Limit properties of the monotone rearrangement for density and regression function estimation

Dragi Anevski, Anne-Laure Fougères

Limit properties of the monotone rearrangement for density and regression function estimation

Dragi Anevski and Anne-Laure Fougères
Göteborg University and Université Paris X - Nanterre

25th October 2007

Abstract

The monotone rearrrangement algorithm was introduced by Hardy, Littlewood and Pólya as a sorting device for functions. Assuming that x is a monotone function and that an estimate x_n of x is given, consider the monotone rearrangement $\hat{x}_n$ of x_n. This new estimator is shown to be uniformly consistent. Under suitable assumptions, pointwise limit distribution results for $\hat{x}_n$ are obtained. The framework is general and allows for weakly dependent and long range dependent stationary data. Applications in monotone density and regression function estimation are detailed.

Keywords: Limit distributions, density estimation, regression function estimation, dependence, monotone rearrangement.

1 Introduction

Assume that $(t_i, x(t_i))_{i=1}^n$, for some points $t_i \in [0, 1]$ (e.g. $(t_i = i/n)$), are pairs of data points. The (decreasing) sorting of the points $x(t_i)$ is then an elementary operation and produces the new sorted sequence of pairs $(t_i, y(t_i))$ where $y = sort(x)$ is the sorted vector. Let # denote the counting measure of a set. Then we can define the sorting y of x by

$$\begin{aligned} z(s) &= \#\{t_i : x(t_i) \geq s\} \\ y(t) &= z^{-1}(t), \end{aligned}$$

where z^{-1} denotes the inverse of a function (if the points $x(t_i)$ are not unique it denotes the generalized inverse).

The "sorting" of a function $\{x(t), t \in [0,1]\}$ can then analogously be defined by the monotone rearrangement (cf. Hardy et al. [21]),

$$\begin{aligned} z(s) &= \lambda\{t \in [0,1] : x(t) \geq s\}, \\ y(t) &= z^{-1}(t), \end{aligned}$$

where the counting measure $\#$ has been replaced by the Lebesgue measure λ, and z^{-1} denotes the generalized inverse.

The monotone rearrangement algorithm of a set or a function has mainly been used as a device in analysis, see e.g. Lieb and Loss [22, Chapter 3] or in optimal transportation (see Villani [37, Chapter 3]). Fougères [15] was the first to use the algorithm in a statistical context, for density estimation under order restrictions. Meanwhile, Polonik [28, 29] also developed tools of a similar kind for density estimation for multivariate data. More recently, several authors revisited the monotone rearrangement procedure in the estimation context under monotonicity; see Dette et al. [11], and Chernozhukov et al. [7].

We introduce the following two-step approach for estimating a monotone function. Assume that x is a monotone function on an interval $I \subset \mathbb{R}$. Assume also that we already have an estimate x_n of x, but that this estimate is not necessarily monotone. We then propose to use the monotone rearrangement $\hat{x}_n$ of x_n as an estimate of x.

Under the assumption that we have process limit distribution results for (a localized version of) the stochastic part of x_n and that the deterministic part of x_n is asymptotically differentiable at a fixed point t_0, with strictly negative derivative, we obtain pointwise limit distribution results for $\hat{x}_n(t_0)$. The framework is general and allows for weakly dependent as well as long range dependent data. This is the topic for Section 3.

Possible applications of the general results are to monotone density and regression function estimation, which we explore in more detail in Section 4.

These are the problems of estimating f and m respectively in

(i) $t_1, \ldots, t_n$ stationary observations with marginal decreasing density f on $\mathbb{R}^+$,

(ii) (t_i, y_i) observations from $y_i = m(t_i) + \epsilon_i$, $t_i = i/n, i = 1, \ldots, n, m$ decreasing on $[0, 1]$, $\{\epsilon_i\}$ stationary sequence with mean zero.

The standard approaches in these two problems have been isotonic regression for the regression problem, first studied by Brunk [6], and (nonparametric) Maximum Likelihood estimation (NPMLE) for the density estimation problem, first introduced by Grenander [18]. A wide literature exists for regression and density estimation under order restrictions. One can refer e.g. to Mukerjee [26], Ramsay [31], Mammen [23], Hall and Huang [19], Mammen et al. [24], Gijbels [17], Birke and Dette [5], Dette and Pilz [12], Dette et al. [11] for the regression context. Besides, see Eggermont and Lariccia [14], Fougères [15], Hall and Kang [20], Meyer and Woodroofe [25], Polonik [28], Van der Vaart and Van der Laan [35], among others, for a focus on monotone (or unimodal) density estimation. Anevski and Hössjer [2] gave a general approach unifying both contexts.

Using kernel estimators as preliminary estimators of f and m on which the monotone rearrangement is then applied, we are able to derive limit distribution results for quite general dependence situations, demanding essentially stationarity for the underlying random parts $\{t_i\}$ and $\{\epsilon_i\}$ respectively. The results are however stated in a form that allows for other estimators than the kernel based as starting points, e.g. wavelet or splines estimators.

The paper is organized as follows: In Section 2 we define the monotone rearrangement algorithm and derive some simple properties that will be used in the sequel. In particular our definition differs slightly from Hardy, Littlewood and Polya's original definition [21]; the difference is motivated by the fact that we will use localization and restriction. The most important properties for the algorithm that are derived are the equivariance under addition of constants, the continuity of the map and a certain localization property, cf. Lemma 3, Theorem 1 and Theorem 2 below. Furthermore we state conditions that allow for the extension of the map to unbounded intervals.

In Section 3 we define the generic estimator of the monotone function, and state the consistency and limit distribution properties for the estimator.

The limit distribution is given in Theorem 4 and is of the general form

$$d_n^{-1}\left[\hat{x}_n(t_0) - x(t_0)\right] \stackrel{\mathcal{L}}{\to} T\left(A\cdot + \tilde{v}(\cdot;t_0)\right)(0) + \Delta,$$

where T is the monotone rearrangement map, $\Delta = \lim_{n\to\infty} d_n^{-1}[\mathbb{E}\{x_n(t_0 + sd_n)\} - x(t_0)]$ is the asymptotic local bias of the preliminary estimator and $\tilde{v}(s;t_0) \stackrel{\mathcal{L}}{=} \lim_{n\to\infty} d_n^{-1}\left[x_n(t_0+sd_n) - \mathbb{E}\{x_n(t_0+sd_n)\}\right]$ is the weak local limit of the process part of the preliminary estimator; here $d_n \downarrow 0$ is a deterministic sequence that is determined by the dependence structure of the data.

In Section 4 we apply the obtained results in Section 3 to regression function estimation and density estimation under order restrictions, and derive the limit distributions for the estimators. This gives rise to some new universal limit random variables, such as e.g. in the regression context $T(s + B(s))(0)$ with T the monotone rearrangement map and B standard two sided Brownian motion for independent and weakly dependent data, or $T(s + B_{1,\beta}(s))(0)$ with $B_{1,\beta}$ fractional Brownian motion with self similarity parameter β, when data are long range dependent. The rate of convergence d_n is e.g. for the regression problem the optimal $n^{-1/3}$ in the i.i.d. and weakly dependent data context and of a non-polynomial rate in the long range dependent context, similarly to previously obtained results in isotonic regression for long range dependent data, cf. Anevski and Hössjer [2].

In the appendix we derive some useful but technical results on maximal bounds on the rescaled process parts in the density and regression estimation problems, i.e. for the local partial sum process and empirical processes, for weakly dependent as well as long range dependent data.

2 The monotone rearrangement algorithm

Consider an interval $I \subset \mathbb{R}$, and let $\mathcal{B}(I) = \{f : f(I) \text{ bounded}\}$ and $\mathcal{D}(I) = \{f : f \text{ decreasing on } I\}$. For each Borel set A of $\mathbb{R}$, denote $\lambda(A)$ the Lebesgue measure of A on $\mathbb{R}$. In a first step, the monotone rearrangement will be defined for finite intervals I, and some extensions for infinite I will be discussed in a second step.

2.1 Definition and properties for finite intervals

Definition 1. *Let $I \subset \mathbb{R}$ be a finite interval, and assume $f \in \mathcal{B}(I)$. Let $r_{f,I}$ be the right continuous map from $f(I)$ to $\mathbb{R}^+$, called "upper level set function"* *of f and defined for each $u \in f(I)$ by*

$$r_{f,I}(u) \ := \ \lambda\{t \in I : \ f(t) > u\} = \lambda\{I \cap f^{-1}(u,\infty)\}.$$

The monotone rearrangement map $T_I : \mathcal{B}(I) \ni f \mapsto T_I(f) \in \mathcal{D}(I)$ is defined up to a translation as the (right continuous) generalized inverse of the upper level set function

$$T_I(f)(t) \ := \ \inf\{u \in f(I) : r_{f,I}(u) \leq t - \inf I\}, \tag{1}$$

for $t \in I$.

The following lemmas are listing some simple and useful properties of the maps $u \mapsto r_{f,I}(u)$, $f \mapsto r_{f,I}$ and $f \mapsto T_I(f)$ respectively.

Lemma 1. *Assume $I \subset \mathbb{R}$ is a finite interval, and $f \in \mathcal{B}(I)$. Then*

- (i) *If f has no flat regions on I, i.e. $\lambda\{I \cap f^{-1}(u)\} = 0$ for all $u \in f(I)$, then $r_{f,I}$ is continuous,*
- (ii) *If there is a $u_0 \in f(I)$ such that $\lambda\{I \cap f^{-1}(u_0)\} = c > 0$ then $r_{f,I}$ has a discontinuity at u_0 of height c,*
- (iii) *If f has a discontinuity at $t_0 \in I$ and f is decreasing, then $r_{f,I}$ admits a flat region with level t_0.*

Proof Assertions (i) and (ii) are both consequences of the fact that

$$\begin{aligned}\lim_{u \to u_0} |r_{f,I}(u) - r_{f,I}(u_0)| \ &= \ \lim_{u \to u_0} \lambda\{t \in I : \max(u,u_0) \geq f(t) > \min(u,u_0)\} \\ &= \ \lambda\{I \cap f^{-1}(u_0)\},\end{aligned}$$

which is equal to 0 in (i), and to c in (ii). Finally, assertion (iii) arises from writing that $r_{f,I}(u) = r_{f,I}(f(t_0^-)) = t_0$ for each $u \in (f(t_0^+), f(t_0^-))$. $\square$

Lemma 2. *Let $I \subset \mathbb{R}$ be a finite interval, and assume $f \in \mathcal{B}(I)$. Then*

- (i) *If c is a constant then $r_{f+c,I}(u) = r_{f,I}(u-c)$, for each $u \in f(I)+c$.*
- (ii) *$r_{cf,I}(u) = r_{f,I}(u/c)$ if $c > 0$, for each $u \in cf(I)$.*
- (iii) *$f \leq g \Rightarrow r_{f,I} \leq r_{g,I}$.*
- (iv) *Let $f_c(t) = f(tc)$. Then $c\, r_{f_c,I} = r_{f,I}$.*
- (v) *Let $f_c(t) = f(t+c)$. Then $r_{f_c,I} = r_{f,I}$.*

Proof (i)-(iii) follow from the definition; indeed, for each $u \in f(I)+c$, $r_{f+c,I}(u) = \lambda\{t \in I : f(t)+c > u\} = r_{f,I}(u-c)$, and for each $u \in cf(I)$, $r_{cf,I}(u) = \lambda\{t \in I : cf(t) > u\} = r_{f,I}(u/c)$ if $c > 0$. As for (iii), $\{t \in I : f(t) > u\} \subset \{t \in I : g(t) > u\}$, for each fixed u, if $f \leq g$. Statement (iv) follows from $r_{f_c,I}(u) = \lambda\{t \in I/c : f(ct) > u\} = \lambda\{s/c \in I/c : f(s) > u\} = r_{f,I}(u)/c$, for each $u \in f(I)$. Statement (v) is a consequence of $r_{f_c,I}(u) = \lambda\{t \in I-c : f(t+c) > u\} = \lambda\{s-c \in I-c : f(s) > u\} = \lambda\{t \in I : f(t) > u\}$, for each $u \in f(I)$. $\square$

Lemma 3. *Let $I \subset \mathbb{R}$ be a finite interval and assume f, g are functions in $\mathcal{B}(I)$. The monotone rearrangement map T_I satisfies the following:*

$$\begin{array}{rl} (i) & T_I(f+c) = T_I(f) + c, \ \textit{if } c \textit{ is a constant;} \\ (ii) & T_I(cf) = cT_I(f), \ \textit{if } c > 0 \textit{ is a constant;} \\ (iii) & f \leq g \Rightarrow T_I(f) \leq T_I(g); \\ (iv) & \textit{Let } f_c(t) = f(ct); \ \textit{then } T_{I/c}(f_c)(t) = T_I(f)(ct); \\ (v) & \textit{Let } f_c(t) = f(t+c); \ \textit{then } T_{I-c}(f_c)(t) = T_I(f)(t+c). \end{array}$$

Proof Let $I = [a,b]$; each assertion is a consequence of its counterpart in Lemma 2. Let $t \in I$; statement (i) follows from $T_I(f+c)(t) = \inf\{u \in f(I)+c : r_{f,I}(u-c) \leq t-a\} = T_I(f)(t)+c$, whereas (ii) comes from $T_I(cf)(t) = \inf\{u \in cf(I) : r_{f,I}(u/c) \leq t-a\} = cT_I(f)(t)$. To show (iii), note that $f \leq g \Rightarrow r_{f,I} \leq r_{g,I} \Rightarrow T_I(f) \leq T_I(g)$. Assertion (iv) follows from the fact that for each $t \in I/c$, $T_{I/c}(f_c)(t) = \inf\{u \in f(I) : r_{f,I}(u) \leq ct-a\} = T_I(f)(ct)$. Finally, statement (v) follows since for each $t \in I-c$, $T_{I-c}(f_c)(t) = \inf\{u \in f(I) : r_{f,I}(u) \leq t+c-a\} = T_I(f)(t+c)$. $\square$

The previous result implies that the map T_I is continuous, as stated in the following theorem.

