CONSISTENT ESTIMATION OF A CONVEX DENSITY AT THE ORIGIN [1]

Fadoua Balabdaoui[2]

,

Centre de Recherche en Mathématiques de la Décision
Université Paris-Dauphine, Paris, France

Abstract: Motivated by Hampel's birds migration problem, Groeneboom, Jongbloed, and Wellner (2001b) established the asymptotic distribution theory for the nonparametric Least Squares and Maximum Likelihood estimators of a convex and decreasing density, g_0, at a fixed point $t_0 > 0$. However, estimation of the distribution function of the birds' resting times involves estimation of g_0' at 0, a boundary point at which the estimators are not consistent.

In this paper, we focus on the Least Squares estimator, $\tilde{g}_n$. Our goal is to show that consistent estimators of both $g_0(0)$ and $g_0'(0)$ can be based solely on $\tilde{g}_n$. Following the idea of Kulikov and Lopuhaä (2006) in monotone estimation, we show that it suffices to take $\tilde{g}_n(n^{-\alpha})$ and $\tilde{g}_n'(n^{-\alpha})$, with $\alpha \in (0, 1/3)$. We establish their joint asymptotic distributions and show that $\alpha = 1/5$ should be taken as it yields the fastest rates of convergence.

Key words and phrases: Asymptotic distribution, Brownian motion, Convex density, Hampel's birds problem, Inconsistency at the boundaries, Least Squares estimation.

1 Introduction

Suppose that we are interested in estimating the distribution function of resting periods of migrating birds in a certain habitat area. We denote the true distribution by F_0. The birds are marked individually so that they can be identified, and some of them are captured in mist nets with a certain probability.

Hampel (1987) assumed that the birds are caught according to a Poisson process with a rate λ small enough that the probability of catching the same bird three times or more is negligible, and that F_0 has a density, f_0, such that

[1]This work was completed while the author was a postdoctoral fellow at the Institute of Mathematical Stochastics, Göttingen, Germany.

[2]*E-mail address:* fadoua@ceremade.dauphine.fr

$\lim_{t\to\infty} f_0(t) = 0$. Those birds that are caught exactly twice yield a distribution of observed (minimum) resting periods.

The question is: What is the relationship between the observed resting period distribution to the true one? Assuming that f_0 admits a finite second moment, and if g_0 is the density of the observed spans between two capture dates, Hampel (1987) showed that this relationship is given by

$$g_0(s) = \frac{2}{\mu_2} \int_0^\infty (t-s)_+ f_0(t)dt, \quad s > 0 \tag{1.1}$$

where $\mu_2 = \int_0^\infty t^2 f_0(t)dt$. Inverting (1.1) yields at any continuity point $t > 0$ of F_0

$$F_0(t) = 1 - \frac{g_0'(t)}{g_0'(0)},$$

where $g_0'(0)$ denotes here the right derivative of g_0 at 0. See also Anevski (2003) for an explicit derivation.

Groeneboom, Jongbloed, and Wellner (2001b) considered two nonparametric estimators of g_0: The Least Squares and Maximum Likelihood estimators, denoted thereafter by LSE and MLE. Under the assumption that g_0'' is continuous in the neighborhood of a fixed point t_0 with $g_0''(t_0) \neq 0$, they showed that these estimators are piecewise linear, and

$$\begin{pmatrix} n^{\frac{2}{5}}\left(\bar{g}_n(t_0) - g_0(t_0)\right) \\ n^{\frac{1}{5}}\left(\bar{g}_n'(t_0) - g_0'(t_0)\right) \end{pmatrix} \to_d \begin{pmatrix} c_0(t_0)\ H^{(2)}(0) \\ c_1(t_0)\ H^{(3)}(0) \end{pmatrix} \tag{1.2}$$

where $\bar{g}_n$ denotes either one of the estimators, H is the "invelope" of the process

$$Y(t) = \begin{cases} \int_0^t W(s)ds + t^4, & t \geq 0 \\ \int_t^0 W(s)ds + t^4 & t < 0 \end{cases} \tag{1.3}$$

in the following sense: H is an almost surely unique random process satisfying:

(i) $H^{(2)}$ is convex (H is 4-convex) on $\mathbb{R}$

(ii) $H(t) \geq Y(t),\ t \in \mathbb{R}$

(iii) $H(t) = Y(t)$, if $H^{(2)}$ changes slope at t. In view of (ii), the condition (iii) can be also rewritten as

$$\int_{-\infty}^{\infty} (H(t) - Y(t))dH^{(3)}(t) = 0$$

(see GROENEBOOM, JONGBLOED, AND WELLNER (2001A)). Here, W is a standard two-sided Brownian motion starting at 0. The constants $c_0(t_0)$ and $c_1(t_0)$ are given by

$$c_0(t_0) = \left(\frac{1}{24}(g_0(t_0))^2 g_0''(t_0)\right)^{\frac{1}{5}}, c_1(t_0) = \left(\frac{1}{24^3}g_0(t_0)(g_0''(t_0))^3\right)^{\frac{1}{5}}. \tag{1.4}$$

The asymptotic result in (1.2) does not apply anymore when t_0 is replaced by 0. To study the behavior of the estimators near this boundary point, we focus in this paper on the LSE. The MLE can be handled very similarly, but it involves much more cumbersome calculations, due mainly to the nonlinear nature of the characterization of this estimator.

The LSE enjoys the property of preserving the shape restrictions of the estimated density g_0, and any consistent estimator of $g_0'(0)$ will yield an estimator of F_0, that is a genuine distribution function ($-\tilde{g}_n'$ is nondecreasing) and pointwise consistent if g_0 is differentiable on $(0, \infty)$. This follows from Lemma 3.1 of GROENEBOOM, JONGBLOED, AND WELLNER (2001B).

To achieve consistency at 0, methods based on boundary corrections of kernel density estimators have been suggested (see e.g. JONES (1993) and the references therein). On the other hand, CHENG (1997) showed how local linear density estimators adjust automatically for boundary effects without requiring any further correction. Consistent estimation of higher derivatives can be achieved using local polynomial density estimators as it was shown by CHENG, FAN AND MARRON (1997). For estimating $g_0'(0)$, where g_0 is a compactly supported density on $[0, 1]$ assumed to be twice differentiable, their Theorem 1 implies that $n^{-1/5}$ is the optimal rate of convergence, and is achieved by their local quadratic smoother provided that the kernel function and its derivative are bounded on $[0, 1]$.

In this paper, the goal is to answer the following question: Would it be possible, using *only* the already computed LSE, to construct a consistent estimator of $g_0'(0)$? This is possible by following the same idea of KULIKOV AND LOPUHAÄ (2006) of "staying away of 0" which they used to achieve consistency at 0 in monotone estimation. More precisely, we show that for $\alpha \in (0, 1/3)$, $\tilde{g}_n(n^{-\alpha})$ and $\tilde{g}_n'(n^{-\alpha})$ are consistent estimators of 0, and that $\alpha = 1/5$ yields the fastest rates of convergence.

The paper will be structured as follows: We recall in the next section that the LSE fails to be consistent at 0, and show that its first derivative is not even bounded in probability. In Section 3, we review the characterization of the LSE, give preliminary results that we will use in the main proof and explain via heuristic arguments where the main difficulty lies. In Section 4, we present the three joint asymptotic regimes of $\tilde{g}_n(n^{-\alpha})$ and $\tilde{g}_n'(n^{-\alpha})$ depending on whether $\alpha \in (0, 1/5)$, $\alpha = 1/5$ or $\alpha \in (1/5, 1/3)$. The optimal rates of convergence, $n^{-2/5}$ and $n^{-1/5}$, are attained by our *modified LSE estimator*, which is obtained by replacing the original LSE, on the shrinking neighborhood $[0, n^{-1/5}]$, by tangential extrapolation from the point $n^{-1/5}$ to the left. To illustrate the theory, we present in Section 5 numerical results based on simulations, and finish off with conclusions and some open questions. The proof of our main result is given in Section 6. As the cases $\alpha = 1/5$ and $\alpha \in (1/5, 1/3)$ share many similarities in the proof, we show the result only for $\alpha = 1/5$ and indicate at the end of Section 6 how the proof can be readapted in the other case. We also give heuristic arguments for $\alpha \in (0, 1/5)$, for which the limiting distribution depends on the same process in the estimation problem considered by GROENEBOOM, JONGBLOED, AND WELLNER (2001B) at an interior point.

Throughout the paper, we assume that the true convex density g_0 is twice continuously differentiable at 0, and that $g_0''(0) > 0$.

2 Inconsistency of the Least Squares estimator

Given $X_1, \cdots, X_n$ n i.i.d. observations from a convex and nonincreasing density g_0 and $\mathbb{G}_n$ their corresponding empirical distribution, the LSE $\tilde{g}_n$ is defined as the unique minimizer of the quadratic criterion

$$\Phi_n(g) = \frac{1}{2} \int_0^\infty g^2(t)dt - \int_0^\infty g(t)d\mathbb{G}_n(t) \tag{2.5}$$

over the class of square integrable convex and nonincreasing functions on $(0, \infty)$.

GROENEBOOM, JONGBLOED, AND WELLNER (2001B) established that the LSE exists, is piecewise linear, and that it is a density although the minimization is not restricted to the space of densities. Furthermore, if $\tilde{H}_n(t) = \int_0^t (t-s)\tilde{g}_n(s)ds$ and $\mathbb{Y}_n(t) = \int_0^t (t-s)d\mathbb{G}_n(s)$, then the nonincreasing and convex piecewise linear function $\tilde{g}_n$ is the LSE of g_0 if and only if

$$\tilde{H}_n(t) \begin{cases} \geq \mathbb{Y}_n(t), & \text{for all } t > 0 \\ = \mathbb{Y}_n(t), & \text{if } t \text{ is a jump point of } \tilde{g}_n'. \end{cases} \tag{2.6}$$

By the inequality condition in (2.6), any jump location τ is a point where $\tilde{H}_n - \mathbb{Y}_n$ reaches its minimum. Therefore, the derivatives of $\tilde{H}_n$ and $\mathbb{Y}_n$ at τ have to be equal, that is

$$\tau \text{ is a jump point} \Rightarrow \tilde{H}_n(\tau) = \mathbb{Y}_n(\tau) \text{ and } \tilde{G}_n(\tau) = \mathbb{G}_n(\tau) \tag{2.7}$$

where $\tilde{G}_n$ is the cumulative distribution function corresponding to $\tilde{g}_n$.

We show now that $\tilde{g}_n'(0)$ is not a consistent estimator of $g_0'(0)$. This is a consequence of the following proposition.

Proposition 2.1 *$\tilde{g}_n'(0)$ is not bounded in probability.*

To prove Proposition 2.1, we recall the following result due to Groeneboom, Jongbloed, and Wellner (2001b) (see page 1673).

Lemma 2.1 *If Y_1 and Y_2 are two independent standard Exponentials, then*

$$\liminf_{n\to\infty} P(\tilde{g}_n(0) \geq 2g_0(0)) \geq P\left(\frac{Y_1}{(Y_1+Y_2)^2} \geq 1\right) > 0$$

Hence, $\tilde{g}_n(0)$ is not consistent.

