Journal of Multivariate Analysis 170 (2019) 95–114

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Data depth for measurable noisy random functions

Stanislav Nagy [a,*], Frédéric Ferraty [b]

[a] Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic

[b] Institut de Mathématiques de Toulouse, Université Paul-Sabatier, 118, route de Narbonne, F-31062 Toulouse Cedex 9, France

ARTICLE INFO

Article history:
Received 16 November 2017
Available online 19 November 2018

Keywords:
Asymptotics
Data depth
Functional data
Measurement error
Rate of convergence
Smoothing

AMS subject classifications:
Primary
62G20
62G35
Secondary
62G08

ABSTRACT

In the literature on data depth applicable to random functions, it is usually assumed that the trajectories of all the random curves are continuous, known at each point of the domain, and observed exactly. These assumptions turn out to be unrealistic in practice, as the functions are often observed only on a finite grid of time points, and in the presence of measurement errors. In this work, we provide the necessary theoretical background enabling the extension of the statistical methodology based on data depth to measurable (not necessarily continuous) random functions observed within the latter framework. It is shown that even if the random functions are discontinuous, observed discretely, and contaminated with additive noise, many common depth functionals maintain the fine consistency properties valid in the ideal case of completely observed noiseless functions. For the integrated depth for functions, we provide uniform rates of convergence over the space of integrable functions.

1. Introduction

In functional data analysis, it is usually assumed that the random curves are observed perfectly, i.e., that the functional values are known precisely at all points of their domain (also referred to as time); see the monographs of Ramsay and Silverman [56], Ferraty and Vieu [20], or Horváth and Kokoszka [27], or the recent surveys of Cuevas [10] and Goia and Vieu [24]. While this assumption is appropriate for the development of basic methodology for dealing with functions, in practice it is seldom satisfied. On the contrary, real functional datasets are never observed continuously in time, and often a substantial amount of additive noise, introduced by the measurement procedure, prevents from observing these values accurately. If the functions are observed densely enough to permit the reconstruction of their original trajectories, data analysis is routinely performed after a preprocessing step. Here the discrete, inaccurately observed values are smoothed to approximate the unobservable, supposedly regular function. Nonetheless, such data imputation may affect the subsequent statistical analysis. In the present contribution, we consider the field of statistical depth for functional data, and provide theoretical justification for the use of pre-smoothed functional data in combination with the analysis based on depth.

Statistical depth is a nonparametric instrument applicable to multivariate (and infinite-dimensional) data. For these, it aims to introduce a generalization of quantiles. By identifying points in the sample space located "centrally" with respect to a given probability distribution P, it is capable of distinguishing "typical" and "potentially outlying" observations by assigning to them very high, or very low values of depth, respectively. For a general account on data depth in the classical setup of finite-dimensional spaces, see [67] and the references therein.

* Corresponding author.
E-mail address: nagy@karlin.mff.cuni.cz (S. Nagy).

https://doi.org/10.1016/j.jmva.2018.11.003
0047-259X/

The first extension of depth to functional data analysis is due to Fraiman and Muniz [21]. Since then, numerous depth functionals have been proposed, mostly in the space of continuous functions over a compact interval. Let us mention [7,8,13,38,39,42,50]. Most of these diverse approaches can be classified into three groups of depths following similar ideas, as discussed in [23,49]: (i) integrated depths, (ii) band depths, and (iii) infimal depths.

In some fields of functional data analysis, noisy random functions have already been studied intensively; a body of literature on this topic encompassing functional linear regression, functional principal components, continuous additive models, etc. can be found in [9,28,32,34,44,52,55,64–66], among others. In those works, noisy functional data are typically either (i) pre-smoothed in the first step, and then statistical analysis is performed for the smoothed approximants of the data curves; or (ii) the statistical procedure in question is revisited, and suitably applied directly to the available discrete noisy observations. The latter approach is usually less versatile, and requires a specific type of data such as random curves in Hilbert spaces decomposable into their functional principal components. In this paper, we thus follow the former method, and explore its potential in the field of functional data depth.

For functional data depth, no theoretical results of this kind are available. Nevertheless, also here the pre-smoothing step is very common, and known to give good results in practice; see [8,11,12,29,30,40,41,50,59,60]. In the present paper we bridge the concepts of noisy functional data and data depth, and provide theoretical properties of depth computed from pre-smoothed functions. To this end, we focus on the main categories of depth functionals (integrated depths, band depths, and infimal depths), and establish the mathematical background necessary for depth-based analysis of discretely observed and contaminated curves.

In Section 2 we gather the preliminaries and notation. Then, the first result establishing mild conditions for the weak convergence of the empirical measures based on the reconstructions of the noisy curves is given. This is a non-trivial extension of the classical theorem of Varadarajan [62] to functional data. Considering noisy (hence discontinuous) curves, it is necessary to abandon the framework of depths defined in the space of continuous functions. In Section 3, extensions of the main representatives of integrated and band depths into the space of integrable functions are given, and their properties are examined. Some new consistency results for these are provided. In particular, it is shown that under suitable conditions both integrated, and (adjusted) band depths retain the fine consistency properties valid in the space of perfectly observed continuous functions. In Section 4 we focus on integrated depths. For the first time, we provide the rate of convergence of a depth functional over the whole functional space. Using an extension of the theory from Section 2, we derive the rate of convergence of the integrated depth in the setup of noisy random functions, and demonstrate that if the noisy curves are observed densely in time, then the same rate of convergence as for perfectly observed data can be achieved. Some interesting extensions of our theory are discussed in Section 5. A simulation study illustrating the performance of the proposed methodology is given in Section 6. Its complete results can be found in the Online Supplement accompanying the paper. The proofs are provided in the Appendix.

2. Weak convergence for noisy functional data

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be the probability space on which all random elements are defined. For a measurable space S, $\mathcal{P}(S)$ is the set of all probability measures defined on S, and $X \sim P \in \mathcal{P}(S)$ stands for a random variable X in S with distribution P. For $P \in \mathcal{P}(S)$, $\omega \in \Omega$ and any integer $n \geq 1$, $P_n(\omega) = P_n \in \mathcal{P}(S)$ denotes the empirical measure of the random sample $X_1, \ldots, X_n$ from P corresponding to ω. The weak convergence, as $\nu \to \infty$, of a sequence $\{P_\nu : \nu \in \mathbb{N}\} \subset \mathcal{P}(S)$ to $P \in \mathcal{P}(S)$ is denoted by $P_\nu \xrightarrow{w} P$. The space of continuous functions over $[0, 1]$ is denoted by $\mathcal{C}[0, 1]$. $L^2[0, 1]$ is the space of measurable functions $x\colon [0, 1] \to \mathbb{R}$ for which x^2 is Lebesgue integrable. For λ the Lebesgue measure on $\mathbb{R}$, the norm of $x \in L^2[0, 1]$ is denoted by

$$\|x\| = \sqrt{\int_0^1 x(t)^2 \,\mathrm{d}\,\lambda(t)}. \tag{1}$$

For $X \sim P \in \mathcal{P}\{L^2[0, 1]\}$, $P_t \in \mathcal{P}(\mathbb{R})$ stands for the marginal distribution of $X(t)$ for $t \in [0, 1]$, and F_t for its distribution function. The marginal distribution of the empirical measure P_n at $t \in [0, 1]$ is denoted by $P_{n,t} \in \mathcal{P}(\mathbb{R})$, and its distribution function by $F_{n,t}$.

Consider a random function $X \sim P \in \mathcal{P}\{L^2[0, 1]\}$, i.e., a mapping

$$X\colon [0, 1] \times \Omega \to \mathbb{R}\colon (t, \omega) \mapsto X(t, \omega), \quad \{\mathcal{B}([0, 1]) \times \mathcal{F}\}\text{-measurable},$$

such that $X(\cdot, \omega) \in L^2[0, 1]$ for P-almost all $\omega \in \Omega$. $\mathcal{B}[0, 1]$ stands for the Borel sets in $[0, 1]$. Note that we assume that X is measurable with respect to the product σ-algebra. This technical assumption is needed to ensure proper measurability of all the considered quantities, especially with respect to the integrated depth for discontinuous functions; see [46]. When not necessary, the argument of the random element ω in X will be omitted.

Let $X_1, \ldots, X_n$ be independent random functions distributed as X. Assume that the complete set of functional values $\{X_i(t)\colon t \in [0, 1]\}$ with $i \in \{1, \ldots, n\}$ cannot be observed directly. Instead, for each i, only the finite vector

$$(X^*_{i,1}, \ldots, X^*_{i,m_i}) = \big(X_i(T_{i,1}) + \varepsilon_{i,1}, \ldots, X_i(T_{i,m_i}) + \varepsilon_{i,m_i}\big) \in \mathbb{R}^{m_i}, \tag{2}$$

is known to the observer. Here, $m_i \geq 1$ is an integer and $\left(T_{i,1}, \ldots, T_{i,m_i}\right) \in [0,1]^{m_i}$ is a random sample of size m_i from a distribution $T \sim P_T \in \mathcal{P}([0,1])$. $\varepsilon_{i,1}, \ldots, \varepsilon_{i,m_i}$ are independent, centred random variables such that $\operatorname{var} \varepsilon_{i,j} = \sigma^2(T_{i,j})$, where $\sigma^2 \colon [0,1] \to [0,\infty)$ is a deterministic function. In what follows, all the variables $\varepsilon_{i,j}$, $T_{i,j}$ and X_i are assumed to be independent; a more general scenario will be discussed in Section 5. The choice $\sigma^2 \equiv 0$ covers the case when the random functions are observed discretely without noise. This setup was considered by Nagy et al. [48] in the space $\mathcal{C}[0,1]$, and in the present contribution we extend these results substantially to discontinuous functions contaminated with measurement errors, using very different proof techniques.

Though the number of the observed functional values m_i in (2) can be considered fixed for a single function X_i, it will typically grow as the sampling process from P continues, i.e., $m_i \to \infty$ as $i \to \infty$. Random functions observed within this setup are in the literature often called dense functional data [34], in contrast to sparse functional data [65], where m_i is bounded as i grows.

The assumption of having a non-random m_i in (2) is made for mere notational convenience. In what follows m_i could be replaced by its stochastic counterpart, and where applicable, $m_i \to \infty$ as $i \to \infty$ can always be replaced by $m_i \xrightarrow{\text{a.s.}} \infty$ as $i \to \infty$ for m_i independent of the other random variables. Likewise, it is possible to consider a related model, where in the random sample $X_1, \ldots, X_n$ the function X_i is observed at $m_{i,n}$ points. In this case, one can proceed analogously as pursued here, the only assumption being $m_{i,n} \geq m_i$ for all $i \geq 1$ and $n \geq i$, for some $m_i \to \infty$ as $i \to \infty$. Note that in (2) we do not have to require the usual $m_{i,n} \to \infty$ as $n \to \infty$ for all $i \geq 1$. That is, in our setup, for $i \geq 1$ fixed the ith sampled function is not bound to be observed at more points as the sampling process continues. Therefore, the currently developed theory naturally embeds also the commonly considered setup of increasingly dense functional data, where $\min(m_{1,n}, \ldots, m_{n,n}) \to \infty$ as $n \to \infty$.

For the random vector (2), a reconstruction of the unobserved curve $X_i \in L^2[0,1]$ based on these points is necessary in order to perform statistical inference on the original distribution P. This is often done by nonparametric smoothing. Here, we mainly focus on the kernel estimator of X_i; for other smoothing methods, see Section 5. For a kernel $K \colon \mathbb{R} \to [0,\infty)$ and a bandwidth $h_{m_i} > 0$, we approximate X_i, for $t \in [0,1]$, by

$$\widetilde{X}_i(t) = \frac{\sum_{j=1}^{m_i} X_{i,j}^* K\{(t - T_{i,j})/h_{m_i}\}}{\sum_{j=1}^{m_i} K\{(t - T_{i,j})/h_{m_i}\}}. \tag{3}$$

After all the curves X_i are approximated by $\widetilde{X}_i$, $P_n \in \mathcal{P}\{L^2[0,1]\}$ based on $X_1, \ldots, X_n$ can be estimated by the measure of empirical type $\widetilde{P}_n \in \mathcal{P}\{L^2[0,1]\}$ concentrated in the smoothed functions $\widetilde{X}_1, \ldots, \widetilde{X}_n$. In the following theorem we formulate a weak convergence result for $\widetilde{P}_n$. For this, we need the following assumptions:

(**A1**) $\mathrm{E}\, X(T)^2 < \infty$.
(**A2**) There exist constants $0 < r \leq R$ and $0 < b < B$ such that for the kernel K,

$$\forall_{u \in \mathbb{R}} \quad B\, \mathbf{1}(-R \leq u \leq R) \geq K(u) \geq b\, \mathbf{1}(-r \leq u \leq r).$$

(**A3**) P_T has a density over $[0,1]$ bounded from below by $c_T > 0$.
(**A4**) $\sup_{t \in [0,1]} \sigma^2(t) \leq \sigma^2$ for some $\sigma^2 < \infty$.
(**A5**) $h_{m_i} \to 0$ and $m_i h_{m_i} \to \infty$ as $i \to \infty$.

Theorem 1. *Under Assumptions (A1)–(A5),* $\mathsf{P}(\widetilde{P}_n \xrightarrow{w} P) = 1$, *i.e., the sequence* $\widetilde{P}_1(\omega), \widetilde{P}_2(\omega), \ldots$ *of empirical measures converges to* P *weakly in* $\mathcal{P}\{L^2[0,1]\}$ *for* P*-almost all* $\omega \in \Omega$.

