-2)}(u)\right| < D_k$$

with probability greater than $1-\epsilon$ and where $t_{p_k} = t_0 + p_k M$.

By repeating the arguments above, we can find $\xi_{k,1} \in [t_0, t_{p_k}]$ and and $\xi_{k,2} \in [t_{p_k} + M, t_{2p_k} + M]$ (maybe at the cost of increasing t_0) such that

$$\left|g_k^{(k-2)}(\xi_{k,1}) - f_{0,k}^{(k-2)}(\xi_{k,1})\right| = \inf_{u\in[t_0,t_{p_k}]}\left|g_k^{(k-2)}(u) - f_{0,k}^{(k-2)}(u)\right|$$

and

$$\left|g_k^{(k-2)}(\xi_{k,2}) - f_{0,k}^{(k-2)}(\xi_{k,2})\right| = \inf_{u\in[t_{p_k}+M,t_{2p_k}+M]} \left|g_k^{(k-2)}(u) - f_{0,k}^{(k-2)}(u)\right|.$$

On the other hand, we can assume (at the cost of increasing t_0) that $t_0 - M > A$. By assumption, G_k is $2k$-convex and hence $g_k^{(k-2)}$ is convex. It follows that, for $t \in [t_0 - M, t_0]$, we have

$$\begin{aligned} g_k^{(k-1)}(t) &\leq \frac{g_k^{(k-2)}(\xi_{k,2}) - g_k^{(k-2)}(\xi_{k,1})}{\xi_{k,2} - \xi_{k,1}} \\ &\leq \frac{f_{0,k}^{(k-2)}(\xi_{k,2}) - f_{0,k}^{(k-2)}(\xi_{k,1}) + 2D_k}{\xi_{k,2} - \xi_{k,1}} \\ &\leq f_{0,k}^{(k-1)}(\xi_{k,2}) + \frac{2D_k}{M}, \end{aligned}$$

where $g_k^{(k-1)}$ is either the left or left $(k-1)$-st derivative. Therefore,

$$\begin{aligned} g_k^{(k-1)}(t) - f_{0,k}^{(k-1)}(t) &\leq f_{0,k}^{(k-1)}(\xi_{k,2}) - f_{0,k}^{(k-1)}(t) + \frac{2D_k}{M} \\ &= k!(\xi_{k,2} - t) + \frac{2D_k}{M} \\ &= k!(\xi_{k,2} - t_0 + t_0 - t) + \frac{2D_k}{M} \\ &\leq k!(p_k + 1)\ M + \frac{2D_k}{M}. \end{aligned}$$

Similarly, at the cost of increasing t_0 or D_k (or both), we can find t_{-p_k}, and $\xi_{k,-2} < \xi_{k,-1}$ to the left of $t_0 - M$ such that

$$\left|g_k^{(k-2)}(\xi_{k,-1}) - f_{0,k}^{(k-2)}(\xi_{k,-1})\right| = \inf_{u\in[t_{-p_k},t_0]} \left|g_k^{(k-2)}(u) - f_{0,k}^{(k-2)}(u)\right| < D_k$$

and

$$\left|g_k^{(k-2)}(\xi_{k,-2}) - f_{0,k}^{(k-2)}(\xi_{k,-2})\right| = \inf_{u\in[t_{-2p_k},t_{-p_k}-M]} \left|g_k^{(k-2)}(u) - f_{0,k}^{(k-2)}(u)\right| < D_k.$$

It follows that,

$$\begin{aligned} g_k^{(k-1)}(t) &\geq \frac{g_k^{(k-2)}(\xi_{k,-1}) - g_k^{(k-2)}(\xi_{k,-2})}{\xi_{k,-1} - \xi_{k,-2}} \\ &\geq \frac{f_{0,k}^{(k-2)}(\xi_{k,-2}) - f_{0,k}^{(k-2)}(\xi_{k,-1}) - 2D_k}{\xi_{k,-1} - \xi_{k,-2}} \\ &\geq f_{0,k}^{(k-1)}(\xi_{k,-2}) - \frac{2D_k}{M} \end{aligned}$$

and therefore,

$$\begin{aligned}
g_k^{(k-1)}(t) - f_{0,k}^{(k-1)}(t) &\geq f_{0,k}^{(k-1)}(\xi_{k,-2}) - f_{0,k}^{(k-1)}(t) - \frac{2D_k}{M} \\
&= k!(\xi_{k,-2} - t) - \frac{2D_k}{M} \\
&= -k!(-\xi_{k,-2} + (t_0 - M) - (t_0 - M) + t) - \frac{2D_k}{M} \\
&\geq -k!(p_k + 1)\ M - \frac{2D_k}{M}.
\end{aligned}$$

It follows that

$$\|g_k^{(k-1)} - f_{0,k}^{(k-1)}\|_{[t_0-M,t_0]} \leq k!(p_k+1)\ M + \frac{2D_k}{M}$$

with probability greater than $1-\epsilon$.

By applying the same arguments above (maybe at the cost of increasing either p_k or t_0), we can find a constant $C_k > 0$ depending only on M and ϵ such that

$$\|g_k^{(k-1)} - f_{0,k}^{(k-1)}\|_{[t_{-p_k}-M,t_{p_k}+M]} < C_k.$$

But, we can write

$$g_k^{(k-2)}(t) - f_{0,k}^{(k-2)}(t) = g_k^{(k-2)}(\xi_{k,-1}) - f_{0,k}^{(k-2)}(\xi_{k,-1}) + \int_{\xi_{k,-1}}^{t} (g_k^{(k-1)}(s) - f_{0,k}^{(k-1)}(s))ds$$

for all $t \in [t_{-p_k} - M, t_{p_k} + M]$. It follows that

$$\begin{aligned}
\left| g_k^{(k-2)}(t) - f_{0,k}^{(k-2)}(t) \right| &\leq D_k + (t - \xi_{k,-1})C_k \\
&\leq D_k + 2M(1+p_k)C_k
\end{aligned}$$

for $t \in [t_{-p_k} - M, t_{p_k} + M]$, or

$$\|g_k^{(k-2)} - f_{0,k}^{(k-2)}\|_{[t_{-p_k}-M,t_{p_k}+M]} < D_k + 2M(1+p_k)C_k$$

with probability greater than $1-\epsilon$. By induction, we can prove that there exists $K_k > 0$ depending only on M and ϵ such that

$$\|g_k^{(j)} - f_{0,k}^{(j)}\|_{[t_{-p_k}-M,t_{p_k}+M]} < K_k$$

for $j = 0, \cdots, k-3$.

Now to prove (i) for $j = k-1$, we consider again $[t_0, t_1]$ and $\xi \in (t_0, t_1)$ given by (3.17). We write

$$\begin{aligned} G_k^{(k-1)}(t) - Y_k^{(k-1)}(t) &= G_k^{(k-1)}(\xi) - Y_k^{(k-1)}(\xi) + \int_\xi^t d(G_k^{(k-1)}(s) - Y_k^{(k-1)}(s)) \\ &= G_k^{(k-1)}(\xi) - Y_k^{(k-1)}(\xi) + \int_\xi^t (g_k(s) - f_{0,k}(s))ds + W(t) - W(\xi), \end{aligned}$$

for $t \in [t_0, t_1]$. It follows that

$$\begin{aligned} \|G_k^{(k-1)}(t) - Y_k^{(k-1)}\|_{[t_0,t_1]} &\leq \frac{2}{M} + K(t-\xi) + C \\ &\leq \frac{2\delta}{M} + KM + C, \end{aligned}$$

with probability greater than $1-\epsilon$, where K is the constant given in (i) and $C > 0$ satisfies

$$P(|W(u)| > C,\ u \in [0, M]) \leq \epsilon.$$

For $0 \leq j \leq k-2$, the result follows using induction. ■

When $G_k \equiv H_k$, then we can prove a result that is stronger than that of Lemma 3.5.1:

Lemma 3.5.2 *Let H_k be the stochastic process constructed in the proof of Theorem 3.2.1. Let $f_{0,k}$ be again the function defined on $\mathbb{R}$ by*

$$f_{0,k}(t) = t^k,$$

and $a < b$ in $\mathbb{R}$. Then for any fixed $0 < \epsilon < 1$):

(i) *There exists an $M = M_\epsilon$ independent of t such that*

$$P(t - \tau^- > M, \tau^+ - t > M) < \epsilon$$

where τ^- and τ^+ are respectively the last point of touch of H_k and Y_k before t and the first point of touch after t.

(ii) *There exists an M depending only on $b-a$ and ϵ such that for $j = 0, \cdots, k-1$*

$$P(\|H_k^{(j)} - Y_k^{(j)}\|_{[a,b]} > M) < \epsilon, \tag{3.22}$$

(iii) *There exists an M depending only on $b-a$ and ϵ such that for $j = k, \cdots, 2k-1$*

$$P(\|H_k^{(j)} - f_{0,k}^{(j)}\|_{[a,b]} > M) < \epsilon, \tag{3.23}$$

where $H_k^{(2k-1)}$ denotes either the left or the right $(2k-1)$-th derivative of H_k. When $j = k$, (3.23) specializes to

$$P(\|f_k - f_{0,k}\|_{[a,b]} > M) < \epsilon,$$

where $f_k = H_k^{(k)}$.

To prove the above lemma, we need the following result:

Lemma 3.5.3 *Let $\epsilon > 0$ and $x \in \mathbb{R}$. We can find $M > 0$, $K > 0$, $D > 0$ independent of x and $(k+1+j)$ points of touch of H_k with respect to Y_k, $x < \tau_1 < \cdots < \tau_{k+1+j} < x + K$ such that $\tau_{i'} - \tau_i > M$, $1 \leq i < i' \leq k+1+j$, and the event*

$$\inf_{t \in [\tau_1, \tau_{k+1+j}]} |f_k^{(j)}(t) - f_{0,k}^{(j)}(t)| \leq D$$

occurs with probability greater than $1-\epsilon$ for all $j = 0, \cdots, k-1$ (for $j = k-1$, $f_k^{(k-1)}$ should be read either as the left or right $(k-1)$-st derivative).

Proof. We restrict ourselves to the case of k even. We start by proving the same result for $f_{c,k}$, the solution of the LS problem.

Let $j = 0$. For ease of notation, we omit the subscripts k in $f_{c,k}$ and $f_{0,k}$. Fix $x > 0$ (the case $x < 0$ can be handled similarly) and let $c > 0$ be large enough so that we can find $(k+1)$ points of touch after the point x, $\tau_{1,c}, \cdots, \tau_{k+1,c}$, that are separated by at least M from each other. Consider the event

$$\inf_{t \in [\tau_{1,c}, \tau_{k+1,c}]} |f_c(t) - f_0(t)| \geq D \tag{3.24}$$

and let B be the B-spline of order $k-1$ with support $[\tau_{1,c}, \tau_{k+1,c}]$; i.e., B is given by

$$B(t) = (-1)^k k \left(\frac{(t-\tau_{1,c})_+^{k-1}}{\prod_{j \neq 1}(\tau_{j,c} - \tau_{1,c})} + \cdots + \frac{(t-\tau_{k,c})_+^{k-1}}{\prod_{j \neq k}(\tau_{j,c} - \tau_{k,c})} \right)$$

(see Lemma 2.5.1 in Chapter 2). Let $|\eta| > 0$ and consider the perturbation function $p = B$. Recall that $p \equiv 0$ on $(-\infty, \tau_{1,c}) \cup (\tau_{k+1,c}, \infty)$. It is easy to check that for $|\eta|$ small enough, the perturbed function

$$f_{c,\eta}(t) = f_c(t) + \eta p(t)$$

is in the class $\mathcal{C}_{\underline{m}_1,\underline{m}_2}$, with

$$\underline{m}_1 = \underline{m}_2 = \left(c^k, \cdots, \frac{k!}{2!}c^2\right).$$

Indeed, p was chosen so that it satisfies $p^{(j)}(\tau_{1,c}) = p^{(j)}(\tau_{k+1,c}) = 0$ for $0 \le j \le k-2$, which guarantees that the perturbed function $f_{c,\eta}$ belongs to $C^{k-2}(-c, c)$. Also, the boundary conditions at $-c$ and c are satisfied since p is equal to 0 outside the interval $[\tau_{1,c}, \tau_{k+1,c}]$. Finally, since p is a spline a degree $k-1$, the function $f_{c,\eta}^{(k-2)}$ is also piecewise linear and one can check that it is nonincreasing and convex for very small values of $|\eta|$. It follows that

$$\lim_{\eta \to 0} \frac{\Phi_c(f_{c,\eta}) - \Phi_c(f_c)}{\eta} = 0$$

which yields

$$\int_{\tau_{1,c}}^{\tau_{k+1,c}} p(t) f_c(t) dt - \int_{\tau_{1,c}}^{\tau_{k+1,c}} p(t)(dW(t) + f_0(t)dt) = 0\,,$$

or equivalently

$$\int_{\tau_{1,c}}^{\tau_{k+1,c}} p(t)(f_c(t) - f_0(t))dt = \int_{\tau_{1,c}}^{\tau_{k+1,c}} p(t) dW(t).$$

For any ω in the event (3.24), we have

$$\left| \int_{\tau_{1,c}}^{\tau_{k+1,c}} p(t) dW(t) \right| \ge D \int_{\tau_{1,c}}^{\tau_{k+1,c}} p(t) dt = D \tag{3.25}$$

where in (3.25), we used the fact that B integrates to 1. But we can find $D > 0$ large enough such that the probability of the previous event is very small. Indeed, let $\mathcal{G}_{x_0,M,K}$ be the class of functions g such that

$$g(t) = \left\{ \frac{(t-y_1)_+^{k-1}}{\prod_{j \ne 1}(y_j - y_1)} + \cdots + \frac{(t-y_1)_+^{k-1}}{\prod_{j \ne k}(y_j - y_k)} \right\} 1_{[y_1, y_{k+1}]}(t),$$

where $x_0 \le y_1 < \cdots < y_{k+1} \le x_0 + K$ and $y_j - y_i \ge M$ for $1 \le i < j \le k+1$ and M and K are two positive constants independent of x_0. Define

$$W_g = \int_{-\infty}^{\infty} g(t) dW(t), \qquad \text{for } \ g \in \mathcal{G}_{x_0,M,K}\,.$$

The process $\{W_g : \ g \in \mathcal{G}_{x_0,M,K}\}$ is a mean zero Gaussian process, and for any g and h in the class $\mathcal{G}_{x_0,M,K}$, we have

$$Var\,(W_g - W_h) \;=\; E\,(W_g - W_h)^2 = \int_{-\infty}^{\infty} (g(t) - h(t))^2\, dt.$$

and therefore, if we equip the class $\mathcal{G}_{x_0,M,K}$ with the standard deviation semi-metric d given by

$$d^2(g,h) = \int (g(t) - h(t))^2\, dt,$$

the process $(W_g, g \in \mathcal{G}_{x_0,M,K})$ is sub-Gaussian with respect to d; i.e., for any g and h in $\mathcal{G}_{x_0,M,K}$ and $x \ge 0$

$$P(|W_g - W_h| > x) \le 2e^{-\frac{1}{2}x^2/d^2(g,h)}.$$

In the following, we will get an upper bound of the covering number $N(\epsilon, \mathcal{G}_{x_0,M,K}, d)$ for the class $\mathcal{G}_{x_0,M,K}$ when $\epsilon > 0$. For this purpose, we first note that for any g and h in $\mathcal{G}_{x_0,M,K}$

$$d^2(g,h) \le \int_{x_0}^{x_0+K} (g(t) - h(t))^2\, dt = K \int_{x_0}^{x_0+K} (g(t) - h(t))^2\, dQ(t)$$

where Q is the probability measure corresponding to the uniform distribution on $[x_0, x_0+K]$; i.e.,

$$dQ(t) = \frac{1}{K} 1_{[x_0, x_0+K]}(t) dt,$$

and therefore, it suffices to find an upper bound for the covering number of the class $\mathcal{G}_{x_0,M,K}$ with respect to $L_2(Q)$.

Any function in class $\mathcal{G}_{x_0,M,K}$ is a sum of functions of the form

$$g_j(t) = \left\{ \frac{(t - y_j)_+^{k-1}}{\prod_{j' \ne j} (y_{j'} - y_j)} \right\} 1_{[y_1, y_{k+1}]}(t),$$

over $j \in \{1, \cdots, k\}$. Denote by $\mathcal{G}_{x_0,M,K,j}$ the class of functions g_j. Taking $\psi(t) = t_+^k$, we have by Lemma 2.6.16 in van der Vaart and Wellner (1996) that the class of functions $\{t \mapsto \psi(t - y_j), y_j \in \mathbb{R}\}$ is VC-subgraph with VC-index equal to 2 and therefore the class of functions $\{t \mapsto \psi(t - y_j),\ t, y_j \in [x_0, x_0 + K]\}$, $\mathcal{G}^1_{x_0,M,K,j}$ say, is also VC-subgraph with VC-index equal 2 and admits K^{k-1} as an envelope. Therefore, by Theorem 2.6.7 of van der Vaart and Wellner (1996), there exists $C_1 > 0$ and $K_1 > 0$ (here $K_1 = 2$) such that for any $0 < \epsilon < 1$ and for all $j \in \{1, \cdots, k\}$

$$N(\epsilon, \mathcal{G}^1_{x_0,M,K,j}, L_2(Q)) \le C_1 \left(\frac{1}{\epsilon}\right)^{K_1}.$$

where C_1 and K_1 are independent of x_0. On the other hand, since $y_j - y_i \ge M$, the functions

$$t \mapsto \frac{1}{\prod_{j' \neq j}(y_{j'} - y_j)} 1_{[y_1, y_{k+1}]}(t)$$

indexed by the y_j's are all bounded by the constant $1/M^k$ and form a VC-subgraph class with a VC-index that is smaller than 5 and more importantly that is independent of x_0. Denote this class by $\mathcal{G}^2_{x_0,M,K,j}$. By the same theorem of van der Vaart and Wellner (1996), there exist $C_2 > 0$ and K_2 (here $K_2 \le 8$) also independent of x_0 such that

$$N(\epsilon, \mathcal{G}^2_{x_0,M,K,j}, L_2(Q)) \le C_2 \left(\frac{1}{\epsilon}\right)^{K_2}$$

for $0 < \epsilon < 1$. By Lemma 16 of Nolan and Pollard (1987), it follows there exists $C_3 > 0$ and $K_3 > 0$ independent of x_0 such that

$$N(\epsilon, \mathcal{G}_{x_0,M,K}, L_2(Q)) \le C_3 \left(\frac{1}{\epsilon}\right)^{K_3}$$

for all $0 < \epsilon < 1$ and therefore

$$N(\epsilon, \mathcal{G}_{x_0,M,K}, d) \le C_3 K^{K_3/2} \left(\frac{1}{\epsilon}\right)^{K_3}.$$

Using the fact that the packing number $D(\epsilon, \mathcal{G}_{x_0,M,K}, d) \le N(\epsilon/2, \mathcal{G}_{x_0,M,K}, d)$ and Corollary 2.2.8 of van der Vaart and Wellner (1996), it follows that there exists a constant $C > 0$, $D > 0$, and a (the diameter of the class) independent of x_0 such that for

$$E \sup_{g \in \mathcal{G}_{x_0,M,K}} |W_g| \le E|W_{g_0}| + C \int_0^a \sqrt{1 + D \log\left(\frac{1}{\epsilon}\right)} d\epsilon$$

where the integral on the right side converges and g_0 is any element in the class $\mathcal{G}_{x_0,M,K}$ and we can take, e.g.,

$$g_0(t) = \frac{1}{M^k}\left((t-x_0)_+^{k-1} + (t-x_0-M)_+^{k-1} + \cdots + (t-x_0-(k-1)M)_+^{k-1}\right)1_{[x_0,x_0+kM]}(t)$$

where $y_1 = x_0, y_2 = x_0 + M, \cdots, y_{k+1} = x_0 + kM$. By a change of variable, we have

$$E|W_{g_0}| = \frac{1}{M^k}E\left|\int_0^{kM}\left(t_+^{k-1} + \cdots + (t-(k-1)M)_+^{k-1}\right)dW(t)\right|$$

which is clearly independent of x_0. Now, we can write

$$\begin{aligned} P(|W_p| > \lambda) &\leq P(\sup_{g\in\mathcal{G}_{x_0,M,K}} |W_g| > \lambda) \\ &\leq E\sup_{g\in\mathcal{G}_{x_0,M,K}} |W_g|/\lambda, \quad \text{by Markov's inequality} \\ &\leq \left(E|W_{g_0}| + C\int_0^a \sqrt{1 + D\log\left(\frac{1}{\epsilon}\right)}d\epsilon\right)/\lambda \quad \to 0 \text{ as } \lambda \to \infty. \end{aligned}$$

Now, let $c(n) = n$ and f_n, and $\tau_{1,n}, \cdots, \tau_{k+1,n}$ are the LS solution on $[-n, n]$ and $(k+1)$ points of touch to the left of x. Also, let $\xi_n \in [\tau_{1,n}, \tau_{k+1,n}]$ the point where the infimum of the function $f_n - f_0$ is attained. By tightness of the points of touch, we can find subsequences $(\tau_{1,n_l}, \cdots, \tau_{k+1,n_l})$ and (ξ_{n_l}) that converge to $(\tau_1, \cdots, \tau_{k+1})$ and ξ respectively. By the same arguments used in the construction of H_k, there exists a further subsequence (f_{n_p}) which converges to f_k in the supremum norm on the space of continuous functions on $[-K, K]$. On the other hand, it is easy to see that $\tau_1, \cdots, \tau_{k+1}$ are points of touch of H_k with respect to Y_k that are to the right of x and to the left of $x + K$. Furthermore, $\tau_{i'} - \tau_i \geq M$, for $1 \leq i < i' \leq k+1$. For ease of notation, we replace n_p by n. We have

$$\begin{aligned} |f_k(\xi) - f_0(\xi)| &\leq |f_n(\xi_n) - f_0(\xi_n)| + |f_0(\xi_n) - f_0(\xi)| \\ &\quad + |f_n(\xi_n) - f_k(\xi_n)| + |f_k(\xi_n) - f_k(\xi)|. \end{aligned}$$

By the arguments used above, we know that there exists $D > 0$ independent of x that bounds the first term from above with large probability as $n \to \infty$. To control the second and fourth terms, we use the fact that $\xi_n \to \xi$ and continuity of f_0 and f_k. Therefore, we can find an integer $N_1 > 0$ that might depend on x such that for all $n \geq N_1$, we have

$$\max\{|f_k(\xi_n) - f_k(\xi)|, |f_0(\xi_n) - f_0(\xi)|\} \leq D.$$

Finally, using the fact that $\xi_n \in [-K, K]$ and that f_n converges uniformly to f_k on $[-K, K]$, we can find an integer $N_2 > 0$ that might depend on x such that for all $n \geq N_2$, we have

$$|f_n(\xi_n) - f_k(\xi_n)| \leq D.$$

It follows that with large probability, there exists $\xi \in [\tau_1, \tau_{k+1}]$ such that

$$|f_k(\xi) - f_0(\xi)| \leq 3\,D,$$

or equivalently

$$\inf_{t \in [\tau_1, \tau_{k+1}]} |f_k(t) - f_0(t)| \leq 3\,D.$$

For $j > 1$, we take the perturbation function p_j to be

$$p_j = q_j^{(j)},$$

where $q_j = B_j$, the B-spline of degree $k-1+j$ with $k+1+j$ knots taken to be points of touch that are at least M distant from each other; i.e.,

$$\begin{aligned} q_j(t) &= B_j(t) \\ &= (-1)^{k+j}(k+j)\left(\frac{(t-\tau_{1,n})_+^{k+j-1}}{\prod_{j\neq 1}(\tau_{j,n}-\tau_{1,n})} + \cdots + \frac{(t-\tau_{k+j,n})_+^{k+j-1}}{\prod_{j\neq k+j}(\tau_{j,n}-\tau_{k+j,n})}\right). \end{aligned}$$

The function p_j is a valid perturbation function and therefore we have

$$\int_{\tau_{1,n}}^{\tau_{k+1+j,n}} p_j(t)(f_n(t) - f_0(t))dt = \int_{\tau_{1,n}}^{\tau_{k+1+j,n}} p_j(t)dW(t).$$

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

$$\int_{\tau_{1,n}}^{\tau_{k+1+j,n}} (-1)^j q_j(t)(f_n^{(j)}(t) - f_0^{(j)}(t))dt = \int_{\tau_{1,n}}^{\tau_{k+1+j,n}} p_j(t)dW(t).$$

The proof follows from arguments which are similar to those used for $j = 0$. ∎

Proof of Lemma 3.5.2 Fix $\epsilon > 0$ small. (i) follows from tightness of the points of touch of $H_{c,k}$ and Y_k and the construction of H_k. Indeed, there exists $M > 0$ independent of t and two points of touch τ_n^- and τ_n^+ between the processes $H_{n,k}$ and Y_k such that

$\tau_n^- \in [t-3M, t-M]$ and $\tau_n^+ \in [t+M, t+3M]$ with probability greater than $1-\epsilon$. Then, we can find a subsequence n_j such that $\tau_{n_j}^- \to \tau^-$, $\tau_{n_j}^+ \to \tau^+$, $\|H_{n_j,k} - H_k\|_{[t-3M,t+3M]} \to 0$. Therefore, we have

$$H_{n_j,k}(\tau_{n_j}^-) \to H_k(\tau^-), \qquad \text{and} \qquad H_{n_j,k}(\tau_{n_j}^+) \to H_k(\tau^+)$$

as $n_j \to \infty$. But by continuity of Y_k, we have

$$Y_k(\tau_{n_j}^-) \to Y_k(\tau^-) \qquad \text{and} \qquad Y_k(\tau_{n_j}^+) \to Y_k(\tau^+).$$

It follows that $H_k(\tau^-) = Y_k(\tau^-)$ and $H_k(\tau^+) = Y_k(\tau^+)$; i.e., τ^- and τ^+ are points of touch of H_k and Y_k occurring before and after t respectively. Furthermore, we have $t - 3M \leq \tau^- \leq t - M < t + M \leq \tau^+ \leq t + 3M$. These points of touch might not be successive but it is clear that (i) will hold for successive points of touch.

Let $[a,b] \subset \mathbb{R}$ be a finite interval. We prove (ii) and (iii) only when k is even as the arguments are very similar for k odd. We start with proving (iii) and for that we fix $t \in [a,b]$. Using the same type of arguments used in proof of Lemma 3.5.3, we can find $D > 0$ independent of t and a point $\xi_1 > b$ such that

$$|f_k^{(k-2)}(\xi_1) - f_0^{(k-2)}(\xi_1)| \leq D.$$

with large probability. Using again the same kind of arguments, we can find another point ξ_2 such that $\xi_2 - \xi_1 \geq M$ and

$$|f_k^{(k-2)}(\xi_2) - f_0^{(k-2)}(\xi_2)| \leq D$$

maybe at the cost of increasing D and where $M > 0$ is a constant that is independent of t. By tightness of the points of touch, we know that there exists $K > 0$ such that $0 \leq \xi_1 - b \leq \xi_2 - b \leq K$ with large probability. By convexity of $f_k^{(k-2)}$, we have

$$\begin{aligned}
f_k^{(k-1)}(t) &\leq \frac{f_k^{(k-2)}(\xi_2) - f_k^{(k-2)}(\xi_1)}{\xi_2 - \xi_1} \\
&\leq \frac{f_0^{(k-2)}(\xi_2) - f_0^{(k-2)}(\xi_1) + 2D}{\xi_2 - \xi_1} \\
&\leq f_0^{(k-1)}(\xi_2) + \frac{2D}{M},
\end{aligned}$$

where $f_k^{(k-1)}$ is either the left or right $(k-1)$st derivative. Therefore,

$$\begin{aligned}
f_k^{(k-1)}(t) - f_0^{(k-1}(t) &\leq f_0^{(k-1)}(\xi_2) - f_0^{(k-1)}(t) + \frac{2D}{M} \\
&= k!(\xi_2 - t) + \frac{2D}{M} \\
&= k!(\xi_2 - b + b - t) + \frac{2D}{M} \\
&\leq k!\,(K + b - a) + \frac{2D}{M}.
\end{aligned}$$

Similarly, we can find two points ξ_{-2} and ξ_{-1} this time to the left of a such that the events $\xi_{-1} - \xi_{-2} \geq M$, $\max\{|f_k^{(k-2)}(\xi_{-2}) - f_0^{(k-2)}(\xi_{-2})|, |f_k^{(k-2)}(\xi_{-1}) - f_0^{(k-2)}(\xi_{-1})|\} \leq D$ and $a - K \leq \xi_{-2} < \xi_{-1} <\leq a$ occur with very large probability maybe at the cost of increasing one of the constants M, K or D. Then it follows that

$$\begin{aligned}
f_k^{(k-1)}(t) &\geq \frac{f_k^{(k-2)}(\xi_{-1}) - f_k^{(k-2)}(\xi_{-2})}{\xi_{-1} - \xi_{-2}} \\
&\geq \frac{f_0^{(k-2)}(\xi_{-1}) - f_0^{(k-2)}(\xi_{-2}) - 2D}{\xi_{-1} - \xi_{-2}} \\
&\geq f_0^{(k-1)}(\xi_{-2}) - \frac{2D}{M},
\end{aligned}$$

and hence

$$\begin{aligned}
f_k^{(k-1)}(t) - f_0^{(k-1)}(t) &\geq f_0^{(k-1)}(\xi_{-2}) - f_0^{(k-1)}(t) - \frac{2D}{M} \\
&= k!(\xi_{-2} - t) - \frac{2D}{M} \\
&= -k!(t - a + a - \xi_{-2}) - \frac{2D}{M} \\
&\geq -k!\,(b - a + K) - \frac{2D}{M}.
\end{aligned}$$

It follows that with large probability we have for all $t \in [a, b]$

$$|f_k^{(k-1)}(t) - f_0^{(k-1)}(t)| \leq k!\,(K + b - a) + \frac{2D}{M}$$

and it is clear that the bound in the inequality depends only on $b - a$. Thus by applying a similar argument on $[a, b + K]$, we can find a constant $C > 0$ depending only on $b - a$ and K such that

$$\|f_k^{(k-1)} - f_0^{(k-1)}\|_{[a,b+K]} < C.$$

Now, by writing

$$(f_k^{(k-2)}(t) - f_0^{(k-2)}(t)) - (f_k^{(k-2)}(\xi_1) - f_0^{(k-2)}(\xi_1)) = \int_{\xi_1}^{t} \left(f_k^{(k-1)}(s) - f_0^{(k-1)}(ds) \right) ds.$$

It follows that

$$\begin{aligned} |f_k^{(k-2)}(t) - f_0^{(k-2)}(t)| &\leq |f_k^{(k-2)}(\xi_1) - f_0^{(k-2)}(\xi_1)| + (\xi_1 - t)\|f_k^{(k-1)} - f_0^{(k-1)}\|_{[a,b+K]} \\ &\leq D + (K+b-a)C. \end{aligned}$$

Using induction and Lemma 3.5.3, we can show (iii) for $j = 0, \cdots, k-3$.

