

## Accepted Manuscript

Weak convergence of discretely observed functional data with applications

Stanislav Nagy, Irène Gijbels, Daniel Hlubinka

PII: S0047-259X(15)00149-9

DOI: <http://dx.doi.org/10.1016/j.jmva.2015.06.006>

Reference: YJMVA 3948

To appear in: *Journal of Multivariate Analysis*

Received date: 17 December 2014

Please cite this article as: S. Nagy, I. Gijbels, D. Hlubinka, Weak convergence of discretely observed functional data with applications, *Journal of Multivariate Analysis* (2015), <http://dx.doi.org/10.1016/j.jmva.2015.06.006>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# Weak Convergence of Discretely Observed Functional Data with Applications

Stanislav Nagy<sup>a,b,c,\*</sup>, Irène Gijbels<sup>a,b</sup>, Daniel Hlubinka<sup>c</sup>

<sup>a</sup>KU Leuven, Department of Mathematics, Statistics Section, Celestijnenlaan 200b - box 2400, 3001 Leuven, Belgium

<sup>b</sup>KU Leuven, Leuven Statistics Research Centre (LStat), Celestijnenlaan 200b - box 5307, 3001 Leuven, Belgium

<sup>c</sup>Charles University in Prague, Department of Probability and Math. Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic

## Abstract

A general result on weak convergence of the empirical measure of discretely observed functional data is shown. It is applied to the problem of estimation of functional mean value, and the problem of consistency of various types of depth for functional data. Counterexamples illustrating the fact that the assumptions as stated cannot be dropped easily are given.

**Keywords:** consistency, data depth, functional data, functional moments, weak convergence  
**2000 MSC:** 62G20, 62G05, 62H12

## 1. Introduction: Complete and Discrete Design of Functional Data

By functional data we usually understand an outcome of an experiment that can be represented as a set of continuous functions, whose domain is a common compact interval. The domain can be referred to as time. The value of the outcoming function can then be interpreted as the value of the measured characteristic, evolving within the given time frame. Depending on the nature of the experiment, two distinct approaches towards the observation design of functional data can be found in the literature.

The first, more traditional approach, assumes that all the functions involved are observed completely in time. In other words, the functional values of all the functions resulting from the experiment are known at each time point of their domain. We call this setup **complete design**, for brevity. Under this complete design assumption there exists a wide variety of statistical procedures enabling the analysis of functional data. For exposition see, for example, Ramsay and Silverman (2002, 2005), Ferraty and Vieu (2006) and Horváth and Kokoszka (2012).

The second, perhaps more realistic, approach to the observation design of functions arises when assuming that each random function is observed only at a finite grid of time points. These points can be either deterministic and preset by the experimenter, or occurring at random. Here, we focus on nonparametric statistical analysis of such partially observed data.

\*Corresponding author. Phone: +32 16 37 23 86.

Email addresses: stanislav.nagy@wis.kuleuven.be (Stanislav Nagy), irene.gijbels@wis.kuleuven.be (Irène Gijbels), hlubinka@karlin.mff.cuni.cz (Daniel Hlubinka)

Typically, to make valid nonparametric inference exploiting the functional nature of data, it should be assumed that the number of points at which random functions are observed tends to be larger when the sample size increases, and that the largest span between two adjacent points vanishes as the sampling process continues to infinity. This setup will be called **discrete design**.

The majority of papers in functional data analysis consider the data as being observed at each point of the domain. See Cuevas (2014, for example Section 4.2) for some discussion on the nature and treatment of functional data. In a variable selection problem Aneiros and Vieu (2014) consider the functions to be observed through a fine grid, and look into asymptotics when the distances between the grid points diminish to zero. Other recent contributions taking data as discretized can be found in a book edited by Bongiorno et al. (2014).

To accomplish statistical analysis for functions observed within the discrete design setting, it is necessary to perform an initial step of representing the discretely observed functions by elements of the space of complete functions. This is usually done by approximating, or interpolating, the discretely observed values of functions. After doing this, the resulting approximating (complete) functions can be utilized as if they were the original, continuously unobservable, set of curves.

Such preprocessing of functional data is customarily considered to be imperative when encountering a discretely observed functional data set (Ramsay and Silverman, 2002). Though often discussed, few theoretical results investigate the effect of this fundamental data imputation on the statistics involved.

