HAL open science

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

Dragi Anevski, Anne-Laure Fougères

▶ To cite this version:

Dragi Anevski, Anne-Laure Fougères. Limit properties of the monotone rearrangement for density and regression function estimation. Bernoulli, 2019, 25 (1), pp.549-583. 10.3150/17-BEJ998 . hal-00182165v2

HAL Id: hal-00182165
https://hal.science/hal-00182165v2

Submitted on 21 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

D. Anevski*, A.-L. Fougères†

November 2017

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 as soon as x_n is. 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. Asymptotics for rearrangement estimators with vanishing derivatives are also obtained in these two contexts.

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

*Centre for Mathematical Sciences, Lund University, BOX 118, 221 00 Lund, Sweden

†Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne cedex, France

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 [23, Chapter 3] or in optimal transportation (see Villani [38, Chapter 3]). Fougères [15] was the first to use the algorithm in a statistical context, for density estimation under order restrictions. Meanwhile, Polonik [29, 30] 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. [12], Neumeyer [28], Chernozhukov et al. [9], Birke and Dette [5], Jankowski and Wellner [22], Volgushev and Dette [41], Birke et al. [6], Volguchev [39]. Some tests of monotonicity have also been recently introduced, see e.g. Volguchev et al. [40] and Birke 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. The procedure adopted in this paper is 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, where we also explore in more detail the applications of our general results to monotone density and regression function estimation. These are the problems of estimating f and m respectively in

$$\begin{array}{ll} (i) & t_1, \ldots, t_n \text{ stationary observations with marginal} \\ & \text{decreasing density } f \text{ on } \mathbb{R}^+, \\ (ii) & (t_i, y_i) \text{ observations from } y_i = m(t_i) + \epsilon_i, \\ & t_i = i/n, i = 1, \ldots, n, m \text{ decreasing on } [0,1], \\ & \{\epsilon_i\} \text{ stationary sequence with mean zero.} \end{array}$$

The standard approaches in these two problems have been isotonic regression for the regression problem, first studied by Brunk [8], 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 [27], Ramsay [32], Mammen [24], Hall and Huang [19], Mammen et al. [25], Gijbels [17], Birke and Dette [4], Dette and Pilz [13], Dette et al. [12] for the regression context. Besides, see Eggermont and Lariccia [14], Fougères [15], Hall and Kang [20], Meyer and Woodroofe [26], Polonik [29], Van der Vaart and Van der Laan [36], among others, for a focus on monotone (or unimodal) density estimation. Anevski and Hössjer [1] gave a general approach unifying both contexts. In their introduction, Birke and Dette [5] provide nice references in which physical or economical arguments justify the assumption of monotonicity. Our approach is similar in spirit to the general methods studied in Anevski and Hössjer [1] and first introduced in the regression estimation setting by Mammen [24]: 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.

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. [10] for quantile estimation. Chernozhukov et al. [10] rearrange the empirical quantile function, and use the fact that the rearrangement map is Hadamard differentiable together with Donsker type results, to obtain general statements about the final estimator. This approach is however not applicable in our case. In fact, our preliminary estimators are all local estimators, and they do not converge weakly as processes. Therefore the Hadamard differentiability of T has no implication in our estimation problems; we need to make a more detailed reasoning, assuming local limit process results for the preliminary estimator, together with properties of the map T. These two features may be seen as replacements for the Donsker result and the Hadamard differentiability result of T, that are used in [10].

An approach similar to ours for local estimates is given in Dette et al. [12], 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.

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 present the monotone rearrangement algorithm as classically defined, and we derive some simple properties that will be used in the sequel. Then we define the generic estimator of the monotone function of interest, and state the consistency for the estimator. In Section 3 the pointwise limit distribution properties are considered. For this purpose, we need to generalise the monotone rearrangement map for some specific functions, as will be done in Sections 3.1 and 3.2. The limit distribution given in Theorem 4 is of the general form

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

where T is the monotone rearrangement map;

$$\Delta \quad = \quad \lim_{n\to\infty} d_n^{-1}[\mathbb{E}\{x_n(t_0+sd_n)\} - x(t_0)]$$

is the asymptotic local bias of the preliminary estimator; A is the uniform limit, in s over compact intervals,

$$d_n^{-1}\{x_{b,n}(t_0+sd_n) - x_{b,n}(t_0)\} \quad \to \quad A,$$

(typically with $A = x'(t_0)$ in our applications); and

$$\tilde{v}(s;t_0) \quad \stackrel{\mathcal{L}}{=} \quad \lim_{n\to\infty} d_n^{-1}\left[x_n(t_0+sd_n) - \mathbb{E}\{x_n(t_0+sd_n)\}\right]$$

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. We then apply the obtained results 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 [1].

In Section 4 we derive limit distribution results for the proposed estimator in the case when the estimand has q vanishing derivatives while its $(q+1)$st derivative is strictly negative. The limit results are given in a general setting, and applied to both the density function and regression function estimation cases, and in similar dependence settings as for the "regular" case, that are derived in Sections 3.3 and 3.4. The limit distribution results are now of the form

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

where T is the monotone rearrangement map, $A_x(s)$ is a function that is given as a uniform limit over compact intervals,

$$d_n^{-(q+1)}(x_{b,n}(t_0 + sd_n) - x(t_0)) \quad \to \quad A_x(s)$$

as $d_n \to 0$, which (for symmetric kernels) is a convolution of a degree $q+1$ monomial with the kernel, while $\tilde{v}$ is the limit process that turns up in the above. The rate d_n is now different: It is slower than above, and e.g. for independent data, in both the density estimation and regression function estimation contexts, it is $d_n = n^{-1/(3+2q)}$.

In Appendix A we give some proofs for the results of Section 2. In Appendix B we state a general result on maximal bounds on the rescaled process part. In the supplementary material, we prove the statement in Appendix B as well as derive further 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. Furthermore in the supplementary material we present a simulation study that illustrates the finite sample behaviour of our estimator, and compare it to other estimators that are considered in the paper of Birke and Dette [5].

In this context it may be instructive to compare our results with previously obtained results, for similar procedures. The estimator defined by Dette et al. [12], is a two step procedure similar to ours for regression estimation problems, consisting of first defining a smooth estimate of the estimand and next do the monotone rearrangement of that estimator. We would like to point out that the assumptions in the two approaches are somewhat different: In our paper we use a fixed design setting, which enables us to use the

Gasser-Müller estimator as the first step estimator, while the results in Dette et al. [12] are derived in a random design setting. In [12] is used however an extra smoothing procedure in the second step, and therefore their obtained estimator is *not* the same as ours, and in fact their estimator may be seen as a smooth monotone rearrangement of the preliminary estimator. More interestingly, the two estimators give qualitatively very different results, with different rates and different limit random variables. Within the class of continuously differentiable monotone functions, the estimator considered in [12] is not rate optimal (for the independent data case, which is the only case they consider), they get a slower rate than the optimal, cf. Theorem 3.2 in [12]. Furthermore their limit random variable is Gaussian, with the Gaussian distribution turning up due to the over smoothing in the second extra step, whereas ours converge to the above defined new universal random variable. We would like to also emphasize that we are able to state our results for also dependent data, covering both weak and strong dependence. In Neumeyer [28] the same estimator as ours is treated, for general estimands and thus treating both regression and density estimation problem; the consistency of the resulting estimator is derived (see Neumeyer [28, Theorem 3.1]).

2 The monotone estimation procedure

2.1 Monotone rearrangement: first definitions

Monotone rearrangements were originally defined by Hardy et al. [21, Chapter 10.12] for non negative and integrable functions on $[0, 1]$. In Lieb and Loss [23, Chapter 3], the definition is extended to Borel measurable functions from $\mathbb{R}^n$ into $\mathbb{C}$ that vanish at infinity. We use their definition for Borel measurable functions from $\mathbb{R}$ into $\mathbb{R}^+$ that vanish at infinity, in the sense that for each positive u

$$r_f(u) := \lambda\{t \in \mathbb{R} : f(t) > u\} < +\infty \ , \tag{1}$$

where $\lambda(A)$ denotes the Lebesgue measure of any Borel set A on $\mathbb{R}$. Note that this definition holds in particular for integrable functions like densities on $\mathbb{R}^+$ as considered in Fougères [15].

Definition 1. *Let f be a positive function defined on $\mathbb{R}^+$ satisfying (1). The monotone rearrangement of f is defined as the (right continuous) generalized*

inverse of r_f*, namely*

$$T(f)(t) := \inf\{u \in \mathbb{R}^+ : r_f(u) \leq t\} \ , \tag{2}$$

for each positive t*.*

This rearrangement satisfies several properties that will be listed later in this section, see Lemmas 2 and 3 and Theorem 1. Note that related results were obtained by Chernozhukov et al. [10].

A particular class of functions for which (1) is satisfied is the set of bounded functions defined on a finite interval $I \subset \mathbb{R}$. Denote $\mathcal{B}(I) = \{f : f(I) \text{ bounded}\}$ and $\mathcal{D}(I) = \{f : f \text{ decreasing on } I\}$. Let $r_{f,I}$ be the right continuous map from $f(I)$ to $\mathbb{R}^+$, defined for each $u \in f(I)$ by

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

as the (right continuous) generalized inverse of $r_{f,I}$

$$T_I(f)(t) \ := \ \inf\{u \in f(I) : r_{f,I}(u) \leq t - \inf I\} \ . \tag{3}$$

The following lemmas and theorem 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. The proofs are straightforward and relegated to Appendix A for more clarity.

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*.*

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* $r_{f_c,I} = r_{f,I}$*.*

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:*

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

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

Theorem 1. *Let $||\cdot||$ be the supremum 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.

Note that Lemma 2 holds (with identical proof) for T as defined in (2), and Lemma 3 follows from that. Thus Theorem 1 also holds in this case for the supremum norm over $\mathbb{R}^+$, with identical proof.