Now to show (ii), we start with $j = k-1$; i.e., for $t \in [a,b]$ and $\epsilon > 0$,we want to show that we can find $M = M(\epsilon) > 0$ such that

$$P(\|H_k^{(k-1)}(t) - Y_k^{(k-1)}(t)\|_{[a,b]} > M) \leq \epsilon.$$

But, we know that we can find $M_1 > 0$ and $K > 0$ independent of any $t \in [a,b]$ and two points $\xi_1 \leq \xi_2$ to the right of b such that $\xi_2 - \xi_1 \geq M_1$, $b \leq \xi_1 < \xi_2 \leq b + K$ and

$$H_k^{(k-2)}(\xi_1) = Y_k^{(k-2)}(\xi_1) \qquad \text{and} \qquad H_k^{(k-2)}(\xi_2) = Y_k^{(k-2)}(\xi_2).$$

The existence of such points follows from applying the mean value theorem repeatedly to a number of points of touch and also using tightness. Using again the mean value theorem, we can find $\xi \in (\xi_1, \xi_2)$ such that

$$H_k^{(k-1)}(\xi) = Y_k^{(k-1)}(\xi).$$

Now, we can write for any $t \in [a,b]$

$$\begin{aligned} & H_k^{(k-1)}(t) - Y_k^{(k-1)}(t) \\ &= \left(H_k^{(k-1)}(t) - Y_k^{(k-1)}(t) \right) - \left(H_k^{(k-1)}(\xi) - Y_k^{(k-1)}(\xi) \right) \\ &= \int_{\xi}^{t} d(H_k^{(k-1)}(s) - Y_k^{(k-1)}(s)) \\ &= \int_{\xi}^{t} (f_k(s) - f_0(s))ds - \int_{\xi}^{t} dW(s) \\ &= \int_{\xi}^{t} (f_k(s) - f_0(s))ds - (W(t) - W(\xi)). \end{aligned}$$

By stationarity of the increments of W and since $0 \leq \xi - t \leq b - a + K$, the second term can be bounded with large probability by a constant dependent of on K and $b-a$. As for the first term, we know by (iii) that there exists M_2 depending only on $b-a$ such that $\|f_k - f_0\|_{[a,b+K]} < M_2$ with large probability. Therefore,

$$\left| \int_\xi^t (f_k(s) - f_0(s))ds \right| \quad \leq \quad M_2(\xi - t) \leq M_2(b - a + K).$$

It follows that, with large probability, we can find a constant $C > 0$, depending only on $b-a$ and K such that

$$\|H_k^{(k-1)} - Y_k^{(k-1)}\|_{[a,b+K]} < C.$$

Now, by writing

$$\begin{aligned} H_k^{(k-2)}(t) - Y_k^{(k-2)}(t) &= H_k^{(k-2)}(t) - Y_k^{(k-2)}(t) - (H_k^{(k-2)}(\xi_1) - Y_k^{(k-2)}(\xi_1)) \\ &= \int_{\xi_1}^t (H_k^{(k-1)}(s) - Y_k^{(k-1)}(s))ds, \end{aligned}$$

it follows that

$$\|H_k^{(k-2)} - Y_k^{(k-2)}\|_{[a,b]} \leq (b - a + K)C.$$

For $0 \leq j \leq k-3$, we use induction together with tightness of the distance between points of touch and the mean value theorem. ■

Now we use Lemma 3.5.1 to complete the proof of Theorem 3.2.1 by showing that H_k determined by (i) - (iv) of Theorem 3.2.1 is unique. Suppose that there exists another process G_k that satisfies the properties (i) - (iv) of Theorem 3.2.1. As the proof follows along similar arguments for k odd, we only focus here on the case where k is even. Fix $n > 0$ and let $a_{-n,2} < a_{-n,1}$ be two points of touch between H_k and Y_k to the left of $-n$, such that $a_{-n,1} - a_{-n,2} > M$. Also, consider $b_{n,1} < b_{n,2}$ to be two points of touch between H_k and Y_k to the right of n such that $b_{n,2} - b_{n,1} > M$. There exists $K > 0$ independent of n such that $-n - K < a_{-n,2} < a_{-n,1} < -n$ and $n < b_{n,1} < b_{n,2} < n + K$ with large probability. For a k-convex function f and real arbitrary points $a < b$, we define $\phi_{a,b}(f)$ by

$$\phi_{a,b}(f) = \frac{1}{2} \int_a^b f^2(t)dt - \int_a^b f(t)dX_k(t).$$

For ease of notation, we omit the subscript k in H_k and G_k. Let $h = H^{(k)}$, $g = G^{(k)}$ and $a < b$ be two points of touch between H and Y_k. Then we have

$$\begin{aligned}
&\phi_{a,b}(g) - \phi_{a,b}(h) \\
&= \frac{1}{2}\int_a^b (g(t)-h(t))^2 dt + \int_a^b (g(t)-h(t))h(t)dt - \int_a^b (g(t)-h(t))dX_k(t) \\
&= \frac{1}{2}\int_a^b (g(t)-h(t))^2 dt + \int_a^b (g(t)-h(t))d(H^{(k-1)} - Y_k^{(k-1)}).
\end{aligned}$$

This yields, using successive integrations by parts,

$$\begin{aligned}
&\phi_{a,b}(g) - \phi_{a,b}(h) \\
&= \frac{1}{2}\int_a^b (g(t)-h(t))^2 dt \\
&\quad + \Big((H^{(k-1)}(b) - Y_k^{(k-1)}(b))(g(b)-h(b)) \\
&\qquad - (H^{(k-1)}(a) - Y_k^{(k-1)}(a))(g(a)-h(a))\Big) \\
&\quad - \Big((H^{(k-2)}(b) - Y_k^{(k-2)}(b))(g'(b)-h'(b)) \\
&\qquad - (H^{(k-2)}(a) - Y_k^{(k-2)}(a))(g'(a)-h'(a))\Big) \\
&\quad \vdots \\
&\quad + \Big((H'(b) - Y_k'(b))(g^{(k-2)}(b)-h^{(k-2)}(b)) \\
&\qquad - (H'(a) - Y_k'(a))(g^{(k-2)}(a)-h^{(k-2)}(a))\Big) \qquad (3.26) \\
&\quad - \Big((H(b) - Y_k(b))(g^{(k-1)}(b-)-h^{(k-1)}(b-)) \\
&\qquad - (H(a) - Y_k(a))(g^{(k-1)}(a+)-h^{(k-1)}(a+))\Big) \qquad (3.27) \\
&\quad + \int_a^b (H(t)-Y_k(t))d(g^{(k-1)}(t) - h^{(k-1)}(t))
\end{aligned}$$

where the terms in (3.26) and (3.27) are equal to 0 and last term can be rewritten as

$$\int_a^b (H(t)-Y_k(t))d(g^{(k-1)}(t) - h^{(k-1)}(t)) = \int_a^b (H(t)-Y_k(t))dg^{(k-1)}(t) \geq 0$$

using the characterization of H. Now, if we take c and d to be arbitrary points (not necessarily points of touch of H and Y_k), we get

$$\phi_{c,d}(h) - \phi_{c,d}(g)$$

$$
\begin{aligned}
&= \frac{1}{2}\int_c^d (h(t)-g(t))^2 dt \\
&\quad + \Big((G^{(k-1)}(d) - Y_k^{(k-1)}(d))(h(d)-g(d)) - (G^{(k-1)}(c) - Y_k^{(k-1)}(c))(h(c)-g(c)) \Big) \\
&\quad - \Big((G^{(k-2)}(d) - Y_k^{(k-2)}(d))(h'(d)-g'(d)) - (G^{(k-2)}(c) - Y_k^{(k-2)}(c))(h'(c)-g'(c)) \Big) \\
&\quad \vdots \\
&\quad + \Big((G(d) - Y_k(d))(h^{(k-1)}(d) - g^{(k-1)}(d)) - (G(c) - Y_k(c))(h^{(k-1)}(c) - g^{(k-1)}(c)) \Big) \\
&\quad + \int_c^d (G(t) - Y_k(t)) dh^{(k-1)}(t).
\end{aligned}
$$

Now, let $a = a_{-n,1}$, $b = b_{n,1}$, $c = a_{-n,2}$ and $b = b_{n,2}$ and let $J_n = [a_{-n,1}, a_{-n,2}]$ and $K_n = [b_{n,1}, b_{n,2}]$. Then, we have

$$
\begin{aligned}
& \phi_{a_{-n,1},b_{n,1}}(g) - \phi_{a_{-n,1},b_{n,1}}(h) + \phi_{a_{-n,2},b_{n,2}}(h) - \phi_{a_{-n,2},b_{n,2}}(g) \qquad (3.28)\\
&\quad \geq \frac{1}{2}\int_{a_{-n,1}}^{b_{n,1}} (g(t)-h(t))^2 dt + \frac{1}{2}\int_{a_{-n,2}}^{b_{n,2}} (g(t)-h(t))^2 dt \\
&\qquad + \sum_{j=2}^{k-1} \Big[\Big(H^{(j)}(t) - Y_k^{(j)}(t)\Big)\Big(g^{(j-2)}(t) - h^{(j-2)}(t)\Big) \Big]_{a_{-n,1}}^{b_{n,1}} \\
&\qquad + \sum_{j=2}^{k-1} \Big[\Big(G^{(j)}(t) - Y_k^{(j)}(t)\Big)\Big(h^{(j-2)}(t) - g^{(j-2)}(t)\Big) \Big]_{a_{-n,2}}^{b_{n,2}}.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
& \phi_{a_{-n,1},b_{n,1}}(g) - \phi_{a_{-n,1},b_{n,1}}(h) + \phi_{a_{-n,2},b_{n,2}}(h) - \phi_{a_{-n,2},b_{n,2}}(g) \qquad (3.29)\\
&= \frac{1}{2}\int_{J_n\cup K_n} \left(g^2(t) - h^2(t)\right) dt - \int_{J_n\cup K_n} (g(t)-h(t))\, dX_k(t) \\
&= \frac{1}{2}\int_{J_n\cup K_n} (g(t)-h(t))\,(g(t)-f_0(t))\, dt \\
&\quad + \frac{1}{2}\int_{J_n\cup K_n} (g(t)-h(t))\,(h(t)-f_0(t))\, dt - \int_{J_n\cup K_n} (g(t)-h(t))\, dW(t)
\end{aligned}
$$

where $f_0(t) = t^k$.

As in Groeneboom, Jongbloed, and Wellner (2001a), we first suppose that

$$
\lim_{n\to\infty} \int_{-n}^{n} (g(t)-h(t))^2\, dt < \infty. \tag{3.30}
$$

This implies that

$$
\lim_{|t|\to\infty} (g(t)-h(t)) = 0.
$$

Since g and h are at least $(k-2)$ times differentiable, $g-h$ is a function of uniformly bounded variation on J_n and K_n. Therefore, using the fact that the respective lengths of J_n and K_n are $O_p(1)$ which follows from Lemma 3.5.2 (i), and the same arguments in page 1640 of Groeneboom, Jongbloed, and Wellner (2001a), we get that

$$\liminf_{n\to\infty} \int_{J_n \cup K_n} (g(t)-h(t))\, dW(t) = 0$$

almost surely. The hypothesis in (3.30) implies that

$$\lim_{n\to\infty} \int_{a_{-n,1}}^{a_{-n,2}} (g(t)-h(t))^2\, dt \to 0, \qquad \text{as } n \to \infty.$$

On the other hand, we can write using integration by parts,

$$\begin{aligned} &\int_{a_{-n,1}}^{a_{-n,2}} \left(g'(t)-h'(t)\right)^2 dt \\ &= \Big[(g(t)-h(t)) \left(g'(t)-h'(t)\right) \Big]_{a_{-n,1}}^{a_{-n,2}} - \int_{a_{-n,1}}^{a_{-n,2}} (g(t)-h(t)) \left(g''(t)-h''(t)\right) dt \end{aligned}$$

and therefore

$$\begin{aligned} &\int_{a_{-n,1}}^{a_{-n,2}} \left(g'(t)-h'(t)\right)^2 dt \\ &\leq\ 2\|g-h\|_{[a_{-n,1},a_{-n,2}]} \times \|g'-h'\|_{[a_{-n,1},a_{-n,2}]} \\ &\quad +(a_{-n,2}-a_{-n,1})\|g-h\|_{[a_{-n,1},a_{-n,2}]} \times \|g''-h''\|_{[a_{-n,1},a_{-n,2}]} \end{aligned}$$

which converges to 0 as $n \to \infty$ with arbitrarily high probability since the length of $J_n = [a_{-n,1}, a_{-n,2}]$,

$$\|g'-h'\|_{[a_{-n,1},a_{-n,2}]} \quad \text{and} \quad \|g''-h''\|_{[a_{-n,1},a_{-n,2}]}$$

are $O_p(1)$ uniformly in n by Lemma 3.5.1 (ii).

Consider now the sequence of functions $(\psi_n)_n$ defined on $[0,1]$ as

$$\psi_n(t) = g'((a_{-n,2}-a_{-n,1})t + a_{-n,1}) - h'((a_{-n,2}-a_{-n,1})t + a_{-n,1}), \qquad 0 \leq t \leq 1.$$

Using the same arguments above, it is easy to see that $\|\psi_n\|_{[0,1]}$ and $\|\psi_n'\|_{[0,1]}$ are $O_p(1)$ and therefore, by Arzelà-Ascoli's theorem, we can find a subsequence (n') and ψ such that

$$\|\psi_{n'} - \psi\|_{[0,1]} \to 0, \qquad \text{as } n \to \infty.$$

But $\psi \equiv 0$ on $[0,1]$. Indeed, first note that

$$\int_0^1 \psi_n^2(t)dt = \frac{1}{a_{-n,2}-a_{-n,1}} \int_{a_{-n,1}}^{a_{-n,2}} \left(g'(t)-h'(t)\right)^2 dt \to 0, \quad \text{as } n \to \infty.$$

Therefore, since

$$\int_0^1 \psi^2(t)dt \leq \liminf_{n\to\infty} \int_0^1 \psi_n^2(t)dt$$

it follows that

$$\int_0^1 \psi^2(t)dt = 0$$

and $\psi \equiv 0$, by continuity. We conclude that from every subsequence $(\psi_{n'})_{n'}$, we can extract a further subsequence $(\psi_{n''})_{n''}$ that converges to 0 on $[0,1]$. Thus, $\lim_{n\to\infty} \|\psi_n\|_{[0,1]} = 0$. It follows that

$$\|g'-h'\|_{[a_{-n,1},a_{-n,2}]} \to 0, \quad \text{as } n \to \infty$$

with large probability. If $k \geq 5$, we can show by induction that for all $j = 4, \cdots, k-1$ we have

$$\lim_{n\to\infty} \|g^{(j-2)} - h^{(j-2)}\|_{[a_{-n,1},a_{-n,2}]} = 0$$

with large probability, and the same thing holds when $(a_{-n,1}, a_{-n,2})$ is replaced by $(b_{n,2}, b_{n,1})$. On the other hand, by Lemma 3.5.1 (i), we know that there exists $D > 0$ such that

$$\max\left\{ \|H^{(j)} - Y_k^{(j)}\|_{[a_{-n,1},a_{-n,2}]}, \|G^{(j)} - Y_k^{(j)}\|_{[a_{-n,1},a_{-n,2}]} \right\} \leq D$$

with arbitrarily high probability, for $j = 0, \cdots, k-1$. To see that, consider the first term (the second term is handled similarly) and fix $\epsilon > 0$. There exist $K > 0$ (maybe different from the one considered above) independent of n such that have

$$P([a_{-n,1}, a_{-n,2}] \subseteq [-n-K, -n]) \geq 1 - \epsilon/2$$

and $D > 0$ depending only on K (and therefore independent of n) such that

$$P(\|H^{(j)} - Y_k^{(j)}\|_{[-n-K,-n]} \leq D) \geq 1 - \epsilon/2.$$

It follows that

$$
\begin{aligned}
&P(\|H^{(j)} - Y_k^{(j)}\|_{[a_{-n,1}, a_{-n,2}]} > D) \\
&\quad = \; P(\|H^{(j)} - Y_k^{(j)}\|_{[a_{-n,1}, a_{-n,2}]} > D, [a_{-n,1}, a_{-n,2}] \subseteq [-n-K, -n]) \\
&\qquad + P(\|H^{(j)} - Y_k^{(j)}\|_{[a_{-n,1}, a_{-n,2}]} > D, [a_{-n,1}, a_{-n,2}] \nsubseteq [-n-K, -n]) \\
&\quad \leq \; P(\|H^{(j)} - Y_k^{(j)}\|_{[-n-K, -n]} > D) + P([a_{-n,1}, a_{-n,2}] \nsubseteq [-n-K, -n]) \\
&\quad < \; \epsilon/2 + \epsilon/2 \\
&\quad = \; \epsilon.
\end{aligned}
$$

Using similar arguments, we can show

$$
\max\left\{ \|H^{(j)} - Y_k^{(j)}\|_{[b_{n,2}, b_{n,1}]}, \|G^{(j)} - Y_k^{(j)}\|_{[b_{n,2}, b_{n,1}]} \right\} = O_p(1)
$$

uniformly in n. Therefore, we conclude that with large probability, we have

$$
\sum_{j=0}^{k-1} \left[\left(H^{(j)}(t) - Y_k^{(j)}(t) \right) \left(g^{(j-2)}(t) - h^{(j-2)}(t) \right) \right]_{a_{-n,1}}^{b_{n,1}} \to 0,
$$

and

$$
\sum_{j=0}^{k-1} \left[\left(G^{(j)}(t) - Y_k^{(j)}(t) \right) \left(h^{(j-2)}(t) - g^{(j-2)}(t) \right) \right]_{a_{-n,2}}^{b_{n,2}} \to 0
$$

as $n \to \infty$. Finally, by the same arguments used in Groeneboom, Jongbloed, and Wellner (2001a), we have

$$
\liminf_{n\to\infty} \int_{J_n \cup K_n} (g(t) - h(t)) (g(t) - f_0(t)) \, dt = 0,
$$

and

$$
\liminf_{n\to\infty} \int_{J_n \cup K_n} (g(t) - h(t)) (h(t) - f_0(t)) \, dt = 0.
$$

almost surely. From (3.28) and (3.29), we have

$$
\frac{1}{2} \int_{a_{-n,1}}^{b_{n,1}} (g(t) - h(t))^2 \, dt + \frac{1}{2} \int_{a_{-n,2}}^{b_{n,2}} (g(t) - h(t))^2 \, dt \to 0, \quad \text{as } n \to \infty,
$$

which implies that

$$
\frac{1}{2} \int_{a_{-n,1}}^{b_{n,1}} (g(t) - h(t))^2 \, dt + \frac{1}{2} \int_{a_{-n,2}}^{b_{n,2}} (g(t) - h(t))^2 \, dt \geq \int_{-n}^{n} (g(t) - h(t))^2 \, dt \to 0
$$

as $n \to \infty$. But the latter is impossible if $g \neq h$.

Now, suppose that

$$\lim_{n\to\infty} \int_{-n}^{n} (g(t) - h(t))^2 \, dt = \infty.$$

We can write

$$\begin{aligned}
&\int_{J_n \cup K_n} (g(t) - h(t)) \, dW(t) \\
&\quad = \int_{J_n \cup K_n} ((g(t) - f_0(t)) - (h(t) - f_0(t))) \, dW(t)
\end{aligned}$$

and by Lemma 3.5.1 (ii), we have

$$\liminf_{n\to\infty} \int_{J_n \cup K_n} (g(t) - h(t)) \, dW(t) < \infty$$

almost surely. By the same result and using the same techniques as in GROENEBOOM, JONGBLOED, AND WELLNER (2001A), we have

$$\liminf_{n\to\infty} \left\{ \int_{J_n \cup K_n} (g(t) - h(t)) \, (g(t) - f_0(t)) \, dt \right\}^2 < \infty$$

and

$$\liminf_{n\to\infty} \left\{ \int_{J_n \cup K_n} (g(t) - h(t)) \, (h(t) - f_0(t)) \, dt \right\}^2 < \infty.$$

Finally, we have

$$\begin{aligned}
&\sum_{j=0}^{k-1} \left[\left(H^{(j)}(t) - Y_k^{(j)}(t)\right) \left(g^{(j-2)}(t) - h^{(j-2)}(t)\right) \right]_{a_{-n,1}}^{b_{n,1}} \\
&\quad = \sum_{j=0}^{k-1} \left[\left(H^{(j)}(t) - Y_k^{(j)}(t)\right) \left(\left(g^{(j-2)}(t) - f_0^{(j-2)}(t)\right) - \left(h^{(j-2)}(t) - f_0^{(j-2)}(t)\right) \right) \right]_{a_{-n,1}}^{b_{n,1}}
\end{aligned}$$

is tight and the same thing holds if we replace H by G and $(a_{-n,1}, b_{n,1})$ by $(a_{-n,2}, b_{n,2})$. This implies that

$$\lim_{n\to\infty} \int_{-n}^{n} (g(t) - h(t))^2 \, dt < \infty$$

which is in contradiction with the assumption made above.

We conclude that for arbitrarily large n, $g \equiv h$ on $[-n, n]$ and hence $g \equiv h$ on $\mathbb{R}$. Using condition (iv) satisfied by both processes H and G, the latter implies that $H \equiv G$ on $\mathbb{R}$. Indeed, since $H^{(k)} \equiv G^{(k)}$, there exist α and β such that

$$H^{(k-2)}(t) - G^{(k-2)}(t) = \alpha + \beta t, \text{ for t} \in \mathbb{R}.$$

But by condition (iv), $\lim_{|t|\to\infty}(H^{(k-2)}(t) - G^{(k-2)}(t)) = 0$ which implies that $\alpha = \beta = 0$ and hence $H^{(k-2)} \equiv G^{(k-2)}$. The result follows by induction.

■

Chapter 4

COMPUTATION: ITERATIVE SPLINE ALGORITHMS

4.1 Introduction

The *iterative* $(2k-1)$*-st spline algorithm* is an extension of the *iterative cubic spline algorithm*, a term that was coined by GROENEBOOM, JONGBLOED, AND WELLNER (2001A). The latter was used to compute the "invelope" H of two-sided Brownian motion + t^4 that is involved in the limiting distribution of the LSE and MLE of a non-increasing and convex density on $(0, \infty)$ (see GROENEBOOM, JONGBLOED, AND WELLNER (2001A)). The algorithm is described briefly in pages 1643 and 1644 of their article. However, more details about how this algorithm works can be found in GROENEBOOM, JONGBLOED, AND WELLNER (2003). Here, we try to give a full description about how the *iterative spline algorithms* are implemented to compute the LSE and MLE of a k-monotone density on $(0, \infty)$ for an arbitrary integer $k \geq 2$, and also to approximate the envelopes ("invelopes") of the $(k-1)$-fold integral of two-sided Brownian motion + $(k!/(2k)!)\, t^{2k}$ when k is odd (even) on a finite interval $[-c, c]$. These algorithms belong to the family of *vertex direction algorithms* (see GROENEBOOM, JONGBLOED, AND WELLNER (2003)). They were around for many decades and their development was motivated by problems in D-optimal design (see FEDOROV (1972), WYNN (1970), BÖHNING (1986)), estimation of random coefficients in regression models (see e.g. MALLET (1986)), and nonparametric estimation in mixture models (see SIMAR (1976), BÖHNING (1982), LESPERANCE AND KALBFLEISCH (1992), GROENEBOOM, JONGBLOED, AND WELLNER (2003)), which will be the focus here. In mixture models, nonparametric estimation of the mixing distribution or the mixed density yields a constrained, infinite dimensional optimization (e.g. minimization) problem. Thus, an efficient computational method is needed. GROENEBOOM, JONGBLOED, AND WELLNER (2003) extended the algorithm that was implemented by SIMAR (1976) to compute the MLE of a compound (mixed) Poisson distribution.

Groeneboom, Jongbloed, and Wellner (2003) referred to this extension as the *support reduction algorithm*. The same authors developed and used the *iterative cubic spline algorithm* to compute the LSE of a non-increasing and convex density on $(0, \infty)$ and also to approximate the process H. However, the authors seem to reserve the term only for the second estimation problem.

In the *support reduction algorithms*, the support reduction step is very crucial and it is the only step where it is ensured that one "stays" in the class of functions considered in the optimization problem. In this chapter, we explain in detail why in our estimation problems, such a step is always possible and we hope that this will shed more light on how the *iterative cubic spline algorithm* works. In the following, we present the general set-up. Let ϕ be a convex functional to be minimized over the class of functions

$$\mathcal{C} = \left\{ g = \int_\Theta f_\theta d\mu(\theta), \;\; \mu \text{ is a positive measure} \right\}.$$

The directional derivative of ϕ at the point g in the direction of f_θ is denoted by $D_\phi(f_\theta, g)$ and defined by

$$D_\phi(f_\theta, g) = \lim_{\epsilon \searrow 0} \frac{\phi(g + \epsilon f_\theta) - \phi(g)}{\epsilon}.$$

Suppose that ϕ admits a unique minimizer, $\text{argmin}_{g \in \mathcal{C}} \phi(g)$. Under the assumptions A1, A2' and A3, Groeneboom, Jongbloed, and Wellner (2003) showed that the *support reduction algorithm* converges to $\text{argmin}_{g \in \mathcal{C}} \phi(g)$. In the current estimation problems, these assumptions are satisfied. The chapter will be organized as follows: In the first two sections, we describe the *iterative* $(2k-1)$*-st spline algorithm* and explain how it works for calculating the LSE of a k-monotone density and for approximating the stochastic process H_k. The last section is reserved for calculating the MLE of a k-monotone density. In this case, the algorithm is different as it involves a linearization step that is not required in the first two estimation problems. However, the algorithm shares with the *iterative* $(2k-1)$*-st spline algorithm* the same basic structure.

Based on two samples of size $n = 100$ and $n = 1000$, the MLE and LSE of the Exponential density, viewed respectively as a k-monotone density with $k = 3$ and $k = 6$, are computed. For the same values of k, approximations of the process H_k and some of its derivatives, on the interval $[-4, 4]$, are calculated.

4.2 Computing the LSE of a k-monotone density

Let $X_1, \cdots, X_n$ be n i.i.d. random variables from a k-monotone density g_0 on $(0, \infty)$ and let $\mathbb{G}_n$ denote their empirical distribution function. We know from Chapter 2 that the functional

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

defined on the space of square integrable k-monotone functions on $(0, \infty)$ admits a unique minimizer $\tilde{g}_n$. From Proposition 2.2.3, Chapter 2, we know that $\tilde{g}_n$ is a finite scale mixture of $Beta(1, k)$'s ; i.e., there exist an integer m, $\tilde{\theta}_1, \cdots, \tilde{\theta}_m$ and $\tilde{w}_1, \cdots, \tilde{w}_m$ such that for all $t > 0$

$$\tilde{g}_n(t) = \tilde{w}_1 \frac{k(\tilde{\theta}_1 - t)_+^{k-1}}{\tilde{\theta}_1^k} + \cdots + \tilde{w}_m \frac{k(\tilde{\theta}_m - t)_+^{k-1}}{\tilde{\theta}_m^k}$$

where the weights $\tilde{w}_1, \cdots, \tilde{w}_m$ do not necessarily sum up to one for $k > 2$ (see Balabdaoui (2004)). The directional derivative of the functional ϕ at a point g in the class

$$\mathcal{C} = \left\{ g : g(t) = \int_0^\infty \frac{k(\theta - t)_+^{k-1}}{\theta^k} d\mu(\theta), \;\; \mu \text{ is a positive measure} \right\}$$

in the direction of $f_\theta(t) = \frac{k(\theta-t)_+^{k-1}}{\theta^k}, \theta \in \Theta = (0, \infty)$ is given by

$$\begin{aligned} D_\phi(f_\theta, g) &= \int_0^\infty \frac{k(\theta - t)_+^{k-1}}{\theta^k} g(t)dt - \int_0^\infty \frac{k(\theta - t)_+^{k-1}}{\theta^k} d\mathbb{G}_n(t) \\ &= \frac{k}{\theta^k}\left(H(\theta, g) - \mathbb{Y}_n(\theta)\right) \end{aligned}$$

where $H(\cdot, g)$ and $\mathbb{Y}_n$ are respectively the k-fold integral of g and $(k-1)$-fold integral of the empirical distribution function $\mathbb{G}_n$. When $g = \tilde{g}_n$, then $H(\cdot, g)$ is nothing but $\tilde{H}_n$ defined in Chapter 2. It follows from the characterization of $\tilde{g}_n$ that $D_\phi(f_\theta, \tilde{g}_n) \geq 0$ for all $\theta \in (0, \infty)$ and equal to zero if and only if θ belongs to the support of the mixing measure $\tilde{\mu}_n$ associated with the LSE $\tilde{g}_n$. The support reduction algorithm consists of the following steps:

1. Given the current iterate $g \in \mathcal{C}$ with support $S = \{\theta_1, \cdots, \theta_p\}$, we find the minimizer of $\theta \mapsto D_\phi(f_\theta, g)$ over $(0, \infty)$. If $D_\phi(f_\theta, g) \geq 0$ for all $\theta \in (0, \infty)$, then we conclude that g is the LSE $\tilde{g}_n$. Otherwise, we denote the minimizer by θ_{p+1}. Since the rank

of θ_{p+1} in the set $\{\theta_1, \cdots, \theta_p\}$ is not important for the description of the algorithm, we can assume, without loss of generality, that $\theta_{p+1} \geq \max(S)$. Thus, the new set of support points is $S_{new} = \{\theta_1, \cdots, \theta_p, \theta_{p+1}\}$.