Theorem 1. *Let $||\cdot||$ be an arbitrary norm on $\mathcal{B}(I)$. Then the map T_I is a contraction, i.e. $||T_I(f)-T_I(g)|| \leq ||f-g||$. In particular, T_I is a continuous map, i.e. for all $f_n, f \in \mathcal{B}(I)$,*

$$||f_n - f|| \to 0 \quad \Rightarrow \quad ||T_I(f_n) - T_I(f)|| \to 0,$$

as n tends to infinity.

Proof Let f, g be functions in $\mathcal{B}(I)$. Clearly $g(u) - ||f-g|| \leq f(u) \leq g(u) + ||f-g||$, which by Lemma 3 ($i$) and ($iii$) implies that $T_I(g)(u) - ||f-g|| \leq T_I(f)(u) \leq T_I(g)(u) + ||f-g||$, so that $|T_I(f) - T_I(g)|(u) \leq ||f-g||$, for each u. Since the right hand side is independent of u, the absolute value on the left hand side can be replaced by the norm, which implies the statement of the theorem. □

Remark 1. *One can also refer to Lieb and Loss [22, Theorem 3.5] for a proof of the contraction property (the "non expansivity" property of the map T_I), for the L^p-norms.* □

2.2 Extension to infinite intervals

It is not possible to define a monotone rearrangement on an infinite interval I or on $\mathbb{R}$ for *any* function $\varphi \in \mathcal{B}(\mathbb{R})$. This can however be done for positive functions f such that for each $u > 0$, $r_f(u) := \lambda\{t \in \mathbb{R} : f(t) > u\} < +\infty$, defining in this situation

$$T(f)(t) := \inf\{u \in \mathbb{R}_+ : r_f(u) \leq t\},$$

for each positive t. Such a definition is precisely the definition considered by Hardy et al. [21, Chapter 10.12], and is in particular valid for densities $f \in \mathcal{B}(\mathbb{R})$ (see also Lieb and Loss [22, Chapter 3] and Fougères [15]).

If it remains impossible to define $T(\varphi)$ for any function φ for which $r_\varphi(u)$ is possibly infinite for some positive u, such a definition can be given *locally* around a fixed point $x \in I_0$, where I_0 is a finite interval, as soon as the function φ satisfies the following property:

> There exists a constant $M < \infty$ and a finite interval I_1 including I_0 such that

$$\inf_{t \in (\inf I_1, \sup I_0)} \varphi(t) \geq -M \quad \text{and} \quad \sup_{t \in (\sup I_1, \infty)} \varphi(t) \leq -M, \tag{2}$$

$$\inf_{t \in (-\infty, \inf I_1)} \varphi(t) \geq +M \quad \text{and} \quad \sup_{t \in (\inf I_0, \sup I_1)} \varphi(t) \leq +M. \tag{3}$$

Theorem 2. *Let I_0 be a finite and fixed interval, and let $\varphi \in \mathcal{B}(\mathbb{R})$ such that (2) and (3) are satisfied. Then for any finite interval J containing I_1, one has $T_J(\varphi) \equiv T_{I_1}(\varphi)$ on I_0.*

Proof Define $y_1 := \inf\{y \in I_1 : \forall x \in [\inf J, y[\ \ \varphi(x) > \varphi(y)\}$ and $z_0 := \inf\{x \in J : \varphi(x) \in \varphi(I_0)\}$. If follows from those definitions, from the left part of (3) and from the continuity of φ that $y_1 \in I_1$, $z_0 \in I_1$ and $y_1 < z_0 \leq \inf I_0$. As a consequence, one has:

$$\begin{aligned} r_{\varphi,J}\{\varphi(y_1)\} &= \lambda\{t \in J : \varphi(t) > \varphi(y_1)\} \\ &= y_1 - \inf J + \lambda\{t \in I_1 \cap (y_1, \infty) : \varphi(t) > \varphi(y_1)\}, \end{aligned}$$

where the second equality comes from splitting J into $J \cap (-\infty, y_1)$ and $J \cap (y_1, \infty)$, and using the right part of (2). Similarly, one has

$$r_{\varphi,I_1}\{\varphi(y_1)\} = y_1 - \inf I_1 + \lambda\{t \in I_1 \cap (y_1, \infty) : \varphi(t) > \varphi(y_1)\},$$

so that the following equality holds:

$$r_{\varphi,J}\{\varphi(y_1)\} + \inf J = r_{\varphi,I_1}\{\varphi(y_1)\} + \inf I_1. \tag{4}$$

Now, define

$$y_\star := r_{\varphi,J}\{\varphi(y_1)\} + \inf J = r_{\varphi,I_1}\{\varphi(y_1)\} + \inf I_1.$$

It follows from this definition that $T_J(\varphi)(y_\star) = \varphi(y_1) = T_{I_1}(\varphi)(y_\star)$. Besides, $y_\star \leq \inf I_0$. To prove this, note that $T_J(\varphi)(\inf I_0) \leq M$ because of the right parts of (2) and (3); so $y_\star \leq \inf I_0$ will follow as soon as $T_J(\varphi)(y_\star) \geq M$, since $T_J(\varphi)$ is a decreasing function. This last inequality can be proved easily by contradiction, using jointly that $T_J(\varphi)(y_\star) = \varphi(y_1)$ and the left part of (3).

Finally, let us check that if both functions $T_J(\varphi)$ and $T_{I_1}(\varphi)$ cross at one point (say, $y_\star$), then they will coincide for each point $\sup I_0 \geq x \geq y_\star$: Under the hypothesis that they cross at $y_\star$, it is equivalent to show that for each $-M \leq u \leq \varphi(y_1)$, one gets

$$r_{\varphi,J}(u) + \inf J = r_{\varphi,I_1}(u) + \inf I_1. \tag{5}$$

Let $u \in [-M, \varphi(y_1)]$, and write on one hand

$$\begin{aligned} r_{\varphi,J}(u) &= \lambda\{t \in J : \varphi(t) > u\} \\ &= \lambda\{t \in J : \varphi(t) > \varphi(y_1)\} + \lambda\{t \in J : \varphi(y_1) \geq \varphi(t) > u\} \\ &= r_{\varphi,J}\{\varphi(y_1)\} + \lambda\{t \in J \cap (y_1, \infty) : \varphi(y_1) \geq \varphi(t) > u\} \\ &= r_{\varphi,I_1}\{\varphi(y_1)\} + \inf I_1 - \inf J \\ &\quad + \lambda\{t \in I_1 \cap (y_1, \infty) : \varphi(y_1) \geq \varphi(t) > u\}, \end{aligned}$$

where the last equality follows from (4) and the right part of (2). On the other hand,

$$\begin{aligned} r_{\varphi,I_1}(u) &= r_{\varphi,I_1}\{\varphi(y_1)\} + \lambda\{t \in I_1 : \varphi(y_1) \geq \varphi(t) > u\} \\ &= r_{\varphi,I_1}\{\varphi(y_1)\} + \lambda\{t \in I_1 \cap (y_1, \infty) : \varphi(y_1) \geq \varphi(t) > u\}, \end{aligned}$$

so that equality (5) holds, and this concludes the proof of Theorem 2. □

Theorem 2 implies that an extension of the definition of T_I to $I = \mathbb{R}$ can be given for any continuous function $\varphi \in \mathcal{B}(\mathbb{R})$ such that (2) and (3) hold. Indeed, for any finite interval J big enough, $T_J(\varphi)(t)$ does not depend anymore on J, so that one can define, for each $t \in I_0$:

$$T(\varphi)(t) := T_{I_1}(\varphi)(t).$$

A straightforward consequence of this definition is that both Lemma 3, Theorem 1 and Theorem 2 hold for T:

Corollary 1. *Let $I_0 \subset \mathbb{R}$ be a finite and fixed interval. Assume φ is continuous and satisfies (2) and (3). Then*

- (*i*) *T satisfies Lemma 3 with the equalities and inequalities assumed to hold on I_0,*
- (*ii*) *T satisfies Theorem 1 with norm $||\cdot||$ defined on the set of functions on I_0,*
- (*iii*) *Theorem 2 holds with T_J replaced by T.*

3 The monotone estimation procedure

3.1 Definition and first properties

Let x be a function of interest (such as a density function, or a regression function) and assume x is non increasing. Consider an estimator x_n of x constructed from n observations, which is not supposed to be monotone. Typically, x_n can be an estimator based on kernel, wavelets, splines, etc.

Definition 2. *We define as a new estimator of x the monotone rearrangement of x_n, namely $T(x_n)$. This is a non increasing estimator of x.*

Theorem 3. *(i). Assume that* $\{x_n\}_{n\geq1}$ *is a uniformly consistent estimator of* x *(in probability, uniformly on a compact set* $B \subset \mathbb{R}$*). If* x *is non increasing, then* $\{T(x_n)\}_{n\geq1}$ *is a uniformly consistent estimator of* x *(in probability, uniformly on* B*).*

(ii). Assume that $\{x_n\}_{n\geq1}$ *is an estimator that converges in probability in* $\mathbb{L}^p$ *norm to* x*. If* x *is non increasing, then* $\{T(x_n)\}_{n\geq1}$ *converges in probability in* $\mathbb{L}^p$ *norm to* x*.*

Proof Both (i) and (ii) follow from the fact that $||x|| = \sup_{t\in K}|x(t)|$ is a norm, and T a contraction with respect to $||\cdot||$, by Theorem 1. Moreover $T(x) = x$ if x is non increasing. □

Remark 2. *The strong convergence in* $\mathbb{L}^p$*-norm of* $T(f_n)$ *to* f*, as a consequence of the corresponding result for* f_n*, was first established in Fougères [15, Theorem 5] in the case when* f_n *is the kernel estimator of a density function* f*. Chernozhukov et al. [7] give a refinement of the non expansivity property, see their Proposition 1, part 2, providing a bound for the gain done by rearranging* f_n *and examining the multivariate framework as well.*

3.2 Limit distribution results

Let $J \subset \mathbb{R}$ be a finite or infinite interval, and $C(J)$ the set of continuous functions on J. Let x_n be a stochastic process in $C(J)$ and let t_0 be a fixed interior point in J. In this section limit distribution results for the random variable $T(x_n)(t_0)$ will be derived, where T is the monotone rearrangement map. The proof of these results are along the lines of Anevski and Hössjer [2], and their notation will be used for clarity.