Remark 1. *One can also refer to Lieb and Loss [23, Theorem 3.5] for a proof of the contraction property (the "non expansivity" property of the map T_I), for the L^p-norms for functions f and g vanishing at infinity.*

Finally, observe that when the function f is replaced by a stochastic process x defined almost surely, then for almost every realisation of x, one can define r_x (resp. $r_{x,I}$) and thereafter its generalized inverse. Thus one can define the monotone rearrangement almost surely for every stochastic process with finite support or satisfying (1) almost surely. We will make use of this last concept to define new estimators in the next section.

2.2 The new estimators: definition and first properties

Let x be a function of interest such as a density function, a regression function, or a spectral density, for example. 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. Let $\aleph_n$ denote the support of x_n. Assume that x_n is such that it is possible to define either $T(x_n)$ as in (2) (when $\aleph_n$ is infinite) or $T_{\aleph_n}(x_n)$ as in (3) (when $\aleph_n$ is finite). This will in particular be the case as soon as x_n is a density of $\mathcal{B}(\mathbb{R}^+)$, or x_n is a regression function on $[0,1]$. For sake of simplicity, a unique notation T will be used in the following to refer equally to T or $T_{\aleph_n}$.

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 2. *(i). Assume that $\{x_n\}_{n\geq 1}$ is a uniformly consistent estimator of x (in probability, uniformly on a compact set $B \subset \mathbb{R}$). If x is non increasing, then $\{T(x_n)\}_{n\geq 1}$ is also a uniformly consistent estimator of x (in probability, uniformly on B).*

(ii). Assume that $\{x_n\}_{n\geq 1}$ is an estimator that converges in probability in $\mathbb{L}^p$-norm to x, and that $\{x_n\}_{n\geq 1}$ and x are vanishing at infinity. If x is non increasing, then $\{T(x_n)\}_{n\geq 1}$ also converges in probability in $\mathbb{L}^p$-norm to x.

Proof Part (i) follows from the fact that $||x|| = \sup_{t\in K} |x(t)|$ is a norm (for every compact $K \subset \mathbb{R}$), that $T(x) = x$ if x is non increasing, and that T is a contraction with respect to $||\cdot||$, by Theorem 1. To get (ii), assume that $||x_n - x||_{L^p} \xrightarrow{P} 0$, and then note that

$$||T(x_n) - x||_{L^p} = ||T(x_n) - T(x)||_{L^p} \leq ||x_n - x||_{L^p} ,$$

thanks to Lieb and Loss [23, Theorem 3.5]. □

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; the assumption that the functions $\{f_n\}, f$ vanish at infinity are then naturally satisfied. Chernozhukov et al. [9] 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 Limit distribution results

Simple monotone estimators have been defined in Section 2.2 that satisfy several desired consistency properties. The aim of this section is to go further into asymptotic properties, focusing on pointwise limit distribution results for these estimators. To this purpose, we will first provide a more general definition of monotone rearrangement for some specific functions (and processes), and then state in Section 3.2 the main result of the paper in a general setting. The particular results obtained for density regression function estimations will next be developed.

3.1 Extension of monotone rearrangement algorithm

If φ is a function for which $r_\varphi(u)$ is possibly infinite for some positive u, a definition of $T(\varphi)$ 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:

> Let I_0 and the function φ be given. Assume there exists a constant $M = M(\varphi, I_0) < \infty$ and a finite interval $I_1 = I_1(\varphi, I_0) \supset I_0$ such that
>
> $$\inf_{t \in (\inf I_1, \sup I_0)} \varphi(t) > -M \quad \text{and} \quad \sup_{t \in (\sup I_1, \infty)} \varphi(t) < -M, \tag{4}$$
>
> $$\inf_{t \in (-\infty, \inf I_1)} \varphi(t) > +M \quad \text{and} \quad \sup_{t \in (\inf I_0, \sup I_1)} \varphi(t) < +M. \tag{5}$$

Theorem 3. *Let I_0 be a finite and fixed interval, and let φ be a continuous function $\mathbb{R}$ such that* (4) *and* (5) *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 The proof consists of the following three steps: (i) We construct a point y_* on which T_J and T_{I_1} agree, i.e. such that $T_J(\varphi)(y_*) = T_{I_1}(\varphi)(y_*)$. (ii) We show that $\inf I_1 \leq y_* \leq \inf I_0$. (iii) We show that if $T_J(\varphi)$ and $T_{I_1}(\varphi)$ agree at y_*, they will coincide on $[y_*, \sup I_o] =: \tilde{I}$.

(i) Define $y_1 := \inf\{y \in I_1 : \varphi(y) < \varphi([\inf J, y))\}$. Then by (5) we have that $\varphi(y_1) \geq +M$, which by the definition of y_1 implies that also $\inf \varphi([\inf J, y_1)) \geq +M$. Furthermore, from the left part of (4) and the right

part of (5) it follows that $\varphi(I_0) \subset (-M, +M)$. This implies that

$$\begin{aligned} y_1 &< \inf\{y \in J : \varphi(y) \in (-M, M)\} \\ &\leq \inf\{y \in J : \varphi(y) \in \varphi(I_0)\} \\ &=: z_0, \end{aligned}$$

where the first inequality follows from continuity of φ, the definition of y_1 and the theorem of intermediate values. Thus $y_1 \in I_1, z_0 \leq \inf I_0$ and $y_1 < z_0$ and so also $z_0 \in I_1$.

As a consequence, one has

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

where the first two terms in the last equality follow since from the definition of y_1, $\varphi(t) > \varphi(y)$ for all $t \in J \cap (y_1, \infty)$, while the last term in the last equality follows from $y_1 \in I_1$. Similarly, one has

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

Thus the following equality holds

$$r_{\varphi,J}\{\varphi(y_1)\} + \inf J = r_{\varphi,I_1}\{\varphi(y_1)\} + \inf I_1 := y_\star \tag{6}$$

Then $T_J(\varphi)(y_\star) = \inf\{u \in \varphi(J) : r_{\varphi,J}(u) \leq r_{\varphi,J}(\varphi(y_1))\}$, which since $r_{\varphi,J}$ is a decreasing function, is equal to $\varphi(y_1)$. Similarly, $\varphi(y_1) = T_{I_1}(\varphi)(y_\star)$, and we have shown that $T_J(\varphi)(y_\star) = \varphi(y_1) = T_{I_1}(\varphi)(y_\star)$, and thus the two maps agree at $y_\star$.

(ii): From the right hand parts of (4) and (5) follow that

$$T_J(\varphi)(\inf I_0) \leq M. \tag{7}$$

Furthermore

$$T_J(\varphi)(y_*) = \varphi(y_1) \geq M. \tag{8}$$

Since $T_J(\varphi)$ is a decreasing function (7) and (8) imply that $y_* \leq \inf I_0$, and (6) implies that $y_* \leq \inf I_1$.

(iii): Finally we prove that if $T_J(\varphi)$ and $T_{I_1}(\varphi)$ coincide at $y_\star$, they will coincide on $\tilde{I} = [y_*, \sup I_0]$. Let $u \in [-M, \varphi(y_1)] =: \tilde{Y}$ be arbitrary, 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 next to last equality follows since by the definition of y_1, $\varphi(t) \leq \varphi(y_1)$ only for $t > y_1$, and the last equality follows from (6) and since $y_1 \in I_1$. 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

$$r_{\varphi,J}(u) + \inf J = r_{\varphi,I_1}(u) + \inf I_1. \tag{9}$$

Next, since $T_J(\varphi)$ and $T_{I_1}(\varphi)$ are decreasing, $T_J(\varphi)(y_*) = T_{I_1}(\varphi)(y_*) = \varphi(y_1)$, $\inf I_1 \leq y_* \leq \inf I_0$, and $\inf \varphi([y_\star, \sup I_0]) > -M$, we obtain

$$\begin{aligned} T_J(\varphi)(\tilde{I}) &\subset \tilde{Y}, \\ T_{I_1}(\varphi)(\tilde{I}) &\subset \tilde{Y}. \end{aligned} \tag{10}$$

Therefore, for $t \in \tilde{I}$,

$$\begin{aligned} T_J(\varphi)(t) &= \inf\{u \in \varphi(J) : r_{\varphi,J}(u) \leq t - \inf J\} \\ &= \inf\{u \in \tilde{Y} : r_{\varphi,J}(u) \leq t - \inf J\} \\ &= \inf\{u \in \tilde{Y} : r_{\varphi,I_1}(u) \leq t - \inf I_1\} \\ &= \inf\{u \in \varphi(I_1) : r_{\varphi,I_1}(u) \leq t - \inf I_1\} \\ &= T_{I_1}(\varphi)(t)\ . \end{aligned}$$

The second equality above holds since $T_J(\varphi)(y_*) = \varphi(y_1) = \sup \tilde{Y}$, $\inf \tilde{I} = y_*$, $T_J(\varphi)$ is decreasing and because of the first part of (10), the third equality

follows from (9), and the next to last equality is similar to the second (with T_{I_1} replacing T_J and using the second part of (10)). $\square$

Note also that the interval I_1 can without loss of generality be taken to be symmetric around 0 e.g. as $I_1 = [-k, k]$: In fact assuming that (4) and (5) hold with some I_1 and M we can replace I_1 with $[-k, k]$, with $k = \max(|\inf I_1|, |\sup I_1|)$, then (4) and (5) will hold for $I_1 = [-k, k]$ and with the same M.