Proof of Proposition 2.1: By convexity of $\tilde{g}_n$, we have $\tilde{g}_n(0) \leq \tilde{g}_n(t) - t\tilde{g}_n'(0) = \tilde{g}_n(t) + t|\tilde{g}_n'(0)|$ for all $t > 0$. Hence,

$$P\left(\frac{\tilde{g}_n(0)}{g_0(0)} \geq 2\right) \leq P\left(\frac{\tilde{g}_n(t)}{g_0(0)} \geq \frac{3}{2}\right) + P\left(\frac{|\tilde{g}_n'(0)|}{g_0(0)} \geq \frac{1}{2t}\right).$$

By consistency of $\tilde{g}_n(t)$, the middle term converges to 0 as $n \to \infty$. Using the previous inequality and Lemma 2.1, we conclude that for all $M > 0$

$$\liminf_{n\to\infty} P\left(|\tilde{g}_n'(0)| \geq M\right) > 0$$

and that $\tilde{g}_n'(0)$ is unbounded in probability. $\square$

3 Local behavior of $\tilde{g}_n$

3.1 Local perturbation functions

The characterization of $\tilde{g}_n$ given in (2.6) follows from taking the directional derivative of the quadratic functional Φ_n defined in (2.5) in the direction of the perturbation function $s \mapsto (t-s)_+$, where z_+ is the usual notation of the positive part of a z. To make sure that the resulting function $\tilde{g}_n(s) + \epsilon(t-s)_+$ for small ϵ belongs to class of convex functions, ϵ is only allowed to take positive values when t is not necessarily a knot of $\tilde{g}_n$. This leads to the inequality part of the characterization in (2.6). When t is a knot, then small negative values of ϵ are allowed and the directional derivative is equal to 0. This in turn yields the equality condition in (2.6).

The knots of $\tilde{g}_n$, or equivalently the jump points of $\tilde{g}_n'$, are defined only implicitly in the sense that there exists no explicit formula which pins down their locations exactly. However, these knots can be obtained numerically as limits of an iterative algorithmic procedure (see Groeneboom, Jongbloed, and Wellner (2003) and also Mammen and van de Geer (1997) in the context of nonparametric regression). To establish the limiting theory, it is necessary to exploit the characterization of the LSE and make use of appropriate perturbation functions to be able to track down the asymptotic behavior of these knots and hence that of the estimator.

Consider $\tau_1 < \tau_2$ to be two jump points of $\tilde{g}_n'$. Here we omit the subscript n baring in mind dependence of these points on the sample size. We can always write

$$\tilde{H}_n(\tau_2) = \tilde{H}_n(\tau_1) + (\tau_2 - \tau_1)\tilde{H}_n'(\tau_1) + \int_{\tau_1}^{\tau_2} (\tau_2 - t)\tilde{g}_n(t)dt.$$

Using (2.7), this yields

$$\int_{\tau_1}^{\tau_2} (\tau_2 - t)\tilde{g}_n(t)dt = \mathbb{Y}_n(\tau_2) - \mathbb{Y}_n(\tau_1) - (\tau_2 - \tau_1)\mathbb{G}_n(\tau_1) = \int_{\tau_1}^{\tau_2} (\tau_2 - t)d\mathbb{G}_n(t),$$

or equivalently

$$\boldsymbol{\int_{\tau_1}^{\tau_2} (\tau_2 - t)(\tilde{g}_n(t) - g_0(t))dt = \int_{\tau_1}^{\tau_2} (\tau_2 - t)d(\mathbb{G}_n(t) - G_0(t)).} \tag{3.8}$$

On the other hand, since our estimation problem is local, most interesting perturbation functions are those that vanish except on an interval of the form $[\tau_1, \tau_2]$. Such a function, p say, must satisfy that $\tilde{g}_n + \epsilon p$ is convex and nonincreasing for $\epsilon > 0$ small enough. As $\epsilon \searrow 0$, nonnegativity of the corresponding directional derivative implies that

$$\int_{\tau_1}^{\tau_2} p(t)(\tilde{g}_n(t) - g_0(t))dt \geq \int_{\tau_1}^{\tau_2} p(t)d(\mathbb{G}_n(t) - G_0(t)) \tag{3.9}$$

and this inequality remains valid if we add an *arbitrary* constant to p, using again the fact that $\tilde{G}_n(\tau_j) = \mathbb{G}_n(\tau_j), j = 1, 2$.

The choice of perturbation functions in this problem is actually limited, and the simplest ones to take are of the form of a "reversed hat"; i.e., a convex triangular function on a compact set. Let $s \in (\tau_1, \tau_2)$ and $\bar{\tau}$ be the mid-point of $[\tau_1, \tau_2]$. Consider the following choices of local perturbation functions

$$p_{1,s}(t) = \frac{\tau_1 - t}{s - \tau_1} 1_{[\tau_1, s]}(t) + \frac{t - \tau_2}{\tau_2 - s} 1_{(s, \tau_2]}(t) + \frac{1}{2} 1_{[\tau_1, \tau_2]}(t)$$

$$p_{2,s}(t) = \frac{\tau_1 - t}{s - \tau_1} 1_{[\tau_1, s]}(t) + \frac{t - \tau_2}{\tau_2 - s} 1_{(s, \tau_2]}(t) + \frac{1}{3} 1_{[\tau_1, \tau_2]}(t)$$

and

$$p_0(t) = p_{1,\bar{\tau}}(t) = \frac{\tau_1 - t}{\bar{\tau} - \tau_1} 1_{[\tau_1, \bar{\tau}]}(t) + \frac{t - \tau_2}{\tau_2 - \bar{\tau}} 1_{(\bar{\tau}, \tau_2]}(t) + \frac{1}{2} 1_{[\tau_1, \tau_2]}(t).$$

It can be easily checked that these functions satisfy

$$\int_{\tau_1}^{\tau_2} p_{1,s}(t)dt = 0, \quad \int_{\tau_1}^{\tau_2} p_{1,s}(t)(t - s)dt = \frac{(\tau_2 - s)^2 - (s - \tau_1)^2}{12} = \mathcal{I}_1(s) \tag{3.10}$$
$$\int_{\tau_1}^{\tau_2} p_{1,s}(t)(t - s)^2 dt = \frac{(s - \tau_1)^3 + (\tau_2 - s)^3}{12}$$

$$\int_{\tau_1}^{\tau_2} p_{2,s}(t)dt = -\frac{1}{6}(\tau_2 - \tau_1), \quad \int_{\tau_1}^{\tau_2} p_{2,s}(t)(t - s)dt = 0,$$
$$\int_{\tau_1}^{\tau_2} p_{2,s}(t)(t - s)^2 dt = \frac{(s - \tau_1)^3 + (\tau_2 - s)^3}{36}$$

and

$$\int_{\tau_1}^{\tau_2} p_0(t)dt = 0, \int_{\tau_1}^{\tau_2} p_0(t)(t - \bar{\tau})dt = 0, \quad \int_{\tau_1}^{\tau_2} p_0(t)(t - \bar{\tau})^2 dt = \frac{(\tau_2 - \tau_1)^3}{48}.$$

If τ_1 and τ_2 are successive jump points, then $\tilde{g}_n'$ is constant on (τ_1, τ_2), and $\tilde{g}_n$ is linear on $[\tau_1, \tau_2]$. On the other hand, using Taylor expansion and the assumption that g_0 is twice continuously differentiable to the right of 0, the true density g_0 can be decomposed into a quadratic polynomial plus a lower order remainder. Thus combining this with the inequality in (3.9) and the properties of p_0, $p_{1,s}$ and $p_{2,s}$ (in this order) yields

$$\frac{g_0''(\bar{\tau})}{96}(\tau_2 - \tau_1)^3 \leq \left| \int_{\tau_1}^{\tau_2} p_0(t) d(\mathbb{G}_n(t) - G_0(t)) \right| + o_p((\tau_2 - \tau_1)^3) \tag{3.11}$$

$$\mathcal{I}_1(s)\Big(\tilde{g}_n'(s) - g_0'(s)\Big) \geq - \left| \int_{\tau_1}^{\tau_2} p_{1,s}(t) d(\mathbb{G}_n(t) - G_0(t)) \right| + O_p((\tau_2 - \tau_1)^3) \tag{3.12}$$

$$\tilde{g}_n(s) - g_0(s) \leq \frac{6}{\tau_2 - \tau_1} \left| \int_{\tau_1}^{\tau_2} p_{2,s}(t) d(\mathbb{G}_n(t) - G_0(t)) \right| + O_p((\tau_2 - \tau_1)^2). \tag{3.13}$$

The above identity in (3.8) and the inequalities in (3.11), (3.12) and (3.13) will prove to be very useful for deriving the main result. They will be exploited for different purposes, but note that they share the common feature of linking the LSE to some empirical process term. Lemma 6.1 will give us a way to bound the corresponding rates of convergence of these empirical processes, which is essential for establishing the uniform tightness result of Proposition 6.1. This is explained in more detail in the next subsection, where we also point to the main difficulty encountered in establishing the main result.

3.2 The crucial role of the distance between jump points

To give some insight into the proof of Theorem 4.1 and to describe the main difficulty of the problem, we focus here on $\alpha = 1/5$ for which our proof is given in detail. The key argument in proving the weak convergence is to show the following uniform tightness result:

$$\sup_{t \in [M_1 n^{-1/5}, M_2 n^{-1/5}]} n^{2/5} |\tilde{g}_n(t) - g_0(t)| = O_p(1)$$

and

$$\sup_{t\in[M_1 n^{-1/5}, M_2 n^{-1/5}]} n^{1/5}|\tilde{g}_n'(t) - g_0'(t)| = O_p(1)$$

for all $0 < M_1 < M_2$. Actually, we prove a stronger result where the supremum is taken over the bigger interval $[\tau^-, M_2 n^{-1/5}]$ and τ^- is the last jump point of $\tilde{g}_n'$ located before $M_1 n^{-1/5}$ (see Proposition 6.1). In estimating the convex density at an interior point $t_0 > 0$, Groeneboom, Jongbloed, and Wellner (2001b) did not need to consider this stronger version; it is anyway implied by the $n^{1/5}-$order of the distance between two jump points τ^- and τ^+ in the neighborhood of $t_0 > 0$ as $n \to \infty$; i.e.,

$$\tau^+ - \tau^- = O_p(n^{-1/5}) \tag{3.14}$$

(see their Lemma 4.2). Uniform tightness is the key argument in showing convergence of the second and third derivatives of the "empirical invelope" to the corresponding derivative of the limit invelope, denoted in this manuscript by H_+. The "empirical invelope" is the scaled version of the local process $\tilde{H}_n^{loc}$ which we define in Section 6.