2. We find the minimizer of ϕ over the class

$$\left\{ g : g(t) = \sum_{j=1}^{p+1} \sigma_j \frac{k(\theta_j - t)_+^{k-1}}{\theta_j^k}, \quad \sigma_j \in \mathbb{R}, \ j = 0, \cdots, p+1. \right\}.$$

This means that some of the weights $\sigma_1, \cdots, \sigma_{p+1}$ can be negative. Let g_{min} denote this minimizer.

3. If all the weights σ_j are nonnegative, then we move to the first step. Otherwise, we need to "go back" to the original class of k-monotone functions and this is ensured by finding a coefficient $\lambda \in (0, 1)$ such that the function $(1-\lambda)g + \lambda g_{min}$ is k-monotone.

We will show that there exists always λ such that $(1-\lambda)g + \lambda g_{min}$ is k-monotone. This operation is actually equivalent to *deleting* one point from the new support S_{new}. We find the minimizer of ϕ over the class of k-monotone functions with the new *reduced* support. This reduction is carried on until the obtained minimizer is a k-monotone function; that is, the weights corresponding to its support points are all nonnegative.

Let $S = \{\theta_1, \cdots, \theta_m\}$ be the current set of support points. The following lemma gives the characterization of the minimizer of ϕ in the class of functions g given by

$$g(t) = \sigma_1 \frac{k(\theta_1 - t)_+^{k-1}}{\theta_1^k} + \cdots + \sigma_m \frac{k(\theta_m - t)_+^{k-1}}{\theta_m^k}$$

where $0 < \theta_1 < \cdots < \theta_m$ and $\sigma_1, \cdots, \sigma_m \in \mathbb{R}$. This is also the class of polynomial splines s of degree $k-1$ that are $(k-2)$-times continuously differentiable at the knots $\theta_1, \cdots, \theta_m$ and satisfy the boundary conditions $s^{(j)}(\theta_m) = 0$ for $j = 0, \cdots, k-2$ (for a definition of polynomial splines, see e.g. NÜRNBERGER (1989), Definition 1.15, page 94). We denote this class by $\mathcal{C}'(\theta_1, \cdots, \theta_m)$.

Lemma 4.2.1 *A function g is the minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \cdots, \theta_m)$ if and only if g is the k-th derivative of the polynomial spline P of degree $2k-1$ and knots $\theta_1, \cdots, \theta_m$ that satisfies*

$$P(\theta_i) = \mathbb{Y}_n(\theta_i) \quad \textit{for } i = 1, \cdots, m, \tag{4.1}$$

$$P^{(j)}(0) = 0 \quad \textit{for } j = 0, \cdots, k-1, \tag{4.2}$$

and

$$P^{(l)}(\theta_m) = 0 \quad \textit{for } l = k, \cdots, 2k-2. \tag{4.3}$$

Proof. Let $\epsilon \in \mathbb{R}$ and suppose that g is the minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \cdots, \theta_m)$. We have for all $j = 1, \cdots, m$

$$D_\phi(f_{\theta_j}, g) = \lim_{\epsilon \to 0} \frac{\phi(g + \epsilon f_{\theta_j})) - \phi(g)}{\epsilon} = 0.$$

Conversely, suppose that $g \in \mathcal{C}'(\theta_1, \cdots, \theta_m)$ satisfies $D_\phi(f_{\theta_j}, g) = 0$ for all $j = 1, \cdots, m$. Let h be any arbitrary function in $\mathcal{C}(\theta_1, \cdots, \theta_m)$. By convexity of ϕ, we have

$$\begin{aligned} \phi(h) - \phi(g) &\geq D_\phi(h - g, g) \\ &= D_\phi\left(\sum_{j=1}^m (\sigma_{j,h} - \sigma_{j,g}) f_{\theta_j}, g\right) \\ &= \sum_{j=1}^m (\sigma_{j,h} - \sigma_{j,g}) D(f_{\theta_j}, g) \\ &= 0 \end{aligned}$$

which implies that g is the minimizer.

Now, notice that $D_\phi(f_{\theta_j}, g) = 0,\ j = 1, \cdots, m$, is equivalent to

$$H(\theta_j, g) = \mathbb{Y}_n(\theta_j),\ j = 1, \cdots, m,$$

where

$$H(\theta, g) = \int_0^\theta (\theta - t)^{k-1} g(t) dt.$$

By noticing that $H(\cdot, g)$ is a spline of degree $2k-1$ and knots $\theta_1, \cdots, \theta_m$ and satisfying the boundary conditions in (4.1, 4.2 and 4.3), the results follows. ■

The following lemma ensures that the *reduction step* is always possible.

Lemma 4.2.2 *Let $\{\theta_1, \cdots, \theta_{m-1}\}$ be the set of support points of the current iterate g. Let $\theta_m = argmin_{\theta \in (0,\infty)} D(f_\theta, g)$ and suppose without loss of generality that $\theta_m > \theta_{m-1}$. Let g_{min} be the minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \cdots, \theta_m)$. If g_{min} is not k-monotone, then there exists $\lambda \in (0,1)$ such that the function*

$$(1-\lambda)g + \lambda g_{min}$$

is k-monotone.

Proof. Since g_{min} minimizes ϕ over a bigger class , it follows that

$$\phi(g_{min}) < \phi(g).$$

The last inequality is strict because $g_{min} \neq g$. Using convexity of ϕ, we can write for any $\epsilon > 0$,

$$\begin{aligned}
\phi\left((1-\epsilon)g + \epsilon g_{min}\right) - \phi(g) &\leq (1-\epsilon)\phi(g) + \epsilon\phi(g_{min}) - \phi(g) \\
&= \epsilon(\phi(g_{min}) - \phi(g)) \\
&< 0.
\end{aligned}$$

Now, there exist $\sigma_{1,g}, \cdots, \sigma_{m-1,g}$ such that $\sigma_{j,g} \geq 0$ for $j = 1, \cdots, m-1$ and $\sigma_{1,g_{min}}, \cdots, \sigma_{m,g_{min}} \in \mathbb{R}$ such that g and g_{min} can be written as

$$g(t) = \sigma_{1,g} k \frac{(\theta_1 - t)_+^{k-1}}{\theta_1^k} + \cdots + \sigma_{m-1,g} k \frac{(\theta_{m-1} - t)_+^{k-1}}{\theta_{m-1}^k}$$

and

$$g(t) = \sigma_{1,g_{min}} k \frac{(\theta_1 - t)_+^{k-1}}{\theta_1^k} + \cdots + \sigma_{m,g_{min}} k \frac{(\theta_m - t)_+^{k-1}}{\theta_m^k}.$$

By passing ϵ to the limit, we obtain

$$\begin{aligned}\lim_{\epsilon\searrow 0}\frac{\phi\left((1-\epsilon)g+\epsilon g_{min}\right)-\phi(g)}{\epsilon} &= D_\phi(g_{min}-g,g)\\ &= \sigma_{m,g_{min}}D_\phi(f_{\theta_m},g)+\sum_{j=1}^{m-1}(\sigma_{j,g_{min}}-\sigma_{j,g})D_\phi(f_{\theta_j},g)\\ &= \sigma_{m,g_{min}}D_\phi(f_{\theta_m},g)\end{aligned}$$

where in the last equality we used the fact that $D(f_{\theta_j},g)=0$ for $j=1,\cdots,m-1$. Since by definition of θ_m, $D_\phi(f_{\theta_m},g)<0$ it follows that $\sigma_{m,g_{min}}>0$. Let λ be in $[0,1]$ and consider g_λ the weighted sum of g and g_{min}:

$$g_\lambda=(1-\lambda)g+\lambda g_{min}.$$

We want to find the largest λ such that g_λ is k-monotone. The parameter λ has to be chosen such that

$$\begin{aligned}(1-\lambda)\sigma_{1,g}+\lambda\sigma_{1,g_{min}} &\geq 0\\ &\vdots\\ (1-\lambda)\sigma_{m-1,g}+\lambda\sigma_{m-1,g_{min}} &\geq 0\\ (1-\lambda)\sigma_{m,g}+\lambda\sigma_{m,g_{min}} &\geq 0.\end{aligned}$$

Note that the last inequality is automatically satisfied since $\sigma_{m,g_{min}}>0$ and hence we only need to worry about the first $m-1$ inequalities (it is implicitly assumed that $m\geq 2$). Let J be the set of integers $j\in\{1,\cdots,m-1\}$ such that

$$\sigma_{j,g_{min}}<0.$$

For $j\in J$, define λ_j by

$$\lambda_j=\frac{\sigma_{j,g}}{\sigma_{j,g}-\sigma_{j,g_{min}}}.$$

Clearly, $\lambda_j\in(0,1)$. Now, if we consider j_0 to be the index of the smallest λ_j; i.e.,

$$j_0=\mathrm{argmin}_{j\in J}\lambda_j,$$

then it is easy to verify that for all $j \in J$

$$(1-\lambda_{j_0})\sigma_{j,g} + \lambda_{j_0}\sigma_{j,g_{min}} \geq 0$$

with equality if and only if $j = j_0$ (we assume here that j_0 is unique). To see that, notice that if $\lambda \in (0,1)$ satisfies

$$(1-\lambda)\sigma_{j,g} + \lambda\sigma_{j,g_{min}} \geq 0, \quad \text{for all } \ j \in J \tag{4.4}$$

then

$$\lambda \leq \lambda_j, \quad \text{for all } \ j \in J.$$

It follows that $\lambda \leq \min_{j \in J} \lambda_j = j_{j_0}$ and that the maximal value of $\lambda \in (0,1)$ satisfying the inequality in (4.4) is equal to λ_{j_0}.

Since $(1-\lambda_{j_0})\sigma_{j_0,g} + \lambda_{j_0}\sigma_{j_0,g_{min}} = 0$, the knot θ_{j_0} is deleted from the set of knots $S = \{\theta_1, \cdots, \theta_m\}$. The next step is to compute the $(2k-1)$-th spline with the new set of knots $S \backslash \{\theta_{j_0}\}$. Notice that by moving from the previous step to the new one, the monotonicity of the algorithm is maintained. Indeed, using again the convexity of ϕ, we have

$$\begin{aligned}
\phi(g_{\lambda_{j_0}}) &= \phi((1-\lambda_{j_0})g + \lambda_{j_0} g_{min}) \\
&\leq (1-\lambda_{j_0})\phi(g) + \lambda_{j_0}\phi(g_{min}) \\
&< (1-\lambda_{j_0})\phi(g_{min}) + \lambda_{j_0}\phi(g_{min}) \\
&= \phi(g_{min}).
\end{aligned}$$

Therefore, if g_{j_0} is the minimizer of ϕ over the class of functions $\mathcal{C}(S \backslash \{\theta_{j_0}\})$, we should have

$$\phi(g_{j_0}) \leq \phi(g_{\lambda_{j_0}})$$

which implies that $\phi(g_{j_0}) < \phi(g_{min})$. ■

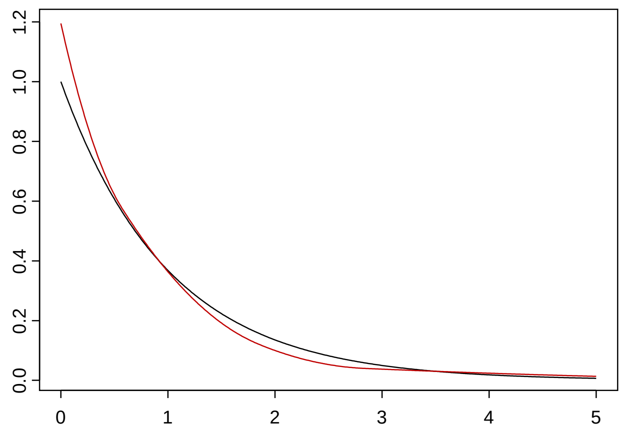

Figure 4.1: The exponential density (in black) and the Least Squares estimator of the (mixed) k-monotone density based on $n = 100$ and $k = 3$ (in red).

To start the algorithm, we fix some initial value $\theta^{(0)} > X_{(n)}$ and minimize the functional ϕ over the cone

$$\mathcal{C}^{(0)} = \left\{ g : g(t) = C\, \frac{k(\theta^{(0)} - t)^{k-1}}{(\theta^{(0)})^k},\ \ C > 0 \right\}.$$

For this purpose, we need to find the value $C^{(0)}$ that minimizes the quadratic function

$$C \mapsto \frac{k^2}{2(2k-1)\theta^{(0)}} C^2 - \frac{1}{n}\sum_{j=1}^{n} k \frac{(\theta^{(0)} - X_{(j)})^{k-1}}{(\theta^{(0)})^k} C$$

which yields

$$C^{(0)} = \left(\frac{2k-1}{k}\right) \frac{1}{n} \sum_{j=1}^{n} \frac{(\theta^{(0)} - X_{(j)})^{k-1}}{(\theta^{(0)})^{k-2}}.$$

As in Groeneboom, Jongbloed, and Wellner (2003), we used an "alternative" directional derivative. Using their notation, the "usual" directional derivative at a point g in the direction of f_θ, denoted before by $D_\phi(f_\theta, g)$, is equal to $c_1(\theta)$, where

$$\phi(g + \epsilon f_\theta) = \phi(g) + \epsilon c_1(\theta) + \frac{\epsilon^2}{2} c_2(\theta)$$

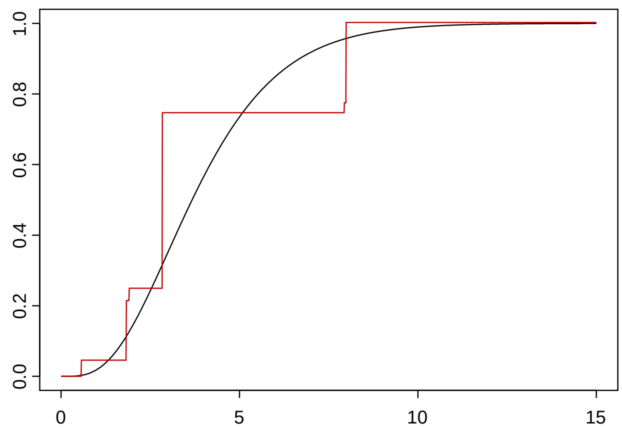

Figure 4.2: The cumulative distribution function of a $Gamma(4,1)$ (in black) and the Least squares estimator of the mixing distribution based on $n = 100$ and $k = 3$ (in red).

with

$$c_2(\theta) = \int_0^\infty f_\theta^2(t)dt = \frac{k^2}{(2k-1)\theta}.$$

The "alternative" directional derivative is given by

$$\tilde{D}_\phi(f_\theta, g) = \frac{D_\phi(f_\theta, g)}{\sqrt{c_2(\theta)}} = k\frac{H(\theta, g) - \mathbb{Y}_n(\theta)}{\theta^{k-1/2}}.$$

Remark 4.2.1 *It should be mentioned here that the "gridless" step that was implemented by* Groeneboom, Jongbloed, and Wellner (2003) *was not considered here. In practice, we only consider a finite grid over which we minimize the directional derivative. The obtained LSE is the minimizer of ϕ over the class of k-monotone functions whose support points belong to the finite grid. The purpose of the "gridless" implementation is to obtain a numerical solution that is closest to the theoretical one by perturbing the support points of the solution. By performing this fine tuning, one can run the algorithm once again considering the new grid and obtain a new minimizer. This step is repeated until the gradient of*

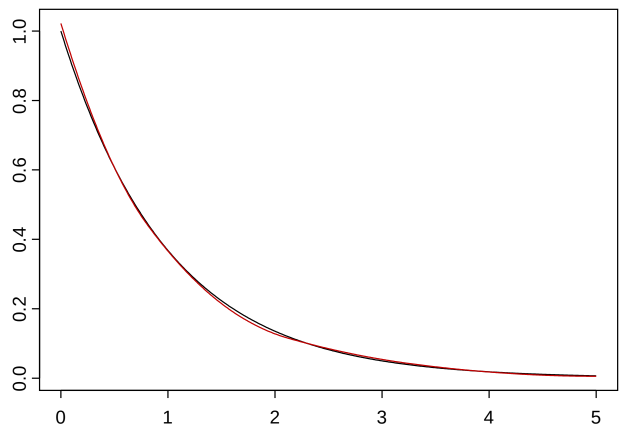

Figure 4.3: The exponential density (the true mixed density), in black and its Least Squares estimator based on $n = 1000$ and $k = 3$, in red.

the functional ϕ is sufficiently small.

Now we describe the preliminary simulations that we have performed. From a standard Exponential, we simulated two samples of respective sizes $n = 100$ and $n = 1000$. The Exponential density is completely monotone and therefore is k-monotone for all integers $k \geq 1$. This is actually the motivation behind considering nonparametric estimation of k-monotone densities (see Chapter 1 for more details). The code of the algorithm was written in S and can be found in Appendix C. To illustrate the asymptotic distribution theory developed in Chapter 2 for any integer $k \geq 2$, we computed the LSE based on $n = 100$ and $n = 1000$ in two different cases: $k = 3$ and $k = 6$.

Note that if θ is a support point of the minimizing measure, then $\theta > X_{(1)}$. This follows from the simple fact that for all $\theta \in (0, X_{(1)})$, $(\theta - X_{(j)})_+^{k-1} = 0$ for $j = 1, \cdots, n$. Therefore, adding $\theta \in (0, X_{(1)})$ to the set of support points does not effect the value of the sum

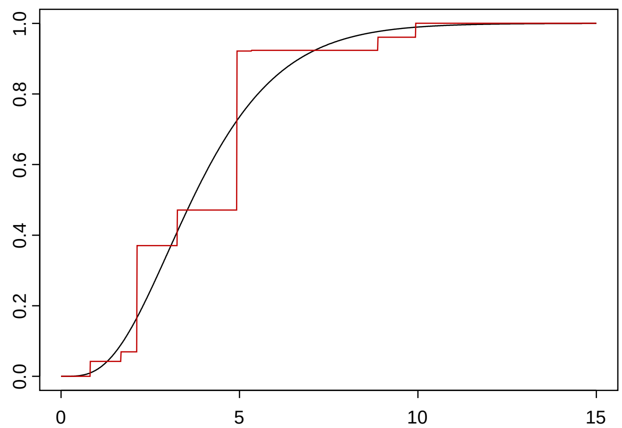

Figure 4.4: The cumulative distribution function of a $Gamma(4, 1)$ (the true mixing distribution), in black and the Least Squares estimator of the mixing distribution based on $n = 1000$ and $k = 3$, in red.

$n^{-1} \sum_{j=1}^{n} g(X_j)$ whereas it increases the value of the integral $\int_0^\infty g^2(t)dt$. The minimization was performed on a finite grid such that, for given n and k, the maximal distance between its points is taken to be 10^{-2}. In practice, we found that it is enough to take $2kX_{(n)}$ as an upper bound for the largest support point as we obtained similar results with larger bounds. The obtained estimates can be found in Table 4.1.

For $k = 3$, the plots in Figure 4.1 and Figure 4.3 show the LSE of the Exponential density based on $n = 100$ and $n = 1000$ respectively. The "alternative" directional derivative $\tilde{D}_\phi(f_\theta, \tilde{g}_n)$, for $n = 1000$, is plotted in Figure 4.5. In the inverse problem, plots of the LSE of the true mixing distribution are shown in Figure 4.2 and Figure 4.4. In general, the true mixing distribution that corresponds to a standard Exponential when viewed as a k-monotone density is a $Gamma(k + 1, 1)$. Indeed, note that

$$\int_x^\infty \frac{1}{\Gamma(k)} (t - x)^{k-1} e^{-(t-x)} dt = 1$$

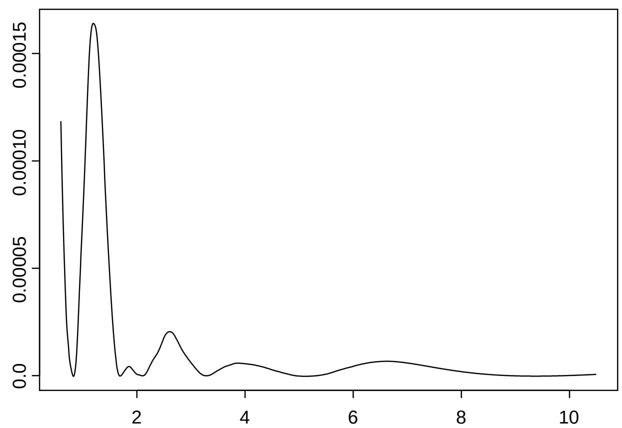

Figure 4.5: The directional derivative for the Least Squares estimator of the Exponential density based on $n = 1000$ and $k = 3$.

for all $x > 0$. It follows that,

$$\begin{aligned}
\exp(-x) &= \int_x^\infty \frac{(t-x)^{k-1}}{(k-1)!} e^{-t} dt \\
&= \int_0^\infty \frac{(t-x)_+^{k-1}}{(k-1)!} e^{-t} dt \\
&= \int_0^\infty k \frac{(t-x)_+^{k-1}}{t^k} \frac{1}{k!} t^k e^{-t} dt \\
&= \int_0^\infty k \frac{(t-x)_+^{k-1}}{t^k} f_k(t) dt \qquad (4.5)
\end{aligned}$$

where f_k is the $Gamma(k+1, 1)$ density.

For $k = 6$, similar plots were produced for $n = 100$ and $n = 1000$: for the direct problem, see Figure 4.6 and Figure 4.8, and for the inverse one, see Figure 4.7 and Figure 4.9.

The figures show consistency of the LSE and it is clear that convergence for estimating the Exponential density is much faster than for estimating the Gamma distribution. This is expected since in the direct problem, the rate of convergence is $n^{-k/(2k+1)}$ whereas it is

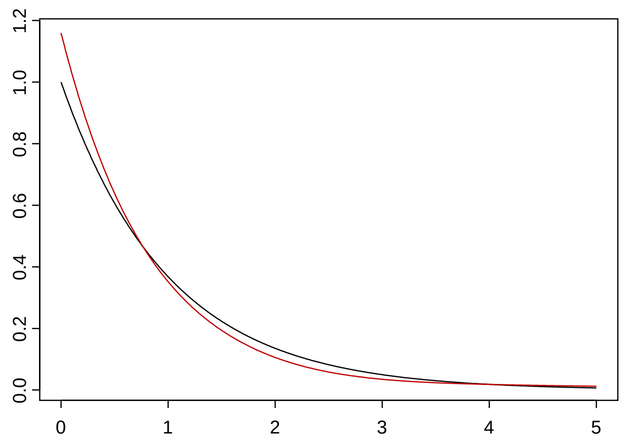

Figure 4.6: The exponential density (the true mixed density), in black and its Least Squares estimator based on $n = 100$ and $k = 6$, in red.

equal to $n^{-1/(2k+1)}$ in the inverse problem. Note also the rate $n^{-1/(2k+1)}$ is slower for larger k and therefore, one should expect to see fewer support points as $k \to \infty$. This fact is confirmed in the numerical examples above (for $n = 1000$, there are 8 support points for $k = 3$ and 4 for $k = 6$, see Table 4.1) and in many other simulations that we performed.

4.3 Approximation of the process H_k on $[-c, c]$

We will focus here on the case when k is even. When k is odd, the steps are very similar. The goal of the algorithm is to find the minimizer of the functional

$$\phi(g) = \frac{1}{2}\int_{-c}^{c} g^2(t)dt - \int_{-c}^{c} g(t)dX_k(t)$$

where

$$dX_k(t) = dW(t) + t^k dt$$

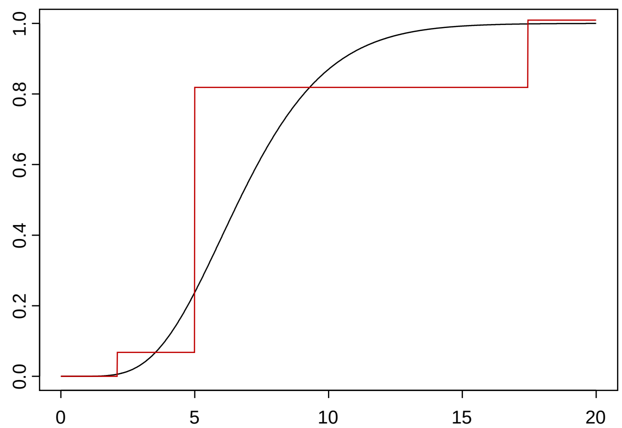

Figure 4.7: The cumulative distribution function of a $Gamma(7,1)$ (the true mixing distribution), in black and its Least squares estimator based on $n = 100$ and $k = 6$, in red.

and W is two-sided Brownian motion starting at 0, over $\mathcal{C}$ the class of functions g that are k-convex; i.e. $g^{(k-2)}$ exists and is convex, and satisfies the boundary conditions

$$\left(g^{(k-2)}(\pm c), \cdots, g^{(2)}(\pm c), g(\pm c)\right) = \left(\frac{k!}{2!}c^2, \cdots, \frac{k!}{(k-2)!}c^{k-2}, c^k\right). \tag{4.6}$$

Recall that if $H_{c,k}$ is the k-fold integral of $g_{c,k}$ determined by

$$H_{c,k}(c) = Y_k(c), H_{c,k}^{(2)}(c) = Y_k^{(2)}(c), \cdots, H_{c,k}^{(k-2)}(c) = Y_k^{(k-2)}(c), \tag{4.7}$$

then $g_{c,k}$ is the minimizer if and only if

$$H_{c,k}(t) \geq Y_k(t), \qquad t \in [-c, c]$$

and

$$\int_{-c}^{c} \left(H_{c,k}(t) - Y_k(t)\right) dg_{c,k}^{(k-1)}(t) = 0,$$

where

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

Table 4.1: Table of the obtained LS estimates for $k = 3, 6$ and $n = 100, 1000$ and the corresponding numbers of iterations N_{it}. A support point is denoted by $\tilde{a}$ and its mass by $\tilde{w}$.

k, n	N_{it}	$(\tilde{a}, \tilde{w})$
$k = 3, n = 100$	13	$(0.569, 0.0459), (1.829, 0.168), (1.909, 0.0347),$
		$(2.839, 0.497), (7.939, 0.027), (7.989, 0.227)$
$k = 3, n = 1000$	14	$(0.814, 0.042), (1.674, 0.027), (2.124, 0.300), (3.254, 0.100),$
		$(4.924, 0.450), (5.334, 0.001), (8.874, 0.037), (9.934, 0.039)$
$k = 6, n = 100$	4	$(2.109, 0.067), (4.999, 0.750), (17.449, 0.190)$
$k = 6, n = 1000$	6	$(2.625, 0.017), (3.615, 0.478), (6.575, 0.478), (11.375, 0.262)$

The above characterization gives a necessary and sufficient condition for a function g in the considered class to be the solution for the minimization problem. But it also implies that this solution cannot have a strictly increasing $(k-1)$-st derivative on a set with nontrivial interior. Indeed, if we assume that there exists an open interval $I \subseteq (-c, c)$ of positive length on which $g_{c,k}^{(k-1)}$ is strictly increasing, then this would imply that $Y_k = H_{c,k}$ on I and that the $(k-1)$-fold integral of Brownian motion is in $C^{2k-2}(I)$. Therefore, the function $g_{c,k}^{(k-1)}$ has to increase on a set of Lebesgue measure zero. We conjecture that this set is finite and consists of the discontinuity points of the monotone function $g_{c,k}^{(k-1)}$. For the particular case of $k = 2$, there is still no proof available for this conjecture (see Groeneboom, Jongbloed, and Wellner (2001a), Section 4). The main difficulty of this problem lies in the fact that in principle, the monotone function $g_{c,k}^{(k-1)}$ could be a Cantor-type function in which case, the set on which it increases is Lebesgue measure zero and is uncountable (see e.g. Gelbaum and Olmsted (1964), example 15, page 96). Based on this conjecture, $H_{c,k}$ is a spline of degree $2k-1$ that stays above Y_k and touches it at the discontinuity points of $g_{c,k}^{(k-1)}$; i.e., those points where $H_{c,k}^{(2k-2)} = g_{c,k}^{(k-2)}$ changes its slope. Therefore, in order to obtain the solution $g_{c,k}$ and its derivatives $g'_{c,k}, \cdots, g_{c,k}^{(k-1)}$, we first find $H_{c,k}$ and then differentiate it $(k+j)$-times for $j = 0, \cdots, k-1$.