The essential contribution of the present paper lies in a theoretical result facilitating the understanding of this phenomenon. In Section 2, we propose a natural and straightforward method of approximating functions from a random sample whose values are observed only at a finite number of points in the domain. Within the discrete design of observations, we show that the probability distribution of these approximations converges weakly towards the original sampling distribution. This enables us to state a Varadarajan type of result (Varadarajan, 1958) for such discretely observed functional data dealing with the weak convergence of empirical measures based on these approximations.

The second part of the paper concerns two applications of the main result. In Section 3 we apply it to the problem of estimation of mean of functional data. We show that by an average based solely on a random sample of discretely observed curves, it is possible to estimate the mean value of a probability distribution in a functional space in an asymptotically unbiased and consistent manner. Finally, in Section 4, the developed theory is applied to a nonparametric tool suitable for functional data — data depth (cf Zuo and Serfling, 2000). There, conditions under which consistency is preserved for functional data depth when functions are observed discretely are explored. Both application sections are completed with a number of examples. These aim to illustrate that the conditions of the theoretical results cannot be dropped generally, and depict pitfalls to keep in mind when replacing a set of discretely observed functional data by complete curves.

The proofs of the theoretical results are provided in a Supplementary Material part accompanying this paper. That part also contains an additional example providing information on the necessity of the conditions.

## 2. Weak Convergence of Discretely Observed Functions

In this section, we state a rather technical, but useful, theorem concerning general weak convergence in the space of continuous functions on a compact interval. We start by introducing

the notation.

For  $K \in \mathbb{N} = \{1, 2, \dots\}$  let  $C^K([0, 1])$  be the Banach space of continuous  $\mathbb{R}^K$ -valued functions on  $[0, 1]$

$$C^K([0, 1]) = \{x: [0, 1] \rightarrow \mathbb{R}^K: x \text{ is continuous on } [0, 1]\}$$

equipped with the uniform norm  $\sup_{t \in [0, 1]} \|x(t)\|$ . Here,  $\|\cdot\|$  denotes the Euclidean norm on the space  $\mathbb{R}^K$ .

For an arbitrary metric space  $\mathcal{M}$  with the  $\sigma$ -algebra of its Borel sets,  $\mathcal{P}(\mathcal{M})$  stands for the collection of all probability measures defined on  $\mathcal{M}$ .

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space on which all random variables will be defined. For  $\mathbb{P} \in \mathcal{P}(C^K([0, 1]))$ , let  $\{X_n\}_{n=1}^\infty \subset C^K([0, 1])$  denote an infinite sequence of independent random functions distributed as  $\mathbb{P}$ . For a fixed random element  $\omega \in \Omega$  we denote the empirical measure defined by the first  $n$  functions from this sequence by  $\mathbb{P}_n(\omega) \in \mathcal{P}(C^K([0, 1]))$ . To put it precisely, if  $\delta_x$  is a Dirac measure on  $C^K([0, 1])$  concentrated at  $x \in C^K([0, 1])$ , then

$$\mathbb{P}_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)} \text{ for } \omega \in \Omega. \quad (1)$$

If no confusion about the underlying random element can arise, the argument  $\omega$  will be dropped.

For a fixed time point  $t \in [0, 1]$  and  $X \sim \mathbb{P}$  we use  $\mathbb{P}_t \in \mathcal{P}(\mathbb{R}^K)$  to denote the marginal distribution of  $X(t)$ . Likewise,  $\mathbb{P}_{n,t}$  stands for the marginal distribution of  $\mathbb{P}_n$  defined in (1) at  $t$ .