As will be seen in Section 3, the Varadarajan-type result displayed in Theorem 1 can be directly applied to the problem of consistency of depth functionals. In Section 4, under slightly stricter assumptions imposed on X, the rates of convergence of the sequence $\widetilde{X}_1, \widetilde{X}_2, \ldots$ of approximating functions to the sequence of independent realizations of X are investigated. As such, these results provide a quantitative version of Theorem 1.

3. Depth functionals in $L^2[0,1]$

In the literature on depth for functional data, numerous approaches to depth computation have been proposed. Though most of these are constructed in the setup of continuous random functions (i.e., for $x, X \in \mathcal{C}[0,1]$), their extensions to the space of measurable functions are rather straightforward. We begin by defining rigorously some representatives of the most widely studied depth functionals, extended to the space $L^2[0,1]$. These include a general univariate integrated depth [49], the h-mode depth, and the adjusted band depth, the latter two taking the form of band depths for functions [23].

Many other depths follow the ideas behind these two general classes of functionals. Namely, the original band depths [38], the corrected band depths [37], the sparse band depths [41], the local band depths [1], various versions of multivariate band depths [30,40], the half-region depths [39], the set band depths [63], and other related functionals all fall logically into the setup of general band depths for functional data. In contrast, the so-called modified versions of all these band depths, the

modified volume depth [22], and the integrated depths of Cuevas and Fraiman [13] and Claeskens et al. [8] are representatives of depths of integrated type. Theoretical properties of all these functionals in $L^2[0,1]$ can be derived from the results given in the present section in an analogous manner. For band depths, however, adjustment as proposed by Gijbels and Nagy [23] is necessary to be introduced into the depth evaluation to avoid consistency issues.

In the definitions below, assume that the depth of the function $x \in L^2[0,1]$ with respect to the distribution of the random function $X \sim P \in \mathcal{P}\{L^2[0,1]\}$ is to be determined. We start with a generic definition of a depth functional of integrated type based on a univariate depth.

Definition (*[13,21,49]*)**.** The integrated depth of x with respect to P is given by

$$aD(x;P) = \int_0^1 D_1(x(t);P_t)\ \mathrm{d}\lambda(t). \tag{4}$$

As for the choice of the univariate depth in the definition of aD, the only assumptions we make are that $D_1\colon \mathbb{R}\times\mathcal{P}(\mathbb{R}) \to [0,1]$: $(u,Q)\mapsto D_1(u;Q)$ is a (depth) mapping that is jointly measurable in its domain, weakly continuous in the distribution argument, and universally consistent; see Sections 4.6 and 5.2 in [49]. The latter two conditions mean that D_1 satisfies

$$\sup_{u\in\mathbb{R}}|D_1(u;Q_\nu) - D_1(u;Q)| \xrightarrow[\nu\to\infty]{} 0$$

whenever $Q_\nu \xrightarrow{w} Q$ as $\nu\to\infty$ in $\mathcal{P}(\mathbb{R})$ such that Q is absolutely continuous, and that

$$\sup_{u\in\mathbb{R}}|D_1(u;Q_n) - D_1(u;Q)| \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

for any $Q\in\mathcal{P}(\mathbb{R})$ and the sequence $Q_1, Q_2, \ldots$ of the associated empirical measures. It is easy to show that these conditions are obeyed for most univariate depths. In particular, writing F_Q for the distribution function of $Q\in\mathcal{P}(\mathbb{R})$, they are true (see Appendix A of [49]) for any of

$$\begin{aligned}
D_1^a(u;Q) &= 1/2 - |1/2 - F_Q(u)|,\\
D_1^b(u;Q) &= \min\{F_Q(u), 1 - \lim_{v\to u-} F_Q(v)\},\\
D_1^c(u;Q) &= F_Q(u)\{1 - \lim_{v\to u-} F_Q(v)\},\\
D_1^d(u;Q) &= \frac{1}{J-1}\sum_{j=2}^{J} \mathsf{P}(u\in[\min(U_1,\ldots,U_j), \max(U_1,\ldots,U_j)]),
\end{aligned} \tag{5}$$

where in the last expression $J\geq 2$ and $U_1,\ldots,U_J$ is a random sample from Q. The depth D_1^a was introduced into (4) by Fraiman and Muniz [21]. If F_Q is continuous it is equivalent with D_1^b, the well-known halfspace depth [61] in $\mathbb{R}$. The depth D_1^c relates to the simplicial depth [35] in $\mathbb{R}$. In (4) it was used by Cuevas and Fraiman [13]. Note that D_1^d forms in (4) the modified band depth for functional data [38]. For $J=2$, D_1^d is equivalent with D_1^c.

In the definitions of band depths that follow, $K_D\colon [0,\infty)\to[0,\infty)$ is a continuous function such that $K_D(t)\to 0$ as $t\to\infty$. It plays the role of a smoothing kernel, and can differ from the kernel K utilized to recover the random functions in Section 2.

Definition (*[11,12]*)**.** The h-mode depth of x with respect to P is given by

$$hD(x;P) = \frac{1}{h(P)}\,\mathrm{E}\,K_D\{\|x-X\|/h(P)\}, \tag{6}$$

where $h\colon \mathcal{P}\{L^2[0,1]\}\to(0,\infty)$ is a bandwidth such that $h(P_\nu)\to h(P)$ as $\nu\to\infty$ whenever $P_\nu \xrightarrow{w} P$ as $\nu\to\infty$ in $\mathcal{P}\{L^2[0,1]\}$.

Definition (*[23]*)**.** The adjusted band depth of order $J\geq 1$ of x with respect to P is given by

$$bD(x;P) = \mathrm{E}\,K_D\left[d\left\{x, B\left(X_1,\ldots,X_J\right)\right\}\right], \tag{7}$$

where $X_1,\ldots,X_J$ are independent random functions from P,

$$\begin{aligned}
B(x_1,\ldots,x_J) &= \bigcup_{\substack{N\subset[0,1]\\ \lambda(N)=1}} \bigcap_{t\in N}\left\{y\in L^2[0,1]\colon \min_{j\in\{1,\ldots,J\}} x_j(t)\leq y(t)\leq \max_{j\in\{1,\ldots,J\}} x_j(t)\right\}\\
&= \left\{y\in L^2[0,1]\colon \min_{j\in\{1,\ldots,J\}} x_j \leq y \leq \max_{j\in\{1,\ldots,J\}} x_j\ \lambda\text{-almost everywhere}\right\}
\end{aligned}$$

is the band of functions $x_1, \ldots, x_J \in L^2[0,1]$ and

$$d\left\{x, B\left(x_1, \ldots, x_J\right)\right\} = \inf\left\{\|x - y\| : y \in B\left(x_1, \ldots, x_J\right)\right\}$$

is the distance between x and the band of functions.

The depth bD for $J = 1$ is the same as hD for the choice $h(P) = 1$ for all $P \in \mathcal{P}\{L^2[0,1]\}$. The sample versions of all the defined depths are formed by substituting the population measure P by the empirical measure P_n in (4), (6), and (7), respectively, if P_n is observed. If the complete curves are not available, the empirical measure of their reconstructions $\widetilde{P}_n$ is used. Note that in the list of depths applicable to integrable functions we omitted any version of the infimal type of depth functionals [23,43] such as the basic infimal depth for functions [42], viz.

$$iD(x; P) = \inf_{t\in[0,1]} D_1\{x(t); P_t\}, \tag{8}$$

for $x \in L^2[0,1]$, $P \in \mathcal{P}\{L^2[0,1]\}$ and D_1 as for aD. Other representative of infimal depths that can be found in the literature is the functional halfspace depth discussed in [17]. Also, the extremal depth [50] is based on a related idea. These depths are not well defined for $x \in L^2[0,1]$, since the functional values of x in (8) are given uniquely only for almost every $t \in [0,1]$. More importantly, even if the definition of the depth was modified accordingly, it can be shown that these functionals suffer from serious consistency issues in the case of discontinuous observations.

To illustrate this point, consider the example of $X \sim P \in \mathcal{P}\{L^2[0,1]\}$ defined such that the functional value $X(0)$ is standard normally distributed, and $X(t) = W(t)$ for $t > 0$, where W is a standard Wiener process on $[0,1]$, independent of $X(0)$. For this random function, say $D_1 = D_1^a$, and $x \equiv 0$, we get $D_1\{x(t); P_t\} = 1/2$ for all $t \in [0,1]$. Nonetheless, for any finite random sample from P and its empirical measure P_n, there exists a subset S of $[0,1]$ of positive Lebesgue measure such that $D_1\{x(t); P_{n,t}\} = 0$ for all $t \in S$; see Example 5 in [23]. Thus, neither the depth iD, nor its version with an essential infimum in (8), can be consistent in $L^2[0,1]$. Furthermore, the consistency difficulties of the infimal depths can be shown to persist also for smoothed versions of the functional data, as can be seen by obvious modifications of Example 5 in [48]. These results are true even though all the marginal distributions P_t are absolutely continuous, which is sufficient for the consistency of iD in $\mathcal{C}[0,1]$; see Theorem 5 in [23].

The next theorem deals with the sample version consistency of all the considered depth functionals (except from iD) in $L^2[0,1]$ in the case when the random functions are observed completely (i.e., when P_n is observable). It provides a unified extension of several consistency results scattered in the literature; see, e.g., [47].

Theorem 2. *Let D be one of the functional depths aD, hD or bD. For any $P \in \mathcal{P}\{L^2[0,1]\}$*

$$\sup_{x\in L^2[0,1]} |D(x; P_n) - D(x; P)| \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

If the distribution of $X \sim P$ is such that X is P-almost surely continuous over $[0,1]$, also the supremum norm (equivalent with the $L^\infty[0,1]$ norm), defined for all $x \in \mathcal{C}[0,1]$, by

$$\|x\|_\infty = \sup_{t\in[0,1]} |x(t)| \tag{9}$$

can be used in both the h-mode and the adjusted band depth to evaluate the depth of functions in $\mathcal{C}[0,1]$. A proof analogous to that of Theorem 2 can be used to show that in this case both hD and bD are universally consistent over $\mathcal{C}[0,1]$. In Section 6 we use both $L^2[0,1]$ and $\mathcal{C}[0,1]$ versions of these two depths to assess their performance when the sample functions are approximated.

While in Theorem 2 we were able to state the universal consistency results without any distributional assumptions, in what follows for aD it is necessary to make one additional assumption:

(**A6**) The marginal distributions P_t of $P \in \mathcal{P}\{L^2[0,1]\}$ are absolutely continuous for all $t \in [0,1]$.

With (A6) we are able to formulate the consistency result also for depths based on noisy observations. This theorem substantially improves on the related consistency results for discretely observed noiseless random functions in $\mathcal{C}[0,1]$ provided in [48].

Theorem 3. *Let $P \in \mathcal{P}\{L^2[0,1]\}$. Then*

$$\sup_{x\in L^2[0,1]} |D(x; \widetilde{P}_n) - D(x; P)| \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

if (A1)–(A5) are satisfied and D is hD or bD. Moreover, if also (A6) is true, then the result holds true also for $D = aD$.

To see that (A6) cannot be dropped for aD, consider $X \sim P \in \mathcal{P}\{L^2[0,1]\}$ with $\mathrm{P}(X \equiv 0) = 1$, $K(u) = \mathbf{1}(|u| \le 1)$, and P_T the uniform distribution on $[0,1]$. Then, using D_1^b, D_1^c or D_1^d for D_1 in (4) one gets for $x \equiv 0$ that $aD(x; P) = 1$. Note that the depth $D_1^a(\cdot; Q)$ is not well suited for $Q \in \mathcal{P}(\mathbb{R})$ whose distribution function is discontinuous, e.g., for Q supported in a

singleton $D_1^a(u; Q) = 0$ for all $u \in \mathbb{R}$. In contrast, for $\widetilde{P}_n$ based on kernel smoothed noisy random functions (with positive noise variance) we have by the usual (univariate) Central Limit Theorem that, for all $t \in [0, 1]$,

$$\lim_{i\to\infty} \mathsf{P}\{\widetilde{X}_i(t) < 0\} = \lim_{i\to\infty} \mathsf{P}\{\widetilde{X}_i(t) > 0\} = 1/2,$$

and thus for P-almost all $\omega \in \Omega$ we have $\limsup_{n\to\infty} aD\{x; \widetilde{P}_n(\omega)\} < 1 = aD(x; P)$.

4. Rates of convergence

In the present section we focus on the integrated depth aD based on the univariate depth $D_1 = D_1^a$ from (5). It constitutes the simplest representative of integrated depths for functions. Extensions of the present results for other univariate depth functions D_1 as considered in Section 3 are possible using analogous techniques. Initially, we state the uniform rate of convergence of the integrated depth over $L^2[0, 1]$. As far as we are aware, it is the first result of this kind for a depth constructed for functional data.

Theorem 4. *For any $P \in \mathcal{P}\{L^2[0, 1]\}$,*

$$\sup_{x\in L^2[0,1]} |aD(x; P_n) - aD(x; P)| = \mathcal{O}_P(n^{-1/2}).$$

4.1. Rate of convergence for smoothed random functions

A theorem concerning the rates of convergence of the reconstructed random curves towards the noiseless random functions will be stated below. For this, a refinement of the condition (A1) is necessary in what follows.