Corollary 1. *Let $I_0 \subset \mathbb{R}$ be a finite and fixed interval. Assume φ is continuous and satisfies (4) and (5). Then for each $t \in I_0$, one can define*

$$T(\varphi)(t) \quad := \quad \lim_{k \to +\infty} T_{[-k,k]}(\varphi)(t). \tag{11}$$

Remark 3. *Even if this definition seems to be dependent on $I_1 = I_1(\varphi, I_0)$, it is* not*, because of Theorem 3 and since we define $T(\varphi)$* locally*, namely only on I_0. So in particular one has $T(\varphi)(t) = T_{I_1}(t)$ for each $t \in I_0$.*

We next state a simple condition that ensures (4) and (5).

Lemma 4. *Let φ be a locally bounded function on $\mathbb{R}$, such that*

$$\lim_{x \to -\infty} \varphi(x) \quad = \quad - \lim_{x \to +\infty} \varphi(x) = +\infty.$$

Then for any interval I_0 there exists a finite interval $I_1 \supset I_0$ and a finite constant M such that (4) and (5) hold.

Proof Let $I_0 = [a, b]$ and put $M = \sup_{x \geq a} \varphi(x)$. Since φ is locally bounded and $\lim_{x \to +\infty} \varphi(x) = -\infty$ it follows that $M < \infty$. Since $\lim_{x \to -\infty} \varphi(x) = +\infty$ there is a $c < a$ such that $\varphi(x) > M$ for all $x \leq c$. Let $c_M := \sup\{x < a : \varphi(t) > M \ \forall \ t \leq x\}$. Define then $m = \inf_{x \in [c_M, b]} \varphi(x)$. Since $\lim_{x \to \infty} \varphi(x) = -\infty$, there exists a $d > b$ such that for all $x \geq d$, $\varphi(x) < \min(m, -M)$. Let $d_{m,M} := \inf\{x > b : \varphi(t) < \min(m, -M) \ \forall \ t > x\}$. Then (4) and (5) hold with $I_1 = [c_M, d_{m,M}]$. $\square$

Note that T as defined in (2) is a continuous map with the metric generated by the supnorm on compact intervals, while T defined as a extension via the local definition in (11) is not. The first statement follows from the fact

that T as defined in (2) is easily seen to satisfy the properties in Theorem 1, with the supnorm over I replaced by the the supnorm metric on compact intervals. The lack of continuity with the respect of uniform convergence on compact intervals is however of no importance for us, in our use of continuity for deriving consistency and limit distributions: We will derive limit distribution results only via local versions $T_{[-c,c]}$, for which we have established continuity in Theorem 1. The consistency is derived using the global map only in the density estimation problem, for which we use definition (2) for T. For the regression problem we apply our results to functions defined on $[0,1]$ and thus there is no need for a global definition then.

3.2 Asymptotic distribution in a general framework

Let $J \subset \mathbb{R}$ be a finite or infinite interval, and $C(J)$ the set of continuous functions on J. Let $\{x_n\}_{n\geq 1}$ be a sequence of stochastic processes in $C(J)$ and let t_0 be a fixed interior point in J. Assume that either J is finite or that x_n satisfies (1) almost surely, so that $T(x_n)$ the monotone rearrangement of x_n can be defined almost surely, as seen in Section 2.1. 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 defined as T_J if J is finite or T if J is infinite. The proof of these results are along the lines of Anevski and Hössjer [1], and their notation will be used for clarity. Decompose in particular x_n into a deterministic part and a stochastic part

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

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 respectively the stochastic and deterministic parts of x_n as

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

This decomposes the rescaling of x_n as

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

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

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

so that

$$y_n(s) \;:=\; g_n(s) + \tilde{v}_n(s;t_0) = d_n^{-1}\{x_n(t_0 + s d_n) - x(t_0)\}. \tag{12}$$

This definition of the rescaled deterministic and stochastic parts is slightly different from the one in Anevski and Hössjer [1], and is due to the fact that we only treat the case where 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 [1] other possibilities occur.

The limit distribution results will be derived using a classical two-step procedure, cf. e.g. Prakasa Rao [31]: 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*

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

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 $c > 0$,

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

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

$$y(s) \;=\; As + \Delta + \tilde{v}(s; t_0). \tag{13}$$

We next give a condition that enables a definition of the monotone rearrangement for processes. 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 $c = c(\delta)$ such that $[-c, c] \supset I_0$ and a finite positive $M = M(\delta)$ such that*

$$\liminf_{n\to\infty} P\left\{ \inf_{s\in(-c,\sup I_0)} z_n(s) > -M, \sup_{s\in(c,\infty)} z_n(s) < -M \right\} \;>\; 1-\delta, \tag{14}$$

and

$$\liminf_{n\to\infty} P\left\{ \inf_{s\in(-\infty,-c)} z_n(s) > +M, \sup_{s\in(\inf I_0,c)} z_n(s) < +M \right\} \;>\; 1-\delta. \tag{15}$$

Note that in the applications typically both $c(\delta) \to \infty$ and $M(\delta) \to \infty$ as $\delta \downarrow 0$. There is no restriction in assuming this, so in the sequel we assume that $\lim_{\delta\to 0} c(\delta) = \infty$ and $\lim_{c\to\infty} \delta(c) = 0$. Denote $T_c = T_{[-c,c]}$. Consider $D_n(\delta) = D_n(\delta(c))$ as the set of ω such that it is possible to define the monotone rearrangement $T(z_n)|_{I_0}$ of z_n on I_0.

Lemma 5. *Let I_0 be a finite and fixed interval in $\mathbb{R}$, and $\{z_n\}$ be a sequence of continuous stochastic processes on $\mathbb{R}$ such that Assumption 2 holds. Then*

$$\lim_{c\to\infty} \liminf_{n\to\infty} P[D_n(\delta(c))] = 1$$
$$\lim_{c\to\infty} \liminf_{n\to\infty} P(\sup_{I_0} |T_c(z_n)(\cdot) - T(z_n)(\cdot)| = 0) = 1.$$

Proof Let $\delta > 0$ be arbitrary. Let $A_n(\delta, c, M)$ and $B_n(\delta, c, M)$ be the sets for which the probabilities are bounded in (14) and (15), respectively, for some finite $c = c(\delta)$. Then, using Theorem 3 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 $J \supset [-c, c]\} := C_n(\delta, c, M)$. Therefore $\liminf_{n\to\infty} P(C_n(\delta, c, M)) \geq 1 - 2\delta$.

Note that $C_n(\delta, c, M)$ is included in the set $D_n(\delta)$ on which it is possible to define $T(z_n)|_{I_0}$, namely as $\lim_{k\to\infty} T_k(z_n)|_{I_0}$, and which is further included in the set $E_n(\delta, c) := \{\sup_I |T_c(z_n) - T(z_n)| = 0\}$. A priori the definition of T depends on δ, so that $T(z_n)|_{I_0} = T^\delta(z_n)|_{I_0}$. We will show, however, that the definition is independent of δ. In fact, consider $\delta_1 < \delta_2$, so that $c_1 = c(\delta_1) > c_2 = c(\delta_2)$ and $M_1 = M(\delta_1) > M_2 = M(\delta_2)$. Then by the triangle inequality

$$\sup_{I_0} |T^{\delta_1}(z_n) - T^{\delta_2}(z_n)| \quad \leq \quad \sup_{I_0} |T^{\delta_1}(z_n) - T_{c_1}(z_n)| + \sup_{I_0} |T_{\tilde{J}}(z_n) - T^{\delta_2}(z_n)|,$$

with $\tilde{J} := [-c_1, c_1]$.

Then the first term on the right hand side is zero on $C_n(\delta_1, c_1, M_1)$, since $C_n(\delta_1, c_1, M_1) \subset E_n(\delta_1)$, and the second term on the right hand side is zero on $C_n(\delta_2, c_2, M_2)$ since $\tilde{J}$ is a compact set containing $[-c_2, c_2]$. Therefore on the set $C_n(\delta_1, c_1, M_1) \cap C_n(\delta_2, c_2, M_2)$ the left hand side is zero, and thus $T^{\delta_1}|_{I_0} = T^{\delta_2}|_{I_0}$ on that set. Note also that $P(D_n(\delta_1) \cap D_n(\delta_2)) \geq P(C_n(\delta_1, c_1, M_1) \cap C_n(\delta_2, c_2, M_2)) \geq 1 - 2\delta_1 - 2\delta_2$. Thus $T^{\delta_1}|_{I_0} = T^{\delta_2}|_{I_0}$ when both definitions exist, and they do with an as high probability as desired. This shows that the definition of T does not depend on δ.

Now, since $\delta > 0$ is arbitrary, letting $\delta \downarrow 0$ and noting that this implies that $c \to \infty$, and using

$$P(E_n(\delta, c)) \quad \geq \quad P(C_n(\delta, c, M)) \geq 1 - 2\delta$$

proves the second statement of the lemma. Noting that $D_n \supset C_n(\delta, c, M)$ proves the first statement of the lemma. □

The next result is the main limit distribution result, stating that the rescaled estimator converges in distribution to a "universal" limit random variable $T(y)(0)$. The existence of the limit r.v. is made explicit in the proof of the theorem, we can for now define it (when it exists, and the proof of the following theorem shows that the limit exists as soon as y satisfies Assumption 2) as a limit in probability

$$T(y)(0) \quad \stackrel{P}{=} \quad \lim_{c\to\infty} T_c(y)(0).$$

In the following theorem $J \subset \mathbb{R}$ will be a (finite or infinite) interval, in our applications $J = [0, 1]$ or $J = \mathbb{R}^+$.

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 $\{y_n\}_{n\geq 1}$ and y are continuous processes and that Assumption 2 holds for both $\{y_n\}_{n\geq 1}$ and y respectively defined by (12) and (13). Then*

$$d_n^{-1}[T_J(x_n)(t_0) - x(t_0)] \xrightarrow{\mathcal{L}} T[A\cdot + \tilde{v}(\,\cdot\,;t_0)](0) + \Delta, \tag{16}$$

as $n \to \infty$.