The difficulty in establishing the uniform tightness result in our boundary problem is that it is not known how the jump points cluster to the right of the origin. For example, for a fixed $K_2 > 0$, we do not know whether the probability that a jump point is found in a neighborhood $[K_1 n^{-1/5}, K_2 n^{-1/5}]$ for some $0 < K_1 < K_2$ increases to 1 as $n \to \infty$. In the interior point problem considered by Groeneboom, Jongbloed, and Wellner (2001b), the event that $\tilde{g}_n'$ will have a jump between $t_0 + Mn^{-1/5}$ and $t_0 - Mn^{-1/5}$ for some $M > 0$ large enough occurs with an increasing probability as $n \to \infty$. This fact follows again from (3.14) and was used by Groeneboom, Jongbloed, and Wellner (2001b) as a very useful "trick". Indeed, this offered some flexibility in choosing jump points, τ^- and τ^+ that are far enough to show that

$$\inf_{t\in[\tau^-,\tau^+]} n^{2/5}|\tilde{g}_n(t) - g_0(t)| = O_p(1). \tag{3.15}$$

By far enough, we mean that the distance $\tau^+ - \tau^- > rn^{1/5}$, where $r > 0$ does not depend on n. Of course, this distance stays bounded with large probability by $Mn^{1/5}$ for some $M > 0$. The result in (3.15) states that with increasing probability, we can find at least a point in $[\tau^-, \tau^+]$ such that the estimation error is of the order $n^{2/5}$. This turned out to be enough to show the uniform $n^{2/5}-$ and $n^{1/5}-$ tightness of the LSE and its derivative on an arbitrary neighborhood $[t_0 - Mn^{-1/5}, t_0 + Mn^{-1/5}]$. This might seem surprising but it was indeed possible using convexity of the estimator and the true density.

Unfortunately, the "trick" cannot be directly used here, and some new idea was needed. Our approach to go around the problem is to consider both the situations: (1) a jump point can be found in a given interval, $[K_1 n^{-1/5}, K_2 n^{-1/5}]$ say, and (2) no jump point can be found. In our proof of uniform tightness, A. Case 2 and C. Case 2 describe exactly the second situation. There, we make use of the inequalities in (3.12) and (3.13) and the empirical process argument of Lemma 6.1.

Localizing and scaling appropriately the processes $\tilde{H}_n(t) = \int_0^t (t-s)\tilde{g}_n(s)ds$ and $\mathbb{Y}_n(t) = \int_0^t \mathbb{G}_n(s)ds$, as it is done in Subsection 6.2, yield $\tilde{H}_n^l$ and $\mathbb{Y}_n^l$. Uniform tightness is then used to show convergence of $\tilde{H}_n^l$ to a limiting process, which has to be necessarily equal almost surely to the invelope of the limit of $\mathbb{Y}_n^l$ as $n \to \infty$.

In the next section, we give the three regimes of the joint asymptotic distributions of the consistent estimators $\tilde{g}_n(n^{-\alpha})$ and $\tilde{g}_n'(n^{-\alpha})$ depending on the value of α. This result can be compared directly with Theorem 4.1 of Kulikov and Lopuhaä (2006) in the case where the first derivative of the true monotone density does not vanish ($k = 1$ in their notation). A corollary follows, where we give the asymptotic distribution of our modified LSE.

4 Consistency at 0 and limiting distributions

Let W be a standard two-sided Brownian motion on $\mathbb{R}$ starting from 0, and Y the Gaussian drifting process defined on $\mathbb{R}$ in (1.3). For all $t > 0$, let

$$Y_0(t) = \int_0^t W(s)ds \quad \text{and} \quad Y_+(t) = Y(t).$$

We consider H_+ and H_0 the invelopes of Y_+ and Y_0, defined in the same sense as the invelope H (see Section 1). Finally, let $c_0(0)$ and $c_1(0)$ the resulting constants when we replace t_0 by 0 in (1.4), $c_3(0) = \sqrt{g_0(0)}$ and

$$k_2 = \frac{c_0(0)}{c_1(0)} = \left(24^2 g_0(0)(g_0''(0))^{-2}\right)^{\frac{1}{5}}. \tag{4.16}$$

Our main result is stated in the following theorem:

Theorem 4.1 *Let $\alpha \in (0, 1/3)$.*

1. *If $\alpha \in (1/5, 1/3)$ and $t > 0$, then*

$$\begin{pmatrix} n^{\frac{1-\alpha}{2}}\left(\tilde{g}_n(n^{-\alpha}\ t) - g_0(n^{-\alpha}\ t)\right) \\ n^{\frac{1-3\alpha}{2}}\left(\tilde{g}_n'(n^{-\alpha}\ t) - g_0'(n^{-\alpha}\ t)\right) \end{pmatrix} \to_d \begin{pmatrix} c_3(0)\ H_0^{(2)}(t) \\ c_3(0)\ H_0^{(3)}(t) \end{pmatrix}.$$

2. *If $\alpha = 1/5$ and $t > 0$, then*

$$\begin{pmatrix} n^{\frac{2}{5}}\left(\tilde{g}_n(n^{-\frac{1}{5}}\ k_2 t) - g_0(n^{-\frac{1}{5}}\ k_2 t)\right) + \frac{1}{2}k_2^2 t^2 g_0''(0) \\ n^{\frac{1}{5}}\left(\tilde{g}_n'(n^{-\frac{1}{5}}\ k_2 t) - g_0'(n^{-\frac{1}{5}}\ k_2 t)\right) + k_2 t g_0''(0) \end{pmatrix} \to_d \begin{pmatrix} c_0(0)\ H_+^{(2)}(t) \\ c_1(0)\ H_+^{(3)}(t) \end{pmatrix}.$$

3. *If $\alpha \in (0, 1/5)$, then*

$$\begin{pmatrix} n^{\frac{2}{5}}\left(\tilde{g}_n(n^{-\alpha}) - g_0(n^{-\alpha})\right) \\ n^{\frac{1}{5}}\left(\tilde{g}_n'(n^{-\alpha}) - g_0'(n^{-\alpha})\right) \end{pmatrix} \to_d \begin{pmatrix} c_0(0)\ H^{(2)}(0) \\ c_1(0)\ H^{(3)}(0) \end{pmatrix}.$$

To keep the manuscript to a reasonable length, we do not present the details of the proof of existence and almost surely uniqueness of the processes H_+ and H_0 . However, we would like to mention that the proof can be constructed along the lines of Groeneboom, Jongbloed, and Wellner (2001a) by defining a stochastic Least Squares problem in the class of convex functions on finite intervals $[0, c]$ and let the intervals grow as $c \to \infty$. For example, the invelope H_+ can be shown to be the limit, in an appropriate sense, of the processes $H_{+,c}$ as $c \to \infty$. Here, $H_{+,c}$ is equal to the second integral of the unique minimizer of

$$\frac{1}{2}\int_0^c g^2(t)dt - \int_0^c g(t)d[W(t) + 4t^3]$$

over the class of convex functions g on $[0, c]$ such that $g(0) = 0$ and $g(c) = 12c^2$, and $H_{+,c}$ satisfies the four boundary conditions $H_{+,c}(0) = Y_+(0) = 0$, $H_{+,c}(c) = Y_+(c)$, $H_{+,c}'(0) = Y_+'(0) = 0$ and $H_{+,c}'(c) = Y_+'(c) = W(c) + 4c^3$. It can be shown that

(i) $H^{(2)}_{+,c}$ is convex on $[0,c]$

(ii) $H_{+,c}(t) \geq Y_+(t), \quad \forall\, t \in [0,c]$

(iii) $\int_0^c (H_{+,c}(t) - Y_+(t))dH^{(3)}_{+,c}(t) = 0,$

and that these conditions are necessary and sufficient for $H^{(2)}_{+,c}$ to be the solution of the above LS problem. The conditions constitute in essence a finite version of the characterization of the limiting process H_+ defined on $[0,\infty)$. For large $c > 0$, $H_{+,c}$ is an approximation of H_+ and we can compute it using the Haar construction (see Rogers and Williams (1994)) and the iterative cubic spline algorithm as described in Groeneboom, Jongbloed, and Wellner (2003). In Figure 1, we present an example where the plots in dotted lines show the restriction on the interval $[0,1]$ of $H^{(j)}_{+,8}$ (computed on $[0,8]$) for $j = 0,1,2,3$.

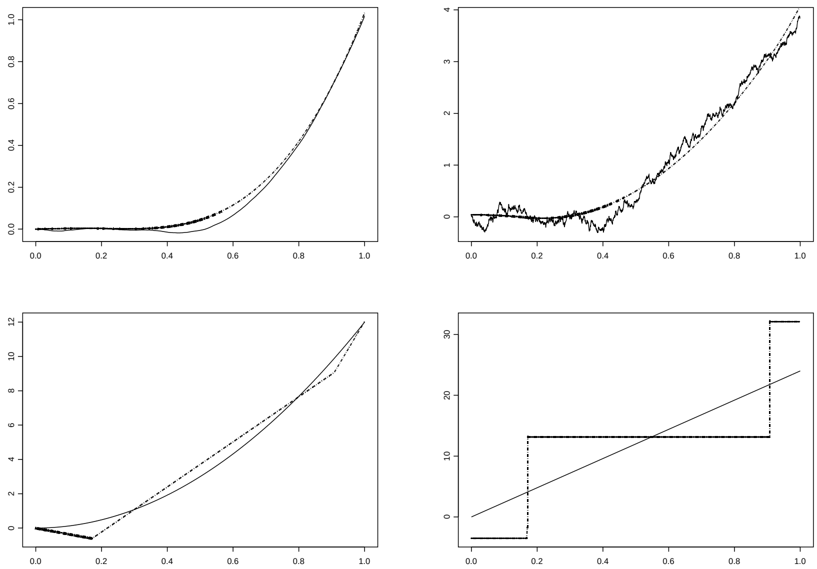

Figure 1: The solid curves in the top left and top right show plots of an approximation of Y_+ and Y'_+ while the dotted lines represent $H_{+,8}$ and $H'_{+,8}$ on $[0,1]$. The solid curves in the bottom left and bottom right show plots of $12t^2$ and $24t$ while the dotted lines represent $H''_{+,8}$ and $H^{(3)}_{+,8}$ on $[0,1]$.

In the top left and top right panels, one can see how the invelope $H_{+,8}$ and its first derivative $H'_{+,8}$ try to follow closely the processes $\int_0^t W(s)ds + t^4$ and $W(t) + 4t^3$ on $[0,1]$. On the other hand, we see in the bottom left and bottom right panels how the convex parabola, $12t^2$, and the increasing straight line, $24t$, are approximated by the piecewise convex function $H''_{+,8}$ and the nondecreasing stepwise function $H^{(3)}_{+,8}$ respectively.