(**A1***) There exist constants $M, L > 0$ and $\beta \in (0, 1]$ such that

$$\sup_{t\in[0,1]} \mathrm{E}\, X(t)^2 \le M, \quad \text{and} \quad \mathsf{P}\{\forall_{s,t\in[0,1]}\ |X(s) - X(t)| \le L\,|s - t|^\beta\} = 1.$$

Theorem 5. *Let conditions (A1*), (A2)–(A5) be satisfied. Then there exists a constant $c > 0$ such that*

$$\mathrm{E}\, \|X_i - \widetilde{X}_i\|^2 \le c\frac{\sigma^2 + M}{m_i h_{m_i}} + L^2 h_{m_i}^{2\beta}.$$

As a corollary of Theorem 5 we obtain that for the optimal choice of bandwidth in our setup [26],

$$h_{m_i} = \mathcal{O}\{m_i^{-1/(2\beta+1)}\} \tag{10}$$

we have for some $C > 0$ and all integers $i \ge 1$,

$$\mathrm{E}\, \|X_i - \widetilde{X}_i\| \le \sqrt{\mathrm{E}\, \|X_i - \widetilde{X}_i\|^2} \le C m_i^{-\beta/(2\beta+1)}. \tag{11}$$

This is the optimal rate of convergence attainable in the nonparametric regression setting as derived by Stone [58] for $\beta \le 1$; see also the discussion in Kohler et al. [33]. In the sequel, for simplicity we restrict ourselves only to the case of the optimal choice of bandwidths (10). All the computations are straightforward to adapt for a general sequence $h_{m_1}, h_{m_2}, \ldots$

4.2. Rate of convergence of depth for noisy functions

Recall that for a random function $X \sim P \in \mathcal{P}\{L^2[0, 1]\}$ and $t \in [0, 1]$, F_t stands for the distribution function of the univariate random variable $X(t)$. Assume that (A1*) and (A6) are satisfied for P. Then we may proceed as in the proof of Theorem 4 in Gijbels and Nagy [23] and show that the collection of all marginal distribution functions $\{F_t : t \in [0, 1]\}$ of P is in fact a collection of uniformly equicontinuous functions, i.e.,

$$\lim_{\varepsilon\to 0+} \sup_{t\in[0,1]} \sup_{|s-s'|\le\varepsilon} |F_t(s) - F_t(s')| = 0.$$

Therefore, it is possible to find a deterministic, continuous and non-decreasing function $\delta_F : [0, \infty) \to [0, \infty)$ depending only on P such that $\delta_F(0) = 0$ and, for all $\varepsilon > 0$,

$$\sup_{t\in[0,1]} \sup_{|s-s'|\le\varepsilon} |F_t(s) - F_t(s')| \le \delta_F(\varepsilon). \tag{12}$$

The function δ_F is called a uniform modulus of continuity of the collection $\{F_t : t \in [0, 1]\}$.

In what follows, we need to manipulate with the Legendre transform of the function $g(\nu) = \delta_F(1/\nu)$ with $\nu > 0$. Recall that for g its Legendre transform g^* is a function defined by

$$g^*(\mu) = \sup_{\nu>0} \{\nu\mu - g(\nu)\} \tag{13}$$

for $\mu \in \mathbb{R}$ such that the right-hand side is finite. Now, we are ready to state the main result of the contribution — the uniform rate of convergence of the integrated depth for noisy random functions.

Theorem 6. *Let $X \sim P \in \mathcal{P}\{L^2[0,1]\}$ be such that (A1*), (A2)–(A6) are satisfied.*

(i) Then

$$\sup_{x\in L^2[0,1]} |aD(x;\widetilde{P}_n) - aD(x;P)| = \mathcal{O}_P\left[\max\left[-g^*\left\{-\frac{1}{n}\sum_{i=1}^n m_i^{-\beta/(2\beta+1)}\right\}, n^{-1/2}\right]\right].$$

(ii) If $\delta_F(\varepsilon) = K\varepsilon^\alpha$ for some $\alpha \in (0,1]$ and $K > 0$, then

$$\sup_{x\in L^2[0,1]} |aD(x;\widetilde{P}_n) - aD(x;P)| = \mathcal{O}_P\left[\left\{\frac{1}{n}\sum_{i=1}^n m_i^{-\beta/(2\beta+1)}\right\}^{\alpha/(1+\alpha)}\right].$$

(iii) If, furthermore, $m_n \sim n^r$ for some $r > 0$, then

$$\sup_{x\in L^2[0,1]} |aD(x;\widetilde{P}_n) - aD(x;P)| = \begin{cases} \mathcal{O}_P[n^{-r\alpha\beta/\{(1+\alpha)(2\beta+1)\}}] & \text{if } r < (2\beta+1)/\beta, \\ \mathcal{O}_P[\{\ln(n)/n\}^{\alpha/(1+\alpha)}] & \text{if } r = (2\beta+1)/\beta, \\ \mathcal{O}_P\{n^{-\alpha/(1+\alpha)}\} & \text{if } r > (2\beta+1)/\beta. \end{cases}$$

The rate in part (iii) of Theorem 6 is the optimal rate $\mathcal{O}_P(n^{-1/2})$ from Theorem 4 if $\alpha = 1$ and $r > (2\beta+1)/\beta$, in all other cases it is slower. For Lipschitz continuous random functions ($\beta = 1$) this corresponds to $r > 3$. So, if m_n is much larger than n, then the contamination does not degrade the rate of convergence; in other words, the high-dimensional setting (i.e., $m_n \gg n$) is a blessing in the contaminated functional data problem. In part (iii) of Theorem 6 we consider only polynomial rate of convergence for m_n; analogous results can be easily obtained under different assumptions. In that case, it is possible to see that if m_n converges fast and the distribution of P is smooth enough ($\alpha = 1$), the oracle rate of convergence from Theorem 4 is attainable also for the reconstructed curves based on observations contaminated by noise.

5. General approximation results

In Sections 2–4, our attention was focused on the analysis using the standard Nadaraya–Watson kernel smoother (3), under the assumptions of independence of the error terms $\varepsilon_{i,1}, \ldots, \varepsilon_{i,m_i}$, and the random design of the observation times $T_{i,j}$ in (2). However, our main results are not restricted to this setting. Indeed, it is not difficult to generalize our findings to many other scenarios, such as those concerning (i) other nonparametric smoothing methods, and (ii) approximation setups with possibly fixed observation points, or correlated errors terms.

Similarly as in Section 2, let $\widetilde{X}_i$ be an approximant of the unobserved functional datum X_i based on some information available about X_i. Denote by $\widetilde{P}_n \in \mathcal{P}\{L^2[0,1]\}$ the probability measure concentrated in the smoothed functions $\widetilde{X}_1, \ldots, \widetilde{X}_n$, corresponding to a random sample $X_1, \ldots, X_n$. The following general version of Theorems 1, 3 and 6 can be devised.

Theorem 7. *Let $X_1, X_2, \ldots$ be independent random functions distributed as $X \sim P \in \mathcal{P}\{L^2[0,1]\}$. Suppose that, as $n \to \infty$,*

$$a_n = \mathrm{E}\,\|X_n - \widetilde{X}_n\| = o(1). \tag{14}$$

Then $\mathsf{P}(\widetilde{P}_n \xrightarrow{w} P) = 1$, and as $n \to \infty$,

$$\sup_{x\in L^2[0,1]} |D(x;\widetilde{P}_n) - D(x;P)| \xrightarrow{\text{a.s.}} 0$$

for D either hD or bD. If, moreover, the random function X is P-almost surely continuous over $[0,1]$ and (A6) is true, then

$$\sup_{x\in L^2[0,1]} |aD(x;\widetilde{P}_n) - aD(x;P)| = \mathcal{O}_P\left[\max\left\{-g^*\left(-\frac{1}{n}\sum_{i=1}^n a_i\right), n^{-1/2}\right\}\right].$$

Finally, if also $\delta_F(\varepsilon) = K\varepsilon^\alpha$ for some $\alpha \in (0,1]$ and $K > 0$, then

$$\sup_{x\in L^2[0,1]} |aD(x;\widetilde{P}_n) - aD(x;P)| = \mathcal{O}_P\left\{\left(\frac{1}{n}\sum_{i=1}^n a_i\right)^{\alpha/(1+\alpha)}\right\}.$$

The proof of Theorem 7 is omitted, as it follows closely the derivations in the proofs of Theorems 1, 3 and 6 in the Appendix.

Under sets of assumptions similar to our conditions (A1)–(A6) for the Nadaraya–Watson kernel smoother, condition (14) can be verified for many other nonparametric smoothing methods. For instance, for the local polynomial fitting procedure and independent errors in model (2), formula (14) can be obtained from the theory provided in [19]. Likewise, if the points $T_{i,j}$ in (2) are not chosen randomly, well-established results in fixed-design nonparametric regression problems provide conditions under which (14) can be verified. If additional smoothness assumptions, such as the existence and boundedness of a higher-order derivative of the random function X can be guaranteed P-almost surely, rates of convergence faster than those in Theorem 5 can be found using Theorem 7.

Finally, some extensions of our theory may be obtained also in the situation when the error terms $\varepsilon_{i,1}, \ldots, \varepsilon_{i,m_i}$ are allowed to be dependent in (2). Then, one deals with the problem of nonparametric regression with correlated errors. Under suitable assumptions on the correlation structure of the random errors, also in that case condition (14) can be recovered. For a general account of results of this type we refer to [36,51], or a recent paper of De Brabanter et al. [14].

6. Simulation study

We complete this paper by illustrating the performance of the proposed method on a small simulation study. It constitutes part of a larger study, including various choices of error variances, and an infinite-dimensional model for the random functions. The complete results of the study, including all its source codes, can be found in the Online Supplement. The study is performed in R 3.4.2 [54] using efficient `Fortran` implementations for kernel smoothing and functional depth computation, developed by the authors. Most of these functions are freely available as part of R package `ddalpha` [53].

Consider the model for densely observed functional data as described in Section 2, in combination with the functional depths introduced in Section 3. Our main aim is to compare possibilities for evaluating depth of discretely observed functional data, and to identify scenarios under which the depth based on complete realizations of functions is well approximated by the depth based on the reconstructions of these. The random variable $X \sim P \in \mathcal{P}\{L^2[0,1]\}$ is given, for all $t \in [0,1]$, by

$$X(t) = (5t)^2(3/4 - t) + A\cos(B\pi t)/2 + C\sin(D\pi t/3),$$

where A, B, C and D are independent standard Gaussian random variables. Consider a random sample $X_1, \ldots, X_n$ from P of size $n = 50$. For the sake of computer processing, in R, the complete function X_i is represented by a vector of size $d = 1001$ of its functional values taken at d equidistant points in $[0,1]$.

In the analysis, X_i is assumed to be observed discretely at $m_i = \lceil 50 + i/5 \rceil$ independent random observation points as discussed in Section 2. These are distributed according to

(**T1**) the uniform distribution over $[0,1]$;
(**T2**) the beta distribution with parameters 2 and 2; and
(**T3**) the beta distribution with parameters 0.7 and 0.9.

In setup (T1) the observation points are spread almost regularly over the domain. This makes the reconstruction of X_i rather simple, at least for larger values of m_i. In (T2) the majority of the observed points concentrates in the centre of the domain, while in (T3) mostly points near the boundaries of the interval are given.

For the noise variance we consider three different scenarios

(**V1**) $\sigma_1^2(t) = \tau^2$;
(**V2**) $\sigma_2^2(t) = \tau^2(1 + t^2)^2$; and
(**V3**) $\sigma_3^2(t) = \tau^2\,\mathbf{1}\{t \in [\min(U_1, U_2), \max(U_1, U_2)]\}$, where U_1 and U_2 are two independent uniformly distributed random variables in $[0,1]$, different for each X_i.

The constant τ is in this section always set to $\tau = 1/2$. In the extended study we consider also the case $\tau = 1/4$. Scenarios (V1) and (V2) correspond to typical homoscedastic and heteroscedastic noise, respectively. Scenario (V3) represents the situation when the noise level varies with the sampling process, and each function X_i is contaminated in a different subset of its domain. Note that in scenario (V3) the variance of the noise term $\sigma_3^2(t)$ is random, and therefore it technically does not satisfy condition (A4). However, it is easy to see that for a random quantity $\sigma^2(t)$ a condition like $\mathrm{E}\sup_{t\in[0,1]}\sigma^2(t) \le \sigma^2 < \infty$ is enough for all the presented results to hold true. In the short simulation study presented here we consider only the situation when the error terms $\varepsilon_{i,j}$ in (2) are independent; in the Online Supplement we provide also some results for the case of correlated errors.

In each of the scenarios above, the sample curves are reconstructed using five methods:

(**S1**) constant only approximation given, for all $t \in [0,1]$, by

$$\widetilde{X}_i(t) = \frac{1}{m_i}\sum_{j=1}^{m_i} X_i^*(T_{i,j})$$

denoting $X_i^*(T_{i,j}) = X_{i,j}^*$;

Table 1
Means and standard deviations (in brackets) of ρ in the simulation study, scenario (T1).