Proof Let c be a positive and finite constant and denote $T_{c,n} = T_{[t_0 - cd_n, t_0 + cd_n]}$. We can decompose

$$\begin{aligned} d_n^{-1}\{T_J(x_n)(t_0) - x(t_0)\} &= d_n^{-1}\{T_J(x_n)(t_0) - T_{c,n}(x_n)(t_0)\} \\ &\quad + d_n^{-1}\{T_{c,n}(x_n)(t_0) - x(t_0)\}. \end{aligned} \tag{17}$$

Let us first consider the second term of the right hand side of (17) and introduce

$$\chi_n(s) := x_n(t_0 + sd_n) = x(t_0) + d_n y_n(s). \tag{18}$$

Applying Lemma 3 leads to

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

which gives

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

Assumption 1 implies that $y_n \xrightarrow{\mathcal{L}} y$ on $C[-c,c]$, with y defined in (13). Applying the continuous mapping theorem on T_c, cf. Theorem 1, proves

$$d_n^{-1}\{T_{c,n}(x_n)(t_0) - x(t_0)\} \xrightarrow{\mathcal{L}} T_c(y)(0) \tag{19}$$

as $n \to \infty$. Lemma 5 via Assumption 2 with $z_n = y$ shows that we can define the limit random variable $T(y)(0)$ as a limit in probability so that, as $c \to \infty$,

$$T_c(y)(0) \xrightarrow{P} T(y)(0). \tag{20}$$

Next we consider the first term of the right hand side of (17). Let ∇ be a positive and finite constant and denote $A_{n,\nabla} = [t_0 - \nabla d_n, t_0 + \nabla d_n]$. From (18) and Lemma 3 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 (12). Using Lemma 5 with $I = [-\nabla, \nabla]$ shows that

$$d_n^{-1}\{T_{c,n}(x_n)(t_0) - T_J(x_n)(t_0)\} \stackrel{P}{\to} 0 \tag{21}$$

when $n \to \infty$.

Let first n tend to infinity in (17), and apply Slutsky's theorem with the use of (19), (21). Note that when $c \to \infty$, (20) gives the result. □

Remark 4. *The approach for deriving the limit distributions is similar to the general approach in Anevski and Hössjer [1] 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 [1].
– 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 [1] is assumed to be approximated

by a convex or concave function.
– 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. □

The rest of this section is to apply the previous result to two nonparametric inference problems: next subsection deals with the estimation of a monotone density function, and the last one with estimating 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.

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.

3.3 Application to monotone density estimation

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

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

where

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

In this section 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 $\{t_i\}_{i\geq 1}$: Let $x_n(t) = (nh)^{-1}\sum_{i=1}^n k\{(t-t_i)/h\}$ be 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 [42]), and define the (monotone) density estimate

$$\hat{f}_n(t) \;=\; T(x_n)(t), \tag{22}$$

where T is the monotone rearrangement map on $\mathbb{R}^+$ as defined in (2). Note that $\hat{f}_n$ is monotone and positive, and integrates to one, cf. equation (4) of Section 3.3. in Lieb and Loss [23].

A straightforward consequence of Theorem 2 and standard convergence results for the kernel density estimate is the following consistency result:

Proposition 1. *The random function $\hat{f}_n$ defined by (22) 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 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 [1].

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

$$\begin{aligned} x_n(t) &= h^{-1}\int k'(u)F_n(t-hu)\,du, \\ x_{b,n}(t) &= h^{-1}\int k'(u)F(t-hu)\,du, \\ v_n(t) &= h^{-1}\int k'(u)F_n^0(t-hu)\,du. \end{aligned} \tag{23}$$

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 5. *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 $\mathbb{E}t_i^4 < \infty$. 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

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

as $n \to \infty$, where $\tilde{v}(s; t)$ is as in (42), 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:

$$\tilde{v}_n(s; t_0) \xrightarrow{\mathcal{L}} c \int k'(u) w(s-u; t_0)\, du =: \tilde{v}(s; t_0), \tag{24}$$

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 [1] state that $w_{n,\delta_n}(s, t_0) \xrightarrow{\mathcal{L}} B(s)$ 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 2 in the supplementary material. Theorem 5 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 $\sigma^2_{n,d_n} \sim nd_nf(t_0)$, 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 [31] and Wright [43] for the independent data cases, and Anevski and Hössjer [1] 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 [1]; note that we obtain the same bias term as in Anevski and Hössjer [1].*

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 [1]. 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 [1]. 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 [11] for a derivation of this limit distribution.*

3.4 Application to monotone regression function estimation

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 $\sigma_n^2 = \mathrm{Var}(\sum_{i=1}^n \epsilon_i)$. The two sided partial sum process w_n is defined by

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

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 $\varphi(n)$ or $\alpha(n) \to 0$ as $n \to \infty$, such that

$$\begin{aligned} \sup_{A \in \bar{\mathcal{F}}_n} |P(A|\mathcal{F}_0) - P(A)| &\leq \varphi(n), \\ \sup_{A \in \mathcal{F}_0, B \in \bar{\mathcal{F}}_n} |P(AB) - P(A)P(B)| &\leq \alpha(n), \end{aligned} \tag{25}$$

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

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

$$y_i = m(t_i) + \epsilon_i,$$

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

$$x_n(t) = h^{-1} \int k((t-u)/h) \bar{y}_n(u)\, du\ , \tag{26}$$

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 $\text{supp}(k) = [-1, 1]$. 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

$$\tilde{m}(t) = T_{[0,1]}(x_n)(t),\ t \in J, \tag{27}$$

where $T_{[0,1]}$ is the monotone rearrangement map on $[0, 1]$. A straightforward application of Theorem 2 and standard consistency results for regression function estimators imply the following consistency result:

Proposition 2. *The random function $\tilde{m}$ defined by (27) 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

$$\begin{aligned} x_{b,n}(t) &= h^{-1} \int k(\frac{t-u}{h}) \bar{m}_n(u)\, du, \\ v_n(t) &= h^{-1} \int k(\frac{t-u}{h}) \bar{\epsilon}_n(u)\, du, \end{aligned} \tag{28}$$

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

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

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

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

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 [3], implying that

$$w_n \xrightarrow{\mathcal{L}} B, \tag{29}$$

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

[b] Define

$$\kappa^2 = \text{Cov}(0) + 2 \sum_{k=1}^{\infty} \text{Cov}(k). \tag{30}$$

Assumption 3. [$\phi -$ **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* (30).

Assumption 4. [$\alpha -$ **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* (30) *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 (29) is valid, cf. Anevski and Hössjer [1] 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

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

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

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

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 [34, 35]), and Taqqu [34] also shows that

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

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

$$l_1(k) \quad = \quad \frac{2}{r!(1-rd)(2-rd)} l_0(k)^r. \tag{31}$$

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

$$\beta \quad = \quad 1 - rd/2. \tag{32}$$

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

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

[34]. From these results follows a two sided version of Taqqu's result stating the behavior of the two sided partial sum process:

$$w_n \stackrel{\mathcal{L}}{\to} B_{r,\beta},$$

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:

$$\begin{aligned} \tilde{v}_n(s;t) &= d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} \int \bar{w}_{\hat{n}}(h^{-1}t + s - u)k'(u)\,du \\ &\stackrel{\mathcal{L}}{=} d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} \int \bar{w}_{\hat{n}}(s - u)k'(u)\,du, \end{aligned} \tag{33}$$

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 [33]).

Theorem 6. *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 (26), 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 (30), 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

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

as $n \to \infty$*, where* $\tilde{m}$ *is defined in (27),*

$$\tilde{v}(s;t) = c\int w(s-u)k'(u)\,du, \tag{34}$$

and respectively

$$\begin{array}{ll} [\mathbf{a}] & w = B\ ;\ c = \sigma a^{-3/2}, \\ [\mathbf{b}] & w = B\ ;\ c = \kappa a^{-3/2}, \\ [\mathbf{c}] & w = B_{r,\beta}\ ;\ c = |\eta_r| a \quad (\textit{where } \beta \textit{ defined in } (32)). \end{array}$$

Proof The conclusions of the theorem follow by an application of Theorem 4 in the context of monotone regression function. Assume first that $d_n = h$ is such that

$$d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} = d_n^{-2}n^{-1}\sigma_{\hat{n}} \to c > 0. \tag{35}$$

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

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

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

$$\tilde{v}_n(s;t) \xrightarrow{\mathcal{L}} \tilde{v}(s;t), \tag{36}$$

where $\tilde{v}(s;t)$ is defined in (43). 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

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

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

Proof that Assumption 2 holds is relegated to the appendix, see Corollary 1 in the supplementary material. An application of Theorem 4 then finishes the proof of Theorem 6. 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 (45) 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

$$d_n^{-2} n^{-1} \sigma_{\hat{n}} = d_n^{-2} n^{-1} |\eta_r| (nd_n)^{1-rd/2} l_1(nd_n)^{1/2} \quad \to \quad |\eta_r| a \tag{37}$$

  if and only if

$$d_n \quad = \quad n^{-rd/(2+rd)} l_2(n; a), \tag{38}$$

  where l_2 is another function slowly varying at infinity, implicitly defined in (37). Thus (45) follows with $c = |\eta_r| a$ and $h = d_n$ given in (47). □

Remark 7. *The present estimator is similar to the estimator first presented by Mammen [24]: 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 [1] 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 Limit results for density and regression function estimators with q vanishing derivatives

The results we have established in the previous section can in fact be extended to the case when the estimand x is monotone and has q vanishing moments at the point of interest t_0, so when $x(t_0) \neq 0, x^{(i)}(t_0) = 0$ for $j = 1, \ldots, q$ and $x^{(q+1)}(t_0) < 0$. We present these results here. We will make an analogous derivation to the case when $x'(t_0) < 0$, and we will highlight when there is a difference to the case already treated.