Corollary 4.1 *The fastest rates of convergence of $\tilde{g}_n(n^{-\alpha})$ and $\tilde{g}'_n(n^{-\alpha})$ to the true values $g_0(0)$ and $g'_0(0)$ are attained for $\alpha = 1/5$. Furthermore, if we define the estimator*

$$\tilde{g}^m_n(t) = \begin{cases} \tilde{g}_n(n^{-\frac{1}{5}}) + (t - n^{-\frac{1}{5}})\tilde{g}'_n(n^{-\frac{1}{5}}) & \textit{if } t \in [0, n^{-\frac{1}{5}}) \\ \tilde{g}_n(t) & \textit{if } t \in]n^{-\frac{1}{5}}, \infty) \end{cases}$$

then

$$\begin{pmatrix} n^{\frac{2}{5}}\Big(\tilde{g}^m_n(0) - g_0(0)\Big) \\ n^{\frac{1}{5}}\Big((\tilde{g}^m_n)'(0) - g'_0(0)\Big) \end{pmatrix} \to_d \begin{pmatrix} c_0(0)\ H^{(2)}_+\Big(\frac{1}{k_2}\Big) - c_1(0)\ H^{(3)}_+\Big(\frac{1}{k_2}\Big) \\ c_1(0)\ H^{(3)}_+\Big(\frac{1}{k_2}\Big) \end{pmatrix}.$$

We will refer to this estimator as the modified LSE estimator.

Proof. *If $\alpha \in (1/5, 1/3)$, the rates $n^{(1-\alpha)/2}$ and $n^{(1-3\alpha)/2}$ are clearly slower than $n^{2/5}$ and $n^{1/5}$ respectively. If $\alpha < 1/5$, the rates $n^{2/5}$ and $n^{1/5}$ are optimal for estimating $g_0(n^{-\alpha})$ and $g'_0(n^{-\alpha})$, but not for estimating $g_0(0)$ and $g'_0(0)$: $n^{-\alpha} >> n^{-1/5}$ is too far from 0. The joint limiting distribution of $(\tilde{g}^m_n(0), (\tilde{g}^m_n)'(0))$ follows immediately from Theorem 4.1.* $\square$

5 Illustrations, conclusions and some open questions

5.1 Simulation results

We simulated $n = 500$ independent random variables from the standard exponential distribution $Exp(1)$ and computed $\tilde{g}_n$ using the support reduction algorithm, as described by Groeneboom, Jongbloed, and Wellner (2003) for Least Squares problems. Inconsistency of the estimator and its first derivative at 0 can be clearly seen on the left panels of Figure 2, whereas the right panels illustrate consistency of the modified estimator and its first derivative. For these simulations we obtained $\tilde{g}_n(0) = 1.757, \tilde{g}^m_n(0) = 0.966$, and $-\tilde{g}'_n(0) =$

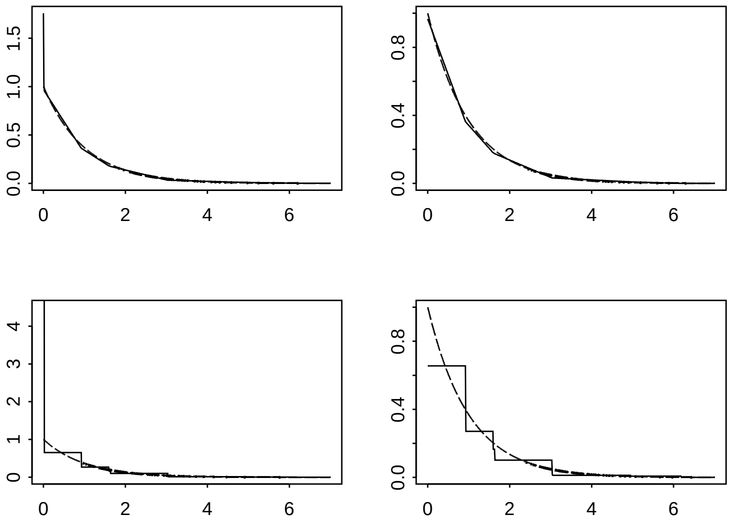

Figure 2: The solid curves in the top left and top right plots depict $\tilde{g}_n$ and $\tilde{g}_n^m$ while the dashed curves represent $g_0(x) = exp(-x), x > 0$. The solid curves in the bottom left and bottom right plots depict $-\tilde{g}_n'$ and $-(\tilde{g}_n^m)'$ while the dashed curves represent $-g_0'(x) = exp(-x), x > 0$.

74.073, $-(\tilde{g}_n^m)'(0) = 0.655$. The fact that the modified estimators converge to the truth with a negative bias is not due to chance. To investigate this more precisely, we generated 100 independent invelopes H_+ on $[0, 8]$, and the results indicate that the random variables $c_0 H_+^{(2)}(1/k_2) - c_1 H_+^{(3)}(1/k_2)$ and $-c_1 H_+^{(3)}(1/k_2)$ are both negative with large probability: The empirical estimates of this probability based on the random sample were found to be 0.94 and 0.97 respectively (here, $1/k_2 = (1/24)^{2/5} \approx 0.280,\ c_0 = (1/24)^{1/5} \approx 0.041,\ c_1 = (1/24)^{3/5} \approx 0.148$). The kernel density estimators shown in Figure 3 indicate clearly that both variables are mainly concentrated on the negative half-line.

5.2 Conclusions and some open questions

Based solely on the LSE $\tilde{g}_n$ of a convex density g_0 on $(0, \infty)$ with g_0'' continuous to the right of 0 and $g_0''(0) \neq 0$, we found that consistent estimation of $g_0'(0)$ is achieved by taking $\tilde{g}_n'(n^{-\alpha})$ with $\alpha \in (0, 1/3)$, and that $\alpha = 1/5$ should be chosen

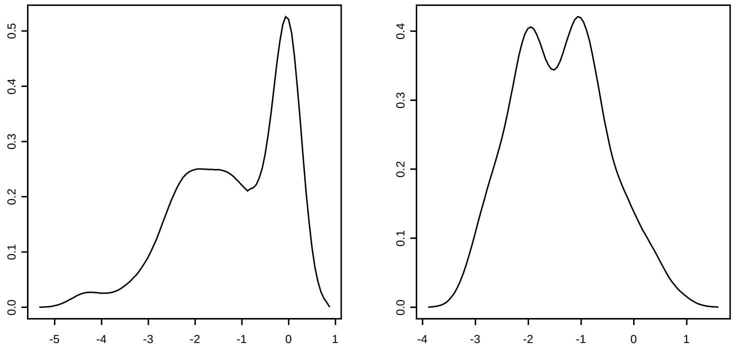

Figure 3: Plots of kernel density estimators of $c_0 H_+^{(3)}(1/k_2) - c_1 H_+^{(3)}(1/k_2)$ (left) and $-c_1 H_+^{(3)}(1/k_2)$ (right). The estimation is based on 100 independent replications of H_+ on $[0, 8]$. Here, $c_0 \approx 0.041$, $c_1 \approx 0.148$ and $1/k_2 \approx 0.280$.

as it yields the fastest rate $n^{-1/5}$. The limiting distribution involves a process H_+, which is the invelope of the Gaussian process $Y_+(t) = \int_0^t W(s)ds + t^4,\ t \geq 0$.

Our idea was inspired by the work of Kulikov and Lopuhaä (2006) who found that taking the Grenander estimator at $n^{-\alpha}$ with $\alpha \in (0, 1)$ ensures consistency in the monotone estimation problem, and that $\alpha = 1/3$ yields the optimal rate $n^{-1/3}$. It is interesting to note that the penalization approach of Woodroofe and Sun (1993) forces rather the data to stay away from 0 with a distance $>> n^{-1}$: Their estimator can be viewed as the Grenander estimator for the transformed data $\lambda_n + \hat{\gamma}_n X_j$, with $\hat{\gamma}_n \to_p 1$ and λ_n is the penalization parameter which must satisfy $\lambda_n\, n \to \infty$, and hence $\lambda_n >> n^{-1}$. For the harder problem of estimating the slope $g_0'(0)$, an alternative approach based on shifting the data (suggested to us by Jongbloed (2006)) would probably require to have even bigger shift to the right of 0. A second approach is to penalize the derivative of the LSE; i.e., to minimize

$$\frac{1}{2}\int_0^\infty g^2(t)dt - \int_0^\infty g(t)d\mathbb{G}_n(t) - \lambda_n g'(0)$$

where g is convex and λ_n is the penalizing parameter. This gives rise to the following open questions: How big λ_n should be chosen to achieve consistency of the first derivative of the estimator? Would λ_n have the same order as the shift in the approach suggested by Jongbloed (2006)? Could we choose λ_n such that the estimator is a density? How do the rates of convergence and limiting distributions depend on λ_n?

In the monotone problem, the proofs of Kulikov and Lopuhaä (2006) make use of the so-called switching relationship introduced the first time by Groeneboom (1985) as a nice geometric interpretation of the Grenander estimator: If $\hat{f}_n(x)$ is the value of this estimator at a point x, and $U_n(a)$ is the location of the maximum of $\mathbb{G}_n(t) - at$ over $[0, \infty)$, then

$$\hat{f}_n(x) \le a \rightleftharpoons U_n(a) \le x.$$

A similar relationship is still lacking in the convex problem. In the latter, the characterization of the estimator is at the level of its second integral. This makes the geometric interpretation of the LSE less obvious. However, this does not imply that one should not explore different ways of viewing this characterization which might enable to simplify many of the arguments used in general in the problem of adaptive estimation of a convex density.

Finally, we would like to note that our modified LSE estimator can be compared to the simple estimator of Kulikov and Lopuhaä (2006). Their adaptive estimator is more efficient as it minimizes the mean square error. In our convex problem, it follows from Theorem 4.1 that

$$n^{\frac{1}{5}}\left(\tilde{g}_n'(n^{-\frac{1}{5}}\ k_2 t) - g_0'(0)\right) \to_d c_1(0) H_+^{(3)}(t)$$

and an adaptive estimator of $g_0'(0)$ can be given by $\tilde{g}_n(n^{-1/5}\hat{k}_2 t^*)$, where t^* is the minimizer of $E[H_+^{(3)}(t^*)]^2$, and $\hat{k}_2$ is a consistent estimator of k_2. We would like to investigate this in a future work as finding an approximate value for t^* would require a more efficient way to generate a sufficient number of invelopes H_+ on a fine grid on $(0, \infty)$.

6 Proof of Theorem 4.1

We prove Theorem 4.1 for $\alpha = 1/5$, which yields the fastest rates of convergence. For $\alpha \in (1/5, 1/3)$, the core of the proof remains exactly the same, except for minor changes that will be indicated at the end of this section. We will also give heuristic arguments for the claimed weak convergence for $\alpha \in (0, 1/5)$.

6.1 Proof of uniform tightness

Proposition 6.1 *For* $0 < M_1 < M_2$, *let* τ^- *be the last jump point of* $\tilde{g}_n'$ *occurring before* $M_1 n^{-1/5}$. *Then,*

$$\sup_{s\in[\tau^-, n^{-1/5}M_2]} |\tilde{g}_n'(s) - g_0'(s)| = O_p(n^{-1/5}) \tag{6.17}$$

and

$$\sup_{s\in[\tau^-, n^{-1/5}M_2]} |\tilde{g}_n(s) - g_0(s)| = O_p(n^{-2/5}) \tag{6.18}$$

The following lemmas will provide the necessary pieces that will go into the proof of Proposition 6.1.