		(S1)	(S2)	(S3)	(S4)	(S5)
(V1)	aD	0.602 (0.094)	0.890 (0.026)	0.932 (0.018)	0.940 (0.017)	0.936 (0.020)
	hD_1	0.585 (0.108)	0.898 (0.029)	0.922 (0.026)	0.933 (0.034)	0.937 (0.030)
	hD_2	0.522 (0.113)	0.788 (0.065)	0.761 (0.074)	0.748 (0.081)	0.783 (0.068)
	bD_1	0.758 (0.110)	0.960 (0.020)	0.966 (0.018)	0.960 (0.056)	0.965 (0.048)
	bD_2	0.630 (0.104)	0.850 (0.054)	0.825 (0.061)	0.796 (0.074)	0.829 (0.063)
(V2)	aD	0.595 (0.104)	0.836 (0.044)	0.903 (0.027)	0.910 (0.026)	0.902 (0.029)
	hD_1	0.599 (0.107)	0.832 (0.044)	0.870 (0.041)	0.880 (0.042)	0.889 (0.036)
	hD_2	0.539 (0.113)	0.615 (0.107)	0.640 (0.097)	0.617 (0.103)	0.668 (0.096)
	bD_1	0.766 (0.100)	0.916 (0.037)	0.935 (0.032)	0.932 (0.045)	0.941 (0.028)
	bD_2	0.640 (0.105)	0.686 (0.094)	0.718 (0.082)	0.692 (0.091)	0.741 (0.083)
(V3)	aD	0.606 (0.103)	0.963 (0.012)	0.971 (0.009)	0.977 (0.008)	0.974 (0.009)
	hD_1	0.604 (0.104)	0.934 (0.024)	0.954 (0.021)	0.972 (0.016)	0.979 (0.008)
	hD_2	0.537 (0.112)	0.707 (0.072)	0.766 (0.083)	0.845 (0.077)	0.912 (0.044)
	bD_1	0.768 (0.088)	0.977 (0.011)	0.983 (0.009)	0.987 (0.015)	0.989 (0.014)
	bD_2	0.637 (0.094)	0.703 (0.063)	0.829 (0.065)	0.866 (0.078)	0.932 (0.039)

(**S2**) piecewise linear interpolation, i.e., for o$t \in [T_{i,(j)}, T_{i,(j+1)})$,

$$\widetilde{X}_i(t) = X_i^*(T_{i,(j)}) + \frac{t - T_{i,(j)}}{T_{i,(j+1)} - T_{i,(j)}}\{X_i^*(T_{i,(j+1)}) - X_i^*(T_{i,(j)})\}$$

where $0 = T_{i,(0)}, 1 = T_{i,(m_i+1)}, T_{i,(1)}, \ldots, T_{i,(m_i)}$ is the ordered sample of the observation points of X_i, $X_i^*(T_{i,(0)}) = X_i^*(T_{i,(1)})$ and $X_i^*(T_{i,(m_i+1)}) = X_i^*(T_{i,(m_i)})$;

(**S3**) kernel smoothing described in detail in Section 2, with the Gaussian kernel

$$K(u) = (2\pi)^{-1/2}\exp(-u^2/2); \tag{15}$$

(**S4**) local linear smoothing [19], with the Gaussian kernel (15); and

(**S5**) local linear smoothing in the presence of correlated errors, described in De Brabanter et al. [14, Section 5], based on a consecutive application of a bimodal kernel function $\overline{K}(u) = 2u^2\exp(-u^2)/\sqrt{\pi}$ and a Gaussian kernel $K(u)$ from (15).

Method (S1) is used only as a benchmark — the results for the remaining methods should be compared taking into account the performance of this most naive reconstruction method. Method (S5) is an adaptation of the local linear smoother (S4) that is capable of dealing with correlated errors. Here, it is used mainly for comparison with (S4); for some simulations where method (S5) bears great relevance see the Online Supplement.

In (S3)–(S5) the bandwidths are taken to be different for each function X_i. In (S3) and (S4) they are selected using the standard leave-one-out cross-validation. For (S5) the complete description of the bandwidth selection procedure can be found in [14]. All these bandwidth selection methods typically provide sufficiently accurate results, in combination with an acceptable computational cost. A fast version of the kernel smoothing method (S3) is taken from the `R` package `ddalpha` [53]; the fast local polynomial smoothers (S4) and (S5) implemented in `R` package `locpol` [6] are considered. For a single random function from models (T1)–(T3), (V1)–(V3), and its piecewise linear interpolation (S2) based on the corresponding contaminated points see Fig. 1.

Five depth functionals are considered in the simulations:

aD: the integrated depth using the univariate depth D_1^q from (5);
hD_1: the h-mode depth with the $L^2([0,1])$ norm (1), the Gaussian kernel K_D as in (15) and $h(\widetilde{P}_n)$ chosen as the 20th percentile of the distribution of $\|\widetilde{X}_i - \widetilde{X}_j\|$, $i \neq j$ in accordance with the choice of Cuevas et al. [12];
hD_2: the h-mode depth with the supremum norm (9), kernel (15), and bandwidth as for hD_1;
bD_1: the adjusted band depth with the L^2 ($[0,1]$) norm (1), $J = 2$ and $K_D(u) = \exp(-u)$;
bD_2: the adjusted band depth with the supremum norm (9), $J = 2$ and $K_D(u) = \exp(-u)$.

In the comparison, these depths are computed with respect to the empirical measure P_n based on the original (noiseless) curves $X_1, \ldots, X_n$, and then also with respect to the empirical measure $\widetilde{P}_n$ based on the reconstructed curves $\widetilde{X}_1, \ldots, \widetilde{X}_n$. A reasonable reconstruction method should provide a good approximation of the depth values $D(X_i; P_n)$ for each i.

To measure the approximation error, we use Pearson's correlation coefficient ρ between $(D(X_1, P_n), \ldots, D(X_n, P_n))$ and $(D(\widetilde{X}_1, \widetilde{P}_n), \ldots, D(\widetilde{X}_n, \widetilde{P}_n))$. Other association/loss measures (such as Spearman's correlation, or MSE) lead to the same conclusions; see the extended simulation study in the Online Supplement.

In Tables 1–3 we see the results of the simulation study for scenarios (T1)–(T3), respectively. The reported values are the mean and the standard deviation (in brackets) of ρ, computed from 100 independent runs. In each setup we clearly see that the naive constant interpolation (S1) provides rather faltering results. While linear interpolation (S2) improves on

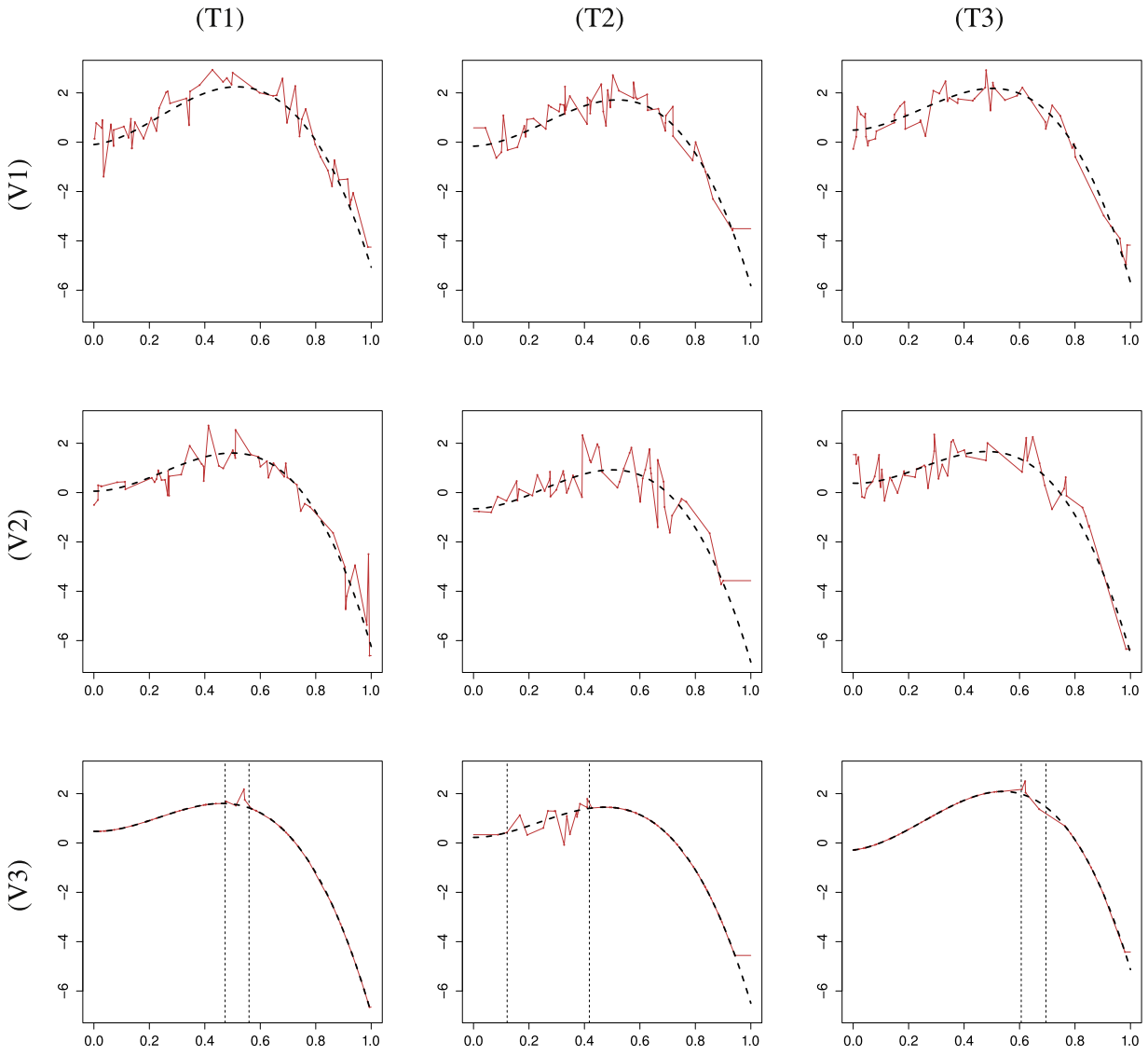

Fig. 1. For each of the scenarios (T1)–(T3), (V1)–(V3), a single random function X_i (thick dashed), and its piecewise linear reconstruction (S2) based on the corresponding $m_i = 50$ discrete contaminated observations (thin solid) are displayed. For (V3), the vertical dashed lines represent the realized values of the variables U_1 and U_2 controlling the noise level.

these, kernel and local polynomial estimators (S3), (S4) and (S5) certainly outperform their competitors in most considered scenarios, for all depth functions. Notable exceptions are scenarios (T2), (V1) and (V2), where especially the local polynomial estimators (S4) and (S5) perform rather poorly for the depths based on the supremum norm (9). This is explained by the fact that the local polynomial estimators are known to break down easily in the extrapolation task. If no points are observed near the boundary of $[0, 1]$, the extrapolants (S4) and (S5) lose their stability, which affects the supremum distance between the curves profoundly.

In most of the scenarios, however, the reconstruction methods (S3)–(S5) perform well. Moreover, the variability of the results typically decreases for (S3)–(S5). This is confirmed also by Fig. 2, and by additional simulations given in the Online Supplement.

To illustrate the "dimensionality blessing" effect mentioned in Section 4.2, we perform another, short simulation study. Here, under the settings as above with (T1) and (V1) we compare three choices for the number of time points

(**M1**) $m_i = \lceil 20 + i/5 \rceil$ for $i \in \{1, \ldots, n\}$;
(**M2**) $m_i = \lceil 50 + i/5 \rceil$ for $i \in \{1, \ldots, n\}$; and
(**M3**) $m_i = \lceil 100 + i/5 \rceil$ for $i \in \{1, \ldots, n\}$.