Assume that $\{x_n\}_{n\geq1}$ is a sequence of stochastic processes in $C(J)$ and write

$$x_n(t) = x_{b,n}(t) + v_n(t), \tag{6}$$

for $t \in J$. Given a sequence $d_n \downarrow 0$ and an interior point t_0 in J define $J_{n,t_0} = d_n^{-1}(J - t_0)$. Then, for $s \in J_{n,t_0}$, it is possible to rescale the deterministic and stochastic parts of x_n as

$$\begin{aligned}\tilde{w}_n(s;t_0) &= d_n^{-1}\{v_n(t_0+sd_n) - v_n(t_0)\},\\ \tilde{g}_n(s) &= d_n^{-1}\{x_{b,n}(t_0+sd_n) - x_{b,n}(t_0)\}.\end{aligned}$$

which decomposes the rescaling of x_n as

$$d_n^{-1}\{x_n(t_0+sd_n)-x_n(t_0)\} \quad = \quad \tilde{g}_n(s)+\tilde{w}_n(s;t_0).$$

However, due to the fact that the final estimator needs to be centered at the estimand $x(t_0)$ and not at the preliminary estimator $x_n(t_0)$, it is more convenient to introduce the following rescaling

$$\begin{aligned} \tilde{v}_n(s;t_0) &= d_n^{-1}v_n(t_0+sd_n) && (7)\\ &= \tilde{w}_n(s;t_0)+d_n^{-1}v_n(t_0), \\ g_n(s) &= d_n^{-1}\{x_{b,n}(t_0+sd_n)-x(t_0)\} && (8)\\ &= \tilde{g}_n(s)+d_n^{-1}\{x_{b,n}(t_0)-x(t_0)\}, \end{aligned}$$

so that

$$y_n(s) \quad := \quad g_n(s)+\tilde{v}_n(s;t_0)=d_n^{-1}\{x_n(t_0+sd_n)-x(t_0)\}. \tag{9}$$

This definition of the rescaled deterministic and stochastic parts is slightly different from the one in Anevski and Hössjer [2], and is due to the fact that we only treat the case when the preliminary estimator and the final estimator have the same rates of convergence, in which case our definition is more convenient, whereas in Anevski and Hössjer [2] other possibilities occur.

The limit distribution results will be derived using a classical two-step procedure, cf. e.g. Prakasa Rao [30] : A local limit distribution is first obtained, under Assumption 1, stating that the estimator $T(x_n)$ converges weakly in a local and shrinking neighbourhood around a fixed point. Then it is shown, under Assumption 2, that the limit distribution of $T(x_n)$ is entirely determined by its behaviour in this shrinking neighbourhood.

Assumption 1. *There exists a stochastic process $\tilde{v}(\cdot;t_0)\neq 0$ such that*

$$\tilde{v}_n(\cdot;t_0) \quad \stackrel{\mathcal{L}}{\to} \quad \tilde{v}(\cdot;t_0),$$

on $C(-\infty,\infty)$ as $n\to\infty$. The functions $\{x_{b,n}\}_{n\geq 1}$ are monotone and there are constants $A<0$ and $\Delta\in\mathbb{R}$ such that for each $c>0$,

$$\sup_{|s|\leq c}|g_n(s)-(As+\Delta)| \quad \to \quad 0, \tag{10}$$

as $n\to\infty$.

In the applications typically

$$
\begin{aligned}
A &= \lim_{n\to\infty} \frac{\tilde{g}_n(s)}{s} = x'(t_0), \\
\Delta &= \lim_{n\to\infty} d_n^{-1}\{x_{b,n}(t_0) - x(t_0)\},
\end{aligned}
$$

so that A is the local asymptotic linear term and Δ is the local asymptotic bias, both properly normalized, of the preliminary estimator x_n.

Define the (limit) function

$$
y(s) \quad = \quad As + \Delta + \tilde{v}(s; t_0). \tag{11}
$$

Let $\{z_n\}$ be an arbitrary sequence of stochastic processes.

Assumption 2. *Let I_0 be a given compact interval and $\delta > 0$. There exists a positive constant c such that $[-c, c] \supset I_0$ and a finite positive M such that*

$$
\liminf_{n\to\infty} P\left\{ \inf_{s\in(-c,\sup I_0)} z_n(s) \geq -M, \sup_{s\in(c,\infty)} z_n(s) \leq -M \right\} \quad > \quad 1-\delta, \tag{12}
$$

and

$$
\liminf_{n\to\infty} P\left\{ \inf_{s\in(-\infty,-c)} z_n(s) \geq +M, \sup_{s\in(\inf I_0,c)} z_n(s) \leq +M \right\} \quad > \quad 1-\delta. \tag{13}
$$

Denote $T_c = T_{[-c,c]}$ and $T_{c,n} = T_{[t_0-cd_n, t_0+cd_n]}$. The truncation result Theorem 2 has a probabilistic counterpart in the following.

Lemma 4. *Let $\{z_n\}$ satisfy Assumption 2. Let I be a finite interval in $\mathbb{R}$. Then for each compact interval $J \supset [-c, c]$*

$$
\lim_{c\to\infty} \limsup_{n\to\infty} P(\sup_I |T_c(z_n)(\cdot) - T_J(z_n)(\cdot)| = 0) \quad = \quad 1.
$$

Proof Let A_n and B_n be the sets for which the probabilities are bounded in (12) and (13), respectively. Then, using Theorem 2 with $I_1 = [-c, c]$ and $I_0 = I$, it follows that $A_n \cap B_n \subset \{\sup_I |T_c(z_n) - T_J(z_n)| = 0\}$ for each compact interval $J \supset [-c, c]$. Since (12) and (13) imply $P(A_n \cap B_n) \geq 1 - 2\delta$ it follows that $\limsup_{n\to\infty} P(\sup_I |T_c(z_n)(\cdot) - T_J(z_n)(\cdot)| = 0) \geq 1 - 2\delta$. Since $\delta > 0$ is arbitrary, taking the limit as $c \to \infty$ of the left hand side of this expression implies the statement of the lemma. $\square$

Note that the previous lemma holds with T_J replaced by T_{I_n} for an arbitrary sequence of intervals I_n growing to $\mathbb{R}$.

Theorem 4. *Let $J \subset \mathbb{R}$ be an interval, and t_0 be a fixed point belonging to the interior of J. Suppose Assumption 1 holds. Assume moreover that Assumption 2 holds for both $\{y_n\}$ and y. Then*

$$d_n^{-1}[T_J(x_n)(t_0) - x(t_0)] \quad \stackrel{\mathcal{L}}{\to} \quad T[A \cdot + \tilde{v}(\,\cdot\,; t_0)](0) + \Delta, \tag{14}$$

as $n \to \infty$.

Proof Let $c > 0$ be fixed. We have

$$\begin{aligned} d_n^{-1}\{T_J(x_n)(t_0) - x(t_0)\} &= d_n^{-1}\{T_J(x_n)(t_0) - T_{c,n}(x_n)(t_0)\} \\ &\quad + d_n^{-1}\{T_{c,n}(x_n)(t_0) - x(t_0)\}. \end{aligned} \tag{15}$$

Let us first consider the second term of the right hand side of (15) and introduce

$$\chi_n(s) \quad := \quad x_n(t_0 + sd_n) = x(t_0) + d_n y_n(s). \tag{16}$$

Applying Lemma 3 (i) and (ii) leads to

$$T_{c,n}(x_n)(t_0 + sd_n) \quad = \quad T_c(\chi_n)(s) = d_n T_c(y_n)(s) + x(t_0),$$

which gives

$$d_n^{-1}\{T_{c,n}(x_n)(t_0) - x(t_0)\} \quad = \quad T_c(y_n)(0).$$

Assumption 1 implies that $y_n \stackrel{\mathcal{L}}{\to} y$ on $C[-c, c]$, with y defined in (11). Applying the continuous mapping theorem on T_c, cf. Theorem 1, proves

$$dn^{-1}\{T_{c,n}(x_n)(t_0) - x(t_0)\} \quad \stackrel{\mathcal{L}}{\to} \quad T_c(y)(0)$$

as $n \to \infty$. Lemma 4 via Assumption 2 with $z_n = y$ implies

$$T_c(y)(0) - T(y)(0) \quad \stackrel{P}{\to} \quad 0$$

as $c \to \infty$.

Next we consider the first term of the right hand side of (15). Let ∇ be a positive and finite constant and denote $A_{n,\nabla} = [t_0 - \nabla d_n, t_0 + \nabla d_n]$. From (16) and Lemma 3 (i, ii) it follows that

$$\sup_{A_{n,\nabla}} d_n^{-1}|T_{c,n}(x_n)(\cdot) - T_J(x_n)(\cdot)| = \sup_{[-\nabla,\nabla]} |T_c(y_n)(\cdot) - T_{J_{n,t_0}}(y_n)(\cdot)|,$$

with y_n as defined in (9). If $J = \mathbb{R}$ (resp. $J \neq \mathbb{R}$), Lemma 4 (resp. note following Lemma 4) can be used with $I = [-\nabla, \nabla]$ to obtain

$$d_n^{-1}\{T_{c,n}(x_n)(t_0) - T_J(x_n)(t_0)\} \quad \stackrel{P}{\rightarrow} \quad 0$$

if we first let $n \to \infty$ and then let $c \to \infty$.

Letting first n and then c tend to infinity in (15), applying Slutsky's theorem and Lemma 3 (i) finishes the proof. □

Remark 3. *The approach for deriving the limit distributions is similar to the general approach in Anevski and Hössjer [2] with a preliminary estimator that is made monotone via the $\mathbb{L}^2$-projection on the space of monotone functions. There are however a few differences:*
– Anevski and Hössjer look at rescaling of an integrated preliminary estimator of the monotone functions, whereas we rescale the estimator directly. Our approach puts a stronger assumption on the asymptotic properties of the preliminary estimator, which is however traded off against weaker conditions on the map T, since we only have to assume that the map T is continuous; had we dealt with rescaling as in Anevski and Hössjer we would have had to prove that the composition $\frac{d}{dt}(\tilde{T})$ (with $\tilde{T}$ defined by $\tilde{T}(F)(t) = \int_0^t T(F')(u)\,du$) is a continuous map, which is generally not true for T equal to the monotone rearrangement map; it is however true, under certain conditions, for $\tilde{T}$ equal to the least concave minorant map (when T becomes the $\mathbb{L}^2$-projection on the space of monotone functions), cf. Proposition 2 in Anevski and Hössjer [2].
– Furthermore, we are able to do rescaling for the preliminary estimator directly since it is a smooth function. On the contrary, for some of the cases treated in Anevski and Hössjer this is not possible, e.g. for the isotonic regression and the NPMLE of a monotone density the rescaled stochastic part is asymptotically white noise. As a consequence our rescaled deterministic function is assumed to be approximated by a linear function, whereas the rescaled deterministic function in Anevski and Hössjer [2] is assumed to be approximated by a convex or concave function.
– Finally, the rescaling is here centered at $x(t_0)$, and not at $x_n(t_0)$, which makes it more convenient to apply the limit distribution result we get. □

4 Applications to nonparametric inference problems

In this section we present estimators of a monotone density function and a monotone regression function. Limit distributions for estimators of a marginal decreasing density f for stationary weakly dependent data with marginal density f as well as of a monotone regression function m with stationary errors, that are weakly or strongly dependent, will be derived.

For the density estimation problem let $\{t_i\}_{i=1}^{\infty}$ denote a stationary process with marginal density function f. Define the empirical distribution function $F_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{t_i \le t\}}$ and the centered empirical process $F_n^0(t) = \frac{1}{n}\sum_{i=1}^{n}(1_{\{t_i \le t\}} - F(t))$. Consider a sequence δ_n such that $\delta_n \downarrow 0, n\delta_n \uparrow \infty$ as $n \to \infty$, and define the centered empirical process locally around t_0 on scale δ_n as

$$\begin{aligned} w_{n,\delta_n}(s;t_0) &= \sigma_{n,\delta_n}^{-1} n\{F_n^0(t_0 + s\delta_n) - F_n^0(t_0)\} \\ &= \sigma_{n,\delta_n}^{-1} \sum_{i=1}^{n} (1_{\{t_i \le t_0 + s\delta_n\}} - 1_{\{t_i \le t_0\}} \\ &\quad -F(t_0 + s\delta_n) + F(t_0)), \end{aligned} \tag{17}$$

where

$$\begin{aligned} \sigma_{n,\delta_n}^2 &= \mathrm{Var}\left[n\left\{F_n^0(t_0+\delta_n) - F_n^0(t_0)\right\}\right] \\ &= \mathrm{Var}\left[\sum_{i=1}^{n}\left\{1_{\{t_0 < t_i \le t_0 + \delta_n\}} - F(t_0+\delta_n) + F(t_0)\right\}\right]. \end{aligned}$$

For the regression function estimation problem let $\{\epsilon_i\}_{i=-\infty}^{\infty}$ be a stationary sequence of random variables with $E(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma^2 < \infty$. Let σ_n^2 =Var($\sum_{i=1}^{n}\epsilon_i$). The two sided partial sum process w_n is defined by

$$w_n(t_i + \frac{1}{2n}) = \begin{cases} \frac{1}{\sigma_n}(\frac{\epsilon_0}{2} + \sum_{j=1}^{i}\epsilon_j), & i = 0, 1, 2, \ldots, \\ \frac{1}{\sigma_n}(-\frac{\epsilon_0}{2} - \sum_{j=i+1}^{-1}\epsilon_j), & i = -1, -2, \ldots, \end{cases}$$

and linearly interpolated between these points. Note that $w_n \in C(\mathbb{R})$.