The symbol  $\mathbb{I}[A]$  stands for the indicator function of  $A$ , i.e. equals 1 if  $A$  holds true, and 0 if not. For a set  $S$  and a sequence of (possibly random) positive integers  $\{m_n\}_{n=1}^\infty \subset \mathbb{N}$ , a triangular array of elements of  $S$  is an arbitrary doubly indexed sequence of elements of the set  $S$

$$\{s_{j,n}\}_{j=1}^{m_n} = \{s_{j,n} \in S: j = 1, \dots, m_n \text{ and } n \in \mathbb{N}\}.$$

In the present section, all random functions are understood as observed within the discrete design. To put this rigorously, let  $\{T_{j,n}\}_{j=1}^{m_n} \subset [0, 1]$  be an arbitrary triangular array of points in  $[0, 1]$ . This array is referred to as the array of observation points. Notice that we do not assume anything about these points, and they can be set either deterministically or randomly. Without loss of generality, assume that they are ordered ascendingly in magnitude

$$0 \leq T_{1,n} \leq \dots \leq T_{m_n,n} \leq 1 \text{ for } n \in \mathbb{N}.$$

Suppose that for each  $n \in \mathbb{N}$  we are able to observe the random function  $X_n$  only at a finite number  $m_n \in \mathbb{N}$  of (random or non-random) discrete points  $T_{1,n}, \dots, T_{m_n,n} \in [0, 1]$ . Thus, sampling functions from  $\mathbb{P}$ , we are in fact being provided only with a sequence of unequally long random vectors of functional values

$$(X_n(T_{1,n})^\top, \dots, X_n(T_{m_n,n})^\top)^\top \in \mathbb{R}^{K \times m_n} \text{ for } n \in \mathbb{N} \quad (2)$$

instead of complete functions  $X_n$  for  $n \in \mathbb{N}$ .

For each such vector define an interpolated version of the discretely observed random function  $X_n$  (see (2)), denoted by  $\tilde{X}_n$ , in the following way:

- Set  $T_{0,n} = 0$  and  $T_{m_n+1,n} = 1$ .
- Define the functional value of  $\tilde{X}_n$  at each point at which  $X_n$  is observed to be the observed functional value

$$\tilde{X}_n(T_{j,n}) = X_n(T_{j,n}) \text{ for } j = 1, \dots, m_n.$$

- At the boundary points  $T_{0,n}$  and  $T_{m_n+1,n}$  set  $\tilde{X}_n$  to be equal to the observed values at the smallest and the largest point, respectively,

$$\begin{aligned}\tilde{X}_n(T_{0,n}) &= X_n(T_{1,n}), \\ \tilde{X}_n(T_{m_n+1,n}) &= X_n(T_{m_n,n}).\end{aligned}$$

- Define the approximating function  $\tilde{X}_n$  at arbitrary  $t \in [T_{j,n}, T_{j+1,n}]$ ,  $j = 0, \dots, m_n$ , as the linear interpolation of the observed points  $\tilde{X}_n(T_{j,n})$  and  $\tilde{X}_n(T_{j+1,n})$ . That is

$$\tilde{X}_n(t) = \frac{\tilde{X}_n(T_{j+1,n}) - \tilde{X}_n(T_{j,n})}{T_{j+1,n} - T_{j,n}} (t - T_{j,n}) + \tilde{X}_n(T_{j,n}) \text{ for } t \in [T_{j,n}, T_{j+1,n}], j = 0, \dots, m_n. \quad (3)$$

The approximating curves  $\tilde{X}_n$  are obviously continuous, piecewise linear, and coincide with the originals  $X_n$  at each point where the original random sample curve was observed

$$\tilde{X}_n(T_{j,n}) = X_n(T_{j,n}) \text{ for } j = 1, \dots, m_n. \quad (4)$$

One such function  $X$  alongside its approximation  $\tilde{X}$  is drawn in Figure 1.

A similar scenario for discrete design, but in a special setting of fixed time points, was used in the proofs section of Claeskens et al. (2014).

Assume now that the discrete design obeys a property of points getting denser in the domain as the sampling process continues, i.e. that

$$\nu(n) = \max_{j=0, \dots, m_n} |T_{j+1,n} - T_{j,n}| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (\mathbf{D})$$

Let us first inspect when the condition **(D)** is satisfied.

If the array of observation points  $\{T_{j,n}\}_{j=1}^{m_n}$ , and the sequence of its sizes  $\{m_n\}_{n=1}^{\infty}$ , are both set non-randomly, then **(D)** reduces to the usual partitioning refinement condition

$$\max_{j=0, \dots, m_n} |T_{j+1,n} - T_{j,n}| \xrightarrow[n \rightarrow \infty]{} 0.$$

The use of random partitioning, however, enables us to do much more. For instance, if each function  $X_n$  is observed at random points sampled independently of each other, an increasing, deterministic, number of observations in time is essentially a sufficient condition for **(D)** to be satisfied.