Lemma 6.1 *Let* G_0 *denote the true cumulative distribution function, and fix* $x > 0, r > 0$. *Consider a VC-subgraph class of functions* $f_{x,y}$ *defined on* $[x, y]$, $x \le y \le x + r$:

$$\mathcal{F}_{x,r} = \{f_{x,y},\ x \le y \le x + r\}$$

admitting an envelope $F_{x,r}$ *satisfying*

$$EF_{x,r}^2(X_1) \le Cr^{2k+1},$$

for some real constant $C > 0$ *and an integer* k *that are independent of* x. *Then, for each* $\epsilon' > 0$, *we have*

$$\sup_{f_{x,y}\in\mathcal{F}_{x,r}} \left| \int f_{x,y} d\left(\mathbb{G}_n - G_0\right) \right| \le \epsilon'(y-x)^{3+k} + O_p\left(n^{-\frac{3+k}{5}}\right).$$

Proof. The argument is very similar to that of Kim and Pollard (1990) for proving their Lemma 4.1 (page 201) except that their grid mesh $[(j-1)n^{-1/3}, jn^{-1/3})$, $1 \le j \le \lfloor R_0 n^{1/3} \rfloor$, is replaced here by $[(j-1)n^{-1/5}, jn^{-1/5}), 1 \le j \le \lfloor rn^{1/5} \rfloor$. To see intuitively why we have the power $(3+k)/5$ instead of their power $2/3$, note that their Maximal inequality 3.1 (i) (see also Theorem 2.14.2 of van der Vaart and Wellner (1996)) implies that for a fixed $\epsilon > 0$ there exists some $M > 0$ independent of j and n such that

$$\begin{aligned}\sup_{f_{x,y}\in\mathcal{F}_{x,jn^{-1/5}}} \left| \int f_{x,y} d\left(\mathbb{G}_n - G_0\right) \right| \le Mn^{-\frac{1}{2}} \sqrt{\left[Cjn^{-\frac{1}{5}}\right]^{2k+1}} &= M(Cj)^{k+\frac{1}{2}} n^{-\frac{2k+6}{10}} \\ &= M(Cj)^{k+\frac{1}{2}} n^{-\frac{3+k}{5}}\end{aligned}$$

with probability greater than $1-\epsilon$. Summation over $1 \le j \le \lfloor rn^{1/5} \rfloor$ yields the result. See also GROENEBOOM, JONGBLOED, AND WELLNER (2001B), page 1677. $\square$

Lemma 6.2 *For arbitrary $0 \le K_1 \le K_2 < K_3 \le K_4$, consider the event*

$$\mathcal{T}_{K_1,K_2,K_3,K_4} = \left\{ \exists\ \tau^-, \tau^+ : K_1 n^{-1/5} \le \tau^- \le K_2 n^{-1/5},\ K_3 n^{-1/5} \le \tau^+ \le K_4 n^{-1/5} \right\},$$

where τ^- and τ^+ are jump points of $\tilde{g}_n'$. Then, for all $\epsilon > 0$, there exists $c > 0$ such that

$$P\Big(\inf_{t \in [\tau^-, \tau^+]} |\tilde{g}_n(t) - g_0(t)| \le cn^{-2/5}, \mathcal{T}_{K_1,K_2,K_3,K_4} \Big) \ge P(\mathcal{T}_{K_1,K_2,K_3,K_4}) - \epsilon$$

for n large enough.

Proof. Suppose that both the events $\mathcal{T}_{K_1,K_2,K_3,K_4}$ and $\inf_{t \in [\tau^-,\tau^+]} |\tilde{g}_n(t) - g_0(t)| > cn^{-2/5}$ occur. Then, we can write

$$\left| \int_{\tau^-}^{\tau^+} (\tau^+ - t)(\tilde{g}_n(t) - g_0(t)) dt \right| > \frac{c}{2}\, n^{-2/5} (\tau^+ - \tau^-)^2 > \frac{c\,(K_3 - K_2)^2}{2}\, n^{-4/5}.$$

Using now the identity (3.8) of Section 3 and Lemma 6.1 with $\mathcal{F}_{x,r} = \{ f_{x,y}(t) = y - t, x < y < x + r \}$ (VC-class with index ≤ 3 by Lemma 2.6.15 of VAN DER VAART AND WELLNER (1996)) and $k = 1$, we can write

$$\begin{aligned} \left| \int_{\tau^-}^{\tau^+} (\tau^+ - t)(\tilde{g}_n(t) - g_0(t)) dt \right| &= \left| \int_{\tau^-}^{\tau^+} (\tau^+ - t) d(\mathbb{G}_n(t) - G_0(t)) dt \right| \\ &= o_p((K_4 - K_1)^4 n^{-4/5}) + O_p(n^{-4/5}), \end{aligned}$$

and hence $O_p(n^{-4/5})$. Thus, we can find $c > 0$ large enough such that

$$P\Big(\inf_{t \in [\tau^-, \tau^+]} |\tilde{g}_n(t) - g_0(t)| > cn^{-2/5}, \mathcal{T}_{K_1,K_2,K_3,K_4} \Big) < \epsilon$$

which implies the claimed result. $\square$

Lemma 6.3 *Let $\epsilon > 0$. For any $K_1 \ge 0$, there exist $K_2 \ge K_1$ such that, for n large enough, the event*

$$\mathcal{T}_{K_1,K_2} = \left\{ \exists \text{ a jump point } \tau : K_1 n^{-1/5} \le \tau \le K_2 n^{-1/5} \right\}$$

occurs with probability greater than $1-\epsilon$. *In particular, this implies that for a given* $K_1 \geq 0$, *there exist* K_2 *and* K_4 *such that* $K_4 \geq K_3 > K_2 \geq K_1$ *and*

$$P(\mathcal{T}_{K_1,K_2,K_3,K_4}) > 1-\epsilon.$$

Proof. Given $K_1 \geq 0$, let τ_1 and τ_2 be the last and first jump points occurring before and after $K_1 n^{-1/5}$ respectively (τ_1 can be equal to 0). Note that for $r_0 > 0$ small enough such that $g_0'' > 0$ on $(0, r_0]$, the event $\tau_2 < r_0$ occurs with probability $\to 1$ since we know that there exists a jump point $\tau < r_0/2$ such that $r_0/2 - \tau = O_p(n^{-1/5})$ with probability $\to 1$ (by Lemma 4.2 of Groeneboom, Jongbloed, and Wellner (2001b)). Using now the inequality (3.11) in Section 3 and Lemma 6.1 (with $k = 0$), we have

$$(\tau_2 - \tau_1)^3 \leq \frac{96}{\inf_{t \in (0,r_0]} g_0''(t)} \left[o_p((\tau_2-\tau_1)^3) + O_p(n^{-3/5}) \right]$$

which implies that

$$\tau_2 - \tau_1 = O_p(n^{-1/5}). \tag{6.19}$$

Note that this is the same upper bound obtained by Groeneboom, Jongbloed, and Wellner (2001b) for the distance between jump points in the neighborhood of an interior point $t_0 > 0$, as we have already mentioned in Subsection 3.2. Thus, there exists $M > 0$ such that $\tau_2 < \tau_1 + M n^{-1/5} \leq (K_1 + M) n^{-1/5}$ with large probability, and hence we can take $K_2 = K_1 + M$. To show the second assertion of the lemma it suffices, for any $K_3 > K_2$, to consider τ' to be the first jump point after $K_3 n^{-1/5}$. Then there exists $K_4 \geq K_3$ such that the probability of the event $n^{-1/5} K_3 \leq \tau' \leq n^{-1/5} K_4$ is greater than $1 - \epsilon$. □

Proof of Proposition 6.1. In all what follows we denote $B_u = \sup_{t \in (0,r_0]} g_0''$, where $(0, r_0]$ is the biggest neighborhood of 0 on which $g_0'' > 0$ and continuous. For convenience, we are going to assume without loss of generality that $M_1 = 6/7$ and $M_2 = 1$. Of course, the same arguments can be used for any other values $0 < M_1 < M_2$ as long as they do not depend on n. Let τ^- be the last jump point occurring before $6/7 n^{-1/5}$ and fix $\epsilon > 0$.

A. Uniform tightness of $n^{1/5}(\tilde{g}_n' - g_0')$ **from below**: We show in the following that there exists $c > 0$ such that

$$P\Big(\forall t \in [\tau^-, n^{-1/5}] \;\; \tilde{g}_n'(t) - g_0'(t) \geq -c\, n^{-1/5}\Big) > 1 - \epsilon \tag{6.20}$$

for n large enough. If we divide $[0, n^{-1/5}]$ into seven equally sized subintervals, then there are only two cases.

<u>Case 1</u>. We can find a jump point in each of the subintervals $[n^{-1/5}/7, 2n^{-1/5}/7]$, $[3n^{-1/5}/7, 4n^{-1/5}/7]$ and $[5n^{-1/5}/7, 6n^{-1/5}/7]$. Let τ_1, τ_2 and τ_3 denote these points, and $\tau_0 = 0$. Using the notation of Lemma 6.2, our assumption implies that the event

$$\mathcal{T} = \mathcal{T}_{0,0,1/7,2/7} \cap \mathcal{T}_{3/7,4/7,5/7,6/7}$$

occurs. We consider the points ξ_{-2} and ξ_{-1} such that

$$|\tilde{g}_n(\xi_{-2}) - g_0(\xi_{-2})| = \inf_{t \in [\tau_0, \tau_1]} |\tilde{g}_n(t) - g_0(t)|$$

$$\text{and} \;\; |\tilde{g}_n(\xi_{-1}) - g_0(\xi_{-1})| = \inf_{t \in [\tau_2, \tau_3]} |\tilde{g}_n(t) - g_0(t)|.$$

By convexity of $\tilde{g}_n$, we can write for all $t \in [\tau^-, n^{-1/5}]$ (note that $t > \xi_{-1} > \xi_{-2}$)

$$\begin{aligned}
\tilde{g}_n'(t) &\geq \frac{\tilde{g}_n(\xi_{-1}) - \tilde{g}_n(\xi_{-2})}{\xi_{-1} - \xi_{-2}} \\
&= \frac{\tilde{g}_n(\xi_{-1}) - g_0(\xi_{-1}) - (\tilde{g}_n(\xi_{-2}) - g_0(\xi_{-2}))}{\xi_{-1} - \xi_{-2}} + \frac{g_0(\xi_{-1}) - g_0(\xi_{-2})}{\xi_{-1} - \xi_{-2}} \\
&\geq -\frac{\inf_{t \in [\tau_0, \tau_1]} |\tilde{g}_n(t) - g_0(t)| + \inf_{t \in [\tau_1, \tau_2]} |\tilde{g}_n(t) - g_0(t)|}{n^{-1/5}/7} + g_0'(\xi_{-2})
\end{aligned}$$

using convexity of g_0. This implies that

$$\tilde{g}_n'(t) - g_0'(t) \geq -\frac{\inf_{t \in [\tau_0, \tau_1]} |\tilde{g}_n(t) - g_0(t)| + \inf_{t \in [\tau_1, \tau_2]} |\tilde{g}_n(t) - g_0(t)|}{n^{-1/5}/7} - B_u n^{-1/5}.$$

Hence, for $c > 0$ large enough, we have

$$\begin{aligned}
&P\Big(\forall t \in [\tau^-, n^{-1/5}] \;\; \tilde{g}_n'(t) - g_0'(t) \geq -c\, n^{-1/5}, \mathcal{T}\Big) \\
&\geq P\Big(\inf_{t \in [\tau_0, \tau_1]} |\tilde{g}_n(t) - g_0(t)| + \inf_{t \in [\tau_2, \tau_3]} |\tilde{g}_n(t) - g_0(t)| \leq \frac{c - B_u}{7}\, n^{-2/5}, \mathcal{T}\Big) \\
&\geq P(\mathcal{T}) - \epsilon,
\end{aligned} \tag{6.21}$$

by arguments that are very similar to those used in proving Lemma 6.2.