Let $\mathrm{Cov}(k) = E(\xi_1\xi_{1+k})$ denote the covariance function of a generic stationary sequence $\{\xi_i\}$, and distinguish between three cases (of which **[a]** is a special case of **[b]**.)

[a] Independence: the ϵ_i are independent.

[b] Weak dependence: $\sum_k |\mathrm{Cov}(k)| < \infty$.

[c] Strong (long range) dependence: $\sum_k |\mathrm{Cov}(k)| = \infty$.

Weak dependence can be further formalized using mixing conditions as follows: Define two σ-algebras of a sequence $\{\xi_i\}$ as $\mathcal{F}_k = \sigma\{\xi_i : i \leq k\}$ and $\bar{\mathcal{F}}_k = \sigma\{\xi_i : i \geq k\}$, where $\sigma\{\xi_i : i \in I\}$ denotes the $\sigma-$algebra generated by $\{\xi_i : i \in I\}$. The stationary sequence $\{\xi_i\}$ is said to be "ϕ-mixing" or "α-mixing" respectively if there is a function $\phi(n)$ or $\alpha(n) \to 0$ as $n \to \infty$, such that

$$\begin{aligned} \sup_{A \in \bar{\mathcal{F}}_n} |P(A|\mathcal{F}_0) - P(A)| &\leq \phi(n), \\ \sup_{A \in \mathcal{F}_0, B \in \bar{\mathcal{F}}_n} |P(AB) - P(A)P(B)| &\leq \alpha(n), \end{aligned} \tag{18}$$

respectively.

Long range dependence is usually formalized using subordination or assuming the processes are linear; we will treat only (Gaussian) subordination.

All limit distribution results stated will be for processes in $C(-\infty, \infty)$ with the uniform metric on compact intervals and the Borel σ-algebra.

4.1 Monotone regression function estimation

In this section we introduce an estimator of a monotone regression function. We derive consistency and limit distributions, under general dependence assumptions.

Assume m is a C^1-function on a compact interval $J \subset \mathbb{R}$, say $J = [0, 1]$ for simplicity; let $(y_i, t_i), i = 1, \cdots, n$ be pairs of data satisfying

$$y_i = m(t_i) + \epsilon_i, \tag{19}$$

where $t_i = i/n$.

Define $\bar{y}_n : [1/n, 1] \mapsto \mathbb{R}$ by linear interpolation of the points $\{(t_i, y_i)\}_{i=1}^n$, and let

$$x_n(t) = h^{-1} \int k((t-u)/h)\bar{y}_n(u)\, du, \tag{20}$$

be the Gasser-Müller kernel estimate of $m(t)$, cf. Gasser and Müller [16], where k is a density in $\mathbf{L}^2(\mathbb{R})$ with compact support, for simplicity take $\mathrm{supp}(k) = [-1,1]$. Let h be the bandwidth, for which we assume that $h \to 0, nh \to \infty$.

To define a monotone estimator of m, we put

$$\tilde{m}(t) \quad = \quad T_{[0,1]}(x_n)(t),\ t \in J, \tag{21}$$

where T is the monotone rearrangement map. A straightforward application of Theorem 3 and standard consistency results for regression function estimators imply the following consistency result:

Proposition 1. *The random function $\tilde{m}$ defined by (21) is a uniformly consistent estimator of m in probability uniformly on compact sets, and in probability in $\mathbb{L}^p$ norm.*

Clearly $x_n(t) = x_{b,n}(t) + v_n(t)$, with

$$\begin{aligned} x_{b,n}(t) &= h^{-1} \int k(\frac{t-u}{h}) \bar{m}_n(u)\, du, \\ v_n(t) &= h^{-1} \int k(\frac{t-u}{h}) \bar{\epsilon}_n(u)\, du, \end{aligned} \tag{22}$$

where the functions $\bar{m}_n$ and $\bar{\epsilon}_n$ are obtained by linear interpolation of $\{(t_i, m(t_i))\}_{i=1}^n$ and $\{(t_i, \epsilon_i)\}_{i=1}^n$ respectively. For the deterministic term $x_{b,n}(t) \to x_b(t) = m(t)$, as $n \to \infty$. Note that $\bar{m}_n$, and thus also $x_{b,n}$, is monotone.

Put

$$\bar{w}_n(t) \quad = \quad \frac{n}{\sigma_n} \int_0^t \bar{\epsilon}_n(u)\, du. \tag{23}$$

Since $\mathrm{supp}(k) = [-1,1]$ and if $t \in (1/n + h, 1-h)$, from a partial integration and change of variable we obtain

$$v_n(t) \quad = \quad \frac{\sigma_n}{nh} \int k'(u) \bar{w}_n(t - uh)\, du.$$

It can be shown that $\bar{w}_n$ and w_n are asymptotically equivalent for all dependence structures treated in this paper. Let us now recall how the two sided partial sum process behaves in the different cases of dependence we consider:

[a] When the ϵ_i are independent, we have the classical Donsker theorem, cf. Billingsley [4], implying that

$$w_n \stackrel{\mathcal{L}}{\to} B, \tag{24}$$

as $n \to \infty$, with B a two sided standard Brownian motion on $C(\mathbb{R})$.

[b] Define

$$\kappa^2 = \mathrm{Cov}(0) + 2\sum_{k=1}^{\infty} \mathrm{Cov}(k). \tag{25}$$

Assumption 3. [ϕ **− mixing**] *Assume $\{\epsilon_i\}_{i\in\mathbf{Z}}$ is a stationary ϕ-mixing sequence with $E\epsilon_i = 0$ and $E\epsilon_i^2 < \infty$. Assume further $\sum_{k=1}^{\infty} \phi(k)^{1/2} < \infty$ and $\kappa^2 > 0$ in (25).*

Assumption 4. [α **− mixing**] *Assume $\{\epsilon_i\}_{i\in\mathbf{Z}}$ is a stationary α-mixing sequence with $E\epsilon_i = 0$ and $E\epsilon_i^4 < \infty$, $\kappa^2 > 0$ in (25) and $\sum_{k=1}^{\infty} \alpha(k)^{1/2-\epsilon} < \infty$, for some $\epsilon > 0$.*

Assumption 3 or 4 imply that $\sigma_n^2 \to \kappa^2$ and that Donsker's result (24) is valid, cf. Anevski and Hössjer [2] and references therein.

[c] We model long range dependent data $\{\epsilon_i\}_{i\geq 1}$ using Gaussian subordination: More precisely, we write $\epsilon_i = g(\xi_i)$ with $\{\xi_i\}_{i\in\mathbf{Z}}$ a stationary Gaussian process with mean zero and covariance function $\mathrm{Cov}(k) = E(\xi_i\xi_{i+k})$ such that $\mathrm{Cov}(0) = 1$ and $\mathrm{Cov}(k) = k^{-d} l_0(k)$, with l_0 a slowly varying function at infinity[1] and $0 < d < 1$ fixed. Furthermore $g : \mathbb{R} \mapsto \mathbb{R}$ is a measurable function with $E\{g(\xi_1)^2\} < \infty$. An expansion $g(\xi_i)$ in Hermite polynomials is available

$$g(\xi_i) = \sum_{k=r}^{\infty} \frac{1}{k!} \eta_k h_k(\xi_i),$$

where equality holds as a limit in $L^2(\varphi)$, with φ the standard Gaussian density function. The functions $h_k(t) = t^{-k}(d/dt)^k(t^k e^{-t^2})$ are the Hermite polynomials of order k, the functions

$$\eta_k = E\{g(\xi_1)h_k(\xi_1)\} = \int g(u)h_k(u)\phi(u)\,du,$$

[1] i.e. $l_0(tk)/l_0(t) \to 1$ as $t \to \infty$ for each positive k.

are the $L^2(\varphi)$-projections on h_k, and r is the index of the first non-zero coefficient in the expansion. Assuming that $0 < dr < 1$, the subordinated sequence $\{\epsilon_i\}_{i\geq 1}$ exhibits long range dependence (see e.g. Taqqu [33, 34]), and Taqqu [33] also shows that

$$\sigma_n^{-1} \sum_{i\leq nt} g(\xi_i) \quad \xrightarrow{\mathcal{L}} \quad z_{r,\beta}(t),$$

in $D[0,1]$ equipped with the Skorokhod topology, with variance $\sigma_n^2 = \mathrm{Var}\ \{\sum_{i=1}^n g(\xi_i)\} = \eta_r^2 n^{2-rd} l_1(n)(1+o(1))$, where

$$l_1(k) \quad = \quad \frac{2}{r!(1-rd)(2-rd)} l_0(k)^r. \tag{26}$$

The limit process $z_{r,\beta}$ is in $C[0,1]$ a.s., and is self similar with parameter

$$\beta \quad = \quad 1 - rd/2. \tag{27}$$

The process $z_{1,\beta}(t)$ is fractional Brownian motion, $z_{2,\beta}(t)$ is the Rosenblatt process, and the processes $z_{r,\beta}(t)$ are all non-Gaussian for $r \geq 2$, cf. Taqqu [33]. From these results follows a two sided version of Taqqu's result stating the behavior of the two sided partial sum process:

$$w_n \quad \xrightarrow{\mathcal{L}} \quad B_{r,\beta}, \tag{28}$$

in $D(-\infty, \infty)$, as $n \to \infty$, where $B_{r,\beta}$ are the two sided versions of the processes $z_{r,\beta}$.

In the sequel, rescaling is done at the bandwidth rate, so that $d_n = h$. For $s > 0$, let consider the following rescaled process:

$$\begin{aligned} \tilde{v}_n(s;t) \quad &= \quad d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} \int \bar{w}_{\hat{n}}(h^{-1}t + s - u)k'(u)\,du \\ &\stackrel{\mathcal{L}}{=} \quad d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} \int \bar{w}_{\hat{n}}(s-u)k'(u)\,du, \end{aligned} \tag{29}$$

with $\hat{n} = [nh]$ the integer part of nh, where the last equality holds due to the stationarity (exactly only for $t = t_i$ and asymptotically otherwise). Note that the right hand side holds also for $s < 0$.

With the bandwidth choice $d_n = h$ we obtain a non-trivial limit process $\tilde{v}$; choosing d_n such that $d_n/h \to 0$ leads to a limit "process" equal to a random variable and $d_n/h \to \infty$ to white noise. In the first case the limit distribution of $T(x_n)$ on the scale d_n will be the constant 0, while in the second case it will (formally) be $T(m'(t_0) \cdot +\tilde{v}(\cdot))(0)$ which is not defined (T can not be defined for generalized functions, in the sense of L. Schwartz [32]).