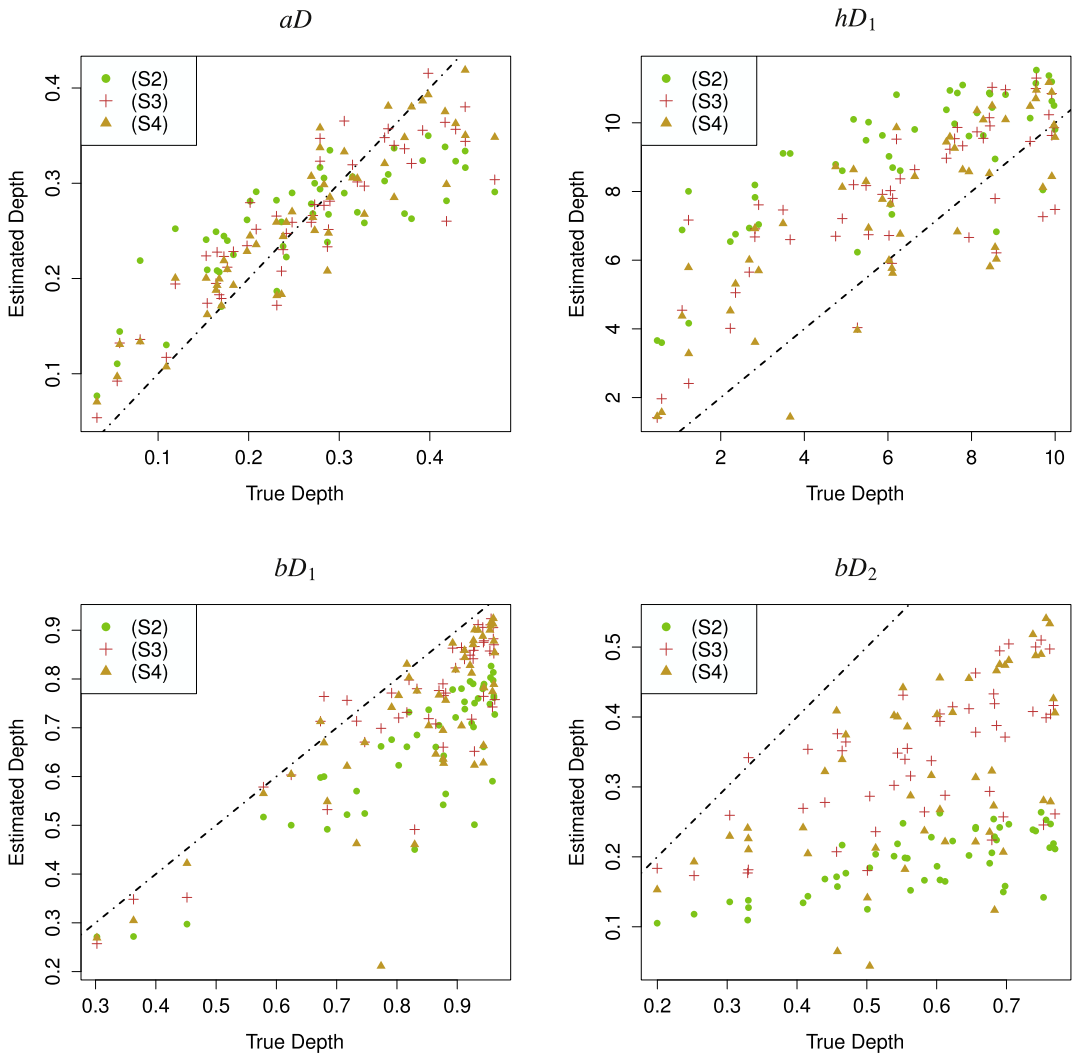

Fig. 2. Plots of the true depth values $D(X_i; P_n)$ against some of the estimated depths $D(\widetilde{X}_i; \widetilde{P}_n)$ for one run of the simulation study, scenario (T2), (V2). The dashed–dotted line represents the axis of the first quadrant in $\mathbb{R}^2$. For curves reconstructed by (S2) the depth hD_1 tends to be biased upwards, while the bD depths are underestimated; for (S3) and (S4), no systematic bias is observed for depths based on the $L^2[0, 1]$ norm.

Table 2
Means and standard deviations (in brackets) of ρ in the simulation study, scenario (T2).

		(S1)	(S2)	(S3)	(S4)	(S5)
(V1)	aD	0.717 (0.074)	0.863 (0.036)	0.916 (0.024)	0.924 (0.021)	0.920 (0.022)
	hD_1	0.700 (0.082)	0.809 (0.060)	0.829 (0.054)	0.768 (0.092)	0.807 (0.074)
	hD_2	0.623 (0.094)	0.655 (0.093)	0.580 (0.107)	0.400 (0.138)	0.458 (0.130)
	bD_1	0.852 (0.081)	0.881 (0.058)	0.891 (0.051)	0.791 (0.116)	0.849 (0.085)
	bD_2	0.724 (0.079)	0.778 (0.067)	0.691 (0.084)	0.508 (0.118)	0.557 (0.109)
(V2)	aD	0.708 (0.083)	0.804 (0.054)	0.886 (0.031)	0.888 (0.033)	0.884 (0.035)
	hD_1	0.693 (0.088)	0.741 (0.084)	0.777 (0.074)	0.689 (0.109)	0.732 (0.097)
	hD_2	0.617 (0.100)	0.566 (0.119)	0.507 (0.119)	0.347 (0.128)	0.400 (0.132)
	bD_1	0.842 (0.071)	0.825 (0.072)	0.846 (0.070)	0.717 (0.126)	0.770 (0.116)
	bD_2	0.711 (0.084)	0.670 (0.096)	0.618 (0.099)	0.464 (0.114)	0.508 (0.113)
(V3)	aD	0.727 (0.075)	0.946 (0.020)	0.960 (0.013)	0.971 (0.010)	0.968 (0.012)
	hD_1	0.711 (0.078)	0.857 (0.050)	0.869 (0.049)	0.890 (0.063)	0.913 (0.056)
	hD_2	0.638 (0.092)	0.604 (0.098)	0.597 (0.109)	0.644 (0.122)	0.713 (0.112)
	bD_1	0.853 (0.055)	0.916 (0.048)	0.922 (0.047)	0.888 (0.107)	0.920 (0.084)
	bD_2	0.725 (0.073)	0.644 (0.090)	0.710 (0.087)	0.676 (0.114)	0.740 (0.108)

Table 3
Means and standard deviations (in brackets) of ρ in the simulation study, scenario (T3).

		(S1)	(S2)	(S3)	(S4)	(S5)
(V1)	aD	0.579 (0.087)	0.885 (0.030)	0.922 (0.024)	0.933 (0.022)	0.927 (0.025)
	hD_1	0.569 (0.109)	0.894 (0.032)	0.912 (0.031)	0.934 (0.031)	0.937 (0.025)
	hD_2	0.507 (0.118)	0.781 (0.068)	0.725 (0.078)	0.786 (0.081)	0.818 (0.069)
	bD_1	0.734 (0.104)	0.957 (0.025)	0.959 (0.025)	0.967 (0.033)	0.968 (0.024)
	bD_2	0.609 (0.107)	0.846 (0.056)	0.796 (0.064)	0.831 (0.077)	0.861 (0.063)
(V2)	aD	0.573 (0.099)	0.833 (0.051)	0.892 (0.039)	0.901 (0.034)	0.889 (0.039)
	hD_1	0.553 (0.111)	0.823 (0.065)	0.850 (0.059)	0.862 (0.054)	0.869 (0.051)
	hD_2	0.487 (0.118)	0.596 (0.123)	0.595 (0.111)	0.594 (0.112)	0.655 (0.098)
	bD_1	0.717 (0.117)	0.903 (0.054)	0.914 (0.051)	0.912 (0.062)	0.919 (0.051)
	bD_2	0.585 (0.110)	0.686 (0.097)	0.686 (0.095)	0.676 (0.100)	0.731 (0.090)
(V3)	aD	0.576 (0.114)	0.959 (0.014)	0.967 (0.013)	0.974 (0.010)	0.970 (0.011)
	hD_1	0.564 (0.128)	0.925 (0.030)	0.941 (0.027)	0.959 (0.035)	0.972 (0.019)
	hD_2	0.500 (0.135)	0.683 (0.093)	0.726 (0.079)	0.829 (0.089)	0.909 (0.054)
	bD_1	0.729 (0.120)	0.970 (0.024)	0.975 (0.021)	0.970 (0.048)	0.983 (0.026)
	bD_2	0.600 (0.121)	0.683 (0.084)	0.796 (0.066)	0.850 (0.092)	0.927 (0.057)

Table 4
Means and standard deviations (in brackets) of ρ in the simulation study, scenarios (M1) and (M3). Corresponding results in scenario (M2) can be found in Table 1, part (V1).

		(S1)	(S2)	(S3)	(S4)	(S5)
(M1)	aD	0.466 (0.120)	0.832 (0.051)	0.850 (0.046)	0.871 (0.042)	0.858 (0.049)
	hD_1	0.469 (0.124)	0.789 (0.079)	0.768 (0.083)	0.799 (0.080)	0.813 (0.074)
	hD_2	0.420 (0.130)	0.588 (0.122)	0.526 (0.124)	0.538 (0.126)	0.590 (0.109)
	bD_1	0.624 (0.144)	0.853 (0.085)	0.839 (0.085)	0.827 (0.115)	0.859 (0.098)
	bD_2	0.515 (0.132)	0.701 (0.096)	0.622 (0.108)	0.606 (0.115)	0.654 (0.104)
(M3)	aD	0.683 (0.085)	0.913 (0.026)	0.961 (0.012)	0.966 (0.011)	0.962 (0.013)
	hD_1	0.661 (0.090)	0.929 (0.022)	0.967 (0.013)	0.974 (0.012)	0.972 (0.011)
	hD_2	0.587 (0.100)	0.854 (0.043)	0.867 (0.050)	0.863 (0.063)	0.876 (0.054)
	bD_1	0.829 (0.071)	0.981 (0.009)	0.987 (0.008)	0.988 (0.008)	0.987 (0.006)
	bD_2	0.691 (0.083)	0.902 (0.036)	0.909 (0.041)	0.898 (0.055)	0.911 (0.046)

Scenario (M2) was considered previously. As confirmed in Table 4 (and the corresponding part of Table 1 for scenario (M2)), the more high-dimensional the discrete observations are, the more accurately the functional data depth can be observed.

Altogether, we may conclude that in accordance with the theoretical developments obtained in Sections 2–4, pre-smoothing of discretely observed noisy random functions is a reasonable method of preprocessing in order to obtain information about functional depth. Moreover, depths defined using the $L^2[0,1]$ norm (1) are generally more computationally stable than the corresponding depths defined via the supremum norm (9).

Acknowledgments

The authors thank the Managing Editor, the Editor-in-Chief and two referees for their valuable comments. This research was carried out during the first author's visit at the Université de Toulouse, supported by the Research Foundation-Flanders. The work is supported by the IAP research network no. P7/06 of the Federal Science Policy (Belgium). The first author gratefully acknowledges support from the Research Foundation-Flanders, the grant 18-00522Y of the Czech Science Foundation, and the PRIMUS/17/SCI/3 project of Charles University.

Appendix. Proofs of the theoretical results

A.1. Proof of Theorem 1

Without loss of generality, assume that in (A2) $B \leq 1$, otherwise continue with the kernel K/B. Using the independence of the random function X, observation points T and the noise ε we can write

$$c_T \, \mathrm{E} \, \|\widetilde{X}_n - X_n\|^2 = \mathrm{E} \int_0^1 \{\widetilde{X}_n(t) - X_n(t)\}^2 c_T \, \mathrm{d}\, \lambda(t) \leq \mathrm{E} \int_0^1 \{\widetilde{X}_n(t) - X_n(t)\}^2 \, \mathrm{d}\, P_T(t). \tag{A.1}$$

Let us first show that the last term above vanishes with $n \to \infty$.

For a fixed $x \in L^2[0,1]$ observed within the design $x_j^* = x(T_j) + \varepsilon_j$ with $j \in \{1,\dots,m\}$, we can write for $w_j(t) = K\{(t-T_j)/h_m\}/\sum_{k=1}^m K\{(t-T_k)/h_m\}$ the following

$$\begin{aligned}
\mathrm{E}_{\varepsilon,T}\int_0^1 \{\widetilde{x}(t)-x(t)\}^2\,\mathrm{d}P_T(t) &\le 2\,\mathrm{E}_{\varepsilon,T}\int_0^1 \widetilde{x}(t)^2\,\mathrm{d}P_T(t) + 2\,\mathrm{E}_{\varepsilon,T}\int_0^1 x(t)^2\,\mathrm{d}P_T(t)\\
&= 2\,\mathrm{E}_{\varepsilon,T}\int_0^1 \left\{\sum_{j=1}^m w_j(t)x_j^*\right\}^2\mathrm{d}P_T(t) + 2\,\mathrm{E}_T\,x(T)^2\\
&\le 2\,\mathrm{E}_{\varepsilon,T}\int_0^1 \sum_{j=1}^m w_j(t)\left(x_j^*\right)^2\mathrm{d}P_T(t) + 2\,\mathrm{E}_T\,x(T)^2\\
&= 2\int_0^1 \sum_{j=1}^m \mathrm{E}_{\varepsilon,T}\,w_j(t)\left\{x(T_j)+\varepsilon_j\right\}^2\mathrm{d}P_T(t) + 2\,\mathrm{E}_T\,x(T)^2\\
&\le 4\int_0^1 \sum_{j=1}^m \mathrm{E}_{\varepsilon,T}\,w_j(t)\left\{x(T_j)^2+\varepsilon_j^2\right\}\mathrm{d}P_T(t) + 2\,\mathrm{E}_T\,x(T)^2\\
&\le 4\left\{\sigma^2 + \int_0^1 \sum_{j=1}^m \mathrm{E}_T\,w_j(t)x(T_j)^2\,\mathrm{d}P_T(t)\right\} + 2\,\mathrm{E}_T\,x(T)^2.
\end{aligned}$$

Now, in the first expression on the right-hand side, we have the kernel estimate for the function x^2 based on the observations $x^2(T_j)$ for $j \in \{1,\dots,m\}$ without noise, i.e., $\sigma^2(t) = 0$ for all $t \in [0,1]$. For this term, it is known (see, e.g., the proof of Theorem 5.1 in [26]) that for some $C > 0$ for any $x \in L^2[0,1]$ such that

$$\mathrm{E}_T\{x(T)^2\} < \infty \tag{A.2}$$

and for any integer $m \ge 1$, we can bound

$$\int_0^1 \sum_{j=1}^m \mathrm{E}_T\,w_j(t)x(T_j)^2\,\mathrm{d}P_T(t) \le C\,\mathrm{E}_T\,x(T)^2,$$

meaning that we can continue the chain of inequalities above by writing

$$\mathrm{E}_{\varepsilon,T}\int_0^1 \{\widetilde{x}(t)-x(t)\}^2\,\mathrm{d}P_T(t) \le 4\sigma^2 + (4C+2)\,\mathrm{E}_T\,x(T)^2. \tag{A.3}$$

Since (A.3) is valid for any $x \in L^2[0,1]$ such that (A.2) is true, we may as well replace the fixed function x by the random function $X \sim P$, apply expectation with respect to X to both sides of (A.3), and get

$$\mathrm{E}\int_0^1 \{\widetilde{X}(t)-X(t)\}^2\,\mathrm{d}P_T(t) \le 4\sigma^2 + (4C+2)\,\mathrm{E}\,X(T)^2, \tag{A.4}$$

where the right-hand side is finite by (A1) and (A4). Note that (A.2) is necessarily true for P-almost all realizations of X by (A1).