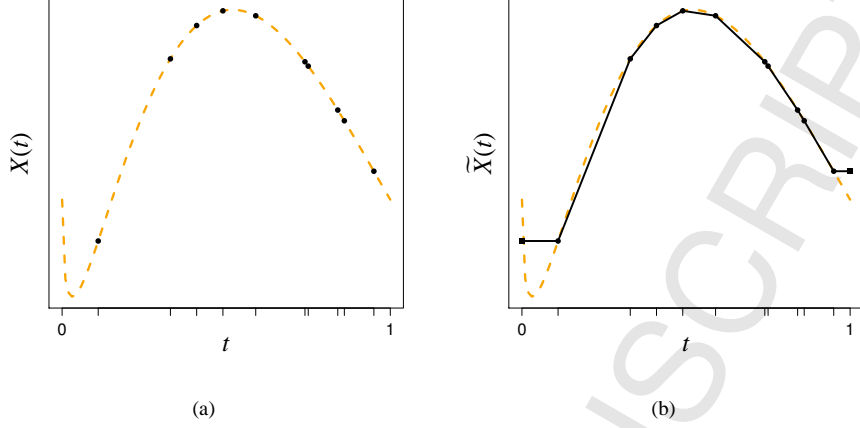


Figure 1: (a) For  $K = 1$ , a single realization of a random function  $X$  (the dashed line). On the horizontal axis the ticks represent  $m = 10$  points at which  $X$  is supposed to be discretely observed. The observed vector  $(X(T_1), \dots, X(T_m))^T$  of functional values of  $X$  is depicted by the black dots. (b) The approximating function  $\tilde{X}$  to  $X$  based on the discrete observations in (a).

**Lemma 1.** Assume that the rows of the triangular array  $\{T_{j,n}\}_{j=1}^{m_n} \subset [0, 1]$  consist of independent random vectors  $(T_{1,n}, \dots, T_{m_n,n})^T$ , where  $T_{1,n}, \dots, T_{m_n,n}$  is an ordered random sample from a distribution  $T$  on  $[0, 1]$  of non-random size  $m_n \in \mathbb{N}$ , for each  $n \in \mathbb{N}$ . Let  $T$  have a density  $f$  in  $[0, 1]$  bounded away from zero, that is for some  $A > 0$

$$\inf_{t \in [0,1]} f(t) \geq A$$

and let  $m_n \xrightarrow{n \rightarrow \infty} \infty$ . Then **(D)** is satisfied for  $\{T_{j,n}\}_{j=1}^{m_n}$ .

*Proof.* Let

$$0 = U_{0,n} \leq U_{1,n} \leq \dots \leq U_{m_n,n} \leq U_{m_n+1,n} = 1$$

be an ordered random sample of size  $m_n$  from the uniform distribution on  $[0, 1]$  appended with the endpoints 0 and 1. The inverse transform sampling method then yields that

$$0 = F^{-1}(U_{0,n}) \leq F^{-1}(U_{1,n}) \leq \dots \leq F^{-1}(U_{m_n,n}) \leq F^{-1}(U_{m_n+1,n}) = 1$$

is an ordered random sample from distribution  $T$  appended with 0 and 1. For the norm of the corresponding partitioning we have by the mean value theorem

$$\nu(n) = \max_{j=0, \dots, m_n} |F^{-1}(U_{j+1,n}) - F^{-1}(U_{j,n})| \leq \frac{1}{A} \max_{j=0, \dots, m_n} |U_{j+1,n} - U_{j,n}|.$$

Using a result due to Lévy which can be found, for instance, in Darling (1953, Section 8), the last quantity here converges to zero in probability provided  $m_n \xrightarrow{n \rightarrow \infty} \infty$ . This completes the proof.  $\square$

Analogous results to Lemma 1 can be proved also under additional, more varied scenarios. These could include setups of time observation distributions varying in  $n \in \mathbb{N}$ , random numbers of observation points (tending properly to infinity), or different combinations of these designs.