The steps of the *support reduction algorithm* are very similar to those described in the

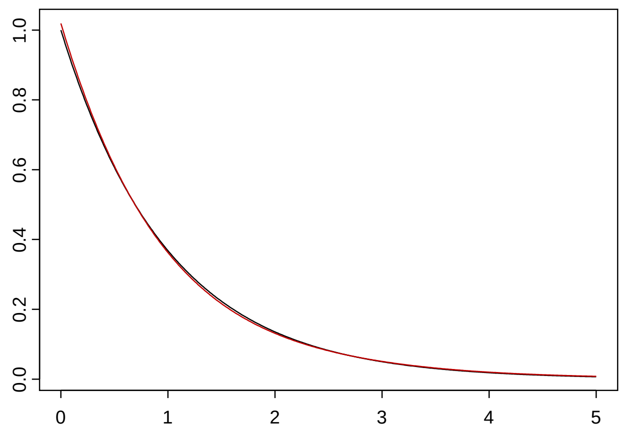

Figure 4.8: The exponential density (the true mixed density), in black and its Least Squares estimator based on $n = 1000$ and $k = 6$, in red.

previous section on calculating the LSE of a k-monotone density. In view of the conjecture, we can restrict ourselves to the class of functions

$$\mathcal{C} = \left\{ g : g(t) = \sum_{j=0}^{k-1} \lambda_j \frac{t^j}{j!} + \mu_1 (t-\theta_1)_+^{k-1} + \cdots + \mu_p (t-\theta_p)_+^{k-1}, \ p \in \mathbb{N} \backslash \{0\} \right\}$$

where $\lambda_j \in \mathbb{R}, \ \mu_j \geq 0$ for $1 \leq j \leq p$ such that g satisfies the constraints in (4.6). Note that any element $g \in \mathcal{C}$ is a spline of degree $k-1$ and simple knots $\theta_1, \cdots, \theta_p$. This means that g is $(k-2)$-times continuously differentiable at these knots. From each iterate $g \in \mathcal{C}$, we can move in the direction of the function

$$f_\theta(t) = \frac{(t-\theta)_+^{k-1}}{(k-1)!} + \alpha_{k-1}(\theta) \frac{(t+c)^{k-1}}{(k-1)!} + \alpha_{k-3}(\theta) + \frac{(t+c)^{k-3}}{(k-3)!} + \cdots + \alpha_1(\theta)(u+c)$$

where

$$\begin{aligned} \alpha_{k-1}(\theta) &= -\frac{(c-\theta)}{2c} \\ \alpha_{k-3}(\theta) &= -\alpha_{k-1}(\theta) \frac{(2c)^3}{3!} - \frac{(c-\theta)^3}{3!} \end{aligned}$$

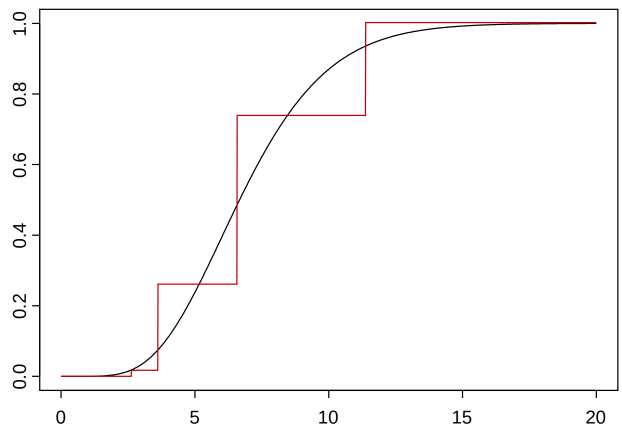

Figure 4.9: The cumulative distribution function of a $Gamma(7,1)$ (the true mixing distribution), in black and its Least squares estimator based on $n = 1000$ and $k = 6$, in red.

$$\vdots$$

$$\alpha_1(\theta) = -\alpha_{k-1}(\theta)\frac{(2c)^{k-1}}{(k-1)!} - \cdots - \alpha_3(\theta)\frac{(2c)^3}{3!} - \frac{(c-\theta)^{k-1}}{(k-1)!}.$$

Indeed, for all $\theta \in [-c, c]$, the function f_θ is a spline of degree $k-1$ with θ as its unique simple knot. Moreover, f_θ satisfies the boundary conditions

$$f_\theta^{(2j)}(\pm c) = 0, \quad \text{for} \ \ j = 0, \cdots, (k-2)/2. \tag{4.8}$$

For an arbitrary $\epsilon > 0$, the function $g + \epsilon f_\theta$ belongs to the class $\mathcal{C}$ and the directional derivative of ϕ at g in the direction of f_θ is given by

$$D_\phi(g, f_\theta) = H(\theta, g) - Y_k(\theta) \tag{4.9}$$

where $H(\cdot, g)$ is the k-fold integral of g determined by the boundary conditions

$$H^{(2j)}(\pm c, g) = Y_k^{(2j)}(\pm c), \quad \text{for} \ \ j = 0, \cdots, (k-2)/2. \tag{4.10}$$

To see the equality in (4.9), note first that $D(f_\theta, g)$ is given by

$$\begin{aligned} D(f_\theta, g) &= \int_{-c}^{c} f_\theta(t) g(t) dt - \int_{-c}^{c} f_\theta(t) dX_k(t) \\ &= \int_{-c}^{c} f_\theta(t) d(H^{(k-1)}(t,g) - Y_k^{(k-1)}(t)) \end{aligned}$$

Thus, using successive integration by parts and the boundary conditions in (4.8) and (4.10), we can write

$$\begin{aligned} & D_\phi(g, f_\theta) \\ &= \left[\left(H^{(k-1)}(t,g) - Y_k^{(k-1)}(t)\right) f_\theta(t)\right]_{-c}^{c} - \int_{-c}^{c} \left(H^{(k-1)}(t,g) - Y_k^{(k-1)}(t)\right) f_\theta'(t) dt \\ &= -\int_{-c}^{c} \left(H^{(k-1)}(t,g) - Y_k^{(k-1)}(t)\right) f_\theta'(t) dt \\ &= -\left[\left(H^{(k-2)}(t,g) - Y_k^{(k-2)}(t)\right) f_\theta'(t)\right]_{-c}^{c} + \int_{-c}^{c} \left(H^{(k-2)}(t,g) - Y_k^{(k-2)}(t)\right) f_\theta''(t) dt \\ &= \int_{-c}^{c} \left(H^{(k-2)}(t,g) - Y_k^{(k-2)}(t)\right) f_\theta''(t) dt \\ & \vdots \\ &= \int_{-c}^{c} \left(H(t,g) - Y_k(t)\right) f_\theta^{(k-1)}(t) dt \\ &= H(\theta, g) - Y_k(\theta). \end{aligned}$$

Note that Y_k plays here a role that is similar to that of the process $\mathbb{Y}_n$. Let $S = \{\theta_1, \cdots, \theta_m\}$ be the set of knots of the current iterate g. The function $H(\cdot, g)$ is a spline of degree $2k-1$ with simple knots $-c, \theta_1, \cdots, \theta_m, c$. If $H(\cdot, g) \geq Y_k$, then $g = H^{(k)}(\cdot, g)$ is the solution of the minimization problem. Otherwise, we add $\theta_{m+1} = \text{argmin}_{\theta \in [-c,c]}(H(\cdot, g)(\theta) - Y_k(\theta))$ to the support S. Without loss of generality, we can assume that $\theta_1 < \cdots < \theta_m < \theta_{m+1}$. Now, let $\mathcal{C}'(\theta_1, \cdots, \theta_{m+1})$ be the class of polynomial splines of degree $k-1$, with simple knots $\theta_1, \cdots, \theta_{m+1}$ satisfying the boundary conditions in (4.6); i.e.,

$$\mathcal{C}'(\theta_1, \cdots, \theta_{m+1}) = \left\{ g : g(t) = \sum_{j=0}^{k-1} \lambda_j \frac{t^j}{j!} + \sigma_1 (t - \theta_1)_+^{k-1} + \cdots + \sigma_{m+1}(t - \theta_{m+1})_+^{k-1} \right\}$$

where $\sigma_j \in \mathbb{R}$ and the λ_j's are different from the ones used in the definition of the class $\mathcal{C}$. Consider H_{min} to be the spline of degree $2k-1$ and simple knots $\theta_1, \cdots, \theta_{m+1}$ satisfying

$$H_{min}(\theta_j) = Y_k(\theta_j), \quad \text{for} \quad j = 1, \cdots, m+1.$$

$$H_{min}^{(2j)}(\pm c) = Y_k^{(2j)}(\pm c), \quad \text{for} \quad j = 0, \cdots, (k-2)/2$$

and

$$H_{min}^{(2j)}(\pm c) = \frac{k!}{(2k-2j)!} c^{2k-2j}, \quad \text{for} \quad j = k, \cdots, (2k-2)/2.$$

The following lemma gives the solution of minimizing ϕ over the class $\mathcal{C}'(\theta_1, \cdots, \theta_{m+1})$.

Lemma 4.3.1 *Let H_{min} be the spline defined above. The function $g_{min} = H_{min}^{(k)}$ is the minimizer of the functional ϕ over the class $\mathcal{C}'(\theta_1, \cdots, \theta_{m+1})$.*

Proof. The arguments are very similar to those used in the proof of Lemma 4.2.2. ■

There exist $\lambda_0, \cdots, \lambda_{2k-1}$, and $\sigma_1, \cdots, \sigma_{m+1}$ such that the spline H_{min} can written as

$$H_{min} = H(t, g_{min}) = \sum_{j=0}^{2k-1} \lambda_j \frac{t^j}{j!} + \sigma_1(t-\theta_1)_+^{2k-1} + \cdots + \sigma_{m+1}(t-\theta_{m+1})_+^{2k-1}.$$

To find the parameters $\lambda_{2k-1}, \cdots, \lambda_1, \lambda_0$ and $\sigma_1, \cdots, \sigma_{m+1}$, we solve a linear system of dimension $(2k+m+1)\times(2k+m+1)$ using the $2k+m+1$ boundary conditions satisfied by H_{min}.

The reduction step is given by the following lemma:

Lemma 4.3.2 *Let g be the current iterate in $\mathcal{C}$ with knots $\theta_1, \cdots, \theta_m$ and $g_{min} = H_{min}^{(k)}$ be new minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \cdots, \theta_{m+1})$. If g_{min} is not in the class $\mathcal{C}'$, then there exists $\lambda \in (0,1)$ such that $(1-\lambda)g + \lambda g_{min} \in \mathcal{C}'$.*

Proof. The arguments are very similar to those used in the proof of Lemma 4.2.2. ■

The steps of the algorithm can be summarized as follows:

1. Given the current iterate g with set of simple knots $S = \{\theta_1, \cdots, \theta_m\}$, we calculate $argmin_{\theta\in[-c,c]} D_\theta(f_\theta, g) = argmin_{\theta\in[-c,c]}(H(\theta, g) - Y_k(\theta))$. If $D_\theta(f_\theta, g) \geq 0$ for all $\theta \in [-c, c]$, then g is the minimizer of ϕ over the class of splines $\mathcal{C}$ and its k-fold

integral $H(\cdot, g)$ is an approximation of the process H_k. Otherwise, we denote $\theta_{m+1} = argmin_{\theta\in[-c,c]}(H(\theta, g) - Y_k(\theta))$. If we assume without loss of generality that $\theta_{m+1} > \theta_m$, then $S_{new} = \{\theta_1, \cdots, \theta_m, \theta_{m+1}\}$ is the new set of knots.

2. We find g_{min} the minimizer of ϕ over the class $\mathcal{C}'(\theta_1, \cdots, \theta_{m+1})$.

3. If $g_{min} \in \mathcal{C}$, we move the Step 1. Otherwise, we find the maximal value of $\lambda \in (0, 1)$ such that $(1-\lambda)g+\lambda g_{min} \in \mathcal{C}$. By finding such a λ, a point θ_j for some $j \in \{1, \cdots, m\}$ will be deleted from the current support. We find the minimizer over $\mathcal{C}'(S_{new}\backslash\{\theta_j\})$. This will be repeated until the minimizer is in the class $\mathcal{C}$.

The algorithm has to start somewhere and the most natural starting spline is the polynomial $H^{(0)}_{c,k}$ that was used in Chapter 3 to prove that $H_{c,k}$ and Y_k have at least a point of touch with probability converging to 1 as $c \to \infty$. Recall that $H^{(0)}_{c,k}$ is the unique polynomial P of degree $2k - 2$ that satisfies (4.6) and (4.7). To be conform with the notation used in Chapter 2, we write the polynomial $H^{(0)}_{c,k}(t)$ as

$$\begin{aligned} H^{(0)}_{c,k}(t) &= \frac{\alpha_{2k-2}}{(2k-2)!}t^{2k-2} + \frac{\alpha_{2k-4}}{(2k-2)!}t^{2k-2} + \cdots + \frac{\alpha_k}{k!}t^k + \frac{\alpha_{k-1}}{(k-1)!}t^{k-1} \\ &+ \frac{\alpha_{k-2}}{(k-2)!}t^{k-2} \cdots + \alpha_0, \end{aligned}$$

where $\alpha_{2k-2}, \cdots, \alpha_k$ are given by

$$\alpha_{2k-2} = \frac{k!}{2!}c^2,$$

$$\alpha_{2k-2j} = \frac{k!}{(2j)!}c^{2j} - \left(\frac{\alpha_{2k-2}}{(2j-2)!}c^{2j-2} + \cdots + \frac{\alpha_{2k-2j+2}}{2!}c^2\right)$$

for $j = 2, \cdots, k/2$, whereas $\alpha_{k-1}, \alpha_{k-2}, \cdots, \alpha_0$ are given by

$$\alpha_{k-1} = \frac{Y_k^{(k-2)}(c) - Y_k^{(k-2)}(-c)}{2c},$$

$$\alpha_{k-2} = \frac{Y_k^{(k-2)}(-c) + Y_k^{(k-2)}(c)}{2} - \left(\frac{\alpha_{2k-2}}{k!}c^k + \cdots + \frac{\alpha_k}{2!}c^2\right),$$

$$\alpha_{k-2j-1} = \frac{Y_k^{(k-2j-2)}(c) - Y_k^{(k-2j-2)}(-c)}{2c} - \left(\frac{\alpha_{k-1}}{(2j+1)!}c^{2j} + \cdots + \frac{\alpha_{k-2j+1}}{3!}c^2\right),$$

and

$$\alpha_{k-2j-2} = \frac{Y_k^{(k-2j-2)}(c) + Y_k^{(k-2j-2)}(-c)}{2} - \left(\frac{\alpha_{2k-2}}{(k+2j)!}c^{k+2j} + \cdots + \frac{\alpha_{k-2j}}{2!}c^2\right)$$

for $j = 1, \cdots, (k-2)/2$.

Example 4.3.1 *For $k = 2$, $H_{c,2}^{(0)}$ is given by*

$$H_{c,2}^{(0)}(t) = \frac{\alpha_2}{2!}t^2 + \alpha_1 t + \alpha_0, \quad t \in [-c, c]$$

with

$$\alpha_2 = c^2, \quad \alpha_1 = \frac{Y_2(c) - Y_2(-c)}{2c}, \quad \alpha_0 = \frac{Y_2(-c) + Y_2(c)}{2} - c^2.$$

Example 4.3.2 *For $k = 4$, $H_{c,4}^{(0)}$ is given by*

$$H_{c,4}^{(0)}(t) = \frac{\alpha_6}{6!}t^6 + \frac{\alpha_4}{4!}t^4 + \frac{\alpha_3}{3!}t^3 + \frac{\alpha_2}{2!}t^2 + \alpha_1 t + \alpha_0, \quad t \in [-c, c]$$

with

$$\alpha_6 = \frac{4!}{2!}c^2, \quad \alpha_4 = \left(1 - \frac{4!}{(2!)^2}\right)c^4, \quad \alpha_3 = \frac{Y_4''(c) - Y_4''(-c)}{2c},$$

$$\begin{aligned}\alpha_2 &= \frac{Y_4''(-c) + Y_4''(-c)}{2} - \left(\frac{\alpha_6}{4!}c^4 + \frac{\alpha_4}{2!}c^2\right)\\ &= \frac{Y_4''(-c) + Y_4''(-c)}{2} - \left(1 - \frac{4!}{(2!)^3}\right)c^6\end{aligned}$$

$$\begin{aligned}\alpha_1 &= \frac{Y_4(c) - Y_4(-c)}{2c} - \frac{\alpha_3}{3!}c^2\\ &= \frac{Y_4(c) - Y_4(-c)}{2c} - \frac{1}{3!}\left(\frac{Y_4''(c) - Y_4''(-c)}{2c}\right)\end{aligned}$$

and

$$\begin{aligned}\alpha_0 &= \frac{Y_4(-c)+Y_4(c)}{2} - \left(\frac{\alpha_6}{6!}c^6 + \frac{\alpha_4}{4!}c^4 + \frac{\alpha_2}{2!}c^2\right)\\ &= \frac{Y_4(-c)+Y_4(c)}{2} - \frac{1}{2!}\left(\frac{Y_4''(-c)+Y_4''(c)}{2}\right)c^2 - \left(\frac{4!}{2!6!} + \frac{1}{4!}\left(1-\frac{4!}{(2!)^2}\right)\right.\\ &\quad \left. - \frac{1}{2!}\left(1 - \frac{4!}{(2!)^3}\right)\right)c^8.\end{aligned}$$

The algorithm was run to obtain an approximation to the process H_k and some of the derivatives $H_k^{(j)}$ for $k = 3$ and $k = 6$ on the interval $[-4, 4]$. Furthermore, for $k = 3$ we obtained similar approximations but on the bigger intervals $[-6, 6]$ and $[-8, 8]$. The purpose of these additional computations was to look at the effect of letting $c \to \infty$ on the locations of the jump points and also on the heights of the jumps. A C program, implementing an approximation to the processes $Y_k, Y_k', \cdots, Y_k^{(k-1)}$ on any interval $[-n, n]$ for $n \in \mathbb{N}\backslash\{0\}$ was developed and can be found in Appendix C. The approximation to Brownian motion and its successive primitives on $[0, 1]$ was based on the Haar function construction (see e.g. Rogers and Williams (1994), Section 1.6). To obtain an approximation of these processes on $[-n, n]$, independent copies were generated on the intervals $[j, j+1]$ for $j = -n, \cdots, n-1$ and pasted "smoothly" at the boundaries. A detailed description of the method and related formulas can be found in Appendix B. For both $k = 3$ and $k = 6$, we took a finite grid with a mesh of size 2^{-11}. The iterative $2k-1$-th spline algorithm was written in S and the corresponding code can be found in Appendix C. The C program was used offline and the obtained approximations to $Y_k, \cdots, Y_k^{(k-1)}$ were stored in a matrix that was thereafter imported and used as an input for the iterative algorithm. For a given interval $[-n, n]$, the output is itself an approximation to the process $H_{n,k}$, the k-fold integral of the LS solution of the Gaussian problem $dX_k(t) = t^k dt + dW(t)$ on $[-n, n]$. An approximation to the derivatives $H_{n,k}', \cdots, H_{n,k}^{(2k-1)}$ can be also obtained on the same chosen grid.

For both $k = 3$ and $k = 6$, the upper left plot in Figure 4.10 and Figure 4.11 shows the difference $-(H_{n,k} - Y_k)$ and $H_{n,k} - Y_k$ on $[-4, 4]$ respectively. The sign of $H_{n,k} - Y_k$ is as expected: nonpositive (nonnegative) when k is odd (even). The curves touch the abscissa axis at the points where the derivative $H_{n,k}^{(2k-2)}$ changes its slope. In the upper right plots

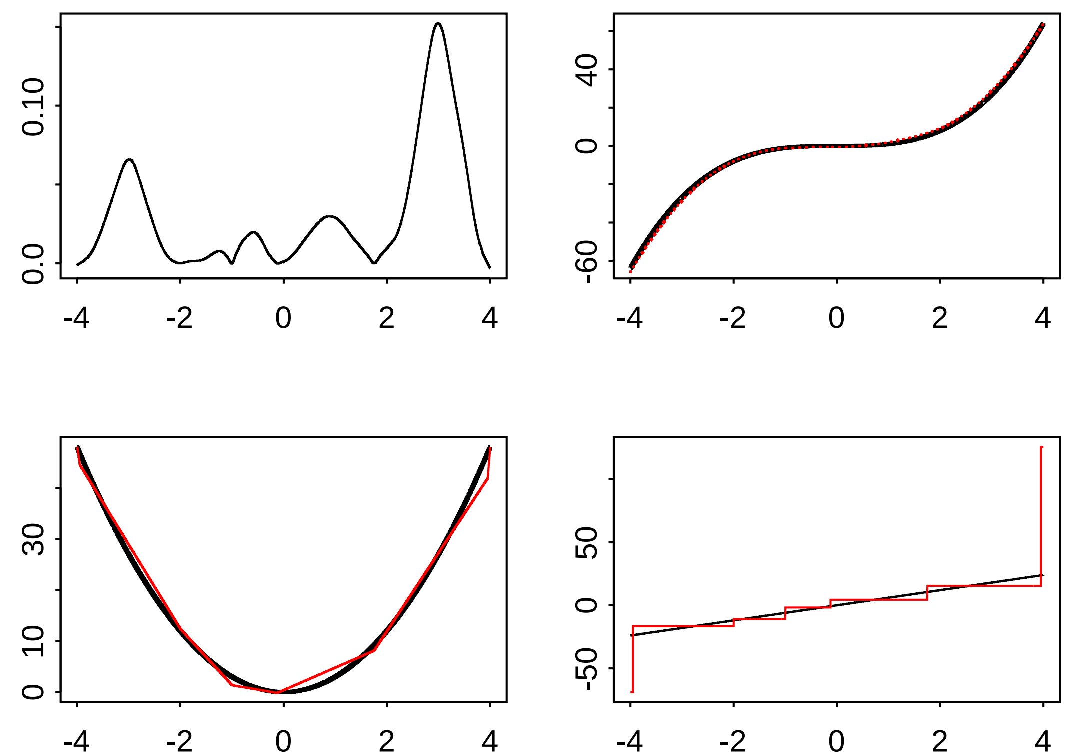

Figure 4.10: Plots of $-(H_{4,3} - Y_3)$, $g_{4,3} = H^{(3)}_{4,3}$ the LS solution (dashed red line) and t^3 (solid black line), $g'_{4,3} = H^{(4)}_{4,3}$ (solid red line) and $3t^2$ (solid black line), and $g''_{4,3} = H^{(5)}_{4,3}$ (solid red line) and $6t$ (solid black line).

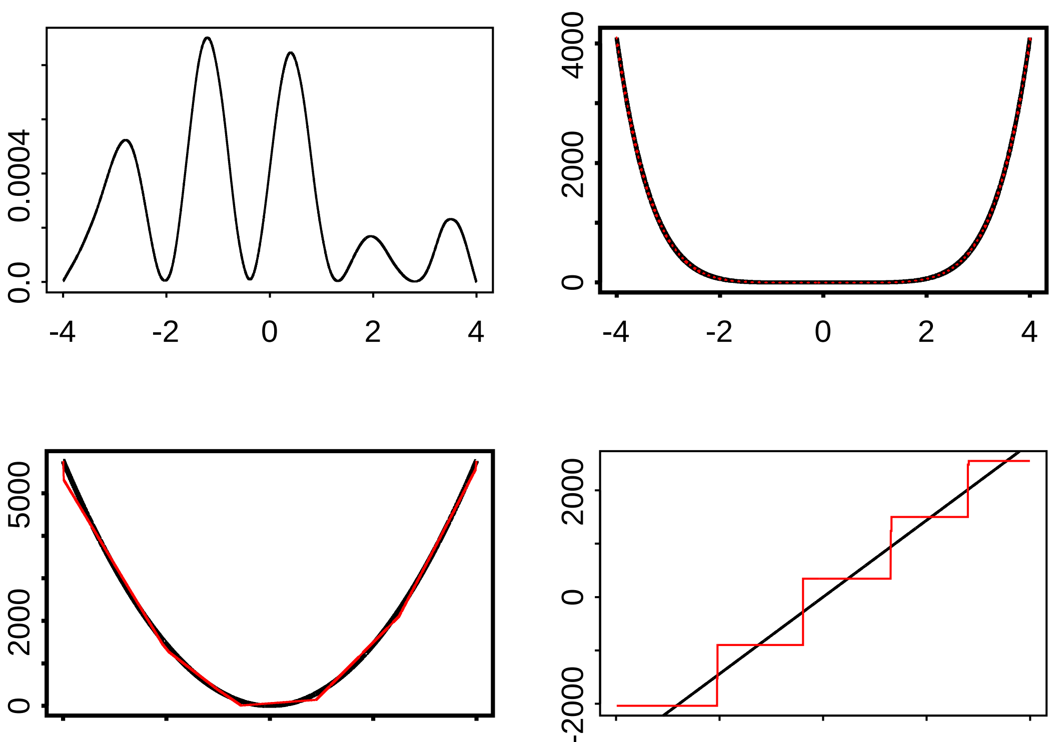

Figure 4.11: Plots of $(H_{4,6} - Y_6)$, $g_{4,6} = H_{4,6}^{(6)}$ the LS solution (dashed red line) and t^6 (solid black line), $g_{4,6}^{(4)} = H_{4,6}^{(10)}$ (solid red line) and $((6!)/2!)\ t^2$ (solid black line), and $g_{4,6}^{(5)} = H_{4,6}^{(11)}$ (solid red line) and $6!\ t$ (solid black line).

are the graphs of $g_{n,k} = H^{(k)}_{n,k}$ (in red) and $g_0(t) = t^k$ (in black). The difference between the graphs is not very visible but the motivation behind plotting the functions instead their difference was to show that the LS solution $g_{n,k}$ has the same "form" as the estimated function g_0. The lower right plots in Figure 4.10 and Figure 4.11 show the convex functions

Table 4.2: Table of set of touch points S between the processes $H_{n,k}$ and Y_k for $k = 3, n = 4, 6, 8$ and $k = 6, n = 4$, the value of the LS solution at the origin $g_{n,k}(0)$ and the corresponding number of iterations N_{it}.

$k, [-n, n]$	N_{it}	S	$g_{n,k}(0)$
$k = 3, [-4, 4]$	19	$\{-3.9501, -2.0004, -2.0000, -1.0000,$ $-0.1250, 1.7500, 3.9511\}$	-0.6016
$k = 3, [-6, 6]$	36	$\{-5.9501, -3.9238, -3.9213, -1.9995,$ -1.0000, -0.1250, 1.7500,4.0097, $4.0107, 4.0112\}$	-0.5990
$k = 3, [-8, 8]$	42	$\{-6.9985, -5.9995, -4.7495, -4.2500,$ -3.9892, -3.9873, -1.9995,-1.7500, -1.0000, -0.1250, 1.7500, 4.0356, $4.0390, 6.3291, 6.6250\}$	-0.6004
$k = 6, [-4, 4]$	37	$\{-3.9941, -2.0478, -2.0385, -0.3886,$ $1.3056, 1.3208, 2.7983, 2.8149,$ $2.8271\}$	-0.8203

$H^{(4)}_{4,3}$ and $H^{(10)}_{4,6}$ (in red) on $[-4, 4]$ for $k = 3$ and $k = 6$ respectively. These derivatives estimate the "true" convex functions $3t^2$ and $(6!/2!)t^2$ (in black) respectively. The jump processes $H^{(5)}_{4,3}$ and $H^{(11)}_{4,6}$ (in red) are shown in the lower left part. They both estimate a linear function and are monotone since the slopes of $H^{(5)}_{4,3}$ and $H^{(11)}_{4,6}$ are increasing by convexity.

The set of points of touch between $H_{n,k}$ and Y_k for $k = 3, n = 4, 6, 8$ and $k = 6, n = 4$ are provided in Table 4.2. For $k = 3$, we generated first the process Y_3 and its derivatives

Y_3' and Y_3'' on the interval $[-8, 8]$. Then, we obtained the envelopes $H_{8,3}, H_{6,3}$ and $H_{4,3}$ using the appropriate boundary conditions at the points $-8, 8$, $-6, 6$ and $-4, 4$ (see Section 2 of Chapter 3 for more details on the construction of the invelope H_k when k is odd). It is clear that the obtained points of touch are different and this fact was already noticed by Groeneboom, Jongbloed and Wellner (2001A) in the problem of estimating a convex function ($k = 2$). The authors also compared the value of the LS solution at the origin and found that it does not change very much as n increases. We notice the same fact for $k = 3$ (compare the values of $g_{n,3}(0)$ in Table 4.2). This stability is expected and follows from the fact that $\lim_{n\to\infty} g_{n,k}(0) = H_3^{(3)}(0)$.