We give proofs for the independent data, the weak dependent and the long range dependent cases. Our results give other rates of convergence and other (new) limit random variables. For instance the limit results that we obtain for the independent data case will be with the (slower rate of convergence) $n^{-1/(q+3)}$, and the limit r.v. will be of the form

$$T(A(s) + \tilde{v}(s))(0),$$

where A is (when k is a symmetric kernel) a convolution of the monomial s^{q+1}, of order $q+1$, with the kernel function $k(s)$ and $\tilde{v}$ is the same process as the one already treated, while T is the monotone rearrangement map.

Thus let x be a function satisfying the above assumption that it has q vanishing derivatives at a point t_0 that is in the interior of its support, while $x^{q+1}(t_0) < 0$. Suppose that x_n is a preliminary estimator of x that we partition as

$$x(t) \quad = \quad x_{b,n}(t) + v_n(t).$$

Let $d_n \downarrow 0$ as $n \to \infty$, be a deterministic sequence. Define the rescaled deterministic and stochastic parts, respectively, by

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

as in Section 3.2 and also

$$\begin{aligned} y_n(s) &= g_n(s) + \tilde{v}_n(s;t_0) && (39) \\ &= d_n^{-(q+1)}(x_n(t_0 + sd_n) - x(t_0)). && (40) \end{aligned}$$

Then Assumption 1 is replaced by

Assumption 5. *There exists a stochastic process $\tilde{v}(\cdot;t_0) \neq 0$ such that*

$$\tilde{v}_n(\cdot;t_0) \xrightarrow{\mathcal{L}} \tilde{v}(\cdot;t_0),$$

on $C(-\infty,\infty)$ as $n \to \infty$. The functions $\{x_{b,n}\}_{n\geq 1}$ are monotone and there exists a function $A(s)$ such that for each $c > 0$

$$\sup_{|s|\leq c} |g_n(s) - A(s)|$$

as $n \to \infty$.

Let us also define the limit process

$$y(s) = A(s) + \tilde{v}(s;t_0). \tag{41}$$

Then we have the following general theorem.

Theorem 7. *Let $J \subset \mathbb{R}$ be an interval, and t_0 a fixed point in the interior of J. Suppose that Assumption 5 holds. Assume furthermore that $\{y_n\}_{n\geq 1}$ and y are continuous processes and that Assumption 2 holds for both $\{y_n\}_{n\geq 1}$ and y, defined in* (40) *and* (41) *respectively. Then*

$$d_n^{-(q+1)}\left[T_J(x_n(t_0)) - x(t_0)\right] \xrightarrow{\mathcal{L}} T[A(\cdot) + \tilde{v}(\cdot;t_0)](0),$$

as $n \to \infty$.

Proof The proof is completely analogous to the proof of Theorem 4: At appropriate places simply replace the factor d_n with d_n^{q+1} and d_n^{-1} with $d_n^{-(q+1)}$. □

4.1 Monotone density function estimation

We now apply the above general result to the estimation of a monotone density f function such that $f^{(j)}(t_0) = 0$ for $j = 1, \ldots, q$ and $f^{(q+1)}(t_0) < 0$.

Recall the definition of the empirical distribution function F_n and centered empirical distribution function F_n^0 in Section 3.3. Let δ_n be a sequence such that $\delta_n \downarrow 0, n\delta_n \uparrow \infty$ as $n \to \infty$, and define the centered empirical process w_{n,δ_n} locally around t_0 on the scale δ_n, exactly as in Section 3.3, with the same definition of normalising factor σ^2_{n,δ_n}. Furthermore, define the kernel function x_n as the preliminary estimator of f as in Section 3.3, and note that we can write the kernel estimator as $x_n = x_{b,n} + v_n$, with $x_n, x_{b,n}, v_n$ given in Equation (23) in Section 3.3. Again we will choose the bandwidth and the local scale to be the same $d_n = h$.

We now however obtain the rescaled process

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

with a (different) scale factor $c_n = d_n^{-(q+1)}(nh)^{-1}\sigma_{n,d_n}$.

Theorem 8. *Let $\{t_i\}_{i\geq 1}$ be a stationary sequence with a monotone marginal density function f such that*

(i) $f^{(j)}(t_0) = 0$ *for* $j = 1, \ldots, q$,

(ii) $\sup_{t \in I_{t_0}} f^{(q+1)}(t) < 0$,

(iii) $f \in C^{q+1}(I_{t_0})$,

for an open interval $I_{t_0} \ni t_0$ where $t_0 > 0$. Assume that $\mathbb{E}t_i^4 < \infty$. 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 $h = an^{-1/(3+2q)}$ and $a > 0$ an arbitrary constant, we obtain

$$n^{1/(3+2q)}\{\hat{f}_n(t_0) - f(t_0)\} \xrightarrow{\mathcal{L}} T(A_f(\cdot) + \tilde{v}_n(\cdot; t_0))(0),$$

as $n \to \infty$, where $\tilde{v}(s;t)$ is as in (42), with $c = a^{-(q+3/2)} f(t_0)^{1/2}$, the function $A_f(s)$ is defined as

$$A_f(s) = \frac{1}{(q+1)!} f^{(q+1)}(t_0) \int k(u)(s+u)^{q+1}\, du$$

and w a standard two sided Brownian motion.

Proof The first part of Assumption 1 is established as in the proof of Theorem 5, with

$$\tilde{v}(s;t_0) = c \int k'(u) w(s-u;t_0)\, du \tag{42}$$

and w a two sided standard Brownian motion, under the respective assumptions in [**a**] and [**b**].

Again we 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 the expression for $x_{b,n}$ proves the second part of Assumption 1 with limit function

$$A(s) = \frac{1}{(q+1)!} f^{(q+1)}(t_0) \int k(u)(s+u)^{q+1}\, du\ .$$

To check Assumption 2 for the case $q \geq 1$ is much easier than for the case $q = 0$, which is the case treated in Section 3. This can be seen in the statement of Lemma 1 in the supplemental part: the rescaled processes $\tilde{v}_n$ are the same for this new case $q \geq 1$ as for the already treated case $q = 0$. However, the rescaled limit deterministic part $A(s)$ is, for $q \geq 1$, a smoothed out higher order monomial $(s+u)^{q+1}$, which of course has not linear increase/decrease (as when $r = 0$) but polynomial increase. That means that one should compare $\tilde{v}_n$ in A and B of Lemma 1 not with e.g. in A of Lemma 1 the linear function $\tau(s-c)$ but instead with $\tau(s-c)^{q+1}$. But bounds of this form are easier to establish for polynomial increase than for linear increase, and in fact if we have established bounds for linear increase then we automatically get (at least) the same bounds for polynomial increase.

Finally in order to check that we can choose the scale d_n so that $c_n \to c$, we make the following calculations.

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

$$\begin{aligned} c_n &= d_n^{-(q+1)}(nh)^{-1}\sigma_{n,d_n} \\ &\sim d_n^{-(q+3/2)}n^{-1/2}f(t_0)^{1/2}. \end{aligned}$$

This tells us that we should choose $d_n = an^{-1/(3+2q)}$ we get $c = a^{-3/2}f(t_0)^{1/2}$.

(2) Mixing data case [**b**]: Similar to the proof of case [**a**]. □

4.2 Monotone regression function estimation

Next we treat the regression estimation problem for a regression function m such that $m^{(j)}(t_0) =$ for $j = 1, \ldots, q$ and $m^{(q)}(t_0) < 0$ for t_0 the point of interest. We have the same setting for the regression problem as in Section 3.4, so m is a function defined on $[0, 1]$, equidistant design $t_i = i/n$, and data (t_i, y_i) from the model $y_i = m(t_i) + \epsilon_i$ where $\{\epsilon_i\}$ is a stationary sequence of independent, weakly dependent or long range dependent data. Define the two-sided partial sum process w_n and the function $\bar{y}_n$ exactly as in Section 3.4, and the Gasser-Müller kernel estimate as in Equation (26). The monotone estimator of m is defined as $\tilde{m}(t) = T_{[0,1]}(x_n)(t)$. Again we can write $x_n = x_{b,n} + v_n$, with $x_{b,n}$ and v_n given in Equation (28). We note that since the process w_n is defined exactly the same way as in Section 3.4, we have that $w_n \xrightarrow{L}$ a limit process which is Brownian motion or Fractional Brownian motion, or something else, depending on the dependence structure for the data $\{\epsilon_i\}$, as in Section 3.4.

We choose bandwidth equal to the local scale factor $h = d_n$, and rescale the process part as

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

while the deterministic part is rescaled as

$$g_n(s) = d_n^{-(q+1)}(x_{b,n}(t_0 + sd_n) - x(t_0)).$$

Theorem 9. *Assume m is monotone on $[0, 1]$ and*

(i) $m^{(j)}(t_0) = 0$ *for* $j = 1, \ldots, q$,

(ii) $\sup_{t \in I_{t_0}} m^{(q+1)}(t) < 0$,

(iii) $m \in C^{q+1}(I_{t_0})$,

for an open interval $I_{t_0} \ni t_0$, *with* $t_0 \in (0,1)$. *Let* x_n *be the kernel estimate of* m *defined above, 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* $h = an^{-1/(3+2q)}$, *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* (30), $h = an^{-1/(3+2q)}$, *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, $h = l_2(n;a)n^{-rd/(2+q+rd)}$ *with* $a > 0$ *and* $n \mapsto l_2(n;a)$ *is a slowly varying function defined in the proof below.*

Then, correspondingly, we obtain

$$d_n^{-1}\{\tilde{m}(t_0) - m(t_0)\} \xrightarrow{\mathcal{L}} T[A_m(\cdot) + \tilde{v}(\cdot;t_0)](0),$$

as $n \to \infty$, *where*

$$\tilde{v}(s;t) = c\int w(s-u)k'(u)\,du, \tag{43}$$

$$A_m(s) = \frac{m^{(q+1)}(t_0)}{(q+1)!}\int (u+s)^{q+1}k(u)\,du. \tag{44}$$

and respectively

[**a**] $w = B\ ;\ c = \sigma a^{-(q+3/2)}$,

[**b**] $w = B\ ;\ c = \kappa a^{-(q+3/2)}$,

[**c**] $w = B_{r,\beta}\ ;\ c = |\eta_r|a$ (*where* β *defined in* (32)).