Theorem 5. *Assume m is monotone on $[0,1]$ and for some open interval $I_{t_0} \ni t_0$, $m \in C^1(I_{t_0})$ and $\sup_{t \in I_{t_0}} m'(t) < 0$ with $t_0 \in (0,1)$. Let x_n be the kernel estimate of m defined in (20), with a non-negative and compactly supported kernel k such that k' is bounded, and with bandwidth h specified below. Suppose that one of the following conditions holds.*

- [**a**] *$\{\epsilon_i\}$ are independent and identically distributed, $E\epsilon_i = 0$; $\sigma^2 = Var(\epsilon_i) < \infty$, and $h = an^{-1/3}$, for an arbitrary $a > 0$,*
- [**b**] *Assumption 3 or 4 holds, $\sigma_n^2 = \mathrm{Var}(\sum_{i=1}^n \epsilon_i)$, κ^2 is defined in (25), and $h = an^{-1/3}$, with $a > 0$ an arbitrary constant,*
- [**c**] *$\epsilon_i = g(\xi_i)$ is a long range dependent subordinated Gaussian sequence with parameters d and r, $h = l_2(n;a)n^{-rd/(2+rd)}$ with $a > 0$ and $n \mapsto l_2(n;a)$ is a slowly varying function defined in the proof below.*

Then, correspondingly, we obtain

$$h^{-1}\{\tilde{m}(t_0) - m(t_0)\} \quad \xrightarrow{\mathcal{L}} \quad T[m'(t_0) \cdot +\tilde{v}(\cdot;t_0)](0) + m'(t_0) \int uk(u)\, du,$$

as $n \to \infty$, where $\tilde{m}$ is defined in (21),

$$\tilde{v}(s;t) \quad = \quad c \int w(s-u)k'(u)\, du, \tag{30}$$

and respectively

- [**a**] $w = B$; $c = \sigma a^{-3/2}$,
- [**b**] $w = B$; $c = \kappa a^{-3/2}$,
- [**c**] $w = B_{r,\beta}$; $c = |\eta_r| a$ *(where β defined in (27)).*

Proof Theorem 5 is an application of Theorem 4 in the context of monotone regression function. Assume first that $d_n = h$ is such that

$$d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} \quad = \quad d_n^{-2}n^{-1}\sigma_{\hat{n}} \to c > 0. \tag{31}$$

Then $w_n \stackrel{\mathcal{L}}{\to} w$ in $D(-\infty,\infty)$, using the supnorm over compact intervals metric, under the respective assumptions in [**a**], [**b**] and [**c**]. Besides, note that if k' is bounded and k has compact support, the map

$$C(-\infty,\infty) \ni z(s) \quad \mapsto \quad \int z(s-u)k'(u)\,du \in C(-\infty,\infty)$$

is continuous, in the supnorm over compact intervals metric. Thus, under the assumptions that k' is bounded and k has compact support, the continuous mapping theorem implies that

$$\tilde{v}_n(s;t) \quad \stackrel{\mathcal{L}}{\to} \quad \tilde{v}(s;t), \tag{32}$$

where $\tilde{v}(s;t)$ is defined in (30). This yields the first part of Assumption 1. Furthermore

$$\begin{aligned} \tilde{g}_n(s) &= h^{-1}\int \ell(u)\bar{m}_n(t_0-hu)\,du \\ &= h^{-1}\int \ell(u)m(t_0-hu)\,du + r_n(s), \end{aligned}$$

with $\ell(v)=k(v+s)-k(v)$ and r_n a remainder term. Since

$$\int v^\lambda \ell(v)\,dv \quad = \quad \begin{cases} 0, & \text{if } \lambda=0, \\ -s, & \text{if } \lambda=1, \end{cases}$$

it follows by a Taylor expansion of m around t_0 that the first term converges towards As, with $A=m'(t_0)$. The remainder term is bounded for any $c>0$ as

$$\begin{aligned} \sup_{|s|\le c}|r_n(s)| &\le h^{-1}\sup_{|s|\le c}\int |\ell(u)|\,du \sup_{|u-t_0|\le (c+1)h}|\bar{m}_n(u)-m(u)| \\ &= O(n^{-1}h^{-1}) = o(1). \end{aligned}$$

Furthermore

$$d_n^{-1}\{x_{b,n}(t_0)-m(t_0)\} \quad \to \quad m'(t_0)\int uk(u)\,du =: \Delta, \tag{33}$$

as $n\to\infty$, which proves Assumption 1.

Proof that Assumption 2 holds is relegated to the appendix, see Corollary 2 in Appendix A.1. An application of Theorem 4 then finishes the proof of Theorem 5. It only remains to check whether $d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} \to c > 0$ for the three types of dependence.

- Independent case [**a**]: We have $\sigma_{\hat{n}}^2 = \sigma^2 n d_n$. Thus $d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} = \sigma n^{-1/2}h^{-3/2}$, and (31) is satisfied with $c = \sigma a^{-3/2}$ if $d_n = h = an^{-1/3}$.

- Mixing case [**b**]: The proof is similar to the proof of [**a**], replacing σ by κ.

- Long range data case [**c**]: Since $\sigma_{\hat{n}}^2 = \eta_r^2(nd_n)^{2-rd}l_1(nd_n)$, if we choose $d_n = h$ we will have

$$d_n^{-2}n^{-1}\sigma_{\hat{n}} = d_n^{-2}n^{-1}|\eta_r|(nd_n)^{1-rd/2}l_1(nd_n)^{1/2} \quad \to \quad |\eta_r|a \qquad (34)$$

 if and only if

$$d_n \quad = \quad n^{-rd/(2+rd)}l_2(n;a), \qquad (35)$$

 where l_2 is another function slowly varying at infinity, implicitly defined in (34). Thus (31) follows with $c = |\eta_r|a$ and $h = d_n$ given in (35). $\square$

Remark 4. *The present estimator is similar to the estimator first presented by Mammen [23]: Mammen proposed to do isotonic regression of a kernel estimator of a regression function (using bandwidth $h = n^{-1/5}$), whereas we do monotone rearrangement of a kernel estimator. Mammen's estimator was extended to dependent data and other bandwidth choices by Anevski and Hössjer [2] who derived limit distributions for weak dependent and long range dependent data that are analogous to our results; for the independent data case and bandwidth choice $h = n^{-1/3}$ the limit distributions are similar with rate of convergence $n^{1/3}$ and nonlinear maps of Gaussian processes.*

4.2 Monotone density estimation

In this subsection we introduce a (monotone) estimator of a monotone density function for stationary data, for which we derive consistency and limit distributions.

Let $t_1, t_2, \ldots$ denote a stationary process with marginal density function f lying in the class of decreasing density functions on $\mathbb{R}^+$, and define the following estimator of the marginal decreasing density for the sequence $\{t_i\}_{i\geq 1}$: Consider $x_n(t) = (nh)^{-1}\sum_{i=1}^n k\{(t-t_i)/h\}$ the kernel estimator of the density f, with k a bounded density function supported on $[-1,1]$ such that $\int k'(u)du = 0$, and $h > 0$ the bandwidth (cf. e.g. Wand and Jones [38]), and define the (monotone) density estimate

$$\hat{f}_n(t) \;=\; T(x_n)(t), \tag{36}$$

where T is the monotone rearrangement map. Note that $\hat{f}_n$ is monotone and positive, and integrates to one, cf. equation (4) of Section 3.3. in Lieb and Loss [22].

A straightforward consequence of Theorem 3 and standard convergence results for the kernel density estimate is the following consistency result:

Proposition 2. *The random function $\hat{f}_n$ defined by (36) is a uniformly consistent estimator of f in probability uniformly on compact sets, and in probability in $\mathbb{L}^p$ norm.*

In the following, the limit distributions for $\hat{f}_n$ in the independent and weakly dependent cases are derived. We will in particular make use of recent results on the weak convergence $w_{n,\delta_n} \xrightarrow{\mathcal{L}} w$, on $D(-\infty,\infty)$, as $n \to \infty$, for independent and weakly dependent data $\{t_i\}$, derived in Anevski and Hössjer [2].

The kernel estimator can be written $x_n = x_{b,n} + v_n$ with

$$\begin{aligned} x_n(t) &= h^{-1}\int k'(u)F_n(t-hu)\,du, \\ x_{b,n}(t) &= h^{-1}\int k'(u)F(t-hu)\,du, \\ v_n(t) &= h^{-1}\int k'(u)F_n^0(t-hu)\,du. \end{aligned} \tag{37}$$

Rescaling is done on a scale d_n that is of the same asymptotic order as h, so that we put $d_n = h$. The rescaled process is

$$\tilde{v}_n(s;t_0) \;=\; c_n\int k'(u)w_{n,d_n}(s-u;t_0)\,du,$$

with $c_n = d_n^{-1}(nh)^{-1}\sigma_{n,d_n}$.

Theorem 6. *Let $\{t_i\}_{i\geq 1}$ be a stationary sequence with a monotone marginal density function f such that $\sup_{t\in I_{t_0}} f'(t) < 0$ and $f \in C^1(I_{t_0})$ for an open interval $I_{t_0} \ni t_0$ where $t_0 > 0$. Assume that $\mathbb{E}t_i^5 < \infty$. Let x_n be the kernel density function defined above, with k a bounded and compactly supported density such that k' is bounded. Suppose that one of the following conditions holds:*

[**a**] $\{t_i\}_{i\geq 1}$ *is an i.i.d. sequence,*

[**b**] 1) $\{t_i\}_{i\geq 1}$ *is a stationary ϕ-mixing sequence with* $\sum_{i=1}^{\infty} \phi^{1/2}(i) < \infty$;
2) $f(t_0) = F'(t_0)$ *exists, as well as the joint density* $f_k(s_1, s_2)$ *of* (t_1, t_{1+k}) *on* $[t_0 - \delta, t_0 + \delta]^2$ *for some* $\delta > 0$, *and* $k \geq 1$;
3) $\displaystyle\sum_{k=1}^{\infty} M_k < \infty$ *holds, for* $M_k = \displaystyle\sup_{t_0-\delta\leq s_1,s_2\leq t_0+\delta} |f_k(s_1, s_2) - f(s_1)f(s_2)|$.

Then choosing $h = an^{-1/3}$ and $a > 0$ an arbitrary constant, we obtain

$$n^{1/3}\{\hat{f}_n(t_0) - f(t_0)\} \quad \stackrel{\mathcal{L}}{\to} \quad aT[f'(t_0)\cdot + \tilde{v}(\cdot; t_0)](0) + f'(t_0)\, a \int uk(u)\, du,$$

as $n \to \infty$, where $\tilde{v}(s;t)$ is as in (38), with $c = a^{-3/2} f(t_0)^{1/2}$, and w a standard two sided Brownian motion.

Proof If k' is bounded and k has compact support, the continuity of the map

$$C(-\infty,\infty) \ni z(s) \quad \mapsto \quad \int z(s-u)k'(u)\, du \in C(-\infty,\infty)$$

implies that, choosing d_n such that $c_n \to c$ as $n \to \infty$ for some constant c, one gets:

$$\tilde{v}_n(s;t_0) \quad \stackrel{\mathcal{L}}{\to} \quad c\int k'(u)w(s-u;t_0)\, du =: \tilde{v}(s;t_0), \tag{38}$$

on $C(-\infty,\infty)$, as $n \to \infty$, thanks to the continuous mapping theorem. Here w is the weak limit of $\{w_n\}$. Theorems 7 and 8 of Anevski and Hössjer [2] state that $w_{n,\delta_n}(s, t_0) \stackrel{\mathcal{L}}{\to} B(s)$ on $D(-\infty,\infty)$ under the respective assumptions in [**a**] and [**b**], where $B(s)$ is a two sided standard Brownian motion. This establishes the first part of Assumption 1 for both cases [**a**] and [**b**].

Next notice that $x_{b,n}(t) = h^{-1} \int k(\frac{t-u}{h}) f(u)\, du$ is monotone. A change of variable and a Taylor expansion in $x_{b,n}$ prove the second part of Assumption 1 with $A = f'(t_0)$ and

$$d_n^{-1}\{x_{b,n}(t_0) - f(t_0)\} \quad \to \quad f'(t_0) \int uk(u)\, du = \Delta.$$

The statement of Assumption 2 is relegated to the appendix, see Corollary 3 in Appendix A.2. Theorem 6 therefore holds as an application of Theorem 4.

Let us finally check that the scale d_n can be chosen so that $c_n \to c$, as assumed at the beginning of the proof:

- Independent data case [**a**]: We have $\sigma^2_{n,d_n} \sim n d_n f(t_0)$, so that

  $$d_n^{-1}(nh)^{-1}\sigma_{n,d_n} \quad \sim \quad d_n^{-3/2} n^{-1/2} f(t_0)^{1/2}.$$

  Choosing $d_n = an^{-1/3}$ we get $c = a^{-3/2} f(t_0)^{1/2}$.

- Mixing data case [**b**]: Similar to the proof of case [**a**].

□

Remark 5. *The present estimator was first proposed for independent data by Fougères [15], who stated the strong consistency uniformly over $\mathbb{R}^+$ for $T(f_n)$ and derived some partial results for the limit distribution. The results for the monotone density function estimator are similar to the results for the Grenander estimator (the NPMLE) of a monotone density, in that we have cube root asymptotics and a limit random variable that is a nonlinear functional of a Gaussian process, for independent and weak dependent data; see Prakasa Rao [30] and Wright [39] for the independent data cases, and Anevski and Hössjer [2] for the weak dependent data cases. In our case however we obtain one extra term that arises from the bias in the kernel estimator. Our estimator is really closer in spirit to the estimator obtained by projecting the kernel estimator on the space of monotone functions (i.e. kernel estimation followed by isotonic regression) first proposed by Anevski and Hössjer [2]; note that we obtain the same bias term as in Anevski and Hössjer [2].*

Remark 6. *The results for the long range dependence case is similar to the result for the isotonic regression of a kernel estimator, cf. Anevski and Hössjer [2]. In this situation $\tilde{v}_n(s;t_0)$ is asymptotically a linear function of s with a random slope, implying that the monotone rearrangement of $g_n + \tilde{v}_n$ is just $g_n + \tilde{v}_n$ which evaluated at zero is zero. This is due to the fact that for long range dependent data the limit process of the empirical process is a deterministic function multiplied by a random variable, cf. the remark after Theorem 12 in Anevski and Hössjer [2]. Thus the limit distribution for the final estimator for long range dependent data is the same as the limit distribution for the kernel estimator itself, i.e. $n^{d/2}\{\hat{f}_n(t) - f(t)\}$ and $n^{d/2}\{f_n(t) - f(t)\}$ have the same distributional limit. See Csörgö and Mielniczuk [9] for a derivation of this limit distribution.*

5 Conclusions

We considered the feature of estimating an arbitrary monotone function x, via a monotone rearrangement of a "preliminary" estimator x_n of the unknown x. We derived consistency and limit distribution results for the monotonized estimator that hold under rather general dependence assumptions.