4.4 Computing the MLE of a k-monotone density on $(0, \infty)$

Let $X_1, \cdots, X_n$ be n i.i.d random variables from a k-monotone density g_0 and $\mathbb{G}_n$ be their empirical distribution function. Consider the functional

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

where g belongs to $\mathcal{C}$, the class of integrable k-monotone functions on $(0, \infty)$. In Section 2 of Chapter 2, it was established that ϕ admits a minimizer $\hat{g}_n$ of the form

$$\hat{g}_n(t) = \hat{w}_1 \frac{k(\theta_1 - t)_+^{k-1}}{\theta_1^k} + \cdots + \hat{w}_m \frac{k(\theta_m - t)_+^{k-1}}{\theta_m^k}$$

where $m \leq n$ and $\hat{w}_1 + \cdots + \hat{w}_m = 1$, since this minimizer is nothing but

the Maximum Likelihood estimator ($\hat{g}_n$ maximizes $-\phi$). Note that in addition to the log-likelihood term, the functional ϕ is also composed of the "penalty" term $\int_0^\infty g(t)dt$. Without this term, the minimization problem will not be proper since for any nontrivial function $g \in \mathcal{C}$, we would have $\lim_{c\to\infty} \phi(cg) = -\lim_{c\to\infty} \log(c) = -\infty$. In the particular case of $k = 2$, Groeneboom, Jongbloed, and Wellner (2001b) proved that the MLE is unique. For $k > 2$, we were able to prove the MLE is unique when $k = 3$ (see Lemma 2.2.5 in Chapter 2) and we conjecture that this holds true for $k > 3$. Groeneboom, Jongbloed, and Wellner (2003) noticed that the *support reduction algorithm* is more efficient when it is based on a Newton-type procedure instead of applying it directly to the objective function ϕ. This entails an additional linearization step based on the well-known approximation

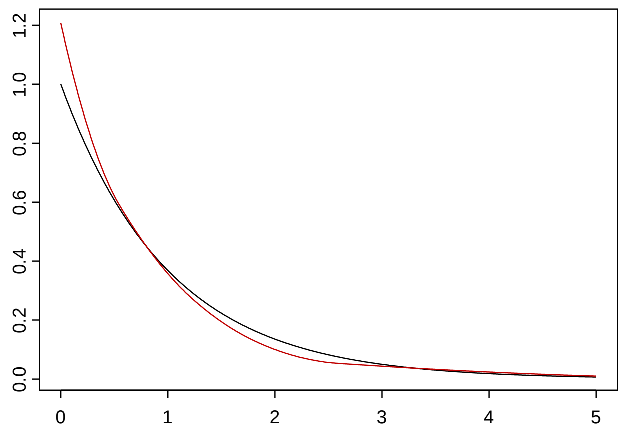

Figure 4.12: The exponential density (the true mixed density), in black and its Maximum Likelihood estimator based on $n = 100$ and $k = 3$, in red.

$$\log(1 + x) \simeq x - \frac{x^2}{2}$$

in the neighborhood of 0. Let $\bar{g}$ be the current iterate and $g \in \mathcal{C}$ such that

$$\frac{g - \bar{g}}{\bar{g}}$$

is very small. Then, we can write

$$\begin{aligned} \phi(g) &= \phi(\bar{g}) + \int_0^\infty -\frac{g(t) - \bar{g}(t)}{\bar{g}(t)} d\mathbb{G}_n(t) \\ &\quad + \int_0^\infty \frac{1}{2} \left(\frac{g(t) - \bar{g}(t)}{\bar{g}(t)} \right)^2 d\mathbb{G}_n(t) + \int_0^\infty (g(t) - \bar{g}(t)) dt. \end{aligned}$$

If we delete the terms that do not depend on f, we can define the following local objective function (see Groeneboom, Jongbloed, and Wellner (2003))

$$\phi_q(g) = -2 \int_0^\infty \frac{g(t)}{\bar{g}(t)} d\mathbb{G}_n(t) + \int_0^\infty \frac{1}{2} \left(\frac{g(t)}{\bar{g}(t)} \right)^2 d\mathbb{G}_n(t) + \int_0^\infty g(t) dt.$$

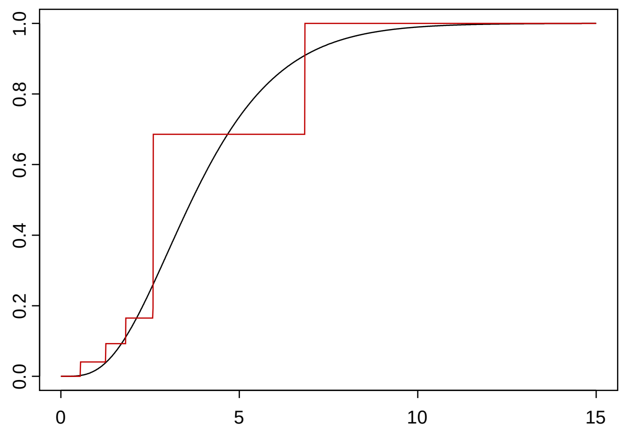

Figure 4.13: The cumulative distribution function of a $Gamma(4, 1)$ (the true mixing distribution), in black and the its Maximum Likelihood estimator based on $n = 100$ and $k = 3$ (in red).

Let $\epsilon > 0$ and $f_\theta(t) = k(t-\theta)_+^{k-1}/\theta^k, \theta > 0$. We have

$$\begin{aligned}\phi_q(g + \epsilon f_\theta) &= \phi_q(g) + \epsilon \left(\int_0^\infty -2 \frac{f_\theta(t)}{\bar{g}(t)} d\mathbb{G}_n(t) + \int_0^\infty \frac{g(t) f_\theta(t)}{(\bar{g}(t))^2} d\mathbb{G}_n(t) + \int_0^\infty f_\theta(t) dt \right) \\ &\quad + \frac{\epsilon^2}{2} \int_0^\infty \left(\frac{f_\theta(t)}{\bar{g}(t)} \right)^2 dt \\ &= \phi_q(g) + \epsilon c_1(\theta, g) + \frac{\epsilon^2}{2} c_2(\theta, g).\end{aligned}$$

The "alternative"directional derivative of ϕ_q at the point g in the direction of f_θ is given by

$$\tilde{D}_{\phi_q}(f_\theta, g) = \frac{c_1(\theta, g)}{\sqrt{c_2(\theta, g)}}.$$

The algorithm consists of an *outer* and *inner* loops. Given a fixed finite grid Θ_f (note that the subsript f is for "finite" and that Θ_f corresponds to Θ_δ used in Groeneboom, Jongbloed, and Wellner (2003)) and the current iterate $\bar{g}$, the *inner* loop is set up to find $\bar{g}_q = argmin\{\phi_q(g) : g \in \text{cone}(f_\theta, \theta \in \Theta_f)\}$. The next iterate is taken to be $(1-\lambda)\bar{g} + \lambda \bar{g}_q$,

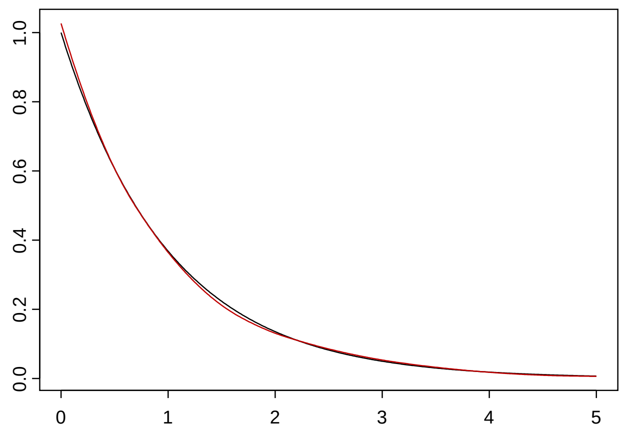

Figure 4.14: The exponential density (the true mixed density), in black and its Maximum Likelihood estimator based on $n = 1000$ and $k = 3$, in red.

where $\lambda \in (0, 1]$ is appropriately chosen to ensure monotonicity of the algorithm. A *reduction* step is needed to construct a starting value $g^{(0)}$ which will depend of course on the current iterate $\bar{g}$. To enter the *outer* loup, the minimal value $\min_{\theta \in \Theta_f} \tilde{D}_{\phi_q}(f_\theta, \bar{g})$ needs to be bigger than some fixed tolerance $-\eta$, otherwise we stop. Let $\bar{S} = \{\bar{\theta}_1, \cdots, \bar{\theta}_p\}$ denote the set of support points of the current iterate $\bar{g}$. We proceed as follows:

1. We calculate $\min_{\theta \in \Theta_f} \tilde{D}_{\phi_q}(f_\theta, \bar{g})$. If it is smaller than $-\eta$, we stop. Otherwise, we move to the second step.

2. We minimize the local objective function ϕ_q (which depends on $\bar{g}$) over the cone

$$\mathcal{C}(\Theta_f) = \left\{ g : g(t) = \int_{\theta \in \Theta_f} f_\theta(t) d\mu(\theta), \quad \text{where } \mu \text{ is a positive measure on } \Theta_f \right\}.$$

   For that, we need to find a starting function $g^{(0)}$. The current iterate $\bar{g}$ is not necessarily a good choice and therefore we need to construct one. This can be done as

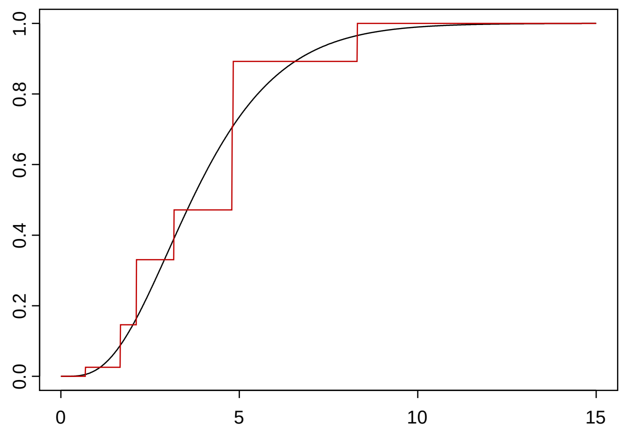

Figure 4.15: The cumulative distribution function of a $Gamma(4,1)$ (the true mixing distribution), in black and the its Maximum Likelihood estimator based on $n = 1000$ and $k = 3$ (in red).

follows: We first minimize the quadratic function

$$\psi(\alpha_1, \cdots, \alpha_p) = \phi_q(\sum_{j=1}^{p} \alpha_j f_{\bar{\theta}_j})$$

where $\alpha_1, \cdots, \alpha_p \in \mathbb{R}$. Finding this minimum is achieved by finding the solution of the linear system

$$(DY)^t DY \underline{\alpha} = 2Y^t \underline{d} - \underline{n_p} \tag{4.1}$$

where $Y = (f_{\bar{\theta}_j}(X_i))_{i,j}$ is a $n \times p$-matrix, D is the $n \times n$ diagonal matrix given by $D_{ii} = 1/\bar{g}(X_i)$, $\underline{d}^t = (1/\bar{g}(X_1), \cdots, 1/\bar{g}(X_n))$, $\underline{n_p}$ and $\underline{\alpha}$ are the $p \times 1$ vectors given by $\underline{n_p}^t = (n, \cdots, n)$ and $\underline{\alpha}^t = (\alpha_1, \cdots, \alpha_p)$ respectively.

Let $g_{min} = \sum_{j=1}^{p} \alpha_{j,min} f_{\bar{\theta}_j}$ be this minimum. Next, if g_{min} is k-monotone; i.e., $\alpha_{j,min} > 0$ for all $j = 1, \cdots, p$, then we take $g^{(0)} = g_{min}$. Otherwise, we find $\lambda \in (0,1)$ such that $(1-\lambda)\bar{g} + \lambda g_{min}$ is k-monotone. Such a $\lambda \in (0,1)$ will always exist and this

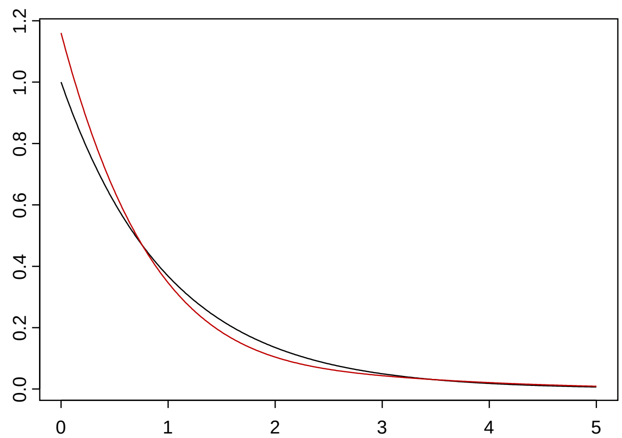

Figure 4.16: The exponential density (the true mixed density), in black and its Maximum Likelihood estimator based on $n = 100$ and $k = 6$, in red.

follows from the same arguments of Lemma 4.2.2. We repeat the reduction and minimization steps till we find a minimizer that is k-monotone. We take this minimizer to be the starting function $g^{(0)}$. The support of $g^{(0)}$ is in general smaller than $\bar{S}$ as a consequence of successive deletions of support points in the reduction steps.

In the *inner* loop, we proceed as we did for computing the LSE and the process $H_{n,k}$ (see the Section 1 and Section 2). Let m be an integer strictly smaller than p and let us denote the current iterate and its support by $\bar{g}_{inner}$ and $\bar{S}_{inner}$. We assume without loss of generality that $\bar{S} = \{\bar{\theta}_1, \cdots, \bar{\theta}_m\}$. Let $\bar{\theta}_{m+1} = argmin_{\theta \in \Theta_f} D_{\phi_q}(f_\theta, \bar{g}_{inner})$. If $D_{\phi_q}(f_{\bar{\theta}_{m+1}}, \bar{g}_{inner}) \leq -\eta$, we stop. Otherwise, we assume without loss of generality that $\bar{\theta}_{m+1} > \bar{\theta}_m$ and find the minimizer of ϕ_q over the class

$$\mathcal{C}'(\bar{\theta}_1, \cdots, \bar{\theta}_{m+1}) = \left\{ g : g = \sum_{j=1}^{m+1} \alpha_j f_{\bar{\theta}_j}, \ \alpha_j \in \mathbb{R}, \ j = 1, \cdots, m+1 \right\}$$

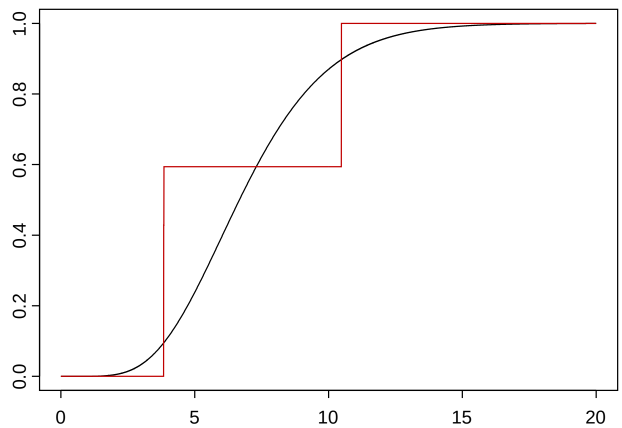

Figure 4.17: The cumulative distribution function of a $Gamma(7, 1)$ (the true mixing distribution), in black and the its Maximum Likelihood estimator based on $n = 100$ and $k = 6$ (in red).

by solving the linear system given in (4.1). If the minimizer, g_{min}, is k-monotone, then we take it as the next iterate. Otherwise, we find $\lambda \in (0, 1)$ such that $(1 - \lambda)\bar{g}_{inner} + \lambda g_{min}$ is k-monotone and take the first minimizer that is k-monotone as the next iterate.

3. Let $g_{min} = argmin\{\phi_q(g) : g \in \mathcal{C}(\Theta_f)\}$ obtained in the previous step. Since there is no guarantee that $\phi(g_{min}) \leq \phi(\bar{g})$, we apply the *Armijo* rule; that is, we find the smallest $\lambda \in (0, 1]$ such that

$$\phi((1 - \lambda)\bar{g} + \lambda g_{min}) \leq \phi(\bar{g}).$$

We take $(1 - \lambda)\bar{g} + \lambda g_{min}$ to be the new iterate for the *outer* loop.

For $k = 3$ and $k = 6$, we calculated the MLE of a standard Exponential based on the same samples of size $n = 100$ and $n = 1000$ used in the Least Squares estimation (see Section 2).

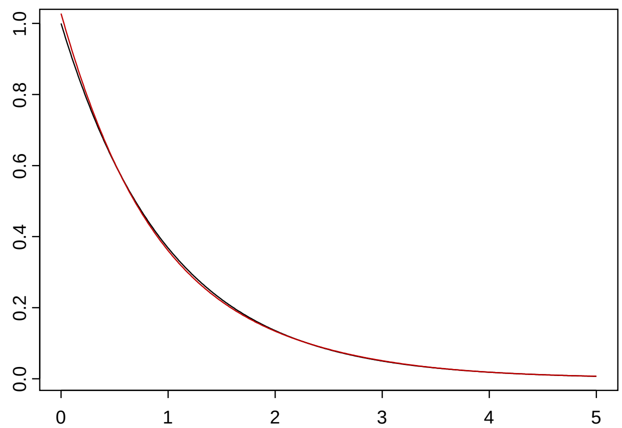

Figure 4.18: The exponential density (the true mixed density), in black and its Maximum Likelihood estimator based on $n = 1000$ and $k = 6$, in red.

The algorithm was coded in S and can be found in Appendix C. To start the algorithm, we calculate $\theta^{(0)}$ the minimizer of the nonlinear function

$$\theta \mapsto -\frac{1}{n}\sum_{j=1}^{n} \log\left(\frac{k(\theta - X_j)^{k-1}}{\theta^k}\right)$$

for $\theta \geq X_{(n)} + a$, where a is some fixed positive number. This minimization can be performed using the S function *nlminb*. Different values of a yield different starting values but the numerical results remained unchanged for many different values which supports our conjecture about uniqueness of the MLE in the general case $k > 3$. As for we did for the LSE, we took a finite grid $\subseteq [X_{(1)}, 2kX_{(n)}]$ with a maximal mesh equal to 0.01. The ML estimation in the direct is illustrated by the plots in Figure 4.12 and Figure 4.14 for $k = 3$, and in Figure 4.16 and Figure 4.18 for $k = 6$. The "alternative" directional derivative $\tilde{D}_\phi(f_\theta, \hat{g}_n)$, for $n = 1000$ and $k = 6$, is plotted in Figure 4.20. For the inverse problem, see Figure 4.13 and Figure 4.15 for $k = 3$, and Figure 4.17 and Figure 4.19 for $k = 6$. Consistency of the MLE is proved in Chapter 2 and it can be clearly seen in these figures. As for the LSE,

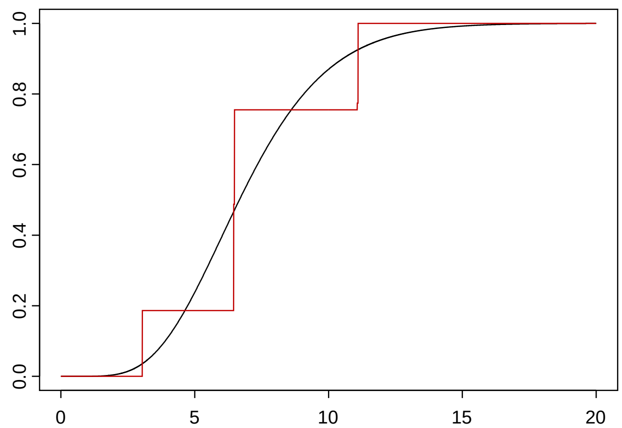

Figure 4.19: The cumulative distribution function of a $Gamma(7, 1)$ (the true mixing distribution), in black and the its Maximum Likelihood estimator based on $n = 1000$ and $k = 6$ (in red).

convergence in the inverse problem is much slower than in the direct one and the difference becomes more pronounced when k is large. Finally, it should be mentioned here that even if the MLE and LSE of the Exponential density show very small visible differences in the direct problem, it can be easily checked by comparing the locations of jump points or the heights of the jumps that these estimators are different (compare Table 4.1 and Table 4.3).

4.5 Future work and open questions

4.5.1 The MLE of a mixture of Exponentials

As it was already mentioned in the introduction, this work was motivated in part by going beyond consistency of the nonparametric Maximum Likelihood estimator of a scale mixture of Exponentials (see Jewell (1982)). As the class of scale mixtures of Exponentials is the intersection of the classes of k-montone densities for $k \geq 1$, a scale mixture of Exponentials

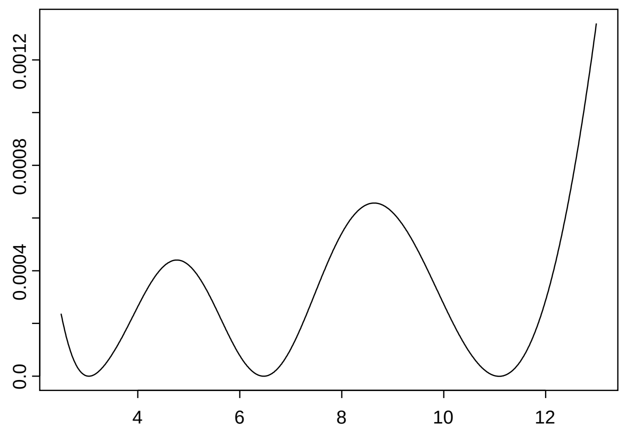

Figure 4.20: The directional derivative for the Maximum Likelihood estimator of the Exponential density based on $n = 1000$ and $k = 6$.

can be viewed as a limit of a sequence of k-monotone densities when $k \to \infty$. More formally, let g be a mixture of Exponentials. There exists a distribution function F such that

$$g(x) = \int_0^\infty t \exp(-xt) dF(t), \quad \text{for all } x > 0.$$

Let g_k be the k-monotone density given by

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

where F_k is a distribution function to be defined. The density g_k can be rewritten

$$\begin{aligned} g_k(x) &= \int_0^\infty \frac{k}{y} \left(1 - \frac{x}{y}\right)_+^{k-1} dF_k(y) \\ &= \int_0^\infty \frac{1}{z} \left(1 - \frac{x}{kz}\right)_+^{k-1} dF_k(kz) \quad \text{by the change of variable } y = kz \\ &\to \int_0^\infty \frac{1}{z} \exp(-x/z) dF^*(z) \end{aligned}$$

if $F_k(k\cdot) \to_d F^*$. By the change of variable $t = 1/z$, we have for all $x > 0$

$$g_k(x) \to -\int_0^\infty t \exp(-xt) dF^*(1/t) = \int_0^\infty t \exp(-xt) d(1 - F^*(1/t)).$$

Table 4.3: Table of the obtained ML estimates for $k = 3, 6$ and $n = 100, 1000$. A support point is denoted by $\hat{a}$ and its mass by $\hat{w}$.

k, n	$(\hat{a}, \hat{w})$
$k = 3, n = 100$	$(0.549, 0.040), (1.259, 0.051), (1.819, 0.072),$
	$(2.579, 0.027), (2.589, 0.492), (6.839, 0.314)$
$k = 3, n = 1000$	$(0.684, 0.025), (1.664, 0.120), (2.114, 0.184),$
	$(3.164, 0.141)$
	$(4.794, 0.236), (4.824, 0.184), (8.304, 0.107)$
$k = 6, n = 100$	$(3.839, 0.428), (3.849, 0.165), (10.479, 0.405)$
$k = 6, n = 1000$	$(3.042, 0.186), (6.452, 0.300), (6.482, 0.267),$
	$(11.072, 0.018), (11.102, 0.226)$

If the distribution functions $F_k, k \in \mathbb{N}$, are chosen such that, for all continuity points $t > 0$ of F, $F_k(kt) \to 1 - F(1/t)$ as $k \to \infty$, then g is the pointwise limit of the sequence $(g_k)_k$.

Based on n i.i.d. random variables from the density g, let the completely monotone density $\hat{g}_n$ be the MLE of g. Recall that the MLE of the mixing distribution $\hat{F}_n$ is discrete with at most n jump points and hence the density $\hat{g}_n$ is a finite mixture of Exponentials with at most n components (see Jewell (1982), Lindsay (1983a), Lindsay (1983b), Lindsay (1995)). Now, for a fixed integer $k \geq 1$, we can also consider $\hat{g}_{n,k}$ to be the MLE of g in the class of k-monotone densities. At any fixed point $x_0 > 0$, the mixed density g satisfies the working assumptions of the asymptotic distribution theory developed in this thesis. Thus, as $n \to \infty$, we have

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

and

$$n^{\frac{1}{2k+1}}(\hat{F}_{n,k}(x_0) - F(x_0)) \rightarrow_d \frac{(-1)^k x_0^k}{k!} c_{k-1}(g) H_k^{(2k-1)}(0)$$

where $\hat{F}_{n,k}$ is the MLE of the mixing distribution corresponding to g viewed as a k-monotone density, H_k is the envelope ("invelope") of the $(k-1)$-fold integral of two-sided Brownian motion $+ ((k!)/(2k)!)\, t^{2k}$ when k is odd (even) and the constants $c_j(g), j = 0, \cdots, k-1$ are given in Theorem 2.7.2.

Under this perspective, the problem of deriving an asymptotic distribution theory for the MLE $\hat{g}_n$ depends not only on the sample size n in the limit, but also on the smoothness parameter k. Here, we list some of the natural questions that we would like to answer in the future:

- For fixed i.i.d. random variables $X_1, \cdots, X_n$ from g, what is the limit of $\hat{g}_{n,k}$ when $k \to \infty$? Do we have

  $$\lim_{k \to \infty} \hat{g}_{n,k}(x) = \hat{g}_n(x), \qquad \text{for } x > 0$$

  for n maybe sufficiently large ?

- If the above does not necessarily hold, but g is completely monotone, can we change the order of the limits on n and k? That is, do we have

  $$g(x) = \lim_{k \to \infty} \lim_{n \to \infty} \hat{g}_{n,k}(x) = \lim_{n \to \infty} \lim_{k \to \infty} \hat{g}_{n,k}(x),$$

  for almost surely all $x > 0$? The first limit follows from the strong consistency of $\hat{g}_{n,k}$ for any fixed $k \geq 1$. Indeed, for $k \geq 1$, the density g is k-monotone and hence by Theorem 2.3.1

  $$\lim_{n \to \infty} \hat{g}_{n,k} = g, \quad \text{uniformly on } [c, \infty),$$

  for $c > 0$. Therefore,

  $$\lim_{k \to \infty} \lim_{n \to \infty} \hat{g}_{n,k} = g, \quad \text{uniformly on } [c, \infty).$$

- What is the rate of convergence of $\hat{g}_n^{(j)}(x_0)$ for a fixed integer $j \geq 0$ and that of $\hat{F}_n(x_0)$? Can these rates be obtained from the rates $n^{-(k-j)/(2k+1)}$ proved for $\hat{g}_{n,k}^{(j)}(x_0), j = 0, \cdots, k-1$ in the direct problems and $n^{-1/(2k+1)}$ for $\hat{F}_{n,k}(x_0)$ in the inverse problem with k fixed?

- Suppose that the limiting distributions of $\hat{g}_n^{(j)}(x_0), j \geq 0$, and $\hat{F}_n$ depend on a process H_∞. How is this process defined? Can it be obtained as the limit (in an appropriate sense) of some scaled version of the sequence $(H_k)_k$? Is it related, as in the k-monotone case, to some Gaussian problem?

4.5.2 Further related problems

But independently of the completely monotone problem, there are still many other problems left in connection with k-monotone densities, for a fixed k. We present in the following some of them that can be investigated in the future:

1. Another mixture form. The integral representation of k-monotone densities, that has been used here, is only one of two possible mixture forms: We can also write a k-monotone density g as

$$g(x) = \frac{1}{\mu_k} \int_0^\infty (t-x)_+^{k-1} dF(t), \quad x > 0 \tag{4.2}$$

where we assume that F is a distribution function with

$$\mu_k = \int_0^\infty t^k dF(t) < \infty.$$

Then F can be given by the following inversion formula:

$$F(x) = 1 - \frac{g^{(k-1)}(x)}{g^{(k-1)}(0)}. \tag{4.3}$$

The integral representation in (4.2) and the inversion formula in (4.3) can be established using similar arguments as in the proof of Theorems 1 and 3 in Williamson (1956). To estimate F of a fixed point x_0, we need to estimate $g^{(k-1)}$ at both the points 0 and x_0. For the special case of monotone densities ($k = 1$), Woodroofe and Sun (1993) showed that the MLE $\hat{g}_n$ is not a consistent estimator at the point 0 and constructed a penalized MLE to obtain consistency. Kulikov (2002) proposed another approach based on $\hat{g}(\alpha_n, 0) = \hat{g}_n(n^{-\alpha})$ as an estimator of $g(0)$, and proved that $\hat{g}(n^{-1/3}, 0)$ has a smaller mean squared error than that of the estimator proposed by Woodroofe and Sun (1993).

We conjecture that the inconsistency problem becomes even more severe for $k \geq 2$. We would like to investigate this fact in the future and generalize the method developed by Woodroofe and Sun (1993) or Kulikov (2002).

2. Estimating a smooth functional. In this thesis, we focused only on estimating a k-monotone density g_0 and its derivatives at a fixed point $x_0 > 0$. If ν_j is the functional defined on $\mathcal{D}_k$ by $\nu_j(g) = g^{(j)}(x_0), g \in \mathcal{D}_k$, then under our working assumptions, the nonparametric MLE of ν_j, $\hat{\nu}_{j,n}$, converges at the rate $n^{-(k-j)/(2k+1)}$, $j = 0, \cdots, k-1$ (see Theorem 2.7.2).