By Theorem 1 in [57], using (A2)–(A5) for any fixed $x_n \in L^2[0,1]$ (here, the subscript n stands for the fact that x is assumed to be observed at $\left(T_{n,1},\dots,T_{n,m_n}\right)$ with the additive noise $\left(\varepsilon_{n,1},\dots,\varepsilon_{n,m_n}\right)$), we find that, as $n \to \infty$,

$$\mathrm{E}_{\varepsilon,T}\int_0^1 \{\widetilde{x}_n(t)-x_n(t)\}^2\,\mathrm{d}P_T(t) \longrightarrow 0;$$

see also Theorems 4.1 and 5.1 in [26]. To allow for the replacement of x by the random X as above, it is enough to apply the expectation with respect to X to this formula. Lebesgue's Dominated Convergence Theorem, as stated in Theorem 4.3.5 of [16], then asserts that, as $n \to \infty$,

$$\mathrm{E}\int_0^1 \{\widetilde{X}_n(t)-X_n(t)\}^2\,\mathrm{d}P_T(t) \longrightarrow 0.$$

The boundedness assumption of the integrand function for the Dominated Convergence Theorem was verified in (A.4).

The rest of the proof is straightforward. By (A.1) and the previous results

$$\lim_{n\to\infty} \mathrm{E}\,\|\widetilde{X}_n - X_n\|^2 = 0.$$

Consequently, by Chebyshev's inequality, §8.3.1 in [16], for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathsf{P}(\|\widetilde{X}_n - X_n\| > \varepsilon) = 0,$$

and by Proposition 9.3.5 in [16], we can conclude that $\widetilde{X}_n$ converges in law to X. The assertion of Theorem 1 then follows by the same argument as Theorem 1, Part (ii) in [48].

A.2. Proof of Theorem 2

The proof is based on the following lemma establishing the uniform qualitative robustness of all the concerned depths over the space $L^2[0,1]$.

Lemma 1. *Let $P_\nu \xrightarrow[\nu\to\infty]{w} P$ in $\mathcal{P}\{L^2[0,1]\}$. Then, as $\nu \to \infty$,*

$$\sup_{x\in L^2[0,1]} |D(x;P_\nu) - D(x;P)| \to 0,$$

if D is hD or bD. Moreover, if also (A6) is true for P, then the result holds true also for $D = aD$.

Proof. As shown in the proof of Theorem 1 in [45],

$$\left\{x_1 \mapsto \frac{1}{h(P)} K_D\left\{\frac{\|x-x_1\|}{h(P)}\right\} : x \in L^2[0,1]\right\}$$

is a set of uniformly equicontinuous, and uniformly bounded functionals $L^2[0,1] \to \mathbb{R}$. Therefore, for hD, the result follows from Corollary 11.3.4 in [16].

For bD, we show a similar continuity result for the set

$$\{(x_1,\dots,x_J) \mapsto K_D\left[d\left\{x, B\left(x_1,\dots,x_J\right)\right\}\right] : x \in L^2[0,1]\}. \tag{A.5}$$

First of all, notice that K_D is uniformly continuous and thus its minimal modulus of continuity

$$\delta_{K_D}(\alpha) = \sup_{|s-t|\le\alpha} |K_D(s) - K_D(t)|$$

vanishes as $\alpha \to 0$. From the definition of δ_{K_D} we have

$$\left|K_D\left[d\left\{x, B\left(x_1,\dots,x_J\right)\right\}\right] - K_D\left[d\left\{x, B\left(y_1,\dots,y_J\right)\right\}\right]\right| \le \delta_{K_D}\left[d\left\{x, B\left(x_1,\dots,x_J\right)\right\} - d\left\{x, B\left(y_1,\dots,y_J\right)\right\}\right]. \tag{A.6}$$

Assume now that for $\varepsilon > 0$ it is true that

$$\forall_{j\in\{1,\dots,J\}} \quad \|x_j - y_j\| \le \varepsilon. \tag{A.7}$$

Denote by $m \in B(x_1,\dots,x_J)$ the function for which $d\left\{x, B(x_1,\dots,x_J)\right\} = \|x-m\|$. Such a function exists because of the construction of the band of functions. In what follows we show that if (A.7) is true, then

$$d\left\{m, B\left(y_1,\dots,y_J\right)\right\} \le 4J\varepsilon, \tag{A.8}$$

meaning that

$$d\left\{x, B\left(y_1,\dots,y_J\right)\right\} \le \|x-m\| + d\left\{m, B\left(y_1,\dots,y_J\right)\right\} \le d\left\{x, B\left(x_1,\dots,x_J\right)\right\} + 4J\varepsilon.$$

Because in the beginning the role of $B\left(x_1,\dots,x_J\right)$ and $B\left(y_1,\dots,y_J\right)$ is symmetric, the last expression implies that if (A.7) is in order, then the right-hand side in (A.6) can be bounded from above by $\delta_{K_D}(4J\varepsilon)$. Because the modulus $\delta_{K_D}(\alpha)$ vanishes as $\alpha \to 0$, the uniform continuity of (A.5) follows.

To show (A.8), consider first $\ell_x \in L^2[0,1]$ defined by $\ell_x(t) = \min_{j=1,\dots,J} x_j(t)$ for $t \in [0,1]$, and $\ell_y \in L^2[0,1]$ given analogously for functions y_j. Partition $[0,1]$ into two disjoint subsets I_x and I_y given by

$$I_x = \{t \in [0,1]: \ell_x(t) \le \ell_y(t)\}, \quad I_y = \{t \in [0,1]: \ell_y(t) < \ell_x(t)\}.$$

Then, partition the set I_x further into $I_{x_j} = \{t \in [0,1]: x_j(t) = \ell_x(t)\}$ and the set I_y accordingly. Assume that the sets $I_{x_1},\dots,I_{x_J}, I_{y_1},\dots,I_{y_J}$ do not overlap; if they do, modify them by assigning the points in the intersections to one of the intersecting sets so that the new partition consists of pairwise disjoint measurable sets whose union is $[0,1]$. This is always possible, since we deal with a finite number of measurable functions.

Writing $\mathbf{1}(I_x)$ for $\mathbf{1}(t \in I_x)$ etc., we can use the construction of the partition above and (A.7) and bound

$$\|(\ell_x - \ell_y)\mathbf{1}(I_{x_j})\| = \|(x_j - \ell_y)\mathbf{1}(I_{x_j})\| \le \|(x_j - y_j)\mathbf{1}(I_{x_j})\| \le \|x_j - y_j\| \le \varepsilon,$$

leading to

$$\|\ell_x - \ell_y\| = \|(\ell_x - \ell_y)\mathbf{1}(I_x)\| + \|(\ell_x - \ell_y)\mathbf{1}(I_y)\| = \sum_{j=1}^{J} \|(\ell_x - \ell_y)\mathbf{1}(I_{x_j})\| + \sum_{j=1}^{J} \|(\ell_x - \ell_y)\mathbf{1}(I_{y_j})\| \le 2J\varepsilon.$$

Analogously, one can prove that also for functions $u_x(t) = \max\{x_1(t), \dots, x_J(t)\}$ and $u_y(t) = \max\{y_1(t), \dots, y_J(t)\}$, one has $\|u_x - u_y\| \le 2J\varepsilon$.

Now, consider $m \in B(x_1, \dots, x_J)$ and similarly as before, define three disjoint sets partitioning $[0, 1]$, viz.

$$I_1 = \{t \in [0, 1]: m(t) < \ell_y(t)\}, \quad I_2 = \{t \in [0, 1]: \ell_y(t) \le m(t) \le u_y(t)\}, \quad I_3 = \{t \in [0, 1]: u_y(t) < m(t)\}.$$

With this notation, one can write, for $m \in B(x_1, \dots, x_J)$,

$$\begin{aligned} d\left\{m, B\left(y_1, \dots, y_J\right)\right\} &\le \|(m - \ell_y)\mathbf{1}(I_1)\| + \|(m - m)\mathbf{1}(I_2)\| + \|(m - u_y)\mathbf{1}(I_3)\| \\ &\le \|(\ell_x - \ell_y)\mathbf{1}(I_1)\| + \|(u_x - u_y)\mathbf{1}(I_3)\| \le 4J\varepsilon. \end{aligned}$$

Hence, by showing (A.8) the uniform continuity of (A.5) is verified.

To conclude the proof for *bD*, for any $P \in \mathcal{P}\{L^2[0, 1]\}$ denote the Jth product measure of P by $P^J \in \mathcal{P}\{L^2[0, 1]^J\}$, where $L^2[0, 1]^J$ is the Jth Cartesian product of the space $L^2[0, 1]$. Now it is enough to use Theorem 2.8 in [3], which implies that if $P_\nu \xrightarrow{w} P$ as $\nu \to \infty$, then $P_\nu^J \xrightarrow{w} P^J$ as $\nu \to \infty$. The assertion of the theorem for *bD* then follows from Corollary 11.3.4 in [16].

For the integrated depth one can modify the proof of Theorem 11 in [49] using the measurability result established Theorem 1 in [46] to obtain the result. Note that for the consistency of the integrated depth, the condition (A6) is not necessary. □

Having established Lemma 1, the assertions for *hD* and *aD* follow directly from an application of Varadarajan's theorem; see, e.g., Theorem 11.4.1 in [16].

For *bD*, the proof is slightly more complicated because the sample version of the depth takes the form of a U-statistic. For this, we need to introduce some additional notation. Recall that for $P \in \mathcal{P}\{L^2[0, 1]\}$ and $J \ge 1$, $P^J \in \mathcal{P}\{L^2[0, 1]^J\}$ stands for the Jth product measure of P. For $P_n = P_n(\omega) \in \mathcal{P}\{L^2[0, 1]\}$ the empirical measure of $X_1, \dots, X_n \in L^2[0, 1]$, define the Jth U-empirical measure, see [18], as

$$L_n^J = \frac{1}{n_{(J)}} \sum_{i_1, \dots, i_J \in I(J, n)} \delta_{\left(X_{i_1}, \dots, X_{i_J}\right)}.$$

Here, $n^{(J)} = \prod_{k=0}^{J-1}(n - k)$, $I(J, n) \subset \{1, \dots, n\}^J$ consists of all J-tuples with pairwise different components, and δ_x is the Dirac measure of $x \in L^2[0, 1]^J$. It is possible to write

$$\begin{aligned} bD(x; P_n) &= \int_{L^2[0,1]^J} K_D\left[d\left\{x, B\left(y_1, \dots, y_J\right)\right\}\right] \mathrm{d}\, L_n^J(y_1, \dots, y_J), \\ bD(x; P) &= \int_{L^2[0,1]^J} K_D\left[d\left\{x, B\left(y_1, \dots, y_J\right)\right\}\right] \mathrm{d}\, P^J(y_1, \dots, y_J). \end{aligned}$$

The measure L_n^J can be also written as

$$L_n^J = \frac{n^J}{n_{(J)}} P_n^J + \left(1 - \frac{n^J}{n_{(J)}}\right) Q_n^J, \tag{A.9}$$

where $Q_n^J = Q_n^J(\omega) \in \mathcal{P}\{L^2[0, 1]^J\}$ is the uniform probability measure over those $\left(X_{i_1}, \dots, X_{i_J}\right) \in L^2[0, 1]^J$ such that $i_j = i_k$ for some $j \ne k$. The next lemma will be used to assert the equivalence of the weak convergence of L_n^J and P_n^J.

Lemma 2. *Let $P_\nu, Q_\nu, P \in \mathcal{P}\{L^2[0, 1]^J\}$ and $\{\alpha_\nu\}_{\nu=1}^\infty \subset [0, 1]$ be such that $\alpha_\nu \to 1$ as $\nu \to \infty$. Then $\alpha_\nu P_\nu + (1 - \alpha_\nu)Q_\nu \xrightarrow{w} P$ as $\nu \to \infty$ if and only if $P_\nu \xrightarrow{w} P$ as $\nu \to \infty$.*

Proof. For the proof, consider the Kantorovich–Rubinstein norm $\|\cdot\|_0$ generating the weak topology on the space of non-negative measures on $L^2[0, 1]^J$; see Chapter 8.3 in [4]. For the "if" part of the proof we can write

$$0 \le \|\alpha_\nu P_\nu + (1 - \alpha_\nu)Q_\nu - P\|_0 \le \alpha_\nu \|P_\nu - P\|_0 + (1 - \alpha_\nu)\|Q_\nu - P\|_0 \le \|P_\nu - P\|_0 + (1 - \alpha_\nu)2,$$

and as $\nu \to \infty$, the right-hand side vanishes.