Proof If we choose $d_n = h$ in such a way so that

$$d_n^{-(q+1)}(nh)^{-1}\sigma_{\hat{n}} \to c > 0, \tag{45}$$

then we obtain the first part of Assumption 1

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

as in the proof of Theorem 6.

For the rescaling of the deterministic part g_n, similarly to as in the proofs of Theorem 6 and Theorem 8, one can show that the second part of Assumption 1 holds, with

$$A_m(s) = \frac{m^{(q+1)}(t_0)}{(q+1)!} \int (u+s)^{q+1} k(u)\, du.$$

The proof that Assumption 2 holds is similar to the reasoning in the proof of Theorem 8.

It only remains to check whether $d_n^{-1}(nh)^{-1}\sigma_{\hat{n}} \to c > 0$ for the three types of dependence.

1. Independent case [**a**]: We have $\sigma_{\hat{n}}^2 = \sigma^2 n d_n$. Thus $d_n^{-(q+1)}(nh)^{-1}\sigma_{\hat{n}} = \sigma n^{-1/2} d_n^{-(q+3/2)}$, and (45) is satisfied with $c = \sigma a^{-(q+3/2)}$ if $d_n = h = an^{-1/(3+2q)}$.

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

3. 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

$$\begin{aligned} d_n^{-(q+1)}(nh)^{-1}\sigma_{\hat{n}} &= d_n^{-(q+1)}(nh)^{-1}|\eta_r|(nd_n)^{1-rd/2} l_1(nd_n)^{1/2} \\ &\to |\eta_r| a \end{aligned} \tag{46}$$

   if and only if

$$d_n = n^{-rd/(q+2+rd)} l_2(n;a), \tag{47}$$

   where l_2 is another function slowly varying at infinity, implicitly defined in (46). Thus (45) follows with $c = |\eta_r| a$ and $h = d_n$ given in (47). □

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.

The work done here has been with the use of kernel based methods for the preliminary estimator x_n of x. Other methods, such as wavelet based ones, are possible. We would like to emphasise that the only assumptions required are given in Assumptions 1 and 2.

The small simulation study that has been performed in the supplementary material deals with independent data, as also done in Dette et al. [12] and Birke and Dette [5]. This independence framework is the only one considered in the density context too. A larger panel of dependence situations in the comparisons would clearly be of interest.

We have studied applications to density and regression function estimation. Other estimation problems that are potentially possible to treat with our methods are e.g. spectral density estimation, considered by Anevski and Soulier [2], and deconvolution, previously studied by van Es et al. [37].

6 Acknowledgements

We thank Philippe Soulier for valuable discussions during numerous occasions, and in particular for pointing out the simple conditions for Lemma 4. The second author is also grateful to Philippe Barbe who provided several precious suggestions for paper [15]. Both authors would like to thank the referee of a previous version of the paper, for pointing out several deficiencies therein. We would also like to thank an Associate Editor and two referees for the many valuable comments, and in particular for enquiring about the q vanishing derivatives case, which has given fruit to the results in Section 4. A special thank also to Mrs Stenshuvud for providing inspiration.

Supplementary material: The supplementary material gives further conditions under which Assumption 2 holds, with application to the density and regression function estimation cases, stated in Appendix B, as well as all proofs. Furthermore it contains a simulation study that illustrates the finite sample behaviour of our estimator, and compare it to other estimators that are considered in the paper of Birke and Dette [5].

References

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

[2] D. Anevski and Ph. Soulier. Estimation of a monotone spectral density. *Ann. Statist.*, 39(1):418–438, 2011.

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

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

[5] M. Birke and H. Dette. A note on estimating a smooth monotone regression by combining kernel and density estimates. *J. Nonparametr. Stat.*, 20(8):679–691, 2008.

[6] M. Birke, N. Neumeyer, and S. Volgushev. The independence process in conditional quantile location-scale models and an application to testing for monotonicity. Submitted, 2013.

[7] M. Birke, N. Neumeyer, and S. Volgushev. The independence process in conditional quantile location-scale models and an application to testing for monotonicity. *arxiv*, arXiv:1609.07696, 2016.

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

[9] V. Chernozhukov, I. Fernandez-Val, and A. Galichon. Improving point and interval estimates of monotone functions by rearrangement. *Biometrika*, 96(3):559–575, 2007.

[10] V. Chernozhukov, I. Fernandez-Val, and A. Galichon. Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125, 2010.

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

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

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

[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] H. K. Jankowski and J. A. Wellner. Estimation of a discrete monotone distribution. *Electron. J. Stat.*, 3:1567–1605, 2009.

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

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

[25] 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.

[26] 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.

[27] H. Mukerjee. Monotone nonparametric regression. *Ann. Statist.*, 16(2):741–750, 1988.

[28] N. Neumeyer. A note on uniform consistency of monotone function estimators. *Statist. Probab. Lett.*, 77(7):693–703, 2007.

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

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

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

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

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

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

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

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

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

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

[39] S. Volguchev. Smoothed quantile regression processes for binary response models. arXiv:1302.5644, 2013.

[40] S. Volguchev, M. Birke, H. Dette, and N. Neumeyer. Significance testing in quantile regression. *Electronic Journal of Statistics*, 7:105–145, 2013.

[41] S. Volgushev and H. Dette. Nonparametric quantile regression for twice censored data. *Bernoulli*, 19(3):748–779, 2013.

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

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

A Proofs of Section 2

Proof of Lemma 1: Assertions (i) and (ii) both follow from 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 $u \in (f(t_0^+), f(t_0^-))$. □

Proof of Lemma 2: (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 $u \in f(I)$. □

Proof of Lemma 3: 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 $t \in I - c$, $T_{I-c}(f_c)(t) = \inf\{u \in f(I) : r_{f,I}(u) \leq t + c - a\} = T_I(f)(t + c)$. □

Proof of Theorem 1: 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 supremum norm, which implies the statement of the theorem. □

B Maximal bounds for rescaled partial sum and empirical processes

In this section we state conditions under which Assumption 2 holds. Further specialisations of these conditions to density and regression function estimation are given in the supplementary material.

Recall that

$$\begin{aligned} \tilde{g}_n(s) &= d_n^{-1}\{x_{b,n}(t_0 + sd_n) - x_{b,n}(t_0)\}, \\ \tilde{v}_n(s) &= d_n^{-1} v_n(t_0 + sd_n). \end{aligned} \tag{48}$$

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.

The proof of the following Lemma is given in the supplementary material.

Lemma 6. *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*

$$x'_{b,n}(t + sd_n) \to x'(t), \tag{49}$$

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

Then (14) *and* (15) *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* $M > 0$

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

(B). For every $\delta > 0$ *and finite* $M > 0$ *there is a finite* C*, not depending on* δ*, such that for each* $c > C$

$$\limsup_{n\to\infty} P\left\{\sup_{s\in d_n^{-1}(0,\ell(n))} \tilde{v}_n(s) > \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \quad < \quad \delta, \tag{50}$$

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

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

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

D. Anevski*, A.-L. Fougères†

November 2017

Abstract

This is the supplementary material for Anevski and Fougères, "Limit properties of the monotone rearrangement for density and regression function estimation", which in this document is referred to as the main article.

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

1 Maximal bounds for rescaled partial sum and empirical processes

In this section we derive conditions under which Assumption 2 in the main article holds, for the density and regression function estimation cases. Recall the definitions of the localised and rescaled bias and process parts

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

*Centre for Mathematical Sciences, Lund University, BOX 118, 221 00 Lund, Sweden

†Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne cedex, France

We recall that it is enough to establish Assumption 2 in the main article for $\tilde{g}_n + \tilde{v}_n$. The main general result that we will prove is the following lemma.

Lemma 1. *Let* $\mathrm{supp}(k) \subset [-1, 1]$. *Suppose that Assumption 1 in the main article 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*

$$x'_{b,n}(t + sd_n) \quad \to \quad x'(t), \tag{2}$$

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

Then (14) *and* (15) *in the main article 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* $M > 0$

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

(B). For every $\delta > 0$ *and finite* $M > 0$ *there is a finite* C, *not depending on* δ, *such that for each* $c > C$

$$\limsup_{n \to \infty} P\left\{\sup_{s \in d_n^{-1}(0, \ell(n))} \tilde{v}_n(s) > \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \quad < \quad \delta, \tag{3}$$

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

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

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 (14) in the main article, we first prove that for each $\delta > 0$ there is a $0 < M < \infty$ and a $0 < c < \infty$, with c independent of δ, such that

$$\liminf_{n\to\infty} P\{\sup_{s\geq c}(\tilde{g}_n + \tilde{v}_n)(s) < -M\} \quad \geq \quad 1-\delta. \tag{4}$$

Let $\tilde{g}_n$ be defined in (1). 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 (2) we obtain

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

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 $\text{supp}(x_n) \subset (\min_{1\leq i\leq n} t_i - h, \max_{1\leq i\leq n} t_i + h)$. Using the rescaling $t = t_0 + sd_n$ this implies that

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

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

$$\liminf_{n\to\infty} P\left\{\sup_{(c,d_n^{-1}\max t_i)} (\tilde{g}_n + \tilde{v}_n)(s) < -M\right\} \quad \geq \quad 1-\delta. \tag{5}$$

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 (5) follows from the two results

$$\liminf_{n\to\infty} P\left\{\sup_{s\in d_n^{-1}(\epsilon,\max t_i)} \tilde{v}_n(s) < \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \;>\; 1-\delta, \qquad (6)$$

$$\liminf_{n\to\infty} P\left[\cap_{s\in(c,d_n^{-1}\epsilon)}\{\tilde{v}_n(s) < \frac{M}{2} - \tau(s-c)\}\right] \;>\; 1-\delta. \qquad (7)$$

The relation (7) is satisfied by assumption (A) and thus we need to treat (6). 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 (6) as soon as for $c > C$

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

i.e. (3).