Can we obtain the rate $n^{-1/2}$ for some other functionals? If yes, can we find a simple characterization for these functionals? If we consider only the k-monotone densities with finite second moment, then the answer for the first question is yes. Indeed, take for example $\nu \equiv \mu$ to be the mean of the mixing distribution F. If $X \sim g_0 \in \mathcal{D}_k$, then there exist two independent random variables Y and Z such that $X = YZ$, $Y \sim Beta(1, k)$ and $Z \sim F$. Therefore, $E(Y)\ (k+1)^{-1} = E(X)$; i.e., $\mu = (k+1)\ E(X)$. Since g_0 was assumed to have a finite second moment, the estimator $(k+1)\overline{X}$ converges at the rate $n^{-1/2}$ by the central limit theorem.

3. Testing problems. Consider the testing problem:

$$\mathcal{H}_0 : g_0(x_0) = \theta_0 \text{ versus } \mathcal{H}_1 : g_0(x_0) \neq \theta_0, \tag{4.4}$$

where g_0 is a monotone density. Banerjee and Wellner (2001a) considered the asymptotic distribution of the log-likelihood ratio statistics in a related monotone function problem under the null hypothesis and also under a fixed alternative. Banerjee and Wellner (2001a) found that, under the null this asymptotic distribution is universal and can be characterized as a functional of standard two-sided Brownian motion with parabolic drift. They conjecture that this similar asymptotic behavior carries over to the testing problem in (4.4).

If g_0 is a k-monotone density, we can consider the more general testing problems

$$\mathcal{H}_{0,j} : g_0^{(j)}(x_0) = \theta_{0,j} \text{ versus } \mathcal{H}_{1,j} : g_0^{(j)}(x_0) \neq \theta_{0,j}, \quad j = 0, \cdots, k-1.$$

If we still consider the log-likelihood ratio as the test statistic, then what is its asymptotic distribution under the null? Under a fixed alternative? Under local aternatives?

4.5.3 Some starting points for the transition to completely monotone

In the previous section, it was stated that if $F_k, k \geq 0$, and F are distribution functions on $(0, \infty)$ such that $\lim_{k\to\infty} F_k(kt) = 1 - F(1/t)$ for any continuity point $t > 0$ of F, then

$$\int_0^\infty \frac{k(t-x)_+^{k-1}}{t^k} dF_k(t) \to \int_0^\infty t \exp(-tx) dF(t), \quad \text{as } k \to \infty$$

for all $x > 0$. But in Section 2, we established that the exponential density is the $Gamma(k+1, 1)$ scale mixture of $Beta(1, k)$'s and hence we can write

$$\begin{aligned} \exp(-x) &= \int_0^\infty \frac{k}{t} \left(1 - \frac{x}{t}\right)_+^{k-1} dF_k(t) dt \\ &= \int_0^\infty \frac{1}{t} \left(1 - \frac{x}{kt}\right)_+^{k-1} dF_k(kt) dt \end{aligned} \tag{4.5}$$

with F_k is $Gamma(k+1, 1)$ distribution function. But note that $F_k(kt) \to 1_{[1,\infty)}(t), t \neq 1$. Indeed, it is known that if $Y_1, \cdots, Y_{k+1}$ are i.i.d. random variables from a standard Exponential, then $S_{k+1} = Y_1 + \cdots + Y_{k+1} \sim Gamma(k+1, 1)$. On the other hand, $S_{k+1}/k \to_p 1$ by the weak law of large numbers. As $F_k(kt)$ is the cumulative distribution of S_{k+1}/k, it follows that $F_k(kt) \to 1_{[1,\infty)}(t)$ for all $t \neq 1$. This fact is not surprising as

$$\lim_{k\to\infty} \frac{1}{t} \left(1 - \frac{x}{kt}\right)_+^{k-1} \to \frac{1}{t} \exp(-x/t)$$

for all $t > 0$ and hence the limit of the sequence $(F_k)_k$ is expected to degenerate at 1 in view of (4.5).

Thus it would be interesting to have a family of distributions to study in which the mixing distribution is nontrivial and has a positive density. For example what happens if we take

$$g(x) = \alpha x^{\alpha-1} \exp(-x^\alpha),$$

the Weibull density with shape parameter $\alpha < 1$; or

$$g(x) = \frac{1}{(1+x)^2}?$$

Example 4.5.1 *It is known that the* $Weibull(1/2, 1)$ *distribution function* G *can be written as*

$$1 - G(x) = \exp(-x^{1/2}) = \int_0^\infty \exp(-yx) f(y) dy$$

where

$$f(y) = \frac{1}{2\sqrt{\pi y^3}} \exp\left(-\frac{1}{4y}\right),$$

and hence the corresponding density can be written as

$$g(x) = \frac{1}{2} x^{-1/2} \exp(-x^{1/2}) = \int_0^\infty y \exp(-yx) f(y) dy.$$

This example is interesting because $\int_0^\infty g^2(x)dx = \infty$, and we might expect the Least Squares estimator to break down or perform badly. (The Weibull densities with $\alpha < 1/2$ should be even worse!) Now by the change of variable $t = 1/y$, $1 - G$ can be rewritten as

$$1 - G(x) = \exp(-x^{1/2}) = \int_0^\infty \exp(-x/t) m(t) dt$$

where

$$m(t) = \frac{1}{2\sqrt{\pi}} t^{-1/2} \exp(-t/4).$$

What is the corresponding sequence $(f_k)_k$ that goes with the kernel $(1 - x/t)_+^k$? That is, f_k would solve

$$\exp(-x^{1/2}) = \int_0^\infty \left(1 - \frac{x}{t}\right)_+^k f_k(t) dt$$

and we should have

$$f_k(x) = \frac{(-1)^k}{k!} x^k G^{(k+1)}(x) = \frac{(-1)^k}{k!} x^k g^{(k)}(x).$$

We can calculate

$$\begin{aligned} f_1(x) &= -xg^{(1)}(x) = x\left(\frac{1}{4x^{3/2}} + \frac{1}{4x}\right)\exp(-x^{1/2}) = \frac{1}{4}(1+x^{-1/2})\exp(-x^{1/2}), \\ f_2(x) &= \frac{x^2}{2}g^{(2)}(x) = \frac{x^2}{2}\left(\frac{3}{8x^{5/2}} + \frac{3}{8x^2} + \frac{1}{8x^{3/2}}\right)\exp(-x^{1/2}), \end{aligned}$$

and so forth. Furthermore, it is the case that

$$kf_k(kx) \to \frac{1}{2\sqrt{\pi}}x^{-1/2}\exp(-x/4) \equiv f_\infty(x) \quad \text{as } k \to \infty.$$

Example 4.5.2 *When*

$$g(x) = \frac{1}{(1+x)^2}$$

we have for all $x \geq 0$

$$1 - G(x) = \int_x^\infty \frac{1}{(1+t)^2}dt = \frac{1}{1+x},$$

and hence

$$\begin{aligned} 1 - G(x) &= \frac{1}{1+x} = \int_0^\infty \exp(-yx)\exp(-y)dy \\ &= \int_0^\infty \exp(-x/t)t^{-2}\exp(-1/t)dt. \end{aligned}$$

Thus $f_\infty(x) = x^{-2}\exp(-1/x), x \geq 0$. Correspondingly for finite k, we have

$$f_k(x) = \frac{(-1)^k}{k!}x^k g^{(k)}(x) = \frac{(-1)^k}{k!}x^k\frac{(k+1)!(-1)^k}{(1+x)^{k+2}} = (k+1)\frac{x^k}{(1+x)^{k+2}},$$

and hence

$$\begin{aligned} kf_k(kx) &= k(k+1)\frac{(kx)^k}{(1+kx)^{k+2}} \\ &= \frac{k(k+1)}{(kx)^2}\frac{(kx)^{k+2}}{(1+kx)^{k+2}} \\ &= \frac{k+1}{k}x^{-2}\frac{1}{(1+1/(kx))^{k+2}} \\ &\to x^{-2}\exp(-1/x) = f_\infty(x) \quad \text{as } k \to \infty. \end{aligned}$$

This example is interesting because g is bounded but heavy-tailed. The f_k's converge to 0 at the origin, but are also heavy-tailed.

BIBLIOGRAPHY

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

Banerjee, M. and Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699 - 1731.

Balabdaoui, F. (2004). A curious fact about k-monotone functions. Technical Report 426, Department of Statistics, University of Washington. Available at: *http://www.stat.washington.edu/www/research/reports/2004/*.

Böhning, D. (1982). Convergence of Simar's algorithm for finding the maximum likelihood estimate of a compound Poisson process. *Ann. Statist.* **10**, 1006 - 1008.

Böhning, D. (1986). A vertex exchange method in D-optimal design theory. *Metrika* **33**, 337 - 347.

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

Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184 - 1186.

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

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

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

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

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

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

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

Durrett, R. (1984). *Brownian Motion and Martingales in Analysis.* Wadsworth and Software.

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

Fedorov, V. V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

Feller, W. (1939). Completely monotone functions and sequences. *Duke Math. J.* **5**, 662 - 674.

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

Gelbaum, B. R. and Olmsted, J. M. (1964). *Counterexamples in Analysis.* Holden-Day, San Francisco.

Ghosal, S. and Van der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29**, 1233 - 1263.

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

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

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

Groeneboom, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* **11**, 1016 - 1027.

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

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

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

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

Groeneboom, P. 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.

Groeneboom, P. and Wellner J.A. (2001). Computing Chernoff's distribution. *Journal of Computational and Graphical Statistics.* **10**, 388-400.

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

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

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

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

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

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

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

Kopotun, K. and Shadrin, A. (2003). On $k-$monotone approximation by free knot splines. *SIAM J. Math. Anal.* **34**, 901 - 924.

Kulikov, V. N. (2002). *Direct and Indirect Use of Maximum Likelihood.* Ph.D. dissertation, Delft University of Technology.

Lachal, A. (1997). Local asymptotic classes for the successive primitives of Brownian motion. *Ann. Prob.* **25**, 1712 - 1734.

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

Lesperance, M. L. and Kalbfleisch, J. D. (1992). An algorithm for computing the nonparametric MLE of a mixing distribution. *Journal of the American Statistical Association* **87**, 120 - 126.

Leurgans, S. (1982). Asymptotic distributions of slope-of-greatest-convex minorant estimators. *Ann. Statist.* **10**, 287 - 296.

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

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

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

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

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

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

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **19**, 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 distributions.* Ph.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. In *Advances in Order Restricted Inference. Lecture Notes in Statist.* **37**, 91 - 124. Springer, New York.

Nolan, D. and Pollard D. (1987). U-Processes: Rates of convergence. *Ann. Statist.* **15**, 780 - 799.

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

Polonik, W. (1995). Density estimation under qualitative assumptions in higher dimensions. *J. Multivariate Anal.* **55**, 61 - 81.

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

Roberts, A. W. and Varberg D. E. (1973). *Convex Function.* Academic Press, New York.

Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov Processes and Martingales.* Wiley, New York.

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

Schumaker, L. L. (1981). *Spline Functions: Basic Theory.* John Wiley and Sons, New York.

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

Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4**, 1200 - 1209.

Stefanski, L.A., and Carroll, R.J. (1990). Deconvoluting Kernel Density Estimators. *Statistics* **21**, 169 - 184.

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

Sun, J. and Woodroofe, M. (1996). Adaptive smoothing for a penalized NPMLE of a non-increasing density. *J. Statist. Plan. Infer.* **52**, 153 - 159.

Ubhaya, V. A. (1989). L_p approximation from nonconvex subsets of special classes of functions. *J. Approx. Theory* **57**, 223 - 238.

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

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

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

Wellner, J. A. (2003). Gaussian white noise models: some results for monotone functions. In *Crossing Boundaries: Statistical Essays in Honor of Jack Hall.* IMS Lecture Notes-Monograph Series **43**, 87 - 104.

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.

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

Wynn, H. P. (1970). The sequential generation of D-optimum experimental designs. *Ann. Math. Statist.* **6**, 1286 - 1301.

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

Appendix A

GAUSSIAN SCALING RELATIONS

Let W be a two-sided Brownian motion process starting from 0, and define the family of processes $\{Y_{k,a,\sigma} : \ a > 0, \ \sigma > 0\}$ for k a nonnegative integer

$$Y_{k,a,\sigma}(t) = \sigma \int_0^t \cdots \int_0^{s_2} W(s_1) ds_1 \cdots ds_{k-1} + at^{2k}$$

when $t \geq 0$ and analogously when $t < 0$. Let $H_{k,a,\sigma}$ be the envelope/invelope process corresponding to $Y_{k,a,\sigma}$. In this paper we have taken $Y_{k,k!/(2k)!,1} \equiv Y_k$ to be the standard or "canonical" version of the family of processes $\{Y_{k,a,\sigma} : \ a > 0, \sigma > 0\}$, and we have defined the envelope or invelope processes H_k in terms of this choice of Y_k. Since the usual choice in the previous literature has been to take $Y_{k,1,1}$ as the canonical process (see e.g. Groeneboom, Jongbloed, and Wellner (2001a) for the case $k = 2$ and Groeneboom (1989) for the case $k = 1$), it is useful to relate the distributions of these different choices of "canonical" via Brownian scaling arguments.

Proposition A.1 (Scaling of the processes $Y_{k,a,\sigma}$ and the invelope or envelope processes $H_{k,a,\sigma}$).

$$Y_{k,a,\sigma}(t) \stackrel{d}{=} \sigma \left(\frac{\sigma}{a}\right)^{\frac{2k-1}{2k+1}} Y_{k,1,1}\left(\left(\frac{a}{\sigma}\right)^{\frac{2}{2k+1}} t\right)$$

as processes for $t \in \mathbb{R}$, and hence also

$$H_{k,a,\sigma}(t) \stackrel{d}{=} \sigma \left(\frac{\sigma}{a}\right)^{\frac{2k-1}{2k+1}} H_{k,1,1}\left(\left(\frac{a}{\sigma}\right)^{\frac{2}{2k+1}} t\right)$$

as processes for $t \in \mathbb{R}$.

Corollary A.1 For the derivatives of the invelope/envelope processes $H_{k,a,\sigma}$ it follows that

$$\begin{aligned} &\left(H_{k,a,\sigma}^{(j)}(t), \ j = 0, \ldots, 2k-1\right) \\ &\stackrel{d}{=} \ \left(\sigma \left(\frac{\sigma}{a}\right)^{\frac{2k-1-2j}{2k+1}} H_{k,1,1}^{(j)}\left(\left(\frac{a}{\sigma}\right)^{\frac{2}{2k+1}} t\right), \ \ j = 0, \ldots, 2k-1\right) . \end{aligned}$$

In particular,

$$\begin{aligned}&\left(H^{(k)}_{k,a,\sigma}(t),\ldots,H^{(2k-1)}_{k,a,\sigma}(t)\right)\\ &\stackrel{d}{=}\ \left(\sigma^{\frac{2k}{2k+1}}a^{\frac{1}{2k+1}}H^{(k)}_{k,1,1}\left(\left(\frac{a}{\sigma}\right)^{\frac{2}{2k+1}}t\right),\ldots,\sigma^{\frac{2}{2k+1}}a^{\frac{2k-1}{2k+1}}H^{(2k-1)}_{k,1,1}\left(\left(\frac{a}{\sigma}\right)^{\frac{2}{2k+1}}t\right)\right).\end{aligned}$$

Corollary A.2 For the particular choice $a = k!/(2k)!$ and $\sigma = 1$,

$$\begin{aligned}&\left(H^{(j)}_{k,k!/(2k)!,\sigma}(t),\ j=0,\ldots,2k-1\right)\\ &\stackrel{d}{=}\ \left(\sigma\left(\frac{(2k)!}{k!}\right)^{\frac{2k-1-2j}{2k+1}}H^{(j)}_{k,1,1}\left(\left(\frac{k!}{(2k)!}\right)^{\frac{2}{2k+1}}t\right),\quad j=0,\ldots,2k-1\right).\end{aligned}$$

Corollary A.3 (i) When $k = 1$ and $j = 1$,

$$H^{(1)}_1(t)\equiv H^{(1)}_{1,1/2,1}(t)\stackrel{d}{=}2^{-1/3}H^{(1)}_{1,1,1}(t/2)\equiv 2^{-1/3}\tilde{H}^{(1)}_1(t/2)$$

where $\tilde{H}_1 \equiv H_{1,1,1}$.

(ii) When $k = 2$, $j = 2, 3$,

$$\begin{aligned}(H^{(2)}_2(t),H^{(3)}_2(t)) &\equiv \left(H^{(1)}_{2,1/12,1}(t),H^{(3)}_{2,1/12,1}(t)\right)\\ &\stackrel{d}{=} \left((12)^{-1/5}H^{(1)}_{2,1,1}((12)^{-2/5}t),\ (12)^{-3/5}H^{(3)}_{2,1,1}((12)^{-2/5}t)\right)\\ &\equiv \left((12)^{-1/5}\tilde{H}^{(1)}_2((12)^{-2/5}t),\ (12)^{-3/5}\tilde{H}^{(3)}_2((12)^{-2/5}t)\right)\end{aligned}$$

where $\tilde{H}_2 \equiv H_{2,1,1}$.

Appendix B

APPROXIMATING PRIMITIVES OF BROWNIAN MOTION ON $[-N,N]$

B.1 *Approximating Brownian motion on* $[0,1]$

Let n be an integer. Consider the functions $h_{nj}, j = 0, \cdots, 2^n - 1$ defined by

$$h_{00}(t) = \begin{cases} t & \text{if } 0 \leq t \leq 1/2 \\ 1-t & 1/2 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$h_{nj}(t) = 2^{-n/2} h\left(2^n t - j\right), \quad \text{for } j = 0, \cdots, 2^n - 1.$$

The functions h_{nj} are called the *Schauder functions*. Let $Z_{nj}, j = 0, \cdots, 2^n - 1$ independent identically distributed standard Gaussians defined on the same probability space $([0,1], \mathcal{B}([0,1], \lambda)$. Now define the processes

$$V_n(t,\omega) = \sum_{j=0}^{2^n-1} h_{nj}(t) Z_{nj}(\omega)$$

and

$$U_m(t,\omega) = \sum_{n=0}^{m} V_n(t,\omega).$$

It can be shown that $U_m(t,\omega)$ converges uniformly as $m \to \infty$ with probability one to the process

$$\mathbb{U}(t,\omega) = \sum_{n=0}^{\infty} V_n(t,\omega).$$

which is a Brownian Bridge. To construct a standard Brownian motion, let Z be an additional standard Gaussian independent of all the $Z_{nj}, j = 0, \cdots, 2^n - 1$ and $n \in \mathbb{N}$. The process $\mathbb{W}$ defined by

$$\mathbb{W}(t,\omega) = \mathbb{U}(t,\omega) + tZ(\omega), \qquad t \in [0,1].$$

is a Brownian motion. For m large enough, the process

$$\begin{aligned} W_m(t,\omega) &= \sum_{n=0}^{m} V_n(t,\omega) + tZ(\omega) \\ &= \sum_{n=0}^{m} \sum_{j=0}^{2^n-1} h_{nj}(t) Z_{nj}(\omega) + tZ(\omega) \end{aligned}$$

is a good approximation to standard Brownian motion on $[0,1]$.

B.2 Approximating the $(k-1)$-fold integral of Brownian motion on $[0,n]$

Let $k \geq 2$ be an integer. Suppose that we want to approximate $I_{k-1}\mathbb{W}(t)$, the $(k-1)$-fold integral of Brownian motion given by

$$I_{k-1}\mathbb{W}(t) = \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} dW(s), \qquad t \in [0,1].$$

Using integration by parts, I_{k-1} can be rewritten

$$I_{k-1}\mathbb{W}(t) = \int_0^t \frac{(t-s)^{k-2}}{(k-2)!} W(s) ds.$$

The *Schauder functions* can be used again to approximate $I_{k-1}\mathbb{W}$. For m large enough, $\mathbb{I}_{k-1}$ can be approximated by

$$I_{k-1}W_m(t) = \sum_{n=0}^{m} \sum_{j=0}^{2^n-1} \int_0^t \frac{(t-s)^{k-2}}{(k-2)!} h_{nj}(s) ds \;\; Z_{nj} + \frac{t^k}{k!} Z \tag{B.1}$$

where $Z_{nj}, j = 0, \cdots, 2^n - 1$ and Z are independent identically distributed $N(0,1)$ defined on the same probability space $([0,1], \mathcal{B}([0,1[), \lambda)$. Thus, I_{k-1} can be given in a closed form once the integrals in the left side of the expression in (B.1) are evaluated analytically.

Lemma B.1 Let $t \in [0,1]$, n an integer and $j = 0, \cdots, 2^n - 1$. If p is an integer larger or equal to 2, then the $(p-1)$-fold integral of the *Schauder function* h_{nj} is given by

$$I_{p-1}h_{nj}(t)$$

$$
\begin{aligned}
&= 0, && \text{if } t \in [0, 2^{-n}j] \\
&= \frac{2^{\frac{n}{2}}}{p!}(t - 2^{-n}j)^p, && \text{if } t \in \left[2^{-n}j, 2^{-n}(j+\frac{1}{2})\right] \\
&= \frac{2^{\frac{n}{2}}}{p!}(2^{-(n+1)p} - (t - 2^{-n}(j+\frac{1}{2}))^p), && \text{if } t \in \left[2^{-n}(j+\frac{1}{2}), 2^{-n}(j+1)\right] \\
&\quad + \frac{2^{-(\frac{n}{2}+1)}}{(p-1)!}(t - 2^{-n}(j+\frac{1}{2}))^{p-1} \\
&= \frac{1}{(p-1)!}2^{-(\frac{n}{2}+1+(n+1)(p-1))}, && \text{if } t \in [2^{-n}(j+1), 1].
\end{aligned}
$$

Proof. The function h_{nj} can be rewritten as

$$
h_{nj}(t) = 2^{-n/2}\begin{cases} 0, & \text{if } t \in [0, 2^{-n}j] \\ 2^n t - j, & \text{if } t \in \left[2^{-n}j, 2^{-n}(j+\frac{1}{2})\right] \\ 1 - (2^n t - j), & \text{if } t \in \left[2^{-n}(j+\frac{1}{2}), 2^{-n}(j+1)\right] \\ 0, & \text{if } t \in [2^{-n}(j+1), 1]. \end{cases}
$$

If $t \in [0, 2^{-n}j]$, it is clear that $I_{k-1}h_{nj}(t) = 0$. If $2^{-n}j \le t \le 2^{-n}(j+1/2)$, we have

$$
\begin{aligned}
&\int_0^t (t-s)^{p-2} h_{nj}(s)ds \\
&\quad = \int_{2^{-n}j}^t (t-s)^{p-2} h_{nj}(s)ds \\
&\quad = \int_{2^{-n}j}^t (t-s)^{p-2}\left(2^{-n/2}(2^n s - j)\right)ds \\
&\quad = 2^{-n/2+n}\int_{2^{-n}j}^t (t-s)^{p-2}(s - 2^{-n}j)ds \\
&\quad = 2^{n/2}\left(-\int_{2^{-n}j}^t (t-s)^{p-1}ds + (t - 2^{-n}j)\int_{2^{-n}j}^t (t-s)^{p-2}ds\right) \\
&\quad = 2^{n/2}\left(-\frac{1}{p}(t - 2^{-n}j)^p + \frac{1}{p-1}(t - 2^{-n}j)^p\right) \\
&\quad = \frac{2^{n/2}}{(p-1)p}(t - 2^{-n}j)^p
\end{aligned}
$$

and hence for all $2^{-n}j \le t \le 2^{-n}(j+1/2)$

$$
I_{p-1}h_{nj}(t) = \frac{1}{(p-2)!}\int_0^t (t-s)^{p-2} h_{nj}(s)ds
$$

$$= \frac{2^{n/2}}{p!}(t - 2^n j)^p.$$

In particular,

$$I_{p-1}h_{nj}(2^{-n}(j+1/2)) = \frac{2^{n/2}}{p!}2^{-(n+1)p}.$$

Now, for $2^{-n}(j+\frac{1}{2}) \le t \le 2^{-n}(j+1)$, we have

$$\begin{aligned} I_{p-1}h_{nj}(t) &= I_{p-1}h_{nj}(2^{-n}(j+1/2)) + \int_{2^{-n}(j+1/2)}^{t} \frac{1}{(p-2)!}(t-s)^{p-2}h_{nj}(s)ds \\ &= \frac{2^{n/2}}{p!}2^{-(n+1)p} + \int_{2^{-n}(j+1/2)}^{t} \frac{1}{(p-2)!}(t-s)^{p-2}h_{nj}(s)ds, \end{aligned} \tag{B.2}$$

where

$$\begin{aligned} &\int_{2^{-n}(j+1/2)}^{t}(t-s)^{p-2}h_{nj}(s)ds \\ &= 2^{-n/2}\int_{2^{-n}(j+1/2)}^{t}(t-s)^{p-2}(1-(2^n s - j))ds \\ &= 2^{-n/2}\left(\int_{2^{-n}(j+1/2)}^{t}(t-s)^{p-2}ds - \int_{2^{-n}(j+1/2)}^{t}(t-s)^{p-2}(2^n s - j)ds\right) \\ &= 2^{-n/2}\left(\frac{1}{p-1}\left(t-2^{-n}(j+1/2)\right)^{p-1} - 2^n\int_{2^{-n}(j+1/2)}^{t}(t-s)^{p-2}(s-j2^{-n})ds\right) \\ &= \frac{2^{-n/2}}{p-1}\left(t-2^{-n}(j+1/2)\right)^{p-1} \\ &\quad - 2^{n/2}\left(-\int_{2^{-n}(j+1/2)}^{t}(t-s)^{p-1}ds + (t-2^{-n}j)\int_{2^{-n}(j+1/2)}^{t}(t-s)^{p-2}ds\right) \\ &= \frac{2^{-n/2}}{p-1}\left(t-2^{-n}(j+1/2)\right)^{p-1} + \frac{2^{n/2}}{p}\left(t-2^{-n}(j+1/2)\right)^{p} \\ &\quad - \frac{2^{n/2}}{p-1}\left(t-2^{-n}j\right)\left(t-2^{-n}(j+1/2)\right)^{p-1} \\ &= \frac{2^{-n/2}}{p-1}\left(t-2^{-n}(j+1/2)\right)^{p-1} + \frac{2^{n/2}}{p}\left(t-2^{-n}(j+1/2)\right)^{p} \\ &\quad - \frac{2^{n/2}}{p-1}\left(t-2^{-n}(j+1/2)\right)\left(t-2^{-n}(j+1/2)\right)^{p-1} \\ &\quad - \frac{2^{n/2}}{p-1}2^{-n-1}\left(t-2^{-n}(j+1/2)\right)^{p-1} \\ &= \frac{2^{-n/2}}{p-1}\left(t-2^{-n}(j+1/2)\right)^{p-1} - \frac{2^{n/2}}{(p-1)p}\left(t-2^{-n}(j+1/2)\right)^{p} \end{aligned}$$

$$
\begin{aligned}
& - \frac{2^{-(n/2+1)}}{p-1}\left(t-2^{-n}(j+1/2)\right)^{p-1} \\
= & \frac{2^{-(n/2+1)}}{p-1}\left(t-2^{-n}(j+1/2)\right)^{p-1} - \frac{2^{n/2}}{(p-1)p}\left(t-2^{-n}(j+1/2)\right)^{p}. \qquad \text{(B.3)}
\end{aligned}
$$

By combining (B.2) and (B.3), we obtain that

$$I_{p-1}h_{nj}(t) = \frac{2^{n/2}}{p!}\left(2^{-(n+1)p} - \left(t-2^{-n}(j+1/2)\right)^{p}\right) + \frac{2^{-(n/2+1)}}{(p-1)!}\left(t-2^{-n}(j+1/2)\right)^{p-1}$$

for all $t \in [2^{-n}(j+1/2), 2^{-n}(j+1)]$. Finally, let $t \in [2^{-n}(j+1), 1]$. We have,

$$
\begin{aligned}
I_{p-1}h_{nj}(t) &= I_{p-1}h_{nj}(2^{-n}(j+1)) + \int_{2^{-n}(j+1)}^{t} \frac{1}{(p-2)!}(t-s)^{p-2}h_{nj}(s)ds \\
&= I_{p-1}h_{nj}(2^{-n}(j+1))
\end{aligned}
$$

since $h_{nj}(t) = 0$ for $t \geq 2^{-n}(j+1)$. Hence,

$$
\begin{aligned}
I_{p-1}h_{nj}(t) &= \frac{2^{n/2}}{p!}\left(2^{-(n+1)p} - (2^{-n}(j+1) - 2^{-n}(j+1/2))^{p}\right) \\
&\quad + \frac{2^{-(n/2+1)}}{(p-1)!}\left(2^{-n}(j+1) - 2^{-n}(j+1/2)\right)^{p-1} \\
&= \frac{2^{-(n/2+1)}}{(p-1)!}2^{-(n+1)(p-1)} \\
&= \frac{1}{(p-1)!}2^{-(\frac{n}{2}+1+(n+1)(p-1))}
\end{aligned}
$$

for all $t \in [2^{-n}(j+1), 1]$. ■

B.3 Approximating the $(k-1)$*-fold integral of Brownian motion on* $[-n, n]$

Let $n > 1$ be an integer. A Brownian motion defined on $[0, n]$ can be obtained by generating n independent copies of standard Brownian motion on the intervals $[i, i+1], i = 0, 1, \cdots, n-1$ and "pasting " them together at the junction points. More explicitly, for $i = 1, \cdots, n$, let $\mathbb{W}_i$ be independent copies of standard Brownian motion on $[0, 1]$, and let $\mathbb{B}_i$ be the resulting Brownian motion on the interval $[0, i]$. We have,

$$\mathbb{B}_1(t) = \mathbb{W}_1(t), \qquad t \in [0, 1]$$

and

$$\mathbb{B}_i(t) = \begin{cases} \mathbb{B}_{i-1}(t), & t \in [0, i-1] \\ \mathbb{B}_{i-1}(i-1) + \mathbb{W}_i(t-(i-1)), & t \in [i-1, i] \end{cases}$$

for $i = 2, \cdots, n$.