For the opposite implication we can find a similar bound, viz.

$$\begin{aligned} 0 \le \|P_\nu - P\|_0 &\le \|P_\nu - \alpha_\nu P_\nu - (1 - \alpha_\nu)Q_\nu\|_0 + \|\alpha_\nu P_\nu + (1 - \alpha_\nu)Q_\nu - P\|_0 \\ &\le (1 - \alpha_\nu)2 + \|\alpha_\nu P_\nu + (1 - \alpha_\nu)Q_\nu - P\|_0, \end{aligned}$$

yielding the result. □

Applying Lemma 2 to the decomposition of L_n^J from (A.9) asserts that for any $\omega \in \Omega$ fixed, $L_n^J(\omega) \xrightarrow{w} P^J$ as $n \to \infty$ is equivalent with $P_n^J(\omega) \xrightarrow{w} P^J$ as $n \to \infty$. By Varadarajan's theorem, Theorem 11.4.1 in [16] and Theorem 2.8 in [3], the latter is true for P-almost all $\omega \in \Omega$. Therefore, we obtain that

$$\mathsf{P}\left(L_n^J \xrightarrow[n\to\infty]{w} P^J\right) = \mathsf{P}\left(P_n^J \xrightarrow[n\to\infty]{w} P^J\right) = 1.$$

This formula together with Lemma 1 provides the assertion for *bD*.

A.3. Proof of Theorem 3

Apply Theorem 1 together with Lemma 1. For bD, modify the proof according to the proof of Theorem 2.

A.4. Proof of Theorem 4

First, notice that

$$\begin{aligned}\sup_{x\in L^2[0,1]} |aD(x;P_n) - aD(x;P)| &= \sup_{x\in L^2[0,1]} \left| \int_0^1 D_1\{x(t);P_{n,t}\} - D_1\{x(t);P_t\}\,\mathrm{d}\,\lambda(t)\right| \\ &\le \int_0^1 \sup_{u\in\mathbb{R}} \Big| \,|1/2 - F_t\,(u)| - \left|1/2 - F_{n,t}\,(u)\right| \Big|\,\mathrm{d}\,\lambda(t) \\ &\le \int_0^1 \sup_{u\in\mathbb{R}} |F_{n,t}(u) - F_t(u)|\,\mathrm{d}\,\lambda(t).\end{aligned}$$

By Lemma 3 given in Appendix A.7 and, for every integer $n \ge 1$,

$$\mathsf{P}\left\{\sup_{x\in L^2[0,1]} |aD(x;P_n) - aD(x;P)| > \frac{\varepsilon}{\sqrt{n}}\right\} \le \mathsf{P}\left\{\int_0^1 \xi_n(t)\,\mathrm{d}\,\lambda(t) > \varepsilon\right\},$$

where ξ_n is defined in (A.15), we obtain the desired result.

A.5. Proof of Theorem 5

By a straightforward modification of Theorem 5.2 and Problem 5.6 in [26], under the conditions assumed in Theorem 5, the rate of convergence as stated holds true for a single fixed function x. Notice that in the proof of Theorem 5.2 in [26], by taking E_X on both sides of the last chain of inequalities in the proof and using (A.1), we see that (in our notation, for $\beta = 1$ and the naive kernel as considered in the proof presented in the book)

$$c_T\,\mathrm{E}\left\|X_n - \widetilde{X}_n\right\|^2 \le 2\sigma^2 \frac{\tilde{c}}{m_n h_{m_n}} + L^2 h_{m_n}^2 + \mathrm{E}_X \int_0^1 X(t)^2 \left\{1 - P_T\left([t-h_{m_n}, t+h_{m_n}]\right)\right\}^{m_n}\,\mathrm{d}\,P_T(t),$$

for some constant $\tilde{c} > 0$. Here, the last summand can be bounded from above using Fubini's theorem, (A1*) and (A3) by

$$\begin{aligned}\int_0^1 \mathrm{E}_X\, X(t)^2 \left\{1 - P_T\left([t-h_{m_n}, t+h_{m_n}]\right)\right\}^{m_n}\,\mathrm{d}\,P_T(t) &\le M \int_0^1 \left\{1 - P_T\left([t-h_{m_n}, t+h_{m_n}]\right)\right\}^{m_n}\,\mathrm{d}\,P_T(t) \\ &\le M \int_0^1 \exp\left\{-m_n P_T\left([t-h_{m_n}, t+h_{m_n}]\right)\right\}\,\mathrm{d}\,P_T(t) \\ &\le M \sup_{u>0}\{u\exp(-u)\} \int_0^1 \frac{1}{m_n P_T\left([t-h_{m_n}, t+h_{m_n}]\right)}\,\mathrm{d}\,P_T(t) \\ &\le \frac{M}{\exp(1)} \frac{1}{2 m_n h_{m_n} c_T}.\end{aligned}$$

Independence of X and the remaining random quantities then enables to proceed as in the proof of Theorem 1 to obtain the result.

A.6. Proof of Theorem 6

A.6.1. Proof of part (i)

For $t \in [0,1]$ and any integer $i \ge 1$, denote $a_i(t) = |X_i(t) - \widetilde{X}_i(t)|$. Using the same argument as in the proof of Theorem 1 in [5], it can be shown that, for any $\varepsilon > 0$,

$$\sup_{u\in\mathbb{R}} |\widetilde{F}_{n,t}(u) - F_t(u)| \le \frac{1}{\varepsilon\, n} \sum_{i=1}^n a_i(t) + \eta_n(t) + \sup_{|s-s'|\le\varepsilon} |F_t(s) - F_t(s')|,$$

where $\eta_n(t) = \sup_{u\in\mathbb{R}} |F_{n,t}(u) - F_t(u)|$ and $\widetilde{F}_{n,t}$ stands for the distribution function of $\widetilde{P}_{n,t} \in \mathcal{P}(\mathbb{R})$. This allows us to write

$$\begin{aligned}\sup_{x\in L^2[0,1]} |aD(x;\widetilde{P}_n) - aD(x;P)| &\le \int_0^1 \sup_{u\in\mathbb{R}} |F_t(u) - \widetilde{F}_{n,t}(u)|\,\mathrm{d}\,\lambda(t) \\ &\le \frac{1}{\varepsilon\, n} \sum_{i=1}^n \int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t) + \int_0^1 \eta_n(t)\,\mathrm{d}\,\lambda(t) + \int_0^1 \sup_{|s-s'|\le\varepsilon} |F_t(s) - F_t(s')|\,\mathrm{d}\,\lambda(t).\end{aligned} \tag{A.10}$$

Because (A.10) holds true for any positive ε, it is also true if we take an infimum over the set $\{\varepsilon > 0\}$ on both sides of the inequality. Bounding further the last summand in (A.10) by (12), we find

$$\sup_{x\in L^2[0,1]} |aD(x;\widetilde{P}_n) - aD(x;P)| \le \inf_{\varepsilon>0}\left\{\frac{1}{\varepsilon\, n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t) + \delta_F(\varepsilon)\right\} + \int_0^1 \eta_n(t)\,\mathrm{d}\,\lambda(t).$$

The first term on the right-hand side relates to the Legendre transform (13), as can be seen from

$$\begin{aligned}\inf_{\varepsilon>0}\left\{\frac{1}{\varepsilon\, n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t) + \delta_F(\varepsilon)\right\} &= -\sup_{\varepsilon>0}\left\{-\frac{1}{\varepsilon\, n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t) - \delta_F(\varepsilon)\right\}\\ &= -\sup_{\nu>0}\left\{-\nu\frac{1}{n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t) - g(\nu)\right\}\\ &= -g^*\left\{-\frac{1}{n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t)\right\}.\end{aligned}$$

Thus, we may write

$$\sup_{x\in L^2[0,1]} |aD(x;\widetilde{P}_n) - aD(x;P)| \le -g^*\left\{-\frac{1}{n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t)\right\} + \int_0^1 \eta_n(t)\,\mathrm{d}\,\lambda(t). \tag{A.11}$$

Using (11) we have, for some $C > 0$,

$$\mathrm{E}\,\frac{1}{n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t) \le \mathrm{E}\,\frac{1}{n}\sum_{i=1}^{n}\sqrt{\int_0^1 a_i(t)^2\,\mathrm{d}\,\lambda(t)} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\,\|X_i - \widetilde{X}_i\| \le \frac{1}{n}\sum_{i=1}^{n} C m_i^{-\beta/(2\beta+1)},$$

yielding

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t) = \mathcal{O}_P\left\{\frac{1}{n}\sum_{i=1}^{n} m_i^{-\beta/(2\beta+1)}\right\}. \tag{A.12}$$

By Lemma 4 from Appendix A.7, (A.12) implies

$$-g^*\left\{-\frac{1}{n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t)\right\} = \mathcal{O}_P\left[-g^*\left\{-\frac{1}{n}\sum_{i=1}^{n} m_i^{-\beta/(2\beta+1)}\right\}\right]. \tag{A.13}$$

For the second term on the right-hand side of (A.11), by Lemma 3 from Appendix A.7 we know that

$$\int_0^1 \eta_n(t)\,\mathrm{d}\,\lambda(t) = \mathcal{O}_P(n^{-1/2}). \tag{A.14}$$

The rate of convergence then follows by combination of (A.11) with (A.13) and (A.14).

A.6.2. Proof of part (ii)

The Legendre transform of $g(\nu) = K\nu^{-\alpha}$ takes the form

$$g^*(\nu) = -K(1+\alpha)\left(-\frac{\nu}{K\alpha}\right)^{\alpha/(1+\alpha)}$$

for $\nu < 0$. Formula (A.11) can then be rewritten as

$$\sup_{x\in L^2[0,1]} |aD(x;\widetilde{P}_n) - aD(x;P)| \le K(1+\alpha)\left\{\frac{\frac{1}{n}\sum_{i=1}^{n}\int_0^1 a_i(t)\,\mathrm{d}\,\lambda(t)}{K\alpha}\right\}^{\alpha/(1+\alpha)} + \int_0^1 \eta_n(t)\,\mathrm{d}\,\lambda(t),$$

which by (A.12) and (A.14) provides the assertion, since the term (A.12) is dominant. To see this, note that for any positive sequence $a_1, a_2, \ldots$ we have that $|(a_1 + \cdots + a_n)/n| \ge a_1/n$, meaning that Cesàro's means never converge faster than $\mathcal{O}(n^{-1})$; see also Lemma 5 in Appendix A.7. Hence, for any $\alpha \in (0, 1]$, for some $C > 0$,

$$-g^*\left\{-\frac{1}{n}\sum_{i=1}^{n} m_i^{-\beta/(2\beta+1)}\right\} = C\left\{\frac{1}{n}\sum_{i=1}^{n} m_i^{-\beta/(2\beta+1)}\right\}^{\alpha/(1+\alpha)} \ge Cn^{-\alpha/(1+\alpha)} \ge Cn^{-1/2}.$$

A.6.3. Proof of part (iii)

Apply Lemma 5 from Appendix A.7 to (A.12) and use part (ii).

A.7. Auxiliary results for the proofs from Section 4

Firstly, we state a useful technical lemma concerning the rate of convergence of an integral of marginal distribution functions of P. Here, for $t \in [0, 1]$ we write $F_{n,t}$ and $\widetilde{F}_{n,t}$ for the distribution functions of the marginal empirical distributions $P_{n,t}$ and $\widetilde{P}_{n,t}$, respectively.

Lemma 3. *For any $P \in \mathcal{P}\{L^2[0, 1]\}$*

$$\int_0^1 \sup_{u\in\mathbb{R}} |F_{n,t}(u) - F_t(u)| \, \mathrm{d}\,\lambda(t) = \mathcal{O}_P(n^{-1/2}).$$

Proof. Denote for $t \in [0, 1]$ and any integer $n \geq 1$,

$$\xi_n(t) = \sqrt{n} \sup_{u\in\mathbb{R}} |F_{n,t}(u) - F_t(u)|. \tag{A.15}$$

For the sequence $\xi_1(t), \xi_2(t), \ldots$ of random processes, the Dvoretzky–Kiefer–Wolfowitz inequality as stated, e.g., in Section 6.5 of [15], gives us that, for any $v > 0$,

$$\sup_{t\in[0,1]} \mathsf{P}\{\xi_n(t) > v\} \leq 2\exp(-2v^2).$$

Using this inequality, we find that, for all $t \in [0, 1]$ and every integer $n \geq 1$,

$$\mathrm{E}\,\xi_n(t)^2 = \int_0^\infty \mathsf{P}\{\xi_n(t)^2 > v\} \, \mathrm{d}\,\lambda(v) \leq 1. \tag{A.16}$$

By Donsker's classical theorem for the empirical cumulative distribution function processes, as stated, e.g., in Theorem 1.8 of [15], we know that for any $t \in [0, 1]$ the sequence of measures $\{\xi_1(t), \xi_2(t), \ldots\} \subset \mathcal{P}(\mathbb{R})$ converges weakly to a measure in $\mathcal{P}(\mathbb{R})$. This, together with (A.16) enables us to use a general result of Grinblat [25, Theorem 3], see also Theorem 1 in [31], stating that for any continuous functional $\phi: L^1[0, 1] \to \mathbb{R}$ the distribution of $\phi(\xi_n)$ converges to the distribution of the corresponding functional of the limit process. Here, of course, $L^1[0, 1]$ stands for the space of measurable functions over $[0, 1]$ whose absolute value is Lebesgue integrable. Applying this result to $\phi(x) = \int_0^1 x(t) \, \mathrm{d}\,\lambda(t)$, we obtain that the sequence of random variables $\int_0^1 \xi_1(t) \, \mathrm{d}\,\lambda(t), \int_0^1 \xi_2(t) \, \mathrm{d}\,\lambda(t), \ldots$ is tight in $\mathcal{P}(\mathbb{R})$, and the assertion follows. □

Lemma 4. *Let the functions δ_F, g and g^* be defined as in Section 4.2, and let $a_1, a_2, \ldots$ and $A_1, A_2, \ldots$ be sequences of positive numbers and random variables, respectively. If $A_n = \mathcal{O}_P(a_n)$, then $-g^*(-A_n) = \mathcal{O}_P\{-g^*(-a_n)\}$.*

Proof. Denote $h(\mu) = -g^*(-\mu)$, and recall that, for any $\mu > 0$,

$$h(\mu) = \inf_{\varepsilon>0} \{\mu/\varepsilon + \delta_F(\varepsilon)\}.$$

First we show that the function h is non-decreasing over $(0, \infty)$. Fix $0 < \mu \leq \mu' < \infty$. Then, for any $\varepsilon > 0$,

$$\mu/\varepsilon + \delta_F(\varepsilon) \leq \mu'/\varepsilon + \delta_F(\varepsilon),$$

and by taking infimum over all $\varepsilon > 0$, we get $h(\mu) \leq h(\mu')$.