Our approach is similar in spirit to the general methods studied in Anevski and Hössjer [2] and first introduced in the regression estimation setting by Mammen [23]: Start with a preliminary estimator and make it monotone by projecting it on the space of monotone functions. The present approach can however at some point be considered preferable: The monotone rearrangement, being basically a sorting, is a simpler procedure than an $\mathbb{L}^2$-projection. Furthermore the consistency and limit distribution results indicate similar properties to Mammen's and Anevski and Hössjer's estimators. Besides, an important advantage of our estimator is the finite sample behavior: Mammen's estimator is monotone but not necessarily smooth; Mammen actually studied two approaches, one with kernel smoothing followed by monotonization and the other approach the other way around, i.e. monotonization followed by kernel smoothing. Mammen showed that the two proposals are first-order equivalent. However, their finite sample size properties are very different: the first resulting estimator is monotone but not necessarily smooth, while the other is smooth but not necessarily monotone, so that one needs to choose which property is more important. This is not the case with our estimator, since if we start with a smooth estimator of the function, e.g. a kernel estimator, the monotone rearrangement will be smooth as well. This can however become a disadvantage for instance when the estimand is discontinuous: then the monotone rearrangement will "oversmooth" since it will give a continuous result, while Mammen's estimator will keep more of the discontinuity intact.

Some simulation studies are available in the literature, which exhibit the small sample size behavior of the rearrangement of a kernel estimator of a density, and compare it to different competitors. See e.g. Fougères [15], Meyer and Woodroofe [25], Hall and Kang [20], Chernozhukov et al. [7]. These references deal with independent data. A larger panel of dependence situations in the comparisons would clearly be of interest, and this will be the object of future work.

Note that our results are geared towards local estimates, i.e. estimates that use only a subset of the data and that are usually estimators of estimands that can be expressed as non-differentiable maps of the distribution function such as e.g. density functions, regression functions, or spectral density functions. This differs from global estimates, as those considered for example by Chernozhukov et al.[8] for quantile estimation.

An approach similar to ours for local estimates is given in Dette et al. [11], using a modified version of the Hardy-Littlewood-Pólya monotone rearrangement: The first step consists of calculating the upper level set function and is identical to ours. However in the second step they use a smoothed version of the (generalized) inverse, which avoids nonregularity problems for the inverse map. The resulting estimator is therefore not rate-optimal, and the limit distributions are standard Gaussian due to the oversmoothing.

Work has been done here using kernel based methods for the preliminary estimator x_n of x. Other methods, such as wavelet based ones, are possible, and let emphasize that the only assumptions required are given in Assumptions 1 and 2.

We have studied applications to density and regression function estimation. Other estimation problems that are possible to treat with our methods are e.g. spectral density estimation, considered by Anevski and Soulier [3], and deconvolution, previously studied by van Es et al. [36] and Anevski [1].

References

[1] D. Anevski. Deconvolution under monotonicity assumptions. Lund Technical Report 17, Centre for Mathematical Sciences, 1999.

[2] D. Anevski and O. Hössjer. A general asymptotic scheme for inference under order restrictions. *Ann. Statist.*, 34(4):1874–1930, 2006.

[3] D. Anevski and Ph. Soulier. Estimation of a monotone spectral density. Preprint, 2007.

[4] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons Inc., New York, 1968.

[5] M. Birke and H. Dette. Estimating a convex function in nonparametric regression. *Scand. J. Statist.*, 34:384–404, 2007.

[6] H. D. Brunk. On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.*, 29:437–454, 1958.

[7] V. Chernozhukov, I. Fernandez-Val, and A. Galichon. Improving estimates of monotone functions by rearrangement. Submitted, 2007.

[8] V. Chernozhukov, I. Fernandez-Val, and A. Galichon. Quantile and probability curves without crossing. Submitted, 2007.

[9] S. Csrgő and J. Mielniczuk. Density estimation under long-range dependence. *Ann. Statist.*, 23:990–999, 1995.

[10] L. de Haan. On regular variation and its application to the weak convergence of sample extremes. Math. Centre Tracts 32, Math.Centre, Amsterdam, 1970.

[11] H. Dette, N. Neumeyer, and K. F. Pilz. A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3):469–490, 2006.

[12] H. Dette and K. F. Pilz. A comparative study of monotone nonparametric kernel estimates. *J. Stat. Comput. Simul.*, 76(1):41–56, 2006.

[13] P. Doukhan. *Mixing. Properties and examples*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994.

[14] P. P. B. Eggermont and V. N. LaRiccia. Maximum likelihood estimation of smooth monotone and unimodal densities. *Ann. Statist.*, 28(3):922–947, 2000.

[15] A.-L. Fougères. Estimation de densités unimodales. *Canad. J. Statist.*, 25(3):375–387, 1997.

[16] T. Gasser and H-G. Muller. Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11:171–185, 1984.

[17] I. Gijbels. Monotone regression. Discussion paper 0334, Institut de Statistique, Université Catholique de Louvain, 2004.

[18] U. Grenander. On the theory of mortality measurements. Part II. *Skand.Akt.*, 39, 1956.

[19] P. Hall and L.-S. Huang. Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.*, 29(3):624–647, 2001.

[20] P. Hall and K.-H. Kang. Unimodal kernel density estimation by data sharpening. *Statist. Sinica*, 15(1):73–98, 2005.

[21] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge, at the University Press, 1952. 2d ed.

[22] E. H. Lieb and M. Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.

[23] E. Mammen. Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.*, 19(2):741–759, 1991.

[24] E. Mammen, J. S. Marron, B. A. Turlach, and M. P. Wand. A general projection framework for constrained smoothing. *Statist. Sci.*, 16(3):232–248, 2001.

[25] M. C. Meyer and M. Woodroofe. Consistent maximum likelihood estimation of a unimodal density using shape restrictions. *Canad. J. Statist.*, 32(1):85–100, 2004.

[26] H. Mukerjee. Monotone nonparametric regression. *Ann. Statist.*, 16:741–750, 1988.

[27] V.V. Petrov. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford Science Publications. Oxford University Press, USA, 1994.

[28] W. Polonik. Density estimation under qualitative assumptions in higher dimensions. *J. Multivariate Anal.*, 55(1):61–81, 1995.

[29] W. Polonik. The silhouette, concentration functions and ML-density estimation under order restrictions. *Ann. Statist.*, 26(5):1857–1877, 1998.

[30] B.L.S. Prakasa Rao. Estimation of a unimodal density. *Sankhya Ser. A*, 31:23–36, 1969.

[31] J.O. Ramsay. Monotone regression splines in action (with comments). *Statist. Sci.*, 3:425–461, 1988.

[32] L. Schwartz. *Théorie des distributions*. Publications de l'Institut de Mathématiques de l'Université de Strasbourg, No. IX-X. Hermann, Paris, 1966.

[33] M.S. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:287–302, 1975.

[34] M.S. Taqqu. Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 50:53–83, 1979.

[35] A. W. van der Vaart and M. J. van der Laan. Smooth estimation of a monotone density. *Statistics*, 37(3):189–203, 2003.

[36] B. van Es, G. Jongbloed, and M. van Zuijlen. Isotonic inverse estimators for nonparametric deconvolution. *Ann. Statist.*, 26(6):2395–2406, 1998.

[37] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[38] M. P. Wand and M. C. Jones. *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London, 1995.

[39] F.T. Wright. The asymptotic behaviour of monotone regression estimates. *Ann. Statist.*, 9:443–448, 1981.

A Maximal bounds for rescaled partial sum and empirical processes

In this section we derive conditions under which Assumption 2 holds, for the density and regression function estimation cases. Recall that

$$\begin{aligned} \tilde{g}_n(s) &= d_n^{-1}\{x_{b,n}(t_0+sd_n)-x_{b,n}(t_0)\}, \\ \tilde{v}_n(s) &= d_n^{-1}v_n(t_0+sd_n). \end{aligned} \tag{39}$$

Since under Assumption 1

$$\begin{aligned} y_n(s)-\{\tilde{g}_n(s)+\tilde{v}_n(s)\} &= d_n^{-1}\{x_{b,n}(t_0)-x(t_0)\} \\ &\to \Delta, \end{aligned}$$

as $n \to \infty$, and $|\Delta| < \infty$, establishing Assumption 2 for the process $\tilde{g}_n + \tilde{v}_n$ implies that it holds also for the process $y_n = g_n + \tilde{v}_n$. Therefore it is enough to establish Assumption 2 for y_n replaced by $\tilde{g}_n + \tilde{v}_n$.

Recall that for the cases that we cover the rescaled processes are of the form

$$\tilde{v}_n(s;t_0) = c_n \int k'(u) z_n(s-u;t_0)\,du,$$

with $z_n = w_{n,d_n}$ the local rescaled empirical process in the density estimation case and $z_n = w_n$ the partial sum process in the regression case. This implies that for the density estimation case the support of $\tilde{v}_n$ is stochastic, since it depends on $\max_{1\le i\le n} t_i$, while for the regression estimation case it does not depend on the data $\{\epsilon_i\}$ and is as a matter of fact compact and deterministic.

Lemma 5. *Let* $\mathrm{supp}(k) \subset [-1,1]$*. Suppose that Assumption 1 holds. Assume that* t_0 *has a neighbourhood* $I = [t_0-\epsilon, t_0+\epsilon]$ *such that* $\tau := \sup_{t\in I} x'(t) < 0$*. Suppose also that*

$$x'_{b,n}(t+sd_n) \to x'(t), \tag{40}$$

as $n \to \infty$*, for all* $t \in I$*.*

Then (12) and (13) written for $z_n = \tilde{g}_n + \tilde{v}_n$ *are implied by the two results: (A). For every* $\delta > 0$ *and* $0 < c < \infty$ *there is a finite* $M > 0$

$$\liminf_{n\to\infty} P\left[\cap_{s\in(c,d_n^{-1}\epsilon)}\{\tilde{v}_n(s) < \frac{M}{2} - \tau(s-c)\}\right] > 1-\delta.$$

(B). For every $\delta > 0$ and finite $M > 0$ there is a finite C such that for each $c > C$

$$\limsup_{n\to\infty} P\left\{\sup_{s\in d_n^{-1}(0,\ell(n))} \tilde{v}_n(s) > \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \quad < \quad \delta, \tag{41}$$

where $\ell(n)$ is a deterministic function which satisfies either of

$$(i) \qquad \liminf_{n\to\infty} P\{\max_{1\le i\le n} t_i < \ell(n)\} = 1,$$

or

$$(ii) \qquad \ell(n) \equiv \max \mathrm{supp}(x_n) \text{ if } \limsup_{n\to\infty} \max \mathrm{supp}(x_n) \le K < \infty.$$

Condition (A) can be seen as boundedness on small sets (i.e. on the sets $(c, d_n^{-1}\epsilon)$), while the conditions in (B) are bounds outside of small sets; the small sets are really compact (of the form $(0,\epsilon)$) on the t-scale, and are increasing due to the rescaling done for the s-scale.

Condition $(B)(ii)$ is appropriate for the regression function estimation case, since then $\limsup_{n\to\infty} \max(\mathrm{supp}(x_n))$ is bounded by $1+\max(\mathrm{supp}(k)) = 2$, while for the density estimation case we will have to invoke the more subtle assumptions in $(B)(i)$.

Proof In order to show (12), we first prove that for each $\delta > 0$ there is a $0 < M < \infty$ and a $0 < c < \infty$ such that

$$\liminf_{n\to\infty} P\{\sup_{s\ge c}(\tilde{g}_n + \tilde{v}_n)(s) < -M\} \quad \ge \quad 1 - \delta. \tag{42}$$

Let $\tilde{g}_n$ be defined in (39). Consider the function

$$k_n(s) \quad = \quad \begin{cases} \tau s, \text{ on } (-\epsilon d_n^{-1}, \epsilon d_n^{-1}), \\ \tau\epsilon d_n^{-1} \text{ on } (\epsilon d_n^{-1}, \infty), \\ -\tau\epsilon d_n^{-1} \text{ on } (-\infty, -\epsilon d_n^{-1}). \end{cases}$$

Then from (40) we obtain

$$\begin{aligned} \tilde{g}_n(s) &\le k_n(s) \text{ on } \mathbb{R}^+, \\ \tilde{g}_n(s) &\ge k_n(s) \text{ on } \mathbb{R}^-, \end{aligned}$$

for all large enough n, since $\tilde{g}_n$ is decreasing (as weighted mean of decreasing functions) and $\tilde{g}_n(0) = 0$.