Now, suppose we want to approximate successive primitives of Brownian motion on $[0, n]$. For example, take $n = 2$ and suppose we want to find an approximation to the first primitive of $\mathbb{B}_2$ on $[0, 2]$. For $t \in [0, 2]$, we have

$$\int_0^t \mathbb{B}_2(s)ds = \begin{cases} \int_0^t \mathbb{W}_1(s)ds, & \text{if } 0 \le t \le 1 \\ \int_0^1 \mathbb{W}_1(s)ds + (t-1)\mathbb{W}_1(1) + \int_0^{t-1} \mathbb{W}_2(s)ds, & \text{if } 1 \le t \le 2 \end{cases}$$

Similarly, for any integer $k \ge 2$, we can establish that the $(k-1)$-fold integral of $\mathbb{B}_2$ on $[0, 2]$ is given by

$$\int_0^t \frac{(t-s)^{k-1}}{(k-1)!} d\mathbb{B}_2(s) = \begin{cases} \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} d\mathbb{W}_1(s), & \text{if } 0 \le t \le 1 \\ \sum_{j=0}^{k-1} \frac{(t-1)^j}{j!} \int_0^1 \frac{(1-s)^{k-1-j}}{(k-1-j)!} d\mathbb{W}_1(s), & \text{if } 1 \le t \le 2 \\ + \int_0^{t-1} \frac{(t-1-s)^{k-1}}{(k-1)!} d\mathbb{W}_2(s). & \end{cases}$$

The last expression also shows that the $(k-1)$-fold integral of $\mathbb{B}_2$ involves the $(k-1)$-fold integral of both the independent processes $\mathbb{W}_1$ and $\mathbb{W}_2$, and the j-fold integral of $\mathbb{W}_2$ at the point $t = 1$ (boundary point), for $j = 0, \cdots, k-1$. This example can be generalized easily to any $n > 1$:

$$\begin{aligned}
&\int_0^t \frac{(t-s)^{k-1}}{(k-1)!} d\mathbb{B}_n(s) \\
&= \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} d\mathbb{W}_1(s), && \text{if } 0 \le t \le 1 \\
&= \sum_{j=0}^{k-1} \frac{(t-1)^j}{j!} \int_0^1 \frac{(1-s)^{k-1-j}}{(k-1-j)!} d\mathbb{B}_1(s) + \int_0^{t-1} \frac{(t-1-s)^{k-1}}{(k-1)!} d\mathbb{W}_2(s), && \text{if } 1 \le t \le 2 \\
&\ \ \vdots \\
&= \sum_{j=0}^{k-1} \frac{(t-(i-1))^j}{j!} \int_0^{i-1} \frac{(i-1-s)^{k-1-j}}{(k-1-j)!} d\mathbb{B}_{i-1}(s) \\
&\quad + \int_0^{t-(i-1)} \frac{(t-(i-1)-s)^{k-1}}{(k-1)!} d\mathbb{W}_i(s), && \text{if } \ i-1 \le t \le i
\end{aligned}$$

$$\begin{aligned}
&\vdots \\
&= \sum_{j=0}^{k-1} \frac{(t-(n-1))^j}{j!} \int_0^{n-1} \frac{(n-1-s)^{k-1-j}}{(k-1-j)!} d\mathbb{B}_{n-1}(s) \\
&\quad + \int_0^{t-(n-1)} \frac{(t-(n-1)-s)^{k-1}}{(k-1)!} d\mathbb{W}_n(s), \qquad \text{if } n-1 \le t \le n.
\end{aligned}$$

The method described above can be used to get an approximation to the $(k-1)$-fold integral of two independent copies of Brownian motion on $[0, n]$. An approximation on $[-n, n]$ is then obtained by “pasting” these copies at the point 0.

Appendix C

PROGRAMS

C.1 *C* code for generating the processes $Y_k, \cdots, Y_k^{(k-1)}$

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <time.h>
#define M_SQRT2 (1.414213562373095)



double SchauderFunc(double);
double IntSchauderFunc(int l, int p, int i, double x);
void IntBrownFunc(double* IntBrown, int K, int m);
void IntBrownOC(double* Output, int K, double C,int m);
double inverse_normal_func(double p);

FILE*ifp;
double normals[256][8];
int half,col=0;
int main(void){

  int i,j;



  int fact;
```

```
  double a,b;
  int K=4,m=12;
  double C=4.0;
  int Lg = (int)pow(2.0,(double)(m))*C+1;
  double* Output = calloc((Lg+1)*(K+1),sizeof(double));

  IntBrownOCdrift(Output, K, C,m);

  return 0;
}

void IntBrownOC(double* Output, int K, double C,int m) {
  int k,y,i,j;
  double val;
  double twoMinp = pow(2.0,(double)(-m));
  int Lg = (int)pow(2.0,(double)(m))+1;
  double* grid = calloc((Lg+1),sizeof(double));
  int* vecBound=calloc(C,sizeof(int));
  double* Wi = calloc((Lg+1)*(K+1),sizeof(double));
  double* Matgrid = calloc((Lg+1)*(K+1),sizeof(double));
  //double* Bi = calloc((Lg+1)*(K+1),sizeof(double));
  double* Bi=Output;
  double* BiMinus1 = calloc((K+1),sizeof(double));
  int stride = (int)pow(2.0,(double)(m))*C+1+1;

  i=1; ////////
  val=0.0;
  while(val<=1) {
    grid[i]=val;
    val += twoMinp;  // equivalent to val = val+twoMinp;
```

```
    i++;  // i=i+1;
  }

  if(C>1)
    for(i=1;i<=C-1;++i)
      vecBound[i]=Lg;

  for(i=1;i<=C;++i) {
    if(half==0) col=C-i;
    else col=C+i-1;
    IntBrownFunc(Wi,K,m);

    /////// for debugging /////////
    /*
      printf("\nIntBrown i= %d \n\n",i);
    for(k=1;k<=K;++k) {
      for(y=1;y<=Lg;++y)
printf("%e ",*(Wi + k*(Lg+1) + y));
      printf("\n");
    }
    */
    ///////////////////////////////
    for(k=1;k<=K;++k) {
      for(j=1;j<=k;++j) {
for(y=1;y<=Lg;++y) {
  Matgrid[j*(Lg+1)+y]=pow(grid[y],(double)(k-j))/factorial(k-j);
}
      }
      matMulAdd(Bi,Wi,BiMinus1,Matgrid,k,Lg,stride);
    }
```

```
    for(k=1;k<=K;++k) {
      BiMinus1[k]=Bi[k*(stride)+Lg];
    }

    Bi+=Lg-1;
  }



  /*
  printf("\nIntBrownOC = \n\n");
  for(k=1;k<=K;++k) {
      for(y=1;y<=stride-1;++y)
printf("%e ",*(Output + k*(stride) + y));
      printf("\n");
  }
  */
}

void IntBrownFunc(double* IntBrown, int K, int m) {

  int i,y,k,n,j;
  double val;
  int twon;
  double twoMinp = pow(2.0,(double)(-m));
  int twomPlus1Min1 = (int) pow(2.0,(double)(m+1))-1;
  int Lg = (int)pow(2.0,(double)(m))+1;
  double* grid = calloc(Lg+1,sizeof(double));
  double* Zv = calloc(twomPlus1Min1+1,sizeof(double));
  double* IntUm = calloc((Lg+1)*(K+1),sizeof(double));
  double Z;
```

```
  i=1; ////////
  val=0;
  while(val<=1) {
    grid[i]=val;
    val += twoMinp;  // equivalent to val = val+twoMinp;
    i++;  // i=i+1;
  }

  //Z=inverse_normal_func(drand48());
    Z=inverse_normal_func((double)rand()/RAND_MAX);
  //Z=normals[255][col];

  for (i = 0; i < twomPlus1Min1+1; i++ ) ///
    //Zv[i] = inverse_normal_func(drand48());
    Zv[i] = inverse_normal_func((double)rand()/RAND_MAX);
    //Zv[i] = normals[i][col];




  // for debugging
  //for (i = 0; i < twomPlus1Min1+1; i++ )
  //  printf("%f ", Zv[i]);
  //printf("\n");
  ////////////////

  for(y=2;y<=Lg;++y)
    for(k=1;k<=K;++k)
      for(n=0;n<=m;++n) {
twon=(int)pow(2.0,(double)n);
```

```
for(j=0;j<=twon-1;++j)
  *(IntUm + k*(Lg+1) + y) += Zv[twon + j] * IntSchauderFunc(k,n,j,grid[y]);
      }

  for(k=1;k<=K;++k)
    for(y=1;y<=Lg;++y)
      *(IntBrown + k*(Lg+1) + y)=  *(IntUm + k*(Lg+1) + y)
+ Z*pow(grid[y],(double)k)/factorial(k);
  //    IntBrown



}

double IntSchauderFunc(int l, int p, int i, double x) {
  double IntSchauder=0.0;
  double twop = pow((double)2,(double)p);
  double twopMin1 = twop -1;
  double twoMinp = pow((double)2,-(double)p);
  double twoHalfp = pow((double)2,(double)p/2.0);
  double twoMinHalfp = pow((double)2,-(double)p/2.0);
  double twoMinpPlus1l = pow((double)2,-(double)(p+1)*l);
  double twoMinpPlus1lMin1 = pow((double)2,-(double)(p+1)*(l-1));
  //double twoMinpIPlus1 = twoMinp*(i+1);
  double twoMinHalfpPlus1 = twoMinHalfp/2.0;
  double factlMin1 = factorial(l-1);
  double factl = factlMin1*l;

  if(i < 0 || i > twopMin1) {
    fprintf(stderr,"i (%d) has to be between 0 and 2^p-1
(%d)\n",i, (int)twopMin1);
```

```
    exit(-1);
  }
  if(l==1) {
    IntSchauder = twoMinHalfp*SchauderFunc((double) twop*x-i);
  }
  else {
    if(x >= twoMinp * i) {
      // Case 1
      if( x <= twoMinp*(i + 1/2.0))
IntSchauder = twoHalfp /factl * pow((double)(x- twoMinp*i),(double)l);
      // Case 2
      else {
// Subcase1
//if( x <=  twoMinpIPlus1 )
if( x <=  twoMinp*(i + 1))
  IntSchauder = twoHalfp/factl*(twoMinpPlus1l - pow((double) x
- twoMinp*(i+1/2.0),(double) l)) +
 (twoMinHalfpPlus1/factlMin1)*pow((double) x-twoMinp*(i+1/2.0),(double) l-1);
// Subcase2
else
  IntSchauder = twoMinHalfpPlus1*twoMinpPlus1lMin1/factlMin1;

      } //else
    } //if(x >= twoMinp * i)
  }//else

  return IntSchauder;
}

double SchauderFunc(double x) {
```

```
  double Schauder= 0.0;
  if(x >= 0 && x <= 0.5)
    Schauder = x;
  else if(x > 0.5 && x <= 1)
    Schauder = 1-x;
  return Schauder ;
}

int factorial(int n) {
  int fact=1;
  int i;

  for(i=2; i <= n; ++i)
    fact *= i;

  return fact;

}

double inverse_error_func(double p) {
/*
           Source: This routine was derived (using f2c) from the
                   FORTRAN subroutine MERFI found in
                   ACM Algorithm 602 obtained from netlib.

                   MDNRIS code contains the 1978 Copyright
                   by IMSL, INC. .  Since MERFI has been
                   submitted to netlib, it may be used with
                   the restriction that it may only be
                   used for noncommercial purposes and that
```

```
                IMSL be acknowledged as the copyright-holder
                of the code.
        */



/* Initialized data */
static double a1 = -.5751703;
static double a2 = -1.896513;
static double a3 = -.05496261;
static double b0 = -.113773;
static double b1 = -3.293474;
static double b2 = -2.374996;
static double b3 = -1.187515;
static double c0 = -.1146666;
static double c1 = -.1314774;
static double c2 = -.2368201;
static double c3 = .05073975;
static double d0 = -44.27977;
static double d1 = 21.98546;
static double d2 = -7.586103;
static double e0 = -.05668422;
static double e1 = .3937021;
static double e2 = -.3166501;
static double e3 = .06208963;
static double f0 = -6.266786;
static double f1 = 4.666263;
static double f2 = -2.962883;
static double g0 = 1.851159e-4;
static double g1 = -.002028152;
```

```
static double g2 = -.1498384;
static double g3 = .01078639;
static double h0 = .09952975;
static double h1 = .5211733;
static double h2 = -.06888301;

/* Local variables */
static double a, b, f, w, x, y, z, sigma, z2, sd, wi, sn;

x = p;

/* determine sign of x */
if (x > 0)
sigma = 1.0;
else
sigma = -1.0;

/* Note: -1.0 < x < 1.0 */

z = fabs(x);

/* z between 0.0 and 0.85, approx. f by a
   rational function in z  */

if (z <= 0.85) {
z2 = z * z;
f = z + z * (b0 + a1 * z2 / (b1 + z2 + a2
    / (b2 + z2 + a3 / (b3 + z2))));

/* z greater than 0.85 */
```

```
} else {
a = 1.0 - z;
b = z;

/* reduced argument is in (0.85,1.0),
   obtain the transformed variable */

w = sqrt(-(double)log(a + a * b));

/* w greater than 4.0, approx. f by a
   rational function in 1.0 / w */

if (w >= 4.0) {
wi = 1.0 / w;
sn = ((g3 * wi + g2) * wi + g1) * wi;
sd = ((wi + h2) * wi + h1) * wi + h0;
f = w + w * (g0 + sn / sd);

/* w between 2.5 and 4.0, approx.
   f by a rational function in w */

} else if (w < 4.0 && w > 2.5) {
sn = ((e3 * w + e2) * w + e1) * w;
sd = ((w + f2) * w + f1) * w + f0;
f = w + w * (e0 + sn / sd);

/* w between 1.13222 and 2.5, approx. f by
   a rational function in w */
} else if (w <= 2.5 && w > 1.13222) {
sn = ((c3 * w + c2) * w + c1) * w;
```

```
sd = ((w + d2) * w + d1) * w + d0;
f = w + w * (c0 + sn / sd);
}
}
y = sigma * f;
return(y);
}


double inverse_normal_func(double p) {
/*
           Source: This routine was derived (using f2c) from the
                   FORTRAN subroutine MDNRIS found in
                   ACM Algorithm 602 obtained from netlib.

                   MDNRIS code contains the 1978 Copyright
                   by IMSL, INC. .  Since MDNRIS has been
                   submitted to netlib it may be used with
                   the restriction that it may only be
                   used for noncommercial purposes and that
                   IMSL be acknowledged as the copyright-holder
                   of the code.
        */

/* Initialized data */
static double eps = 1e-10;
static double g0 = 1.851159e-4;
static double g1 = -.002028152;
static double g2 = -.1498384;
static double g3 = .01078639;
```

```
static double h0 = .09952975;

static double h1 = .5211733;

static double h2 = -.06888301;

static double sqrt2 = M_SQRT2; /* 1.414213562373095; */

/* Local variables */

static double a, w, x;

static double sd, wi, sn, y;

double inverse_error_func(double p);

/* Note: 0.0 < p < 1.0 */

/* assert ( 0.0 < p && p < 1.0 ); */

/* p too small, compute y directly */

if (p <= eps) {

a = p + p;

w = sqrt(-(double)log(a + (a - a * a)));

/* use a rational function in 1.0 / w */

wi = 1.0 / w;

sn = ((g3 * wi + g2) * wi + g1) * wi;

sd = ((wi + h2) * wi + h1) * wi + h0;

y = w + w * (g0 + sn / sd);

y = -y * sqrt2;

} else {

x = 1.0 - (p + p);

y = inverse_error_func(x);

y = -sqrt2 * y;

}
```

```
return(y);
}
```

C.2 S codes for generating the processes $Y_k, \cdots, Y_k^{(k-1)}$

```
SchauderFunc <- function(x){
Schauder <- NULL
if( x < 0 | x > 1)
Schauder <- 0
else{
if(x >= 0 & x <= 1/2)
Schauder <- x
if(x > 1/2 & x <= 1)
Schauder <- 1- x

}
Schauder
}

IntSchauderFunc <- function(l, p, i, x){
if( i < 0 | (i > 2^p -1))
print("i has to be between 0 and 2^p -1")
if(l < 1)
print("l has to be greater or equal to 1")

IntSchauder <- NULL
if(l == 1){
IntSchauder <- 2^{-p/2}*SchauderFunc(2^p *x - i)
```

```
}
else{
  if(x < (2^{- p}* i))
  IntSchauder <- 0
        else {
     if((x >= 2^{- p} * i) & (x <= 2^{- p} * (i + 1/2)))
     IntSchauder <- (2^{p/2}/factorial(l)) * (x - 2^{- p}*i)^{l}
     if((x >= 2^{- p} * (i + 1/2)) & (x <= 2^{- p} * (i + 1)))
     IntSchauder <- (2^{p/2}/factorial(l)) * (2^{-(p + 1) * l}-
(x - 2^{- p} * (i + 1/2))^{l}) + (2^{- (p/2 + 1)}/factorial(l-1))
* (x - 2^{- p} * (i + 1/2))^{l-1}
     if(x > 2^{- p} * (i + 1))
     IntSchauder <- 2^{ - (p/2 + 1 + (l-1) * (p + 1))}/factorial(l-1)
         }
      }
IntSchauder
}

IntBrownFunc <- function(K,m){

grid <- seq(0, 1, 2^{- m})
L.g <- length(grid)
Zv <- rnorm(2^{m + 1} - 1, 0, 1)
Z <- rnorm(1, 0, 1)

IntUm <- matrix(0, nrow=K,ncol=L.g)
IntBrowm <- matrix(0, nrow=K,ncol=L.g)



for(y in 2:L.g) {
```

```
for(k in 1:K){
   for(n in 0:m) {
for(j in 0:(2^n - 1)) {
 IntUm[k,y] <- IntUm[k,y]
+ Zv[2^n + j] * IntSchauderFunc(k, n, j, grid[y])
}
  }
 }
  }

for(k in 1:K){
IntBrowm[k,] <- IntUm[k,] +  Z*( (grid)^{k})/factorial(k)
    }

IntBrowm
}



  IntBrown0C <- function(K,C,m){

grid <- seq(0,1,2^{-m})
L.g <- length(grid)
vec.bound <- NULL
if(C > 1){
vec.bound <- (1:(C-1))*L.g
}
B.iminus1 <- matrix(0,nrow=K,ncol=L.g)
B.i <- matrix(0,nrow=K,ncol=L.g)
Output <- matrix(0,nrow=K,ncol=L.g)
```

```
for(i in 1:C){
  print(i)
  W.i <- IntBrownFunc(K,m)
     for(k in 1:K){
  Matgrid <- rep(0,L.g)
  for(j in 1:k){
 Matgrid <- rbind(Matgrid,grid^{(k-j)}/factorial(k-j))
  }
  Matgrid <- Matgrid[-1,]
B.i[k,] <- W.i[k,] + matrix(B.iminus1[1:k,L.g],nrow=1,ncol=k)%*%Matgrid
}

         B.iminus1 <- B.i
  Output <- cbind(Output,B.i)

}

Output <- Output[,-(1:L.g)]

Output <- Output[,-vec.bound]
Output

}

IntBrownCCdrift <- function(C,m,K){
   # This function calculates the successive integral
   # of a two sided Brownian Motion on [-C,C] + the drift
   # on the specified grid.
```

```
   grid <- seq(-C,C,2^{-m})
   # We generate two independent copies to the right and left of 0.
Output1 <- IntBrown0C(C,m,K)
Output2 <- IntBrown0C(C,m,K)
L.g <- length(grid)

for(k in 1:K){
Output2[k,] <- rev(Output2[k,-1])
    }
    Output <- cbind(Output2,Output1)
   # We add the drift.
for(k in 1:K){
Output[k,] <- Ouput[k,] + (-1)^K *(factorial(2*K)/factorial(K+k))*(grid)^{K+k}
}

Output

    }
```

C.3 S codes for generating the processes $H_{c,k}, \cdots, H_{c,k}^{(2k-1)}$ when k is even

```
# This code calcules an approximation to the process H_K,
# the invelope of Y_K the (k-1)-fold integral of
# two sided Brownian Motion + t^{2K} when K is even (K >=2).
# m is the precision of the Brownian motion approximation using
# the Haar function construction.



IterativeSHk <- function(K=6,C=4,m=11,eps=10^{-7},p=20,p1=10,p2=16365){
```

```
grid <- seq(-C,C,2^{-m})
IntBr <- intbrownk6c4m11
IntBr <- t(IntBr)
Mat0 <- matrix(0,nrow=2*K + p, ncol=2*K+p)
L.g <- length(grid)
# 1 is the location of the successive derivative of Y
# at -C, L.g is that of ...of at C.
Yd <- rbind(IntBr[,1],IntBr[,L.g])

# Select only the even derivatives of Y at -C and C.

Yd <- Yd[,seq(2,K,2)]

# this vector stores in the first row:
# Y^{(k-1)}(-c),Y^{(k-2)}(-c),...,Y(-c)
# and Y^{(k-1)}(c),Y^{(k-2)}(c),...,Y(c).

S0 <- c(-C,C)
Alpha0 <- StartingSplineHk(K,C,Yd)
Coef0 <- Alpha0[1]
H <- EvaluateGrid(K,Alpha0,S0,grid)
Diff <- H - IntBr[K,]

# For later, we need to have the initial conditions
#in the "right form" (as it is required in ComputeSplineHk)
# hence, we need to reverse the components of Yd so that
#we start from Y(-/+ c) and finish with Y^{(k-2)}(-/+ c).

Yd.rev <- Yd
Yd.rev[1,] <- rev(Yd[1,])
```

```
Yd.rev[2,] <- rev(Yd[2,])

# Check whether H >= Y.

min.Diff <- min(Diff[p1:p2])
print(min.Diff)
Count <- 0
while(min.Diff < -eps){
Count <- Count + 1
cat("Main Loup numb = ", Count, "\n")
Diff.sort <- rank(Diff[p1:p2])
min.rank <- min(Diff.sort)
min.pos <- match(min.rank,Diff.sort)
thetamin <- grid[p1:p2][min.pos] # locate t*.
valmin <- Diff[p1:p2][min.pos]

print(c(thetamin,valmin))

# Compute the new spline for the new set of knots.

S <- c(S0,thetamin)
S <- sort(S)
print(S)
#locate the knots in the grid.

positions <- match(S,grid)

Y <- InitialCondHk(K,C,Yd.rev,IntBr,positions)
p <- length(S)-2
Alpha <- ComputeSplineHk(K=K,Y=Y,S=S,Mat0)
```

```
Alpha <- as.numeric(Alpha)
Coef <- c(Alpha[1],Alpha[(2*K+1):(2*K+p)])
Coef <- cumsum(Coef)
min.C <- min(diff(Coef))
count <- 0

   while(min.C < 0){
count <- count+1
cat("Sub loup numb = ",count," of the main loop numb=", Count, "\n")
    index <- IndexFuncHk(S0=S0,S=S,Coef0=Coef0,Coef=Coef)
S <- S[-index]
p <- length(S)-2
positions <- match(S,grid)
Y <- InitialCondHk(K,C,Yd.rev,IntBr,positions)
Alpha <- ComputeSplineHk(K=K,Y=Y,S=S,Mat0)
Alpha <- as.numeric(Alpha)
       Coef <- c(Alpha[1],Alpha[(2*K+1):(2*K+p)])
Coef <- cumsum(Coef)
min.C <- min(diff(Coef))
   }#while min.C < 0

H <- EvaluateGrid(K,Alpha,S,grid)
Diff <- H - IntBr[K,]
min.Diff <- min(Diff[p1:p2])

S0 <- S
Coef0 <- Coef

}#while min.Diff < -eps
print(Alpha)
```

```
print(positions)
Mat.H <- H
for(d in 1:(2*K-2)){
Mat.H <- rbind(Mat.H,EvaluateGridDer(K,Alpha,S,grid,d))
}
Mat.H

}#end of the function



#This code calculates the coefficients of the "starting" spline
#which is of degree 2k-2.
# Yd is a matrix of dimension 2x(K/2) containing
#the derivatives of the (K-1)-integral of a two sided
#Brownian motion (Y) + t^{2K} at the boundary points -C and C.
#It starts with the (K-2)th
#derivative of Y at -C and C, (K-4)th,...,0.

StartingSpline <- function(K,C,Yd){
C <- 2
K <- 4
Yd <- rbind(-1,2)
if((K-2*floor(K/2))!=0){
print("Enter please an even K !")
}

#This part gives the coefficients when K=2.

if(K==2){
Coef <- c(6*C^2,(Yd[2,1]-Yd[1,1])/(2*C),(Yd[2,1]+Yd[1,1])/2 - 6*C^4)
```

```
}

#This part of the code calculates the coefficients when K > 2 (and even).

if(K > 2){
d <- 2*K-2
a.d <- (factorial(2*K)/factorial(2))*C^2
Coef  <- a.d
for(i in (d-1):0){
   p <- 2*K-i
      if(p <= K){
        if((p-2*floor(p/2))!=0){
     Coef <- c(Coef,0)
         }
          else{
          Coef <- c(Coef,(factorial(2*K)/factorial(2*i))*C^{2*i}-
sum(Coef[Coef!=0]*(1/factorial(2*((i-1):1)))*C^{2*((i-1):1)}))
          }
       }


      if(p > K){
     if((p-2*floor(p/2))!=0){
     Coef <- c(Coef, (Yd[2,(p-K+1)/2]-Yd[1,(p-K+1)/2])/(2*C))
        }
        else{
   i <- p/2
          Coef <- c(Coef,(Yd[2,(p-K)/2]+Yd[1,(p-K)/2])/2
- sum(Coef[2*((i-1):1)]*(1/factorial(2*((i-1):1)))*C^{2*((i-1):1)}))
        }
```

```
      }
   }

}
  Coef <- Coef/factorial((2*K-2):0)
   Coef
}



EvaluateGrid <- function(K,Alpha,S,grid){

if(length(S)==2){
 #grid <- seq(-C,C,2^{-m})
 #H <- rep(0,length(grid))
 H <- grid
 for(i in 1:length(H)){
 H[i] <- sum(Alpha*(grid[i])^{(2*K-2):0})
 }

}

if(length(S) > 2){
p <- length(S)-2
Alpha.1 <- Alpha[1:(2*K)]
Alpha.2 <- Alpha[(2*K+1):(2*K+p)]

nr <- length(S) -1
C <- S[length(S)]
pos <- match(S,grid)
```

```
#Seq.1 <- seq(S[1],S[2],2^{-m})
Seq.1 <- grid[pos[1]:pos[2]]
l.1 <- length(Seq.1)
#H.1 <- rep(0,l.1)
H.1 <- Seq.1
for(j in 1:l.1){
H.1[j] <- sum((Alpha.1/factorial((2*K-1):0))*(Seq.1[j])^{(2*K-1):0})
}

H <- H.1[-l.1]
for(i in 2:nr){
#Seq.i <-  seq(S[i],S[i+1],2^{-m})
 Seq.i <-  grid[pos[i]:pos[(i+1)]]
l.i <- length(Seq.i)
#H.i <- rep(0,l.i)
H.i <- Seq.i
       for(j in 1:l.i){
       H.i[j] <- sum((Alpha.1*(Seq.i[j])^{(2*K-1):0})/factorial((2*K-1):0))
+ sum(Alpha.2[1:(i-1)]*(Seq.i[j]-S[2:i])^{2*K-1}/factorial(2*K-1))
        }
    H <- c(H,H.i[-l.i])

}



Lastval <- sum(Alpha.1*C^{(2*K-1):0}/factorial((2*K-1):0) )
 + sum((Alpha.2*(C-S[2:nr])^{2*K-1})/factorial(2*K-1))
H <- c(H,Lastval)
}
```

```
H

}

EvaluateGridDer <- function(K,Alpha,S,grid,d){

if(d > 2*K -2)
print("enter d less than or equal to 2*K-2")
else{

if(length(S)==2){
 grid <- seq(-C,C,2^{-m})
 #H.d <- rep(0,length(grid))
 H.d <- grid
 for(i in 1:length(H.d)){
 H.d[i] <- sum(Alpha[1:(2*K-1-d)]*(grid[i])^{(2*K-2-d):0})
 }

}

if(length(S) > 2){
p <- length(S)-2
Alpha.1 <- Alpha[1:(2*K-d)]
Alpha.2 <- Alpha[(2*K+1):(2*K+p)]

nr <- length(S) -1
C <- S[length(S)]
pos <- match(S,grid)
#Seq.1 <- seq(S[1],S[2],2^{-m})
```

```
Seq.1 <- grid[pos[1]:pos[2]]
l.1 <- length(Seq.1)
#H.1 <- rep(0,l.1)
H.1 <- Seq.1
for(j in 1:l.1){
H.1[j] <- sum((Alpha.1/factorial((2*K-1-d):0))*(Seq.1[j])^{(2*K-1-d):0})
}

H.d <- H.1[-l.1]
for(i in 2:nr){
Seq.i <- grid[pos[i]:pos[(i+1)]]
l.i <- length(Seq.i)
H.i <- Seq.i
       for(j in 1:l.i){
       H.i[j] <- sum((Alpha.1*(Seq.i[j])^{(2*K-1-d):0})/factorial((2*K-1-d):0))
+ sum(Alpha.2[1:(i-1)]*(Seq.i[j]-S[2:i])^{2*K-1-d}/factorial(2*K-1-d))
        }
    H.d <- c(H.d,H.i[-l.i])

}



Lastval <- sum(Alpha.1*C^{(2*K-1-d):0}/factorial((2*K-1-d):0) )
+ sum((Alpha.2*(C-S[2:nr])^{2*K-1-d})/factorial(2*K-1-d))
H.d <- c(H.d,Lastval)
}



}
H.d
```

```
}

InitialCondHk <- function(K,C,Yd.rev,IntBr,positions){

p <- length(positions)
Y.pos <- rep(0,p-2)
for(j in 2:(p-1)){
Y.pos[(j-1)] <- IntBr[K,positions[j]]
}

seq.K <- seq(K,2*K-2,2)
Y1 <- (factorial(K)/factorial(2*K-seq.K))*(-C)^{2*K - seq.K}
Y2 <- (factorial(K)/factorial(2*K-seq.K))*(C)^{2*K - seq.K}

Y <- c(Yd.rev[1,],Y1, Y.pos, Yd.rev[2,], Y2)
Y

}

ComputeSplineHk <- function(K,Y,S,Mat0){
p <- length(S)-2
#Mat <- matrix(0,nrow=2*K + p, ncol=2*K+p)
Mat <- Mat0[1:(2*K+p),1:(2*K+p)]
for(i in 1:K){
```

```
Mat[i,1:(2*K-2*(i-1))] <- (S[1])^{(2*K-1-2*(i-1)):0}
/factorial((2*K-1-2*(i-1)):0)
}
for(i in 2:(p+1)){
Mat[K+i-1,1:(2*K+ i-1)] <- c((S[i])^{(2*K-1):0}
/factorial((2*K-1):0),
   (S[i]-S[2:i])^{2*K-1}/factorial(2*K-1))
}
for(i in 1:K){
Mat[i+K+p,1:(2*K-2*(i-1))] <- (S[p+2])^{(2*K-1-2*(i-1)):0}
/factorial((2*K-1-2*(i-1)):0)
Mat[i+K+p,(2*K+1):(2*K+p)] <- (S[p+2]-S[2:(p+1)])^{2*K-1-2*(i-1)}
/factorial(2*K-1-2*(i-1))
}

rcond.Mat <- rcond.svd.Matrix(svd.Matrix(Mat))
Alpha <- solve.svd.Matrix(svd.Matrix(Mat),Y,tol=rcond.Mat*0.5)
Alpha

}

IndexFuncHk <- function(S0,S,Coef0,Coef){

C0 <- diff(Coef0)
C <- diff(Coef)
L0 <- length(S0)
L <- length(S)
S0 <- S0[-c(1,L0)]
S <- S[-c(1,L)]
```

```
S.merge <- c(S0,S)
           S.merge <- unique(sort(S.merge))
           C0.rep <- rep(0,length(S.merge))
C.rep <- rep(0,length(S.merge))

            for(i in 1:length(S.merge)){
   match.S0 <- match(S.merge[i],S0)
   if (!is.na(match.S0))
   C0.rep[i] <- C0[match.S0]
   else
   C0.rep[i] <- 0
    }

    for(i in 1:length(S.merge)){
  match.S <- match(S.merge[i],S)
  if (!is.na(match.S))
  C.rep[i] <- C[match.S]
  else
  C.rep[i] <- 0

     }



       Lambda <- NULL
for(i in 1:length(C.rep)){
   if(C.rep[i] < 0)
         Lambda <- c(Lambda,C0.rep[i]/(C0.rep[i]-C.rep[i]))
 if(C.rep[i] == 0)
         Lambda <- Lambda
         if(C.rep[i] > 0)
```

```
 Lambda <- c(Lambda,1)

             }

       lambda <- min(Lambda)
       index <- match(lambda,Lambda)
 index <- index +1
index

}
```

C.4 S codes for generating the processes $H_{c,k}, \cdots, H_{c,k}^{(2k-1)}$ when k is odd

Since many of the programs developed for k even can be used with some minor modifications, we include only the S functions that were specifically written for k odd.