Now, consider an integer $M \geq 1$ and take $\mu > 0$ and $\varepsilon > 0$ arbitrary. By a similar argument,

$$\inf_{\varepsilon>0} \{\mu/\varepsilon + \delta_F(\varepsilon)/M\} \leq \inf_{\varepsilon>0} \{\mu/\varepsilon + \delta_F(\varepsilon)\} = h(\mu).$$

Multiplying the last inequality by M, we get, for all $\mu > 0$ and every integer $M \geq 1$,

$$h(M\mu) = \inf_{\varepsilon>0} \{M\mu/\varepsilon + \delta_F(\varepsilon)\} \leq Mh(\mu). \tag{A.17}$$

Now, assume that $A_n = \mathcal{O}_P(a_n)$, i.e., for any $\varepsilon > 0$ there exists $C > 0$ such that, for every integer $n \geq 1$,

$$1 - \varepsilon \leq \mathsf{P}(A_n \leq Ca_n) \leq \mathsf{P}\{A_n \leq \max(C, 1)a_n\}.$$

Since h is non-decreasing and (A.17) is true, we may write for $M = \max(C, 1)$

$$\mathsf{P}(A_n \leq Ma_n) \leq \mathsf{P}\{h(A_n) \leq h(Ma_n)\} \leq \mathsf{P}\{h(A_n) \leq Mh(a_n)\},$$

which enables us to conclude that for every integer $n \geq 1$, $\mathsf{P}\{h(A_n) > Mh(a_n)\} < \varepsilon$. □

Follows a technical lemma stating the rate of convergence of the sequence of Cesàro's means of a given deterministic sequence. This result is used in the proof of Theorem 6.

Lemma 5. *For $r \in \mathbb{R}$ and a sequence $a_n = \mathcal{O}(n^r)$ as $n \to \infty$, consider the sequence $b_n = (a_1 + \cdots + a_n)/n$. Then, as $n \to \infty$,*

$$b_n = \begin{cases} \mathcal{O}(n^r) & \text{if } r > -1, \\ \mathcal{O}\{\ln(n)/n\} & \text{if } r = -1, \\ \mathcal{O}(n^{-1}) & \text{if } r < -1. \end{cases}$$

Proof. By assumptions, there exists $c > 0$ such that for all n large enough $|a_n| \leq cn^r$. We can write

$$|b_n| \leq (|a_1| + \cdots + |a_n|)/n \leq \frac{c}{n} \sum_{j=1}^{n} j^r.$$

Now, it suffices to bound the sum in the last expression, and the desired result follows by application of Euler's summation formula; see Theorem 3.2 in [2]. □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2018.11.003.

References

[1] C. Agostinelli, M. Romanazzi, Local depth, J. Statist. Plann. Inference 141 (2011) 817–830.
[2] T.M. Apostol, Introduction to Analytic Number Theory, Springer-Verlag, New York, 1976.
[3] P. Billingsley, Convergence of Probability Measures, second ed., Wiley, New York, 1999.
[4] V.I. Bogachev, Measure Theory, Vol. II, Springer-Verlag, Berlin, 2007.
[5] A.-K. Bott, L. Devroye, M. Kohler, Estimation of a distribution from data with small measurement errors, Electron. J. Stat. 7 (2013) 2457–2476.
[6] J.L.O. Cabrera, `locpol`: Kernel Local Polynomial Regression, R package version 0.6-0, 2012.
[7] A. Chakraborty, P. Chaudhuri, The spatial distribution in infinite dimensional spaces and related quantiles and depths, Ann. Statist. 42 (2014) 1203–1231.
[8] G. Claeskens, M. Hubert, L. Slaets, K. Vakili, Multivariate functional halfspace depth, J. Amer. Statist. Assoc. 109 (2014) 411–423.
[9] C. Crambes, A. Kneip, P. Sarda, Smoothing splines estimators for functional linear regression, Ann. Statist. 37 (2009) 35–72.
[10] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Statist. Plann. Inference 147 (2014) 1–23.
[11] A. Cuevas, M. Febrero, R. Fraiman, On the use of the bootstrap for estimating functions with functional data, Comput. Statist. Data Anal. 51 (2006) 1063–1074.
[12] A. Cuevas, M. Febrero, R. Fraiman, Robust estimation and classification for functional data via projection-based depth notions, Comput. Statist. 22 (2007) 481–496.
[13] A. Cuevas, R. Fraiman, On depth measures and dual statistics: A methodology for dealing with general data, J. Multivariate Anal. 100 (2009) 753–766.
[14] K. De Brabanter, F. Cao, I. Gijbels, J. Opsomer, Local polynomial regression with correlated errors in random design and unknown correlation structure, Biometrika 105 (2018) 681–690.
[15] R.M. Dudley, Uniform Central Limit Theorems, Cambridge University Press, Cambridge, 1999.
[16] R.M. Dudley, Real Analysis and Probability, Cambridge University Press, Cambridge, 2002, Revised reprint of the 1989 original.
[17] S. Dutta, A.K. Ghosh, P. Chaudhuri, Some intriguing properties of Tukey's half-space depth, Bernoulli 17 (2011) 1420–1434.
[18] S. Eichelsbacher, U. Schmock, Rank-dependent moderate deviations of U-empirical measures in strong topologies, Probab. Theory Related Fields 126 (2003) 61–90.
[19] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Chapman & Hall, London, 1996.
[20] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer, New York, 2006.
[21] R. Fraiman, G. Muniz, Trimmed means for functional data, Test 10 (2001) 419–440.
[22] M.G. Genton, C. Johnson, K. Potter, G. Stenchikov, Y. Sun, Surface boxplots, Stat 3 (2014) 1–11.
[23] I. Gijbels, S. Nagy, Consistency of non-integrated depths for functional data, J. Multivariate Anal. 140 (2015) 259–282.
[24] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, J. Multivariate Anal. 146 (2016) 1–6.
[25] L.Š. Grinblat, A limit theorem for measurable random processes and its applications, Proc. Amer. Math. Soc. 61 (1976) 371–376.
[26] L. Györfi, M. Kohler, A. Krzyzak, H. Walk, A Distribution-free Theory of Nonparametric Regression, Springer-Verlag, New York, 2002.
[27] L. Horváth, P. Kokoszka, Inference for Functional Data With Applications, Springer, New York, 2012.
[28] T. Hsing, R. Eubank, Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators, Wiley, Chichester, 2015.
[29] R.J. Hyndman, H.L. Shang, Rainbow plots, bagplots, and boxplots for functional data, J. Comput. Graph. Statist. 19 (2010) 29–45.
[30] F. Ieva, A.M. Paganoni, Depth measures for multivariate functional data, Comm. Statist. Theory Methods 42 (2013) 1265–1276.
[31] A.V. Ivanov, Convergence of distributions of functionals of measurable random fields, Ukrain. Mat. Zh. 32 (1980) 27–34, 141.
[32] C.-R. Jiang, J.-L. Wang, Covariate adjusted functional principal components analysis for longitudinal data, Ann. Statist. 38 (2010) 1194–1226.
[33] M. Kohler, A. Krzyzak, H. Walk, Optimal global rates of convergence for nonparametric regression with unbounded data, J. Statist. Plann. Inference 139 (2009) 1286–1296.
[34] Y. Li, T. Hsing, Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data, Ann. Statist. 38 (2010) 3321–3351.
[35] R.Y. Liu, On a notion of data depth based on random simplices, Ann. Statist. 18 (1990) 405–414.
[36] X.-H. Liu, Kernel Smoothing for Spatially Correlated Data (Ph.D. thesis), Department of Statistics, Iowa State University, Ames, IA, 2001.
[37] S. López-Pintado, R. Jornsten, Functional analysis via extensions of the band depth, in: Complex Datasets and Inverse Problems, Inst. Math. Statist., Beachwood, OH, 2007, pp. 103–120.
[38] S. López-Pintado, J. Romo, On the concept of depth for functional data, J. Amer. Statist. Assoc. 104 (2009) 718–734.
[39] S. López-Pintado, J. Romo, A half-region depth for functional data, Comput. Statist. Data Anal. 55 (2011) 1679–1695.

[40] S. López-Pintado, Y. Sun, J. Lin, M.G. Genton, Simplicial band depth for multivariate functional data, Adv. Data Anal. Classif. 8 (2014) 321–338.
[41] S. López-Pintado, Y. Wei, Depth for sparse functional data, in: Recent Advances in Functional Data Analysis and Related Topics, in: Contrib. Statist., Physica-Verlag/Springer, Heidelberg, 2011, pp. 209–212.
[42] K. Mosler, Depth statistics, in: C. Becker, R. Fried, S. Kuhnt (Eds.), Robustness and Complex Data Structures, Springer, Heidelberg, 2013, pp. 17–34.
[43] M. Mosler, Y. Polyakova, General notions of depth for functional data, arXiv preprint arXiv:1208.1981, 2012.
[44] H.-G. Müller, Y. Wu, F. Yao, Continuously additive models for nonlinear functional regression, Biometrika 100 (2013) 607–622.
[45] S. Nagy, Consistency of h-mode depth, J. Statist. Plann. Inference 165 (2015) 91–103.
[46] S. Nagy, Integrated depth for measurable functions and sets, Statist. Probab. Lett. 123 (2017) 165–170.
[47] S. Nagy, An overview of consistency results for depth functionals, in: G. Aneiros, E.G. Bongiorno, R. Cao, P. Vieu (Eds.), Functional Statistics and Related Fields, Springer International Publishing, Cham, Switzerland, 2017, pp. 189–196.
[48] S. Nagy, I. Gijbels, D. Hlubinka, Weak convergence of discretely observed functional data with applications, J. Multivariate Anal. 146 (2016) 46–62.
[49] S. Nagy, I. Gijbels, M. Omelka, D. Hlubinka, Integrated depth for functional data: Statistical properties and consistency, ESAIM Probab. Stat. 20 (2016) 95–130.
[50] N.N. Narisetty, V.N. Nair, Extremal depth for functional data and applications, J. Amer. Statist. Assoc. 111 (2016) 1705–1714.
[51] J. Opsomer, Y. Wang, Y. Yang, Nonparametric regression with correlated errors, Statist. Sci. 16 (2001) 134–153.
[52] D. Paul, J. Peng, Consistency of restricted maximum likelihood estimators of principal components, Ann. Statist. 37 (2009) 1229–1271.
[53] O. Pokotylo, P. Mozharovskyi, R. Dyckerhoff, S. Nagy, `ddalpha`: Depth-Based Classification and Calculation of Data Depth, R package version 1.3.1, 2017.
[54] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2015.
[55] P. Radchenko, X. Qiao, G.M. James, Index models for sparsely sampled functional data, J. Amer. Statist. Assoc. 110 (2015) 824–836.
[56] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2015.
[57] C.J. Stone, Consistent nonparametric regression, Ann. Statist. 5 (1977) 595–645.
[58] C.J. Stone, Optimal global rates of convergence for nonparametric regression, Ann. Statist. 10 (1982) 1040–1053.
[59] Y. Sun, M.G. Genton, Functional boxplots, J. Comput. Graph. Statist. 20 (2011) 316–334.
[60] Y. Sun, M.G. Genton, Adjusted functional boxplots for spatio-temporal data visualization and outlier detection, Environmetrics 23 (2012) 54–64.
[61] J.W. Tukey, Mathematics and the picturing of data, in: Proceedings of the International Congress of Mathematicians (Vancouver, BC, 1974), Vol. 2, Canad. Math. Congress, Montréal, QC, 1975, pp. 523–531.
[62] V.S. Varadarajan, On the convergence of sample probability distributions, Sankhyā 19 (1958) 23–26.
[63] R.T. Whitaker, M. Mirzargar, R.M. Kirby, Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles, IEEE Trans. Vis. Comput. Graphics 19 (2013) 2713–2722.
[64] Y. Wu, J. Fan, H.-G. Müller, Varying-coefficient functional linear regression, Bernoulli 16 (2010) 730–758.
[65] F. Yao, H.-G. Müller, J.-L. Wang, Functional data analysis for sparse longitudinal data, J. Amer. Statist. Assoc. 100 (2005) 577–590.
[66] X. Zhang, J.-L. Wang, From sparse to dense functional data and beyond, Ann. Statist. 44 (2016) 2281–2321.
[67] Y. Zuo, R.J. Serfling, General notions of statistical depth function, Ann. Statist. 28 (2000) 461–482.