Let δ be given and suppose part (A) of the assumptions is satisfied, with some M and arbitrary $0 < c < \infty$. We will consider the hypotheses $(B)(i)$ and $(B)(ii)$ separately:

$(B)(i)$ Since the kernel k has support in $[-1, 1]$ one has $\mathrm{supp}(x_n) \subset (\min_{1\le i\le n} t_i - h, \max_{1\le i\le n} t_i + h)$. Using the rescaling $t = t_0 + sd_n$ this implies that

$$\mathrm{supp}(\tilde{g}_n), \mathrm{supp}(\tilde{v}_n) \quad \subset \quad -t_0 + d_n^{-1}(\min t_i - h, \max t_i + h) =: I_n^{(i)}.$$

Since $t_0 > \min t_i$ and h is positive, the supremum over all $s \in I_n^{(i)}$ can be replaced by a supremum over all $s \in (c, d_n^{-1} \max t_i)$, as n tends to ∞, and thus we need to show

$$\liminf_{n\to\infty} P\left\{\sup_{(c,d_n^{-1}\max t_i)} (\tilde{g}_n + \tilde{v}_n)(s) < -M\right\} \quad \geq \quad 1 - \delta. \tag{43}$$

Then for $c \geq 3M/2|\tau|$, we will have $k_n(c) = -3M/2$. This implies that for $c \geq 3M/2|\tau|$,

$$\begin{aligned} P\left\{\sup_{(c,d_n^{-1}\max t_i)} y_n(s) < -M\right\} \quad \geq \quad & P(\cap_{s\in(c,d_n^{-1}\epsilon)}\{\tilde{v}_n(s) < \frac{M}{2} - \tau(s-c)\} \\ & \cap\{\sup_{s\in d_n^{-1}(\epsilon,\max t_i)} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\}), \end{aligned}$$

so that (43) follows from the two results

$$\liminf_{n\to\infty} P\left\{\sup_{s\in d_n^{-1}(\epsilon,\max t_i)} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \quad > \quad 1 - \delta, \tag{44}$$

$$\liminf_{n\to\infty} P\left[\cap_{s\in(c,d_n^{-1}\epsilon)}\{\tilde{v}_n(s) < \frac{M}{2} - \tau(s-c)\}\right] \quad > \quad 1 - \delta. \tag{45}$$

The relation (45) is satisfied by assumption (A) and thus we need to treat (44). Let ℓ be the deterministic function given in assumption $(B)(i)$. Note first that

$$
\begin{aligned}
& P\left\{\sup_{d_n^{-1}(\epsilon,\ell(n))} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \\
\leq\ & P\left\{\sup_{d_n^{-1}(\epsilon,\ell(n))} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c) | \max_{1\leq i\leq n} t_i < \ell(n)\right\} + P\left\{\max_{1\leq i\leq n} t_i > \ell(n)\right\} \\
<\ & P\left\{\sup_{d_n^{-1}(\epsilon,\ell(n))} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c) | \max_{1\leq i\leq n} t_i < \ell(n)\right\} + \delta
\end{aligned}
$$

for all $n \geq N$ for some N, since $\lim_{n\to\infty} P\left\{\max_{1\leq i\leq n} t_i > \ell(n)\right\} = 0$. Therefore, for all $n \geq N$, we have

$$
\begin{aligned}
& P\left\{\sup_{d_n^{-1}(\epsilon,\max t_i)} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \\
\geq\ & P\left\{\sup_{d_n^{-1}(\epsilon,\ell(n))} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c) \mid \max_{1\leq i\leq n} t_i < \ell(n)\right\} P\left\{\max_{1\leq i\leq n} t_i < \ell(n)\right\} \\
\geq\ & \left(P\left\{\sup_{d_n^{-1}(\epsilon,\ell(n))} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} - \delta\right) P\left\{\max_{1\leq i\leq n} t_i < \ell(n)\right\} \\
\geq\ & \left(P\left\{\sup_{d_n^{-1}(0,\ell(n))} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} - \delta\right) P\left\{\max_{1\leq i\leq n} t_i < \ell(n)\right\}.
\end{aligned}
$$

Thus since $\lim_{n\to\infty} P\left\{\max_{1\leq i\leq n} t_i < \ell(n)\right\} = 1$, taking complements leads to (44) as soon as for $c > C$

$$
\limsup_{n\to\infty} P\left\{\sup_{d_n^{-1}(0,\ell(n))} \tilde{v}_n(s) > \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \quad < \quad \delta,
$$

i.e. (41).

$(B)(ii)$. It follows from the definition of K and from $\text{supp}(k) \subset [-1, 1]$ that $\text{supp}(x_n) \subset (-h, K + h)$, so that

$$
\text{supp}(\tilde{g}_n), \text{supp}(\tilde{v}_n) \quad \subset \quad -t_0 + d_n^{-1}(-h, h + K) =: I_n^{(ii)}.
$$

Again this implies that the supremum of $\tilde{v}_n$ over $I_n^{(ii)}$ can be replaced by a supremum over all $s \in (c, d_n^{-1}K)$ and thus (42) will follow as soon as

$$\liminf_{n\to\infty} P\left\{\sup_{(c,d_n^{-1}K)} (\tilde{g}_n + \tilde{v}_n)(s) < -M\right\} \;>\; 1-\delta. \tag{46}$$

For arbitrary M and $c \geq 3M/2$ we have

$$\begin{aligned} & P\left\{\sup_{(c,d_n^{-1}K)} (\tilde{g}_n + \tilde{v}_n)(s) < -M\right\} \\ \geq\;& P(\cap_{s\in(c,d_n^{-1}\epsilon)}\{\tilde{v}_n(s) < \frac{M}{2} - \tau(s-c)\} \\ & \cap\{\sup_{s\in d_n^{-1}(\epsilon,K)} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\}), \end{aligned}$$

so that (46) follows from

$$\liminf_{n\to\infty} P\left\{\sup_{s\in d_n^{-1}(\epsilon,K)} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \;>\; 1-\delta, \tag{47}$$

and (45), which ends the derivation for the case (ii).

Now we prove that with M as above

$$\liminf_{n\to\infty} P\left\{\inf_{\inf I_1 \leq s \leq \sup I_0} y_n(s) \geq -M\right\} \;>\; 1-\delta. \tag{48}$$

Note that with $M = M_c$ corresponding to the bound for c, we have $k_n(c) = -3M_c/2$ and thus $\mathbb{E}y_n(c) \leq -3M_c/2 \leq -M_c - M_c/2$. Since $\tilde{g}_n(s) \to As$ on compact intervals, if n is large enough then we have $\mathbb{E}y_n(s) \geq As - \epsilon$ for each $\epsilon > 0$ arbitrarily small. Thus for $s_M = -M_c/2A$

$$\mathbb{E}y_n(s_M) \;\geq\; -\frac{M_c}{2} - \epsilon,$$

for n large enough. Finally from (42) and by the symmetry of the distribution of $\tilde{v}_n$ around 0, we have that with M replaced by $\max\{M_c, -2A\sup I_0\}$, both (42) and (48) hold, and (44) is proven.

Equation (45) can be proven in a similar way, which yields the lemma. $\square$

Lemma 5 states conditions (A) and (B) as sufficient conditions for Assumption 2. To further simplify condition (B) in Lemma 5, using Boole's inequality and the stationarity of the process $\tilde{v}_n$ we get in both cases (i) and (ii)

$$
\begin{aligned}
& P\left\{\sup_{d_n^{-1}(0,\ell(n))} \tilde{v}_n(s) > \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \\
& \leq d_n^{-1}\ell(n) P\left\{\sup_{(0,1)} \tilde{v}_n(s) > \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\},
\end{aligned} \tag{49}
$$

where $\ell(n)$ is defined for hypothesis (i) and replaced by K when dealing with hypothesis (ii). As a consequence, in Case (i) (resp. Case (ii)) the probability (49) will converge to 0 as soon as

$$
\alpha(n) \ := \ P\left\{\sup_{(0,1)} \tilde{v}_n(s) > \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \to 0
$$

faster than $d_n^{-1}\ell(n) \to \infty$, i.e. that $\alpha(n) = o(d_n\, \ell(n)^{-1})$ as $n \to \infty$ (resp. $\alpha(n) = o(d_n)$). The following conditions are thus respectively sufficient to insure that (49) tends to 0, as $n \to \infty$:

$$
\begin{aligned}
(i) \qquad & P\left\{\max_{1\leq i\leq n} t_i < \ell(n)\right\} \to 1 \text{ and } d_n^{-1}\ell(n)\alpha(n) \to 0, \\
(ii) \qquad & d_n^{-1}\alpha(n) \to 0.
\end{aligned}
$$

Finally, the examination of the convergence of $\alpha(n)$ can be made in two steps via the standard partition

$$
\begin{aligned}
\alpha(n) \ \leq \ & P\left[\sup_{s,s'\in(0,1)} |\tilde{v}_n(s) - \tilde{v}_n(s')| > \frac{1}{2}\left\{\frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\}\right] \\
& + P\left[\tilde{v}_n(0) > \frac{1}{2}\left\{\frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\}\right].
\end{aligned} \tag{50}
$$

In the sequel we will bound the two terms of the right-hand side of (50) separately, for the density and regression estimation problems treated in this paper: See subsections A.1 and A.2.

To further simplify (A) in Lemma 5, note that

$$
\begin{aligned}
& \cap_{s\in(c,d_n^{-1}\epsilon)}\{\tilde{v}_n(s) < \frac{M}{2} - \tau(s-c)\} \\
\supset \quad & \cap_{i\in\mathbb{Z}\cap(c,d_n^{-1}\epsilon)}\{\sup_{s\in[i,i+1)} \tilde{v}_n(s) < \frac{M}{2} - \tau(i-c)\} =: A_n.
\end{aligned}
$$

Thus, taking complements, part (A) of Lemma 5 follows as soon as for every δ and arbitrary $0 < c < \infty$ there is a $0 < M < \infty$ such that $\limsup_{n\to\infty} P(A_n^c) < \delta$. However,

$$
\begin{aligned}
P(A_n^c) &\leq \sum_{i\in\mathbb{Z}\cap(c,d_n^{-1}\epsilon)} P\left\{\sup_{s\in[i,i+1)} \tilde{v}_n(s) > \frac{M}{2} - \tau(i-c)\right\} \\
&\leq \sum_{i=[c]}^{[d_n^{-1}\epsilon]} P\left\{\sup_{s\in[0,1)} \tilde{v}_n(s) > \frac{M}{2} - \tau(i-c)\right\},
\end{aligned}
$$

where the equality follows from the stationarity of $\tilde{v}_n$. In the sequel we will establish maximal inequalities of the form

$$
P\left\{\sup_{s\in[0,1)} \tilde{v}_n(s) > a\right\} \leq Ca^{-p} \tag{51}
$$

for some constant $p > 1$; assume for now that these are established. Then

$$
\begin{aligned}
\limsup_{n\to\infty} P(A_n^c) &\leq \sum_{i=[c]}^{\infty} C\frac{1}{(\frac{M}{2} - \tau(i-c))^p} \\
&\leq \frac{C}{p|\tau|^p}\left(\frac{2}{M}\right)^{p-1} \\
&< \delta,
\end{aligned}
$$

where the next to last inequality holds by an integral approximation of the series and the last by choosing $M = M(\delta) > 2(C/p\delta|\tau|^p)^{1/(p-1)}$. Thus assumption (A) in Lemma 5 follows from (51) with $p > 1$; inequalities of the form (51) will next be treated.

A.1 Maximal bounds for the rescaled partial sum process

Let k be a kernel which is bounded, piecewise differentiable, with a bounded derivative, say $0 \leq |k'| \leq \alpha$. Assume that the sequence $h = h_n$ is such that $nh \to \infty$. We have (see (29))

$$\tilde{v}_n(s, t_0) \stackrel{\mathcal{L}}{=} d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} \int \bar{w}_{\hat{n}}(s-u)k'(u)\, du,$$

where $d_n = h$ is chosen so that $d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} \to 1$ and $\hat{n} = [nh]$. Now $\bar{w}_{\hat{n}}$ is asymptotically equivalent to the piecewise constant partial sum process which we therefore will use for notational simplicity, and which we denote (with a slight abuse of notation) with $w_{\hat{n}}$.