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

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

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 (4) will follow as soon as

$$\liminf_{n\to\infty} P\left\{ \sup_{(c,d_n^{-1}K)} (\tilde{g}_n + \tilde{v}_n)(s) < -M \right\} \quad > \quad 1-\delta. \tag{8}$$

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

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

so that (8) follows from

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

and (7), which ends the derivation for the case (ii).

Now we prove that with M as above

$$\liminf_{n\to\infty} P\left\{ \inf_{\inf I_1 \leq s \leq \sup I_0} y_n(s) \geq -M \right\} \quad > \quad 1-\delta. \tag{9}$$

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 $s_M = -M_c/2A$

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

for n large enough. Finally from (4) 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 (4) and (9) hold, and (6) is proven.

Equation (15) in the main article can be proven in a similar way, which yields the lemma. □

Lemma 1 states conditions (A) and (B) as sufficient conditions for Assumption 2 in the main article.

To further simplify condition (B) in Lemma 1, using Boole's inequality and the stationarity of the process $\tilde{v}_n$ we get in both cases (i) and (ii)

$$\begin{aligned} & P\left\{\sup_{d_n^{-1}(0,\ell(n))} \tilde{v}_n(s) > \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\} \\ & \leq d_n^{-1}\ell(n) P\left\{\sup_{(0,1)} \tilde{v}_n(s) > \frac{M}{2} - \tau(d_n^{-1}\epsilon - c)\right\}, \end{aligned} \tag{10}$$

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 (10) 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 (10) tends to 0, as $n \to \infty$:

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

The examination of the convergence of $\alpha(n)$ will be treated for the density and regression estimation problems, separately in subsections B.1 and B.2.

To further simplify (A) in Lemma 1, note that

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

Thus, taking complements, part (A) of Lemma 1 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

$$P\left\{\sup_{s\in[0,1)} \tilde{v}_n(s) > a\right\} \leq Ca^{-p} \tag{11}$$

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 1 follows from (11) with $p > 1$; inequalities of the form (11) will next be treated.

1.1 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 (32) in the main article)

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

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

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

and thus

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

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 1. *Let $p \geq 2$ be given and assume that the sequence $\{\epsilon_i\}_{i\geq 1}$ satisfies $\mathbb{E}\epsilon_1^p < \infty$. Then under the assumptions of Theorem 6 in the main article*

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

where C is a finite constant, and $\tau = \sup_{t\in I} x'(t)$.

Proof Let $a := M/2 + \tau c$ and $b := -\tau\delta$. One has

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

with

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

[a] If $\{\epsilon_i\}$ is an i.i.d. sequence the moment bound in Theorem 2.9 in Petrov [13] 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 (25) in the main article, then Theorem 1 in Doukhan [7] 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 [2] is satisfied, so that Theorem 12.2 in Billingsley [2] implies

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

where the first inequality follows $\xi\leq|\xi|$ for arbitrary $\xi\in\mathbb{R}$ and the second from Theorem 12.2 in Billingsley [2], and with $C(\hat{n})=c'(\hat{n}+\hat{n}^{p/2})$ for i.i.d. data and $C(\hat{n})=c''\max(\hat{n},\hat{n}^{p/2})$ for mixing data. 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

$$P\left(\sup_{s\in(0,1)}w_{\hat{n}}(s)>a+bh^{-1}\right) \leq Ch^p, \tag{12}$$

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 (31) in the main article, and according to de Haan [5], equation (12.42) in Billingsley [2] 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 [2] then leads to

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

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 (12) holds for $p = 2$. □

Corollary 1. *Suppose the assumptions of Theorem 6 in the main article are satisfied; then Assumption 2 in the main article holds for $y_n = g_n + \tilde{v}_n$ and for y as defined in (13) in the main article in each context* **[a]**, **[b]** *and* **[c]** *listed in Theorem 6 in the main article.*

Proof Note first that if $x_{b,n}$ is defined by (28) in the main article, 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 1 is that

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

Thus, condition $(B)(ii)$ in Lemma 1 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 1 for i.i.d., mixing and subordinated Gaussian long range dependent sequences. Condition (A) of Lemma 1 follows immediately from Proposition 1. Hence Lemma 1 can be applied, so that Assumption 2 holds for $y_n = g_n + \tilde{v}_n$. An analogous result for y defined by (13) in the main article follows easily from the stationarity and finite second moments of $\tilde{v}_n$. □

1.2 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 2. *Under the assumptions of Theorem 5 in the main article, there exists a positive constant C' such that*

$$P\left(\sup_{s\in(0,1)} w_{n,d_n}(s) > M/2 - \tau(\epsilon h^{-1} - c)\right) \;\leq\; C'h^4 \,.$$

Proof Let $a := M/2 + \tau c$ and $b := -\tau\epsilon$. For independent or mixing data satisfying the assumptions of Theorem 5, formula (84) in the proof of Lemma 5 in the tech report version of Anevski and Hössjer [1] says that

$$E|w_{n,d_n}(s;t_0) - w_{n,d_n}(s';t_0)|^4 \;\leq\; \frac{k}{\epsilon}(s-s')^2 \tag{13}$$

with $0 < \epsilon < 1$ chosen such that $\epsilon \leq nd_n|s-s'|$. Similarly to the arguments in that proof, taking p a number which satisfies

$$\frac{\epsilon}{nd_n} \;\leq\; p,$$

and m a finite number, (13) and Theorem 12.2 in Billingsley implies that

$$P(\max_{1\leq i\leq m} |w_{nd_n}(s+ip;t_0) - w_{nd_n}(s;t_0)| \geq \lambda) \;\leq\; \frac{\tilde{k}}{\epsilon\lambda^4} m^2 p^2,$$

with $\tilde{k}$ a constant only depending on the mixing coefficients. Noting that ϵ, p are constants, (satisfying the above restrictions) taking $s = 0$ and noting that $w_{n,d_n}(0; t_0) = 0$, this implies the statement of the Proposition. $\square$

Corollary 2. *Suppose the assumptions of Theorem 5 in the main article are satisfied; then Assumption 2 holds for $y_n = g_n + \tilde{v}_n$ and for y as defined in (13) in the main article, in each context* **[a]** *and* **[b]** *listed in Theorem 5.*

Proof Note first that if $x_{b,n}$ is defined by (23) in the main article, 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 $n \to \infty$. Furthermore, a consequence of Proposition 2 is that

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

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

$$P\{\max_{1\leq i\leq n} t_i \leq \ell(n)\} \to 1 \text{ and } h^{p-1}\ell(n) \to 0, \tag{14}$$

as $n \to \infty$, and with $p \geq 4$. Noting that

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

and that $h = d_n$, we see that the conditions in (14) are implied by

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

which is equivalent to

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

In the case of i.i.d. and mixing data, when $d_n = n^{-1/3}$, this implies that 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 $p \geq 4$ in Proposition 2, implies that the existence of four moments suffices to establish $(B)(i)$ in Lemma 1 for i.i.d. and mixing data. Condition (A) in Lemma 1 is immediate from Proposition 2. Hence Lemma 1 can be applied, so that Assumption 2 in the main article holds for $y_n = g_n + \tilde{v}_n$. An analogous result for y defined by (13) in the main article follows easily from the stationarity and finite second moments of $\tilde{v}_n$. $\square$

2 Illustration of the finite-sample properties

Simulation studies 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, are e.g. Fougères [8], Meyer and Woodroofe [11], Hall and Kang [9], Chernozhukov et al. [4]. In this section we illustrate via a small simulation study the finite sample behavior of the rearrangement estimator in the regression setting, and compare it with the competitors already considered in Birke and Dette [3]. Our aim is to give complementary information to the one collected from Dette et al. [6] and Birke and Dette [3]. We therefore make use of the same simulation setting, which is recalled here for completeness. The only difference is that we consider the framework of fixed design as developed in our paper, whereas they illustrated the random design setting. Indeed, we have also performed the simulation study in the random design case, and since it gave similar results, we prefer to show the fixed design only, for sake of coherence with the model considered in our paper.

Consider the non parametric regression model (ii) as stated in Section 1, with independent and normally distributed errors with standard deviation $\sigma = 0.05$. We follow Birke and Dette [3] and generate samples of size $n = 25$ from (ii) for the two once continuously differentiable regression functions

$$m_1(x) = \begin{cases} -8x^2 + 4x & x \in [0, \frac{1}{4}] \\ \frac{1}{2} & x \in (\frac{1}{4}, \frac{3}{4}] \\ 8x^2 - 12x + 5 & x \in (\frac{3}{4}, 1] \end{cases}$$

and

$$m_2(x) = \begin{cases} \frac{12}{5}x^2 & x \in [0, \frac{1}{4}] \\ -\frac{2}{5}x^2 + \frac{7}{5}x + \frac{7}{40} & x \in (\frac{1}{4}, \frac{3}{4}] \\ \frac{12}{5}x^2 - \frac{14}{5}x + \frac{7}{5} & x \in (\frac{3}{4}, 1] \end{cases}$$

As initial (unconstrained) estimate $\hat{m}$, we use either a local linear regression estimate, as available from the R package `KernSmooth` via the procedure `locpoly`, or a Nadaraya-Watson regression estimate, also available from `KernSmooth` (procedure `ksmooth`). The smooting parameter is in both cases denoted by h_r. Then three competing isotone estimators are implemented: (a) the estimator $\hat{m}_I$ defined by Dette et al. [6] and also studied by Birke and Dette [3], available from the R package `monreg` (procedure `monreg`) ; therein, $h_d = h_r^{1.01}$ is used as the second smoothing parameter, as done in Birke and Dette [3] ; (b) the estimator $\hat{m}_{SI}$ discussed in Mukerjee [12] and Mammen [10]; it consists in monotonizing the initial estimate $\hat{m}$ by the PAVA algorithm, that is available from the R package `Iso` (procedure `pava`); (c) the estimator $T(\hat{m})$ introduced in Definition 2 of the main paper as the increasing rearrangement of $\hat{m}$; recall that the implementation of $T(\hat{m})$ is straightforward, since it only requires to sort the values of the initial estimate $\hat{m}$. Note that we could also have considered the estimator $\hat{m}_{IS}$ as done in Birke and Dette [3], but since the forementioned authors obtained very similar results for $\hat{m}_{SI}$ and $\hat{m}_{IS}$, we decided to skip one of them, keeping only $\hat{m}_{SI}$.