If (D) is satisfied, it is reasonable to ask whether the distribution of approximating random functions has asymptotically the original distribution of the functions which are not continuously observable. The following result provides an affirmative answer. Its proof can be found in the Supplementary Material, Section S.1.

**Theorem 1.** *Let (D) be satisfied and  $\{X_n\}_{n=1}^\infty$  be a sequence of independent random functions from a distribution  $X \sim P \in \mathcal{P}(C^K([0, 1]))$ . Then the following holds true:*

- (i) *The sequence of random functions  $\{\bar{X}_n\}_{n=1}^\infty$  approximating the sequence  $\{X_n\}_{n=1}^\infty$  converges in distribution to  $X$ , and*
- (ii) *the sequence of empirical measures  $\{\bar{P}_n\}_{n=1}^\infty \subset \mathcal{P}(C^K([0, 1]))$ , where*

$$\bar{P}_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{\bar{X}_i(\omega)} \quad \text{for } \omega \in \Omega, \quad (5)$$

*converges to  $P$  in the sense of weak convergence almost surely, that is*

$$P\left(\bar{P}_n \xrightarrow[n \rightarrow \infty]{w} P\right) = 1.$$

In the sequel, we apply this theoretical result to two different problems: functional mean estimation, and functional data depth. Other applications include the functional extensions of regression analysis (for background information see Ramsay and Silverman (2005), Ferraty and Vieu (2006) and Goia and Vieu (2014)), or various statistical testing problems for functional data. Adaptation of the present theory to these settings is, however, not always straightforward, and is part of future research in functional data analysis.

### 3. Application to Functional Mean Estimation

As the first application of the convergence results from Section 2, let us consider the problem of moment estimation for functional data. The issue of establishing a reasonable expectation is crucial for the introduction of statistical methodology to the setting of general functional data sets.

Two intrinsically different approaches towards the definition of functional mean can be found in the literature. The first one, better known from the statistics literature, is a rather straightforward way of defining an expected value of a probability measure  $P \in \mathcal{P}(C^K([0, 1]))$ ,  $X \sim P$ , in the coordinate-wise sense. In this case, the usual finite-dimensional expectation of the marginal distribution  $P_t \in \mathcal{P}(\mathbb{R}^K)$  is evaluated for each  $t \in [0, 1]$ , provided it exists, and then the resulting  $\mathbb{R}^K$ -valued function

$$EX: [0, 1] \rightarrow \mathbb{R}^K: t \mapsto EX(t) \quad (6)$$

is pronounced to be the expectation of the random function  $X$ . This approach has been adopted in the majority of the contemporary literature dealing with functional data — see, for example, Ramsay and Silverman (2002, Section 2.2.2), Ramsay and Silverman (2005, Section 2.3), or Ferraty and Vieu (2006, Section 9.2.1).

A slight complication arising when defining the expectation this way is that even if the function (6) is well defined on  $[0, 1]$ , it may not be continuous, even though the random functions  $X \sim P$  are (Nagy, 2013, Section 2). This failure to share common properties of underlying random functions is a consequence of the fact that the definition is not based upon the notion of integral, as the usual expectation in Euclidean spaces is.

This problem is resolved by the second, more probabilistic-like notion of mean for functional data. It is based upon the idea of integration in general Banach spaces (Diestel and Uhl, 1977) and usually translates into functional data expectation framework in the form of either Bochner, or Pettis mean (cf Grenander (1981) or Araujo and Giné (1980, Section 3.2)). The more general Pettis mean of a probability measure  $P \in \mathcal{P}(C^K([0, 1]))$ ,  $X \sim P$ , is defined as the value  $x \in C^K([0, 1])$  such that for any  $x^*$  element of the dual space of  $C^K([0, 1])$  the equality

$$x^*(x) = \int_{\Omega} x^*(X(\omega)) \, dP(\omega)$$

holds true. Claiming this we silently assume that the right hand side of the previous display is a well defined Lebesgue integral for any  $x^*$ . One sufficient condition for the existence of a unique Pettis mean of  $P$  is that  $P$  is a measure concentrated in a bounded subset of  $C^K([0, 1])$  (cf Winkler (1985, Proposition 1.1.3) or Aliprantis and Border (2006)).