We show the convergence of $\alpha(n)$ in which $\tilde{v}_n(s)$ is replaced by $w_{\hat{n}}(s)$: this will be sufficient since

$$|\tilde{v}_n(s)| \leq \sup_{u\in[-1,1]} |w_{\hat{n}}(s-u)| \int |k'(u)|\, du,$$

and thus

$$\begin{aligned}\sup_{s\in(0,1)} |\tilde{v}_n(s)| &\leq c \sup_{s\in(0,1)} \sup_{u\in[-1,1]} |w_{\hat{n}}(s-u)| \\ &\leq c \sup_{s\in[-1,2]} |w_{\hat{n}}(s)|,\end{aligned}$$

with $c = \int |k'(u)|\, du$, and since the behaviour of the process $w_{\hat{n}}$ on $(0,1)$ and on $(-1,2)$ is qualitatively the same.

Proposition 3. *Let $p \geq 2$ be given and assume that the sequence $\{\epsilon_i\}_{i\geq 1}$ satisfies* $\max(\mathbb{E}\epsilon_1^2, \mathbb{E}\epsilon_1^p) < \infty$. *Then under the assumptions of Theorem 5*

$$P\left(\sup_{s\in(0,1)} w_{\hat{n}}(s) > M/2 - \tau(\delta h^{-1} - C)\right) \leq Ch^p,$$

where C is a finite constant.

Proof Let $a := M/2 + \tau c$ and $b := -\tau\delta$. In a first step, we obtain a majoration of

$$P\left(\sup_{s,s'\in(0,1)} |w_{\hat{n}}(s) - w_{\hat{n}}(s')| > a + bh^{-1}\right)$$

in the 3 dependence situations listed in Theorem 5. One has

$$w_{\hat{n}}(s) - w_{\hat{n}}(s') \;=\; \frac{1}{\sigma_{\hat{n}}} S_n(s,s'), \tag{52}$$

with

$$S_n(s,s') \;=\; \sum_{i=[s'\hat{n}]+1}^{[s\hat{n}]} \epsilon_i.$$

[a] If $\{\epsilon_i\}$ is an i.i.d. sequence the moment bound in Theorem 2.9 in Petrov [27] implies

$$\begin{aligned} \mathbb{E}|S_n(s,s')|^p &\leq c(p)\left(\sum_{i=[s'\hat{n}]+1}^{[s\hat{n}]} \mathbb{E}|\epsilon_i|^p + \left(\sum_{i=[s'\hat{n}]+1}^{[s\hat{n}]} \mathbb{E}(\epsilon_i)^2\right)^{p/2}\right) \\ &\leq c'\left(|s-s'|\hat{n} + |s-s'|^{p/2}\hat{n}^{p/2}\right) \end{aligned}$$

where $c(p)$ depends on p only and $c' = c(p)\cdot\max(||\epsilon_1||_2, \mathbb{E}|\epsilon_1|^p)$.

[b] If $\{\epsilon_i\}$ is a stationary sequence that is α-mixing (and thus also ϕ-mixing) satisfying the mixing condition (18), then Theorem 1 in Doukhan [13] implies

$$\mathbb{E}|S_n(s,s')|^p \;\leq\; \max(\hat{n}|s-s'|M_{p,\epsilon}, \hat{n}^{p/2}|s-s'|^{p/2}M_{p,2}^{p/2}),$$

where $M_{p,\epsilon} = ||\epsilon_i||_{p+\epsilon}^p$, and thus

$$\mathbb{E}|S_n(s,s')|^p \;\leq\; c''\max(\hat{n}|s-s'|^p, \hat{n}^{p/2}|s-s'|^2),$$

with $c'' = \max((\mathbb{E}|\epsilon_i|^{p+\epsilon})^{p/(p+\epsilon)}, (\mathbb{E}|\epsilon_i|^{2+\epsilon})^{2/(2+\epsilon)})$.

Therefore, for both independence and weak dependence cases, equation (12.42) of Billingsley [4] is satisfied, so that Theorem 12.2 in Billingsley implies

$$\begin{aligned} & P\left(\sup_{s,s'\in(0,1)} |w_{\hat{n}}(s) - w_{\hat{n}}(s')| > a + bh^{-1}\right) \\ =\; & P\left(\max_{k\in([s'\hat{n}]+1,[s\hat{n}])} \sigma_{\hat{n}}^{-1} \sum_{i=[s'\hat{n}]+1}^{k} \epsilon_i > a + bh^{-1}\right) \\ \leq\; & \frac{K_p' C(\hat{n})}{\sigma_{\hat{n}}^p (a+bh^{-1})^p} \end{aligned}$$

where $C(\hat{n}) = c'(\hat{n} + \hat{n}^{p/2})$ for i.i.d. data and $C(\hat{n}) = c'' \max(\hat{n}, \hat{n}^{p/2})$ in the mixing case. Since in both cases $\sigma_{\hat{n}} = \hat{n}^{1/2}$ and thus $\hat{n}/\sigma_{\hat{n}}^p = \hat{n}^{1-p/2}$ and $\hat{n}^{p/2}/\sigma_{\hat{n}}^p = 1$, we get the bound

$$P\left(\sup_{s,s'\in(0,1)} |w_{\hat{n}}(s) - w_{\hat{n}}(s')| > a + bh^{-1}\right) \leq Ch^p, \tag{53}$$

if $p \geq 2$.

[c] In the long range dependent case we have

$$E(S_{\tilde{n}}^2) \sim \eta_r^2 l_1(\tilde{n})\tilde{n}^{2\beta},$$

with l_1 as in (26), and according to de Haan [10], equation (12.42) in Billingsley [4] is satisfied, with

$$\begin{aligned} \gamma &= 2, \\ \alpha &= 2\beta, \\ u_l &= \{C_1 \eta_r^2 l_1(\tilde{n})\}^{1/2\beta}, \end{aligned}$$

for some constant $C_1 > 0$. Theorem 12.2 in Billingsley [4] then leads to

$$\begin{aligned} P\left(\max_{k\in([s'\hat{n}]+1,[s\hat{n}])} \sigma_{\hat{n}}^{-1} \sum_{i=[s'\hat{n}]+1}^{k} \epsilon_i > a + bh^{-1}\right) &\leq \frac{K'_{2,2\beta}}{(a+bh^{-1})^2\sigma_{\tilde{n}}^2}\left(\sum_{i=1}^{\tilde{n}} u_i\right)^{2\beta} \\ &= \frac{C}{(a+bh^{-1})^2}, \end{aligned}$$

with $C = C_1 K'_{2,2\beta}$, as $\sigma_{\hat{n}}^2 = \eta_r^2 l_1(\tilde{n})\hat{n}^{2\beta}$. Thus in the long range dependent case (53) holds for $p = 2$.

In a second step, using $w_n(s) - w_n(s') \stackrel{\mathcal{L}}{=} w_n(s - s')$, one can then deduce from (53) that

$$P\left[\sup_{s\in(0,1)} w_n(s) > a + bh^{-1}\right] \leq C''h^p,$$

where $C'' > 0$, which together with (53) and via (50) ends the proof. $\square$

Corollary 2. *Suppose the assumptions of Theorem 5 are satisfied; then Assumption 2 holds for* $y_n = g_n + \tilde{v}_n$ *and for* y *as defined in (11) in each context* **[a]**, **[b]** *and* **[c]** *listed in Theorem 5.*

Proof Note first that if $x_{b,n}$ is defined by (22), if m is a C^1-function, and k is a kernel with compact support, then $x'_{b,n}(t + sd_n) \to m'(t)$ for each t when $n \to \infty$. Besides, a consequence of Proposition 3 is that

$$\limsup_{n\to\infty} P\left[\sup_{s\in(0,h^{-1}K)} \tilde{v}_n(s) < \frac{M}{2} - \tau(h^{-1}\epsilon - c)\right] \;\geq\; 1-\delta.$$

Thus, condition $(B)(ii)$ in Lemma 5 is satisfied as soon as $d_n^{-1}d_n^p = d_n^{p-1} \to 0$, which is equivalent to $p > 1$. Thus the existence of two moments suffices to get condition (B) (ii) of Lemma 5 for i.i.d., mixing and subordinated Gaussian long range dependent sequences. Condition (A) of Lemma 5 follows immediately from Proposition 3. Hence Lemma 5 can be applied, so that Assumption 2 holds for $y_n = g_n + \tilde{v}_n$. An analogous result for y defined by (11) follows easily from the stationarity and finite second moments of $\tilde{v}_n$. □

A.2 Maximal bounds for the rescaled empirical process

The rescaled process is

$$\tilde{v}_n(s;t_0) \;=\; c_n \int k'(u) w_{n,d_n}(s-u;t_0)\,du$$

with $c_n = d_n^{-1}(nh)^{-1}\sigma_{n,d_n}$. Note that, similarly to the regression case, deriving the maximal bound for the process w_{n,d_n} implies the maximal bound for the process $\tilde{v}_n$.

Proposition 4. *Under the assumptions of Theorem 6, there exists a positive constant* C' *such that*

$$P\left(\sup_{s\in(0,1)} w_{n,d_n}(s) > M/2 - \tau(\epsilon h^{-1} - c)\right) \;\leq\; C'h^5,$$

Proof Let $a := M/2 + \tau c$ and $b := -\tau\epsilon$. For independent or mixing data satisfying the assumptions of Theorem 6, Lemma C2 in Anevski and Hössjer [2] implies that

$$
\begin{aligned}
& P\left\{\sup_{s,s'\in(0,1)} |w_{n,d_n}(s;t_0) - w_{n,d_n}(s';t_0)| \geq a + bh^{-1}\right\} \\
& \leq \frac{K}{(a+bh^{-1})^4 + (a+bh^{-1})^5}
\end{aligned}
\tag{54}
$$

for a positive constant K.

From (54) one can deduce the corresponding bound for

$$
P\left\{\sup_{s(0,1)} |w_{n,d_n}(s;t_0)| \geq a + bh^{-1}\right\}
$$

which together with (54) and via (50), implies the statement of the proposition. □

Corollary 3. *Suppose the assumptions of Theorem 6 are satisfied; then Assumption 2 holds for $y_n = g_n + \tilde{v}_n$ and for y as defined in (11) in each context* [**a**] *and* [**b**] *listed in Theorem 6.*

Proof Note first that if $x_{b,n}$ is defined by (37), if f is a C^1-function, and k is a kernel with compact support, then $x'_{b,n}(t + sd_n) \to f'(t)$ for each t when $n \to \infty$. Besides, a consequence of Proposition 4 is that

$$
\liminf_{n\to\infty} P\left(\sup_{s\in(0,h^{-1}\max t_i)} \tilde{v}_n(s) < \frac{M}{2} - \tau(h^{-1}\epsilon - c)\right) \quad \geq \quad 1 - \delta
$$

if there exists a function $\ell(n)$ such that

$$
P\{\max_{1\leq i\leq n} t_i \leq \ell(n)\} \to 1 \text{ and } h^{p-1}\ell(n) \to 0, \tag{55}
$$

as $n \to \infty$, and with $p \geq 5$. Note that

$$
P\left\{\max_{1\leq i\leq n} t_i > \ell(n)\right\} \quad \leq \quad n\frac{\mathbb{E}|t_i|^p}{\ell(n)^p} = \frac{nI}{\ell(n)^p}.
$$

Thus the conditions in (55) are implied by

$$\ell(n)^{-p} n \to 0, \ d_n^{-1} \ell(n) d_n^p \to 0,$$

which is equivalent to

$$n^{1/p} << \ell(n) << d_n^{1-p}.$$

In the case of i.i.d. and mixing data, when $d_n = n^{-1/3}$, p should satisfy

$$\begin{aligned} \frac{1}{p} < \frac{p-1}{3} \quad &\Leftrightarrow \quad (p^2 - p) > 3 \\ &\Leftrightarrow \quad p > \frac{1+\sqrt{13}}{2}. \end{aligned}$$

This, together with the restriction $p \geq 5$ in Proposition 4, implies that the existence of five moments suffices to establish $(B)(i)$ in Lemma 5 for i.i.d. and mixing data. Condition (A) in Lemma 5 is immediate from Proposition 4. Hence Lemma 5 can be applied, so that Assumption 2 holds for $y_n = g_n + \tilde{v}_n$. An analogous result for y defined by (11) follows easily from the stationarity and finite second moments of $\tilde{v}_n$. $\square$