```
StartingSplineHkOdd <- function(K,C,Yd){

if((K-2*floor(K/2))==0)
print("enter K odd")
else{
if(K==3){
Coef5 <- 0
Coef4 <- (factorial(3)/factorial(2))*C^2
Coef3 <- C^3-Coef4*C
Coef2 <- Yd[1,1] - ((Coef4/factorial(2))*C^2-Coef3*C)
Coef1 <- (Yd[2,2]-Yd[2,1])/(2*C) - (Coef3/factorial(3))*C^2
Coef0 <-  (Yd[2,2]+Yd[2,1])/2 - ((Coef4/factorial(4))*C^4
   + (Coef2/factorial(2))*C^2)
```

```
Coef.res <- c(Coef5,Coef4,Coef3,Coef2,Coef1,Coef0)

}

if(K > 3){

Seq <- seq(K+1,2*K-4,2)
Seq <- rev(Seq)
Coef <-  (factorial(K)/factorial(2))*C^2
for(i in 1:length(Seq)){
Seq.new <- seq(2,2*K-Seq[i],2)
Seq.new <- rev(Seq.new)
len <- length(Seq.new)
Coef <- c(Coef,(factorial(K)/factorial(Seq.new[1]))*C^{Seq.new[1]}
-sum((Coef*C^{Seq.new[2:len]})/factorial(Seq.new[2:len])))
}
Seq1.k <- seq(1,K-2,2)
Seq1.k <- rev(Seq1.k)
Coefk <- C^K -sum((Coef*C^{Seq1.k})/factorial(Seq1.k))
Seq2.k <- seq(2,K-1,2)
Seq2.k <- rev(Seq2.k)
Coefkm1 <- Yd[1,1]-sum((Coef*C^{Seq2.k})/factorial(Seq2.k))+Coefk*C
Coef <- c(Coef,Coefkm1)
Seq2 <- seq(K+3,2*K,2)
for(j in 1:length(Seq2)){
Seq.new <- seq(2,Seq2[j]-2,2)
Seq.new <- rev(Seq.new)
Coef <- c(Coef,(Yd[1,j+1] + Yd[2,j+1])/2
 -sum((Coef*C^{Seq.new})/factorial(Seq.new)))
}
```

```
Coef0 <- Coef
Seq3 <- seq(3,K,2)
Coef1 <- Coefk
for(j in 1:length(Seq3)){
Seq.new <- seq(3,Seq3[j],2)
Seq.new <- rev(Seq.new)
Coef1 <- c(Coef1,(Yd[2,j+1] - Yd[1,j+1])/(2*C)
- sum((Coef1*C^{Seq.new-1})/factorial(Seq.new)))
}

Coef.res <- rep(0,2*K)
Coef.res[seq(2,2*K,2)] <- Coef0
Coef.res[seq(K,2*K-1,2)] <- Coef1

}

}
Coef.res/factorial((2*K-1):0)


}



ComputeSplineHkOdd <- function(K,Y,S,Mat0){
p <- length(S)-2
Mat <- Mat0[1:(2*K+p),1:(2*K+p)]
for(i in 1:K){
Mat[i,1:(2*K-2*(i-1))] <- (S[1])^{(2*K-1-2*(i-1)):0}
/factorial((2*K-1-2*(i-1)):0)
}
```

```
for(i in 2:(p+1)){
Mat[K+i-1,1:(2*K+ i-1)] <- c((S[i])^{(2*K-1):0}
 /factorial((2*K-1):0),
(S[i]-S[2:i])^{2*K-1}/factorial(2*K-1))
}
for(i in 1:K){
 if(i == 1+(K-1)/2){
Mat[i+K+p,1:K] <- (S[p+2])^{(K-1):0}/factorial((K-1):0)
Mat[i+K+p,(2*K+1):(2*K+p)] <- (S[p+2]-S[2:(p+1)])^{K-1}/factorial(K-1)
}
else{
Mat[i+K+p,1:(2*K-2*(i-1))] <- (S[p+2])^{(2*K-1-2*(i-1)):0}
/factorial((2*K-1-2*(i-1)):0)
Mat[i+K+p,(2*K+1):(2*K+p)] <- (S[p+2]-S[2:(p+1)])^{2*K-1-2*(i-1)}
/factorial(2*K-1-2*(i-1))
}
}


rcond.Mat <- rcond.svd.Matrix(svd.Matrix(Mat))
Alpha <- solve.svd.Matrix(svd.Matrix(Mat),Y,tol=rcond.Mat*0.5)
Alpha

}



InitialCondHkOdd <- function(K,C,Yd.rev,IntBr,positions){



p <- length(positions)
```

```
Y.pos <- rep(0,p-2)
for(j in 2:(p-1)){
Y.pos[(j-1)] <- IntBr[K,positions[j]]
}

Seq.K <- seq(2,K-1,2)
Seq.K <- rev(Seq.K)
l.K <- length(seq(1,K,2))
Y1 <- (factorial(K)/factorial(Seq.K))*(-C)^{Seq.K}
Y2 <- (factorial(K)/factorial(Seq.K))*(C)^{Seq.K}
Y <- c(Yd.rev[1,],Y1, Y.pos, Yd.rev[2,-l.K],C^K, Y2)
Y


}
```

C.5 *S* codes for calculating the MLE of a k-montone density

```
SuppReducAlgoMLE <- function(K,X,prec,eps,p1,p2){
n <- length(X)
#grid <- round(seq(min(X),theta0,by = prec),digits=6)
theta0 <- nlminb(start=max(X)+0.1,objective=minusloglik,
K=K,X=X,lower=max(X)+0.0001)$parameters
grid <- round(seq(p1*min(X),p2*K*max(X),by = prec),digits=6)
Mat0 <- matrix(0,nrow=n,ncol=20)
Vec0 <- rep(n,20)
print(theta0)
Cbar <- 1
Sbar <- theta0
```

```
Matfbar <- EvaluateMatf(Sbar,K=K,X=X,Mat0)
valfbar <- matrix(0,nrow=length(Sbar),ncol=n)
 if(length(Sbar)==1){
  valfbar <- Matfbar
  }
   else{
valfbar <- apply(Matfbar%*%diag(Cbar),1,sum)
     }
valfbar <- as.vector(valfbar)
ResminOuter <- FindMinimMLE(valfbar,valfbar,K,X,prec,p1,p2,grid)
valminOuter <- ResminOuter[2]
#rm(ResminOuter)
CountOuter <- 0
while(valminOuter < - eps){
CountOuter <- CountOuter +1
cat("Main Outerloup numb = ",CountOuter,"\n")

#Problems can occur since fbar is not necessarily to the solution
# of the LS problem.
#Therefore, we need to apply again the support reduction step.

print(rbind(Sbar,Cbar))
C <- CalculateOptMLE(valfbar,S=Sbar,K=K,X=X,Vec0,Mat0)
C <- as.vector(C)
S <- Sbar
min.C <- min(C)
if(length(Sbar)==1 & min.C < 0)
print("Sbar is of length 1 and min(C) < 0 !")
l.Sbar <- length(Sbar)
l.Cbar <- length(Cbar)
```

```
 while(min.C < 0){
 index <- IndexFuncMLE(S0=Sbar,S,C0=Cbar,C)
 S <- S[-index]
 C <- CalculateOptMLE(valfbar,S=S,K=K,X=X,Vec0,Mat0)
 C <- as.vector(C)
 min.C <- min(C)

}# while(min.C < 0)

Matg <- EvaluateMatf(S,K=K,X=X,Mat0)
valg <- matrix(0,nrow=length(S),ncol=n)
if(length(S)==1){
valg <- Matg
}
 else{
valg <- apply(Matg%*%diag(C),1,sum)
   }
valg <- as.vector(valg)

ResminInner <- FindMinimMLE(valfbar,valg,K,X,prec,p1,p2,grid)
thetaminInner <- ResminInner[1]
print(thetaminInner)
valminInner <- ResminInner[2]
l.S <- length(S)
l.C <- length(C)
print(valminInner)
CountInner <- 0
 while(valminInner < - eps*10){
  countInner <- CountInner + 1
 cat("MainInnerLoup numb = ",CountInner,"of MainOuterLoup numb=",
```

```
CountOuter,"\n")
 thetaminInner <- ResminInner[1]
print(c(thetaminInner,valminInner))
S0 <- S
C0 <- C
S <- c(S,thetaminInner)
S <- sort(S)
C <- CalculateOptMLE(valfbar,S=S,K=K,X=X,Vec0,Mat0)
C <- as.vector(C)
min.C <- min(C)
countInner <- 0
  while(min.C < 0){
countInner <- countInner +1
cat("SubInnerLoup numb = ",countInner,"of the MainInnerLoup numb = ",
    CountInner, "\n")
index <- IndexFuncMLE(S0=S0,S,C0=C0,C)
S <- S[-index]
C <- CalculateOptMLE(valfbar,S=S,K=K,X=X,Vec0,Mat0)
C <- as.vector(C)
min.C <- min(C)
}# while(min.C < 0)
Matg <- EvaluateMatf(S,K=K,X=X,Mat0)
valg <- matrix(0,nrow=length(S),ncol=n)
if(length(S)==1){
valg <- Matg
}
else{
valg <- apply(Matg%*%diag(C),1,sum)
  }
valg <- as.vector(valg)
```

```
ResminInner <- FindMinimMLE(valfbar,valg,K,X,prec,p1,p2,grid)
valminInner <- ResminInner[2]
valminInner
} #while(valminInner < -eps*10)
#Here we need to ensure monotonicity of the algorithm
l.S <- length(S)
l.C <- length(C)
ind <- 0
max.S <- 1
max.C <- 1
ind <- 0
if((l.C==l.Cbar) & (l.S==l.Sbar)){
max.S <- max(abs(S-Sbar))
max.C <- max(abs(C-Cbar))
cat("max.S = ", max.S,"max.C=",max.C,"\n")
if(max.S ==0 & max.C == 0)
ind <- 1
}
if(ind ==1)
break
else{
 likbar <- LoglikFunc(valfbar,Cbar)
 print(likbar)
 Sq <- S
 Cq <- C
 Merge.out <- MergeFunc(S0=Sbar,C0=Cbar,S=Sq,C=Cq)
 S.m <- Merge.out[1,]
 Cbar.m <- Merge.out[2,]
 Cq.m <- Merge.out[3,]
 Cbar <- as.vector(Cbar.m)
```

```
 Cq.m <- as.vector(Cq.m)
 Mat.m <- EvaluateMatf(S.m,K=K,X=X,Mat0)
 valfq.m <- apply(Mat.m%*%diag(Cq.m),1,sum)
 valfq.m <- as.vector(valfq.m)
 valfbar.m <- apply(Mat.m%*%diag(Cbar.m),1,sum)
 valfbar.m <- as.vector(valfbar.m)
 cat("diff in loglik =",likbar - LoglikFunc(valfq.m,Cq.m),"\n")
likfq <- LoglikFunc(valfq.m,Cq.m)
if(abs(likbar-likfq) <= eps*0.1)
break
else{
res.arj <- Armijo(Cq.m,Cbar.m,valfbar.m,valfq.m,likbar,K=K,X=X)
if(res.arj[2] >= 3000)
lam.arj <- 0
else
lam.arj <- res.arj[1]
cat("lambda=",lam.arj,"counts=",res.arj[2],"\n")
 #Here, we obtain the new iterate fbar
Sbar <- S.m
Cbar <- (1-lam.arj)*Cbar.m + lam.arj*Cq.m
Cbar <- as.vector(Cbar)
#print(rbind(Sbar,Cbar))
f.bar <- cbind(Cbar,Sbar)
f.bar <- as.data.frame(f.bar)
names(f.bar) <- c("w","s")
Cbar <- f.bar$w[f.bar$w !=0]
Sbar <- f.bar$s[f.bar$w !=0]
print(rbind(Sbar,Cbar))
Matfbar <- EvaluateMatf(Sbar,K=K,X=X,Mat0)
valfbar <- apply(Matfbar%*%diag(Cbar),1,sum)
```

```
valfbar <- as.vector(valfbar)
ResminOuter <- FindMinimMLE(valfbar,valfbar,K,X,prec,p1,p2,grid)
valminOuter <- ResminOuter[2]
cat("valminOuter", valminOuter, "\n")
        }
  }
    }# while(valminOuter < -eps)

       Output <- cbind(Sbar,Cbar)
       Output



}



##This function calculates f_{theta_i}(Xj) where Xj
##is a data point and theta_i is a support point of the iterate f.
##and hence it retruns a matrix of dimension n = length(X) x m = length(S).

EvaluateMatf <- function(S,K,X, Mat0){

S <- sort(S)
m <- length(S)
n <- length(X)
#Xs <- sort(X)
#matrix(0,nrow=n,ncol=m)
if(m==1){
Matf <- matrix(0,nrow=n,ncol=1)
}
else{
```

```
Matf <- Mat0[1:n,1:m]
}
for(i in 1:n){
Matf[i,] <- (K/S^{K})*ifelse(S >= X[i], (S-X[i])^{K-1},0)

}
Matf

}

#This function finds the minimum of the directional
#derivative for the ML estimation inside the quadratic
# approximation
# of - loglikelihood if we "move" away from the current iterate
#c_1*f_theta1 +...+ c_m*f_thetam.

FindMinimMLE <- function(valfbar,valg,K,X,prec,p1,p2,grid){
#grid <- round(seq(p1*min(X),p2*K*max(X),by = prec),digits=6)
#grid <- round(seq(min(X),theta0,by = prec),digits=6)
l.g  <- length(grid)
DirecDer.vec <- grid
for(i in 1:l.g){
#print(i)
DirecDer.vec[i] <- DirecDerMLE(grid[i],valfbar,valg,K=K,X=X)
}
minval <- min(DirecDer.vec)
min.rank <- min(rank(DirecDer.vec))
index <- match(min.rank,rank(DirecDer.vec))
#print(cbind(DirecDer.vec,rank(DirecDer.vec)))
#cat("index",index,"\n")
```

```
thetamin <- grid[index]
c(thetamin,minval)

}



# This function calculates the directional derivative
#of the quadratic approximation of -loglikelihood
#at some point theta.
#Sbar and Cbar are respectively the set of support points
# and the weights of the current iterate fbar (outside the quadratic
#approximation of -loglikelihood).
# valfbar, valg are respectively the vectors storing
#[fbar(X_(1)),...fbar(X_(n))] and [g(X_(1)),...g(X_(n))]

DirecDerMLE <- function(theta,valfbar,valg,K,X){
C1 <- NULL
C2 <- NULL
#Xs <- sort(X)
n <- length(X)
Vec.theta <- (K/(theta)^K)*ifelse(theta >= X,(theta-X)^{K-1},0)

C1 <- 1- 2*mean(Vec.theta/valfbar) + mean(valg*Vec.theta/valfbar^2)
C2 <- mean((Vec.theta/valfbar)^2)
DirecDer <- C1/sqrt(C2)
DirecDer

}
#This function solves a linear system
#in order to find the minimizer of -loglikelood
```

```
##over a cone generated by a few active vertices.

CalculateOptMLE <- function(valfbar,S,K,X,Vec0,Mat0){

m <- length(S)
n <- length(X)
nm <- Vec0[1:m]
#rep(n,m)
valfbar <- as.vector(valfbar)
valfbar.inv <- 1/valfbar
Dfbar <- diag(valfbar.inv)
MatY <- EvaluateMatf(S=S,K=K,X=X,Mat0)
MatV <- t(Dfbar%*%MatY)%*%(Dfbar%*%MatY)
B <- 2*(t(MatY)%*%valfbar.inv)-nm
#Alpha <- solve.Matrix(MatV,B,tol=rcond.V*0.1)
#Alpha <- solve.Hermitian(MatV,B,tol=0)
rcond.V <-  rcond.svd.Matrix(svd.Matrix(MatV))
cat("rcond=", rcond.V, "\n")
Alpha <- solve.svd.Matrix(svd.Matrix(MatV),B,rcond.V*0.1)

Alpha
}

#This function calculates -loglikelihood at a current iterate
# with set of support points=S and set of weights = C
#valf is a vector storing the values [f(X_(1)),...,f(X_(n))].
LoglikFunc <- function(valf,C){
```

```
Loglik <- -mean(log(valf)) + sum(C)
Loglik
}

MergeFunc <- function(S0=Sbar,C0=Cbar,S=Sq,C=Cq){



S.merge <- c(S0,S)
S.merge <- unique(sort(S.merge))
C0.rep <- rep(0,length(S.merge))
C.rep <- rep(0,length(S.merge))

for(i in 1:length(S.merge)){
match.S0 <- match(S.merge[i],S0)
if (!is.na(match.S0))
C0.rep[i] <- C0[match.S0]
else
C0.rep[i] <- 0

}
for(i in 1:length(S.merge)){
match.S <- match(S.merge[i],S)
if (!is.na(match.S))
C.rep[i] <- C[match.S]
else
C.rep[i] <- 0

}
rbind(S.merge,C0.rep,C.rep)
}
```

```
# This function looks for a lambda between 0 and 1 such
# that fbar + lambda*(fq-fbar) has a larger likelihood than that
#of fbar in order to ensure the monotonicity of the algorithm.
#Cbar is the vector weights of fbar
#(outside the quadratic approximation).
# Cq is the vector weights of fq the minimizer of the
quadratic approximation of -loglikelihood.
#likbar is -loglikelihood of fbar.
# we need to make some arrangements in order to be able use
# the function "LoglikFunc" as it is coded.

Armijo <- function(Cq,Cbar,valfbar,valfq,likbar,K=K,X=X){
lambda <- 1
sumfq <- sum(Cq)
sumfbar <- sum(Cbar)
likq <- LoglikFunc(valfq,Cq)
likfnew <- likq
#if(likfnew == likbar)
#lambda <- 1
count <- 0
while( likfnew >= likbar & count <= 2000){
count <- count +1
lambda <- lambda/2
valfnew <- valfbar + lambda *(valfq - valfbar)
Cfnew  <- Cbar + lambda *(Cq - Cbar)
likfnew <- LoglikFunc(valfnew,Cfnew)
}

lambda
```

```
}
```

C.6 S codes for calculating the LSE of a k-monotone density

```
LSESupReducAlgo <- function(K=3,X=X1000,prec=0.01,eps= 10^{-8},p1=1,p2=1){

#theta0 <- (2*K-1)*max(X)
grid <- round(seq(min(X)*p1,p2*K*max(X),prec),digits=6)
M.alpha <- matrix(0,nrow=K-1,ncol=K-1)
M0 <- matrix(0,nrow=30,30)
B0 <- rep(0,30)
#grid <- round(seq(min(X),2*K*max(X),prec),digits=6)
Rank <- rank(c(max(X),grid))[1]
    theta0 <- grid[Rank]
#theta0 <- grid[length(grid)]
    print(theta0)
C0 <- ((2*K-1)/(K*theta0^{K-1}))*mean((theta0-X)^{K-1})
#print(C0)
S0 <- theta0
Resmin <- FindMinFunc(X=X, S=S0,C=C0,K=K,prec=prec,grid)
valmin <- Resmin[2]
print(valmin)
Count <- 0
while(valmin < -eps){
Count <- Count + 1
cat("Main loup numb = ",Count,"\n")
```

```
thetamin <- Resmin[1]
print(c(thetamin,valmin))
  S <- c(S0,thetamin)
S <- sort(S)
B <- LSEInitialCond(S=S,K=K,X=X,B0)
C <- LSEComputeSpline(S=S,K=K,B=B,M.alpha,M0)
C <- ((-1)^K * S^K * factorial((2*K-1))/factorial(K))*C
print(S)
print(C)
min.C <- min(C)
count <- 0
   while(min.C < 0){
count <- count+1
cat("Sub loup numb = ",count," of the main loop numb=", Count, "\n")
    index <- IndexFunc(S0=S0,S=S,C0=C0,C=C)
S <- S[-index]
if(length(S)==1)
C <- ((2*K-1)/(K*S^{K-1}))*mean((S-X[X <= S])^{K-1})
else{
B <- LSEInitialCond(S=S,K=K,X=X,B0)
C <- LSEComputeSpline(S=S,K=K,B=B,M.alpha,M0)
       C <- ((-1)^K * S^K * factorial((2*K-1))/factorial(K))*C
}
min.C <- min(C)


}# while(min.C < 0)
S0 <- S
           C0 <- C
Resmin <- FindMinFunc(X=X, S=S0,C=C0,K=K,prec=prec,grid)
```

```
valmin <- Resmin[2]


    }# while(valmin < -eps)


       Output <- cbind(S0,C0)
       Output
}

#This function finds the minimum of
#the directional derivative if we "move" away
# from the current iterarte c_1*f_theta1 +...+ c_m*f_thetam.

FindMinFunc <- function(X,S,C,K,prec,grid){
l.g  <- length(grid)
DirecDer.vec <- grid
for(i in 1:l.g){
#print(i)
DirecDer.vec[i] <- DirecDer(grid[i],X,S,C,K)
}
minval <- min(DirecDer.vec)
index <- match(1,rank(DirecDer.vec))
thetamin <- grid[index]
#free(DirecDer.vec)
c(thetamin,minval)

}

#This function calculates the directional
# derivative for the LS criterion.
```

```
# X is an i.i.d. sample of size n generated
from a K-monotone density.
# Theta is the set of knots theta_1,...,theta_m.
# C is the vector of the weights C_1,...,C_m
#corresponding to f_{theta1},...f_{theta2}
#DirecDer <- function(theta,X,S,C,K){

Out <- NULL
J <- 0
for(i in 1:length(S)){
J <- J + C[i]*J.Func(theta,S[i],K)
}

Out <- (1/theta^{K-1/2})*(J-Integr.Fn(theta=theta,K=K,X=X))

Out

}



#This function calculates the (K-1)-fold integral of the function
#f_thetaj(x) = (K/(thetaj)^K)*(thetaj-x)_{+}^{K-1}.

J.Func <- function(theta,thetaj,K){
Out <- NULL
 if(theta <= thetaj){
  Out <- (factorial(K-1)/factorial(2*K-1))*(-1)^{K-1}
* sum(choose(2*K-1,0:(K-1))*(-1)^{0:(K-1)}*thetaj^{2*K-1-(0:(K-1))}
 *theta^{0:(K-1)})
+ (-1)^{K}*(factorial(K-1)/factorial(2*K-1))*(thetaj-theta)^{2*K-1}
```

```
 }
 else
 Out <- (factorial(K-1)/factorial(2*K-1))*(-1)^{K-1}
* sum(choose(2*K-1,0:(K-1))*(-1)^{0:(K-1)}*theta^{2*K-1-(0:(K-1))}
*thetaj^{0:(K-1)})
+(-1)^{K}*(factorial(K-1)/factorial(2*K-1))*(theta - thetaj)^{2*K-1}

 Out <- (K/thetaj^{K})*Out
 Out

}

#This function calculates the (K-1)fold integral
#of the empirical distribution.
Integr.Fn <- function(theta,K,X){

X.s <- sort(X)
n <- length(X)
rank <- rank(c(theta,X.s))
if(rank[1] ==1)
Output <- 0
else
Output <- (1/factorial(K-1))*(1/n)
*sum((theta-X.s[1:(rank[1]-1)])^{K-1})
Output

}

LSEInitialCond <- function(K,S,X,B0){
```

```
m <- length(S)
S0 <- c(0,S)
#B <- rep(0,m)
B <- B0[1:m]
for(i in 1:m){
B[i] <- Integr.Fn(S0[i],K,X)-Integr.Fn(S0[m+1],K,X)
}

B
}

LSEComputeSpline <- function(S,K,B,M.alpha,M0){
m <- length(S)
S0 <- c(0,S)
#M.alpha <- matrix(0,nrow=K-1,ncol=K-1)
for(i in 1:(K-1)){
M.alpha[i,i:(K-1)] <- choose(i:(K-1),i)*(S[m])^{0:(K-i-1)}
}
M.alpha <- matrix(M.alpha,K-1,K-1)
#M.2 <- matrix(0,nrow=K-1,ncol=m)
M.2 <- M0[1:(K-1),1:m]
M.2 <- matrix(M.2,K-1,m)

for(i in 1:(K-1)){
M.2[i,] <- choose(2*K-1,i)*S^{2*K-1-i}
}
#M.1 <- matrix(0,nrow=m,ncol=K-1)
M.1 <- M0[1:m,1:(K-1)]
M.1 <- matrix(M.1,m,K-1)
```

```
for(j in 1:(K-1)){
M.1[,j] <- (S[m]-S0[1:m])^{j}
}
#M.3 <- matrix(0,nrow=m,ncol=m)
M.3 <- M0[1:m,1:m]
M.3 <- matrix(M.3,m,m)

for(i in 1:m){
M.3[i,i:m] <- (S0[(i+1):(m+1)]-S0[i])^{2*K-1}
}
M.alpha.inv <- solve.UpperTriangular(M.alpha)
Mat <- -M.1%*%M.alpha.inv%*%M.2 + M.3
rcond.Mat <- rcond.svd.Matrix(svd.Matrix(Mat))
#print(rcond.Mat)
Res <- solve.svd.Matrix(svd.Matrix(Mat),B,tol=rcond.Mat*0.5)
Res <- as.numeric(Res)
Res
}
```

VITA

Fadoua Balabdaoui was born on October 13, 1975, in Rabat, Morocco. In July 1999 she received a Diplôme d'Ingénieur Civil from the École Nationale Supérieure des Mines de Paris, where she specialized in Geostatistics. From the fall of 1999 until the summer of 2000, she was at the University of Washington working as a visiting scientist at the Center for Studies in Demography and Ecology and the Department of Statistics. In September 2000 she joined the Department of Statistics at the University of Washington in a pursuit of a Ph.D in Statistics, which she received in June 2004.