As a simple illustration of the four competitors described above, Figure 1 shows the results obtained for one sample of $n = 25$ observations for the unconstrained local linear regression estimate and the three isotone estimates, for both regression functions m_1 and m_2. Two different bandwidths ($h = 0.02$ and 0.05) are used for the initial step, illustrating the effect of smoothing. Some patterns (that occur for most of the samples) can be observed on Figure 1: (a) the smaller the smoothing parameter h is, the bigger effect the monotonizing step has; (b) except for very small bandwidths, the three isoton competitors perform rather similarly; (c) the implementation in `monreg` of the density regression estimate $\hat{m}_I$ seems to break down for very small bandwidths; (d) the rearrangement $T(\hat{m})$ looks slightly downwards biased.

In order to evaluate more accurately the comparative performances of the four competitors, Figure 2 shows the L2-error obtained from 500 samples of $n = 25$ observations for each of these estimators. Figure 2 corroborates several features that could be seen on Figure 1. More precisely, the following comments can be formulated from Figure 2: (a) the smaller the smoothing parameter h_r is, the bigger effect the monotonizing step has; (b) except for very

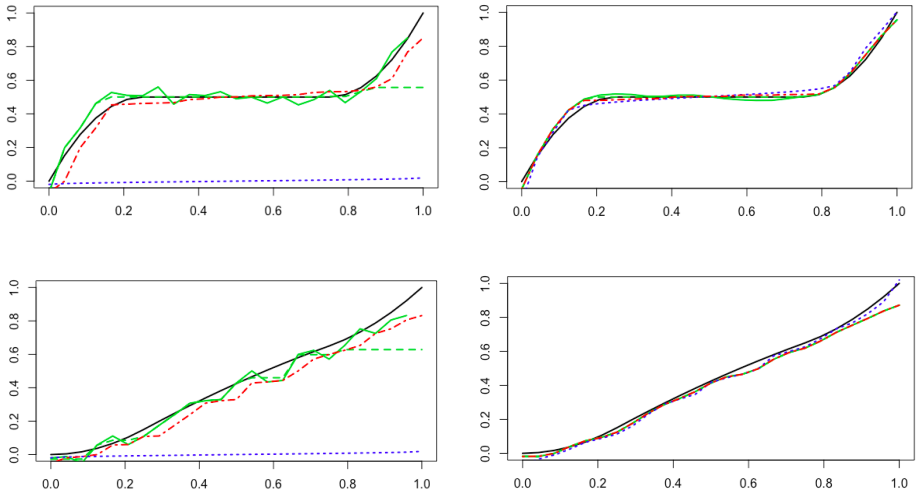

Figure 1: Estimates as a function of $t \in (0,1)$ obtained for one sample of $n = 25$ observations for the unconstrained local linear $\hat{m}(t)$ (green, solid line), the monotonized $\hat{m}_{SI}(t)$ (green, dashed line), the density regression estimate $\hat{m}_I(t)$(blue, dotted line), and the rearrangement $T(\hat{m})(t)$ (red, dot-dashed lines). The true regression function (black, solid line) is respectively m_1 on the first row, and m_2 on the second row. The bandwidth used for the evaluation of $\hat{m}$ is respectively $h_r = 0.02$ on the left hand-side, and $h_r = 0.05$ on the right hand-side.

small bandwidths, both isotone competitors $\hat{m}_{SI}$ and $T(\hat{m})$ perform rather similarly; the third one $\hat{m}_I$ looks less performant. Note that this last point does not completely confirms what can be seen from Figure 1(b) of Birke and Dette [3]. We think that we have reproduced the same experimental design, but were not able to find what can make our findings partially discordant ; (c) the implementation in `monreg` of the density regression estimate $\hat{m}_I$ breaks down for very small bandwidths when starting from the local-linear regression estimate, but this drawback does not occur when starting from the Nadaraya-Watson regression estimate ; (d) as expected, the estimation in case of the regression model m_2 is globally easier than in case of m_1 for all the competitors.

Lastly, the additional information of comparing the two initial estimates (denoted respectively by $\hat{m}_{LL}$ and $\hat{m}_{NW}$) and their rearrangements $T(\hat{m}_{LL})$

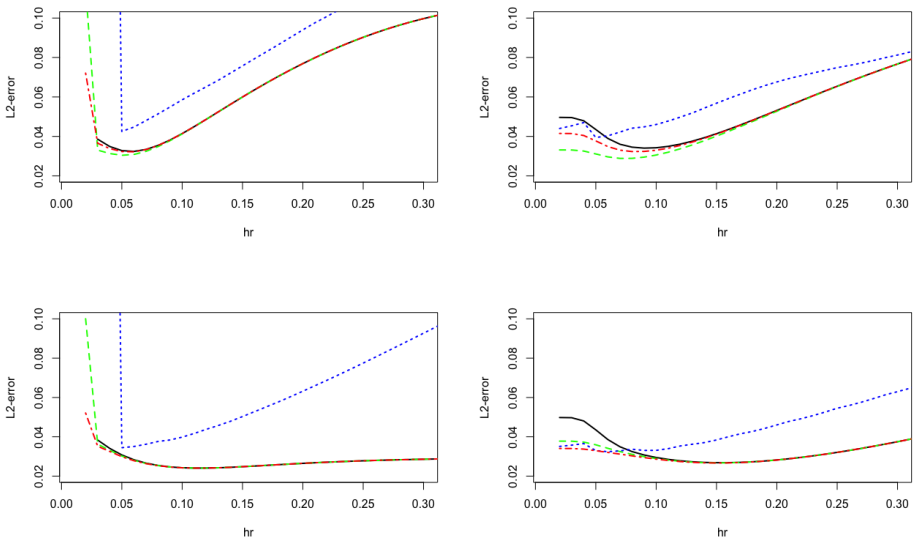

Figure 2: L2-error as a function of the bandwidth $h_r \in [0.01, 0.3]$ obtained from 500 samples of $n = 25$ observations for the unconstrained local linear $\hat{m}$ (black, solid line), the monotonized $\hat{m}_{SI}$ (green, dashed line), the density regression estimate $\hat{m}_I$(blue, dotted line), and the rearrangement $T(\hat{m})$ (red, dot-dashed lines). Initial $\hat{m}$ used is the local-linear estimate on the left hand-side part of the figure, and the Nadaraya-Watson's one on the right hand-side part of the figure. The regression model is respectively m_1 on the first row, and m_2 on the second row.

and $T(\hat{m}_{NW})$ is provided on Figure 3, which shows the L2-error obtained from 500 samples of $n = 25$ observations for both unconstrained estimates and the associate rearranged estimates. One can see from Figure 3 that rearranging the Nadaraya-Watson estimate $\hat{m}_{NW}$ produces a larger improvement than rearranging the local-linear estimate $\hat{m}_{LL}$, especially for small bandwidths. Besides, starting from $\hat{m}_{LL}$ seems to slightly outperform starting from $\hat{m}_{NW}$. Finally, as already said from Figure 2, performing the estimation in case of the regression model m_2 is easier than for m_1.

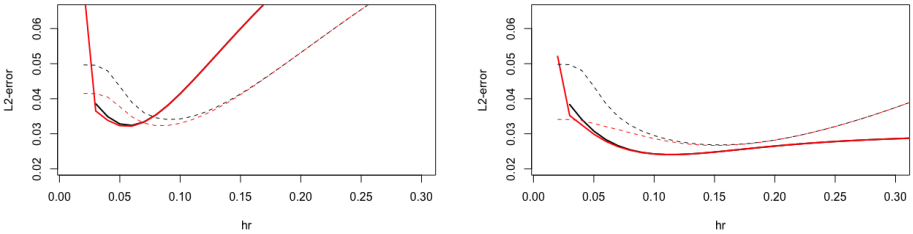

Figure 3: L2-error as a function of the bandwidth $h_r \in [0.01, 0.3]$ obtained from 500 samples of $n = 25$ observations for both unconstrained (in black) estimates (solid line for $\hat{m}_{LL}$ and dashed line for $\hat{m}_{NW}$) and the associate rearranged (in red) estimates $T(\hat{m}_{LL})$ (solid line) and $T(\hat{m}_{NW})$ (dashed line). The regression model is respectively m_1 on the left, and m_2 on the right.

References

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

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

[3] M. Birke and H. Dette. A note on estimating a smooth monotone regression by combining kernel and density estimates. *J. Nonparametr. Stat.*, 20(8):679–691, 2008.

[4] V. Chernozhukov, I. Fernandez-Val, and A. Galichon. Improving point and interval estimates of monotone functions by rearrangement. *Biometrika*, 96(3):559–575, 2007.

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

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

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

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

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

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

[11] 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.

[12] H. Mukerjee. Monotone nonparametric regression. *Ann. Statist.*, 16(2):741–750, 1988.

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