Of course, if  $P$  has a Pettis mean, it must have a coordinate-wise mean (6), and these two coincide. In particular, the mean value is in this case an element of  $C^K([0, 1])$ , and enjoys a number of favourable properties of an operator defined by means of a general integral (Diestel and Uhl, 1977). This observation enables us in the sequel to speak about general functional mean defined in the coordinate-wise sense (6), with only a slight abuse of terminology.

Having discussed the definition of mean value for functional data, the problem of estimation of such an operator given a random sample  $X_1, \dots, X_n$  of functions from  $P$  emerges. If whole trajectories of sample functions are observed, one could utilize (6) and estimate the mean function (assuming it exists, in some sense) by the usual average

$$\bar{X}_n: [0, 1] \rightarrow \mathbb{R}^K: t \mapsto \frac{1}{n} \sum_{i=1}^n X_i(t). \quad (7)$$

This estimator, widely used in the statistical literature (cf Ramsay and Silverman, 2002, 2005), is well justified; at least in the likely case when the Pettis mean of  $P$  exists. However, it relies on the fact that all random functions are observed completely.

In the case of discretely observed functions we may use approximating versions  $\tilde{X}_n$  of sample functions  $X_n$ , plug them into the formula (7), and hope that the resulting average function has the plausible properties of the usual average based on fully observed functions. We shall explore if this is indeed true; in particular, we focus on two crucial properties that an estimator of a mean must satisfy, namely it must be

- (asymptotically) unbiased, and
- strongly consistent.

Herein we show, provided the approximations  $\tilde{X}$  are chosen as introduced in Section 2, that both these conditions can be recovered, under reasonable assumptions. This is established in Theorem 2. The proof of this result is provided in the Supplementary Material, Section S.2.



**Theorem 2.** Let  $m \in \mathbb{N}$  and  $X \sim P \in \mathcal{P}(C^K([0, 1]))$  be such that

$$E(\|X(\cdot)\|^m) : [0, 1] \rightarrow [0, \infty) \text{ is continuous on } [0, 1]. \quad (8)$$

Assume that **(D)** is satisfied for a triangular array  $\{T_{j,n}\}_{j=1}^{m_n} \subset [0, 1]$  independent of  $\{X_n\}_{n=1}^\infty$ , a sequence of independent random functions from  $P$ .

(i) Then  $E\|\tilde{X}_n(t)\|^m \xrightarrow{n \rightarrow \infty} E\|X(t)\|^m$  for all  $t \in [0, 1]$ .

(ii) If, moreover,

$$\sup_{t \in [0, 1]} E(\|X(\cdot)\|^{2m}) < \infty, \quad (9)$$

is satisfied, then

$$\frac{1}{n} \sum_{i=1}^n \|\tilde{X}_i(t)\|^m \xrightarrow[n \rightarrow \infty]{\text{a.s.}} E(\|X(t)\|^m) \text{ for } t \in [0, 1].$$

(iii) In particular, if (9) is true for  $m = 1$ , then

$$\frac{1}{n} \sum_{i=1}^n \tilde{X}_i(t) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} E X(t) \text{ for } t \in [0, 1].$$

For completely observed functions, results corresponding to Theorem 2 are abundant in the literature. See, for instance, Ledoux and Talagrand (2011) for an involved discussion on the law of large numbers in general Banach spaces. The latter theory, however, hinges on the complete observational design of the data curves, and cannot be extended to discretely observed functions.

One may wonder whether the conditions in Theorem 2 can be weakened. In the sequel of this section we provide some examples to illustrate that

- Condition (8) cannot be dropped (see Example 1);
- Condition (9) is needed to guarantee assertions (ii) and (iii) (see Example 2);
- The independence assumption between the random functions  $\{X_n\}_{n=1}^\infty$  and the observation points cannot be dropped (see Example 3).

**Example 1.** We construct a probability measure  $P \in \mathcal{P}(C([0, 1]))$  and a deterministic triangular array of observations

$$\{T_{j,n}\}_{j=1}^n = \left\{ \frac{j}{n} : j = 1, \dots, n, \text{ and } n \in \mathbb{N} \right\} \subset [0, 1] \quad (10)$$

satisfying **(D)** such that the assertions of Theorem 2 do not hold. To this end, we modify a technique for the construction of counterexamples used by Osius (1989) concerning the convergence of moments of weakly convergent random variables.