<u>Case 2</u>. One of the subintervals contains no jump point. If only one interval contains no jump points, for example the subinterval $[5n^{-1/5}/7, 6n^{-1/5}/7]$, then we denote by τ_1 and τ_2 the last jump point before $5n^{-1/5}/7$ and the first jump point after $6n^{-1/5}/7$ respectively. Otherwise, if two or more intervals contain no jump points, then τ_1 and τ_2 will be the last and first jump points occurring before and after the end points of the smallest interval (e.g. τ_1 and τ_2 will be taken as the last and first jump points located before $n^{-1/5}/7$ and after $2n^{-1/5}/7$ if there is no jump point in $[n^{-1/5}/7, 2n^{-1/5}/7]$ nor in $[5n^{-1/5}/7, 6n^{-1/5}/7]$).

Let $\bar{\bar{\tau}} = 3\tau_1/4 + \tau_2/4$. Then, for all $s \in (\tau_1, \bar{\bar{\tau}}]$, $\mathcal{I}_1(s) > (\tau_2 - \tau_1)^2/24$ (see the expression of $\mathcal{I}_1(s)$ given in (3.10)). Thus, the inequality in (3.12) implies that for all $s \in (\tau_1, \bar{\bar{\tau}}]$ we can write

$$\tilde{g}_n'(s) - g_0'(s) \geq -\frac{24}{(\tau_2 - \tau_1)^2} \left| \int_{\tau_1}^{\tau_2} p_{1,s}(t) d(\mathbb{G}_n(t) - G_0(t)) \right| + O_p(\tau_2 - \tau_1). \quad (6.22)$$

On the other hand, applying Lemma 6.1, with

$$\mathcal{F}_{x,r} = \left\{ f_{x,y}(t) = \frac{x-t}{s-x} 1_{[x,s]}(t) + \frac{t-x}{y-s} 1_{(s,y]}(t) + \frac{1}{2} 1_{[x,y]}(t), x \leq y \leq x + r \right\}$$

and $k = 0$, implies that we can find $c > 0$ such that

$$P\left(\forall s \in [\tau_1, \bar{\bar{\tau}}] \ \left| \int_{\tau_1}^{\tau_2} p_{1,s}(t) d(\mathbb{G}_n(t) - G_0(t)) \right| \leq o_p((\tau_2 - \tau_1)^3) + c\, n^{-3/5}, \mathcal{T}^c \right)$$
$$\geq P(\mathcal{T}^c) - \epsilon$$

for n large enough. Using now the stochastic upper bound $\tau_2 - \tau_1 = O_p(n^{-1/5})$ in (6.19), and noting that $\tau_2 - \tau_1 \geq n^{-1/5}/7$, it follows from the inequality in (6.22) that we can find $c' > 0$ such that

$$P\left(\forall s \in [\tau_1, \bar{\bar{\tau}}] \ \ \tilde{g}_n'(s) - g_0'(s) \geq -c'\, n^{-1/5}, \mathcal{T}^c \right) \geq P(\mathcal{T}^c) - \epsilon \quad (6.23)$$

for $n > 0$ large enough.

To finish off, we only need to show that (6.23) is true for $\forall\ s \in [\tau^-, n^{-1/5}]$. We recall again that τ^- is the last jump point occurring before $5n^{-1/5}/7$.

If $\tau^- \geq \bar{\bar{\tau}}$, then by monotonicity of $\tilde{g}_n'$, we have $\tilde{g}_n'(s) \geq \tilde{g}_n'(\bar{\bar{\tau}})$ for all $s \in [\tau^-, n^{-1/5}]$, and this implies that

$$\begin{aligned} \tilde{g}_n'(s) - g_0'(s) &\geq \tilde{g}_n'(\bar{\bar{\tau}}) - g_0'(\bar{\bar{\tau}}) - B_u\, n^{-1/5}, \ \ \forall\ s \in [\tau^-, n^{-1/5}] \\ &\geq -c'\, n^{-1/5} - B_u\, n^{-1/5}. \end{aligned}$$

If $\tau^- < \bar{\bar{\tau}}$, then

$$\begin{cases} \tilde{g}_n'(s) - g_0'(s) \geq -c' \, n^{-1/5}, & \forall \, s \in [\tau^-, \min(\bar{\bar{\tau}}, n^{-1/5})] \subset [\tau_1, \bar{\bar{\tau}}] \\ \tilde{g}_n'(s) - g_0'(s) \geq -c' \, n^{-1/5} - B_u \, n^{-1/5}, & \forall \, s \in (\min(\bar{\bar{\tau}}, n^{-1/5}), n^{-1/5}]. \end{cases}$$

Thus, we conclude that there exists $c > 0$

$$P\left(\forall \, t \in [\tau^-, n^{-1/5}] \;\; \tilde{g}_n'(t) - g_0'(t) \geq -c \, n^{-1/5}, \mathcal{T}^c\right) \geq \; P(\mathcal{T}^c) - \epsilon. \qquad (6.24)$$

Now, combining (6.21) and (6.24) yields (at the cost maybe of increasing $c > 0$)

$$P\left(\forall \, t \in [\tau^-, n^{-1/5}] \;\; \tilde{g}_n'(t) - g_0'(t) \geq -c \, n^{-1/5}\right) > 1 - 2\epsilon$$

for n sufficiently large.

Remark. The class

$$\mathcal{F}_{x,r} = \left\{ f_{x,y}(t) = \frac{x-t}{s-x} 1_{[x,s]}(t) + \frac{t-x}{y-s} 1_{(s,y]}(t) + \frac{1}{2} 1_{[x,y]}(t), x \leq y \leq x + r \right\}$$

is a VC-class because the class of sets of between graphs

$$\mathcal{D} = \left\{ (t, c) : 0 \leq c \leq f_{x,y}(t) \text{ or } f_{x,y}(t) \leq c \leq 0 \right\}$$

is a VC-class (van der Vaart and Wellner (1996), Section 2.6, problem 11). The latter is true since $\mathcal{D} = \mathcal{D}_1 \cap \mathcal{D}_2$, where

$$\mathcal{D}_1 = \left\{ (t, c) : 0 \leq c \leq f_{x,y}(t) 1_{[x,s]}(t) \text{ or } f_{x,y}(t) 1_{[x,s]}(t) \leq c \leq 0 \right\}$$

and

$$\mathcal{D}_2 = \left\{ (t, c) : 0 \leq c \leq f_{x,y}(t) 1_{(s,y]}(t) \text{ or } f_{x,y}(t) 1_{(s,y]}(t) \leq c \leq 0 \right\},$$

which are VC-classes. This follows from the fact that $t \mapsto f_{x,y}(t) 1_{[x,s]}(t)$ and $t \mapsto f_{x,y}(t) 1_{(s,y]}(t)$ are polynomials of degree at most 1. Hence, $\mathcal{D}$ is a VC-class by van der Vaart and Wellner (1996), Lemma 2.6.17 (iii).

B. Uniform tightness of $n^{1/5}(\tilde{g}_n' - g_0')$ from above: We show now that there exists $c > 0$ such that

$$P\Big(\forall t \in [\tau^-, n^{-1/5}] \;\; \tilde{g}_n'(t) - g_0'(t) \leq c \, n^{-1/5}\Big) > 1 - \epsilon \qquad (6.25)$$

for n large enough.

The proof relies mainly on the result of Lemma 6.3. Indeed, we can find $K > 0$ such that the jump points τ_1^+ occurring after $n^{-1/5}$, τ_2^+ after $\tau_1^+ + n^{-1/5}$, τ_3^+ after $\tau_2^+ + n^{-1/5}$ and τ_4^+ after $\tau_3^+ + n^{-1/5}$ are all bounded by $Kn^{-1/5}$ with probability greater than $1-\epsilon$. Using Lemma 6.2, we can find $c' > 0$ such that

$$P\Big(|\tilde{g}_n(\xi_1) - g_0(\xi_1)| \le c'n^{-2/5}, |\tilde{g}_n(\xi_2) - g_0(\xi_2)| \le c'n^{-2/5}\Big) > 1-\epsilon$$

where $|\tilde{g}_n(\xi_1) - g_0(\xi_1)| = \inf_{t\in[\tau_1^+,\tau_2^+]} |\tilde{g}_n(t) - g_0(t)|$ and $|\tilde{g}_n(\xi_2) - g_0(\xi_2)| = \inf_{t\in[\tau_3^+,\tau_4^+]} |\tilde{g}_n(t) - g_0(t)|$. Now, by convexity, we can write for all $t \in [\tau^-, n^{-1/5}]$

$$\begin{aligned}
\tilde{g}_n'(t) &\le \frac{\tilde{g}_n(\xi_2) - \tilde{g}_n(\xi_1)}{\xi_2 - \xi_1} \\
&= \frac{\tilde{g}_n(\xi_2) - g_0(\xi_2) - (\tilde{g}_n(\xi_1) - g_0(\xi_1))}{\xi_2 - \xi_1} + \frac{g_0(\xi_2) - g_0(\xi_1)}{\xi_2 - \xi_1} \\
&\le \frac{\tilde{g}_n(\xi_2) - g_0(\xi_2) - (\tilde{g}_n(\xi_1) - g_0(\xi_1))}{n^{-1/5}} + g_0'(\xi_2), \\
&= g_0'(t) + \frac{\tilde{g}_n(\xi_2) - g_0(\xi_2) - (\tilde{g}_n(\xi_1) - g_0(\xi_1))}{n^{-1/5}} + \big[g_0'(\xi_2) - g_0'(t)\big]
\end{aligned}$$

and therefore with probability greater than $1-\epsilon$

$$\tilde{g}_n'(t) - g_0'(t) \le 2c'n^{-1/5} + B_u K n^{-1/5} = cn^{-1/5}, \quad \forall\ t \in [\tau^-, n^{-1/5}]$$

with $c = 2c' + B_u K$.

C. Uniform tightness of $n^{2/5}(\tilde{g}_n - g_0)$ from above: We show now that there exists $c > 0$ such that

$$P\Big(\forall t \in [\tau^-, n^{-1/5}]\ \ \tilde{g}_n(t) - g_0(t) \le cn^{-2/5}\Big) > 1-\epsilon \tag{6.26}$$

for n large enough.

Case 1. By convexity, we can write for all $t \in [\tau^-, n^{-1/5}]$

$$\tilde{g}_n(t) \le \tilde{g}_n(\xi_{-1}) + \frac{\tilde{g}_n(\xi_1) - \tilde{g}_n(\xi_{-1})}{\xi_1 - \xi_{-1}}(t - \xi_{-1})$$

Above, we used the same notation as before in A. Case 1 and in B: $|\tilde{g}_n(\xi_{-1}) - g_0(\xi_{-1})| = \inf_{t\in[\tau_1,\tau_2]} |\tilde{g}_n(t) - g_0(t)|$, where τ_1 and τ_2 are jump points in $[n^{-1/5}/7, 2n^{-1/5}/7]$

and $[3n^{-1/5}/7, 4n^{-1/5}/7]$ respectively, and $|\tilde{g}_n(\xi_1) - g_0(\xi_1)| = \inf_{t \in [\tau_1^+, \tau_2^+]} |\tilde{g}_n(t) - g_0(t)|$ where τ_1^+ is the first jump point after $n^{-1/5}$, and τ_2^+ is the first jump point after $\tau_1^+ + n^{-1/5}$. Hence, for all $t \in [\tau^-, n^{-1/5}]$, we have

$$\begin{aligned}
&\tilde{g}_n(t) - g_0(t) \\
&\le [\tilde{g}_n(\xi_{-1}) - g_0(\xi_{-1})] + \frac{[\tilde{g}_n(\xi_1) - g_0(\xi_1)] - [\tilde{g}_n(\xi_{-1}) - g_0(\xi_{-1})]}{\xi_1 - \xi_{-1}}(t - \xi_{-1}) \\
&\quad + g_0(\xi_{-1}) - g_0(t) + \frac{g_0(\xi_1) - g_0(\xi_{-1})}{\xi_1 - \xi_{-1}}(t - \xi_{-1}).
\end{aligned}$$

But note that

$$\frac{g_0(\xi_1) - g_0(\xi_{-1})}{\xi_1 - \xi_{-1}} \le g_0'(\xi_1), \quad \text{and} \quad g_0(\xi_{-1}) + g_0'(\xi_{-1})(t - \xi_{-1}) \le g_0(t),$$

and hence

$$\begin{aligned}
&\tilde{g}_n(t) - g_0(t) \\
&\le [\tilde{g}_n(\xi_{-1}) - g_0(\xi_{-1})] + \frac{[\tilde{g}_n(\xi_1) - g_0(\xi_1)] - [\tilde{g}_n(\xi_{-1}) - g_0(\xi_{-1})]}{\xi_1 - \xi_{-1}}(t - \xi_{-1}) \\
&\quad + \left[g_0'(\xi_1) - g_0'(\xi_{-1})\right](t - \xi_{-1}) \\
&\le \inf_{[\tau_1, \tau_2]} |\tilde{g}_n(t) - g_0(t)| + \frac{\inf_{[\tau_1, \tau_2]} |\tilde{g}_n(t) - g_0(t)| + \inf_{[\tau_1^+, \tau_2^+]} |\tilde{g}_n(t) - g_0(t)|}{n^{-1/5}} \, n^{-1/5} \\
&\quad + K B_u \, n^{-2/5}
\end{aligned}$$

since $\exists\, \theta \in [\xi_{-1}, \xi_1]$ such that $g_0'(\xi_1) - g_0'(\xi_{-1}) = (\xi_1 - \xi_{-1}) g_0''(\theta)$. Now, by using Lemma 6.2, we can find $c > 0$ such that

$$P(\forall\, t \in [\tau^-, n^{-1/5}]\ \tilde{g}_n(t) - g_0(t) < c n^{-2/5}, \mathcal{T}) \ge P(\mathcal{T}) - \epsilon \tag{6.27}$$

for n sufficiently large.

<u>Case 2</u>. As before, consider the case where one of the subintervals does not contain any jump point, and let τ_1 and τ_2 be the last and first jump points occurring before and after the end points of the smallest subinterval. Using the inequality in (3.13), Lemma 6.1 and Lemma 6.3 combined with fact that $\tau_2 - \tau_1 > n^{-1/5}/7$ (exactly as in A. Case 2), we can find $c > 0$ such that

$$P(\forall t \in [\tau^-, n^{-1/5}]\ \ \tilde{g}_n(t) - g_0(t) \le c n^{-2/5}, \mathcal{T}^c) \ge P(\mathcal{T}^c) - \epsilon. \tag{6.28}$$

Now, combining (6.27) and (6.28) gives the result.

D. Uniform tightness of $n^{2/5}(\tilde{g}_n - g_0)$ from below: We show finally that there exists $c > 0$ such that

$$P\Big(\forall t \in [\tau^-, n^{-1/5}] \ \ \tilde{g}_n(t) - g_0(t) \geq -cn^{-2/5}\Big) > 1 - \epsilon \qquad (6.29)$$

for n large enough. With ξ_1 as above in C. Case 1; i.e., ξ_1 is the minimizer of $|\tilde{g}_n - g_0|$ in $[\tau_1^+, \tau_2^+]$, where τ_1^+ is the first jump after $n^{-1/5}$ and τ_2^+ is the first jump after $\tau_1^+ + n^{-1/5}$. we have by convexity

$$\begin{aligned}
\tilde{g}_n(t) &\geq \tilde{g}_n(\xi_1) + \tilde{g}_n'(\xi_1)(t - \xi_1) \\
&= g_0(\xi_1) + g_0'(\xi_1)(t - \xi_1) + [\tilde{g}_n(\xi_1) - g_0(\xi_1)] + \left[\tilde{g}_n'(\xi_1) - g_0'(\xi_1)\right](t - \xi_1) \\
&\geq g_0(t) - \frac{1}{2}(t - \xi_1)^2 g_0''(\theta_t) - c'n^{-2/5} - c'n^{-1/5}n^{-1/5}, \quad \theta_t \in [t, \xi_1] \quad (6.30) \\
&\geq g_0(t) - \frac{1}{2}B_u K^2 n^{-2/5} - 2c'n^{-2/5} = g_0(t) - \left(\frac{1}{2}B_u K^2 + 2c'\right) n^{-2/5}
\end{aligned}$$

with probability greater than $1-\epsilon$. It suffices to take $c = \frac{1}{2}B_u K^2 + 2c'$. In (6.30), we used the fact that ξ_1 is bounded with increasing probability by $Mn^{-1/5}$ for some $M > 0$ large enough (this follows from Lemma 6.3), and the uniform $n^{1/5}-$ tightness of $\tilde{g}_n'$ established above. □

6.2 Proof of Theorem 4.1

For $t > 0$, we define the local processes $\mathbb{Y}_n^{loc}$ and $\tilde{H}_n^{loc}$ as

$$\begin{aligned}
\mathbb{Y}_n^{loc}(t) &= n^{4/5} \int_0^{tn^{-1/5}} \left(\mathbb{G}_n(u) - \int_0^u \left[g_0(0) + g_0'(0)s\right] ds\right) du \\
&= n^{4/5} \int_0^{tn^{-1/5}} \left(\left[\mathbb{G}_n(u) - G_0(u)\right] + \int_0^u \left[g_0(s) - g_0(0) - g_0'(0)s\right] ds\right) du,
\end{aligned}$$

$$\tilde{H}_n^{loc}(t) = n^{4/5} \int_0^{tn^{-1/5}} \int_0^u \left[\tilde{g}_n(s) - g_0(0) - g_0'(0)s\right] ds du,$$

and their rescaled counterparts

$$\mathbb{Y}_n^l(t) = k_1 \mathbb{Y}_n^{loc}(k_2 t), \ \ \tilde{H}_n^l(t) = k_1 \tilde{H}_n^{loc}(k_2 t)$$

where k_2 is defined in (4.16) and

$$k_1 = \frac{1}{k_2^2 c_0(0)} = \frac{1}{k_2^3 c_1(0)} = \left(24^{-3} g_0^{-4}(0)(g_0''(0))^3\right)^{\frac{1}{5}}.$$

The scaling considered above is necessary to obtain limiting distributions that are independent of the unknown values $g_0(0)$ and $g_0'(0)$, and is obtained by using the well-known scaling property of Brownian motion W.

Now, from the definition of $\mathbb{Y}_n^l(t)$ and $\tilde{H}_n^l$ it follows that

(i) $(\tilde{H}_n^l)^{(2)}$ is convex on $(0, \infty)$

(ii) $\tilde{H}_n^l(t) \geq \mathbb{Y}_n^l(t), \ t \geq 0$

(iii) $\tilde{H}_n^l(t) = \mathbb{Y}_n^l(t)$ if $(\tilde{H}_n^l)^{(2)}$ changes slope at the point t, or in view of (ii)

$$\int_0^\infty (\tilde{H}_n^l(t) - \tilde{Y}_n^l(t)) d(\tilde{H}_n^l)^{(3)}(t)) = 0.$$

Indeed, we have

$$(\tilde{H}_n^l)^{(2)}(t) = k_1 k_2^2 n^{2/5} \left(\tilde{g}_n(k_2 t n^{-1/5}) - g_0(0) - g_0'(0) k_2 t n^{-1/5}\right)$$

which is convex since it is proportional to the sum of the convex function, $\tilde{g}_n(k_2 t n^{-1/5})$, and a straight line. On the other hand, we can write

$$\tilde{H}_n^l(t) - \mathbb{Y}_n^l(t) = k_1 n^{4/5} \left(\tilde{H}_n(k_2 t n^{-1/5}) - \mathbb{Y}_n(k_2 t n^{-1/5})\right)$$

which, combined with the characterization in (2.6), implies (ii) and (iii).

If we can manage to show that $\mathbb{Y}_n^l$ and $\tilde{H}_n^l$ converge weakly to Y_+ and H_{lim}, then the continuous mapping argument of Groeneboom, Jongbloed, and Wellner (2001b) and almost surely uniqueness of the invelope H_+ can be used to conclude that H_{lim} has to be necessarily equal a.s. to H_+. The continuous mapping is a formal a way to show that the conditions (i), (ii) and (iii) satisfied by $\tilde{H}_n^l$ and $\mathbb{Y}_n^l$ carry over to the limit. Fix two constants $0 < M_1 < M_2$. Then the vector $\left(\mathbb{Y}_n^l, (\mathbb{Y}_n^l)^{(1)}, \tilde{H}_n^l, (\tilde{H}_n^l)^{(1)}, (\tilde{H}_n^l)^{(2)}, (\tilde{H}_n^l)^{(3)}\right)$ is tight in $S[M_1, M_2] = C[M_1, M_2] \times D[M_1, M_2] \times C^3[M_1, M_2] \times D[M_1, M_2]$, where $C[a, b]$ and $D[a, b]$ denote the space of continuous functions on $[a, b]$ and that of right-continuous functions on $[a, b]$ with left limits, equipped with the uniform and Skorohod topologies respectively. Indeed, tightness of $(\mathbb{Y}_n^l, (\mathbb{Y}_n^l)^{(1)})$ follows from weak convergence of $(\mathbb{Y}_n^{loc}, (\mathbb{Y}_n^{loc})^{(1)})$, in the space $C[0, M_2] \times D[0, M_2]$, to

$$\left(\sqrt{g_0(0)} \int_0^t W(s)dt + g_0''(0)\frac{t^4}{24}, \quad \sqrt{g_0(0)}W(t) + g_0''(0)\frac{t^3}{6}\right).$$

The Brownian motion part follows from very standard arguments, e.g. using the Hungarian Embedding of Komlós, Major and Tusnády (1975) and the well known representation of a Brownian Bridge $\mathbb{U}(t) = W(t) - tW(1)$, whereas the drift comes from writing

$$\begin{aligned} & n^{4/5} \int_0^{tn^{-1/5}} \int_0^u (g_0(s) - g_0(0) - sg_0'(0))dsdu \\ &= n^{4/5} \int_0^{tn^{-1/5}} \Big(\frac{1}{3!}g_0''(0)u^3 + o(u^3)\Big)du \to g_0''(0)\frac{t^4}{4!} \qquad (6.31) \end{aligned}$$

as $n \to \infty$. To show tightness of $(\tilde{H}_n^l)^{(j)}$ or equivalently tightness of $(\tilde{H}_n^{loc})^{(j)}$ for $j = 0, 1, 2, 3$, consider again τ^- to be the last jump point of $\tilde{g}_n'$ before $M_1 n^{-1/5}$. We can write

$$\begin{aligned} \tilde{H}_n^{loc}(t) &= n^{4/5} \int_0^{tn^{-1/5}} \int_0^u \Big[\tilde{g}_n(s) - g_0(0) - g_0'(0)s\Big]dsdu \\ &= n^{4/5} \int_0^{\tau^-} \int_0^u \Big[\tilde{g}_n(s) - g_0(0) - g_0'(0)s\Big]dsdu \\ &\quad + n^{4/5} \int_{\tau^-}^{tn^{-1/5}} \int_0^u \Big[\tilde{g}_n(s) - g_0(0) - g_0'(0)s\Big]dsdu \\ &= \tilde{H}_n^{loc}(n^{1/5}\tau^-) + n^{1/5}(tn^{-1/5} - \tau^-)(\tilde{H}_n^{loc})'(n^{1/5}\tau^-) \\ &\quad + n^{4/5} \int_{\tau^-}^{tn^{-1/5}} \int_{\tau^-}^u \Big[\tilde{g}_n(s) - g_0(0) - g_0'(0)s\Big]dsdu \\ &= \mathbb{Y}_n^{loc}(n^{1/5}\tau^-) + n^{1/5}(tn^{-1/5} - \tau^-)(\mathbb{Y}_n^{loc})'(n^{1/5}\tau^-) \\ &\quad + n^{4/5} \int_{\tau^-}^{tn^{-1/5}} \int_{\tau^-}^u \Big[\tilde{g}_n(s) - g_0(0) - g_0'(0)s\Big]dsdu. \end{aligned}$$

Above, we used the fact that $(\tilde{H}_n^{loc})'(n^{1/5}\tau^-) = (\mathbb{Y}_n^{loc})'(n^{1/5}\tau^-)$. This equality, which is equivalent to $\tilde{G}_n(\tau^-) = \mathbb{G}_n(\tau^-)$ where $\tilde{G}_n = \tilde{H}_n'$ and $\mathbb{G}_n = \mathbb{Y}_n'$, follows from (2.7) since τ^- is a jump point of $\tilde{g}_n'$.

Now, by Proposition 6.1, we can find $c > 0$ such that for all $t \in [M_1, M_2]$, the event

$$\left| \int_{\tau^-}^{tn^{-1/5}} \int_{\tau^-}^u \Big[\tilde{g}_n(s) - g_0(0) - g_0'(0)s\Big]dsdu \right| \le c\ (M_2 - M_1)^2\ n^{-\frac{2}{5}} n^{-\frac{2}{5}} = c' n^{-\frac{4}{5}}$$

occurs with probability greater than $1-\epsilon$. By tightness of $\mathbb{Y}_n^{loc}(n^{1/5}\tau^-)$ and $(\mathbb{Y}_n^{loc})'(n^{1/5}\tau^-)$ and the fact that $0 < n^{1/5}(tn^{-1/5}-\tau^-) < M_2$, we conclude that $\tilde{H}_n^{loc}$ and hence $\tilde{H}_n^l$ is tight in $C[M_1, M_2]$. Also, we have

$$(\tilde{H}_n^{loc})'(t) = (\mathbb{Y}_n^{loc})'(n^{1/5}\tau^-) + n^{3/5}\int_{\tau^-}^{tn^{-1/5}} \Big[\tilde{g}_n(u) - g_0(0) - g_0'(0)u\Big] du$$

which can be shown to be tight using similar arguments. Finally, tightness of $(\tilde{H}_n^l)''$ and $(\tilde{H}_n^l)^{(3)}$ in $C[M_1, M_2]$ and $D[M_1, M_2]$ respectively follows directly from the same proposition, and using for $(\tilde{H}_n^l)^{(3)}$ the fact that the subset of $D[M_1, M_2]$ consisting of nondecreasing functions absolutely bounded by some $M < \infty$ is compact in the Skorohod topology. We can then extracted subsequence $\Big(\mathbb{Y}_{n'}^l, (\mathbb{Y}_{n'}^l)^{(1)}, \tilde{H}_{n'}^l, (\tilde{H}_{n'}^l)^{(1)}, (\tilde{H}_{n'}^l)^{(2)}, (\tilde{H}_{n'}^l)^{(3)}\Big)$ which converges weakly to $\Big(Y_+, Y_+^{(1)}, H_{lim}, H_{lim}^{(1)}, H_{lim}^{(2)}, H_{lim}^{(3)}\Big)$ in $S[M_1, M_2]$, and this is for all $0 < M_1 < M_2$. Hence, $H_{lim} =_{a.s.} H_+$, which is also the common limit of any extracted subsequence of $\Big(\mathbb{Y}_n^l, (\mathbb{Y}_n^l)^{(1)}, \tilde{H}_n^l, (\tilde{H}_n^l)^{(1)}, (\tilde{H}_n^l)^{(2)}, (\tilde{H}_n^l)^{(3)}\Big)$. Then, it follows that $(\tilde{H}_n^l)^{(j)}$ converges weakly to $H_+^{(j)}$, and we have

$$\begin{aligned}
(\tilde{H}_n^l)^{(2)}(t) &= k_1 k_2^2\, n^{\frac{2}{5}}\Big(\tilde{g}_n(n^{-\frac{1}{5}}\ k_2 t) - g_0(0) - n^{-\frac{1}{5}} k_2 t g_0'(0)\Big)\\
&= \frac{1}{c_0(0)}\ n^{\frac{2}{5}}\ \Big(\tilde{g}_n(n^{-\frac{1}{5}}\ k_2 t) - g_0(n^{-\frac{1}{5}}\ k_2 t) + \frac{1}{2} n^{-\frac{2}{5}} k_2^2 t^2 g_0''(0) + o_p(n^{-\frac{2}{5}})\Big)\\
&\to_d H_+^{(2)}(t),
\end{aligned}$$

$$\begin{aligned}
(\tilde{H}_n^l)^{(3)}(t) &= k_1 k_2^3\, n^{\frac{1}{5}}\Big(\tilde{g}_n'(n^{-\frac{1}{5}}\ k_2 t) - g_0'(0)\Big)\\
&= \frac{1}{c_1(0)}\ n^{\frac{1}{5}}\ \Big(\tilde{g}_n'(n^{-\frac{1}{5}}\ k_2 t) - g_0'(n^{-\frac{1}{5}}\ k_2 t) + n^{-\frac{1}{5}}\ k_2 t g_0''(0) + o_p(n^{-\frac{1}{5}})\Big)\\
&\to_d H_+^{(3)}(t).
\end{aligned}$$

For $\alpha \in (1/5, 1/3)$, the arguments can be constructed very similarly. A detailed proof can be found in BALABDAOUI (2007). In Proposition 6.1, τ^- should be taken now as the last jump point before $M_1 n^{-\alpha}$, $M_2 n^{-1/5}$ should be replaced by $M_2 n^{-\alpha}$. In Lemma 6.1, if the same conditions hold true, then the power of $(y-x)$ should be replaced by $(1+(2k+1)\alpha)/2\alpha$ and $n^{-(3+k)/5}$ by $n^{-(1+(2k+1)\alpha)/2}$. Also, $n^{(1+3\alpha)/2}$ and $n^{-\alpha}$ are respectively the right rate and size of the shrinking neighborhood around 0 to be considered in the definition of the localized processes

$\mathbb{Y}_n^{loc}$ and $\tilde{H}_n^{loc}$. The scaled localized processes are given by $k_3 \mathbb{Y}_n^{loc}$ and $k_3 \tilde{H}_n^{loc}$, with $k_3 = g_0^{-1/2}(0)$. Finally, we would like to note that the absence of the drift in the limiting process Y_0 follows from a similar calculation as in (6.31), where in this case we have $n^{(1+3\alpha)/2} \cdot n^{-4\alpha} \to 0$ as $n \to \infty$.

For $0 < \alpha < 1/5$, we are in the situation where the jump points are clustering around $n^{-\alpha}$. Indeed, for any jump points τ^- and τ^+ in the neighborhood of 0, we know that $\tau^+ - \tau^- = O_p(n^{-1/5})$ (6.19), and we have in this case $n^{-1/5} = o(n^{-\alpha})$. Hence, $n^{-\alpha}$ is playing a very similar role to that of $t_0 > 0$ in the problem of estimating a convex density at an interior point. Thus, the asymptotics in this case can heuristically deduced by replacing t_0 in (1.2) by $n^{-\alpha}$, and the constants $c_0(t_0)$ and $c_1(t_0)$ by $c_0(0) = \lim_{n\to\infty} c_0(n^{-\alpha})$ and $c_1(0) = \lim_{n\to\infty} c_1(n^{-\alpha})$. $\square$

Acknowledgments

The author would like to thank Jon A. Wellner for very stimulating discussions about the problem during a RiP program at Oberwolfach. The author would like to thank also Enno Mammen and Geurt Jongbloed for their comments when this work was presented at a workshop on Inverse Problems in Göttingen.

References

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

Balabdaoui, F. (2007). Consistent estimation of a convex density at the origin: Back to Hampel's birds problem. Technical report available at: *http://ceremade.dauphine.fr/*

Cheng, M.-Y. (1997). A bandwidth selector for local linear density estimators. *Ann. Statist.* **25**, 1001 - 1013.

Cheng, M.-Y., Fan, J. and Marron, J. S. (1997). On automatic boundary corrections. *Ann. Statist.* **25**, 1691 - 1708.

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

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

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

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.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statist. Comput.* **3**, 135 - 146.

Jongbloed, G. (2006). Personal communication.

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

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

Kulikov, V. and Lopuhaä, H. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist.* **34**, 742-768.

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **25**, 378 - 413.

Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov Processes and Martingales*. Wiley, New York.

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

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

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