For  $K = 1$ , construct a probability measure  $P \in \mathcal{P}(C([0, 1]))$  in a coordinate-wise sense. Initially, suppose that an auxiliary random variable  $Z$  has an exponential distribution with expectation 1 and the marginal  $P_0$  of  $P$  is distributed as  $Z$ .

For a continuous function  $c: [0, 1] \rightarrow [0, 1]$  such that  $c(0) = 0$  and a time point  $t \in (0, 1]$ , define the marginal measure  $P_t$  as a mixture of  $Z$  with mixing proportion  $1 - c(t)$  and a Dirac measure concentrated at a singleton  $1/t$ , that is

$$\text{for } B \subset \mathbb{R} \text{ Borel measurable} \quad \delta_{1/t}(B) = \begin{cases} 1 & \text{if } 1/t \in B, \\ 0 & \text{if } 1/t \notin B, \end{cases} \quad (11)$$

with mixing proportion  $c(t)$ . Notice that the properties of  $c$  entail the weak continuity of the collection of measures  $\{P_t: t \in [0, 1]\} \subset \mathcal{P}(\mathbb{R})$ .

To define a single functional probability  $P$  by means of these marginals  $P_t$ , consider first a function

$$F(x, t) = c(t)\mathbb{I}[x \geq 1/t] + \mathbb{I}[x > 0](1 - c(t))(1 - e^{-x}) \text{ for } x \in \mathbb{R}, t \in [0, 1]. \quad (12)$$

For each  $t \in [0, 1]$ ,  $F(\cdot, t): \mathbb{R} \rightarrow [0, 1]$  constitutes the distribution function of the marginal measure  $P_t$ .

Following the idea of Blumenthal and Corson (1972), consider a probability space  $\Omega = (0, 1)$  equipped with a  $\sigma$ -algebra of its Borel subsets and  $P$  the Lebesgue measure on  $\Omega$ . For  $\omega \in \Omega$ , define a function  $X(t, \cdot)$ , at  $t \in [0, 1]$ , to be the  $\omega$ th quantile of the distribution function  $F(\cdot, t)$ , that is

$$X(t, \omega) = \inf \{x \in \mathbb{R}: F(t, x) \geq \omega\} \text{ for } t \in [0, 1], \omega \in \Omega.$$

Then it is obvious that  $X$  defines a real-valued random process on  $[0, 1]$  with continuous trajectories for all  $\omega \in \Omega$ . These are even possible to be evaluated; using the notation

$$\begin{aligned} a(t) &= \lim_{x \rightarrow \frac{1}{t}^-} F(x, t) = (1 - c(t))(1 - e^{-\frac{1}{t}}), \\ b(t) &= F\left(\frac{1}{t}, t\right) = c(t) + (1 - c(t))(1 - e^{-\frac{1}{t}}), \end{aligned}$$

it is not hard to verify that

$$X(t, \omega) = \begin{cases} \log\left(\frac{1-c(t)}{1-c(t)-\omega}\right) & \text{for } \omega \in (0, a(t)), \\ \frac{1}{t} & \text{for } \omega \in [a(t), b(t)], \\ \log\left(\frac{1-c(t)}{1-\omega}\right) & \text{for } \omega \in (b(t), 1), \end{cases} \quad (13)$$

being a continuous function in  $t \in [0, 1]$  for all  $\omega \in \Omega$ , depicted in Figure 2(b) (for  $c(t) = t/2$ ). The functions  $a(t)$  and  $b(t)$  are plotted in Figure 2(a). Thus, we can define  $P \in \mathcal{P}(C([0, 1]))$  as the distribution of the random function  $X$ .

In this example, we show that for the choice

$$c(t) = \frac{t}{2} \text{ for } t \in [0, 1], \quad (14)$$

satisfying the conditions as imposed above, the condition of the continuity of the coordinate-wise expectation (8) for  $m = 1$  is not satisfied, and at the same time neither of the three parts of Theorem 2 is true. This will illustrate the fact that the coordinate-wise moment continuity condition (8) is an important ingredient in the theorem and cannot be dropped easily.

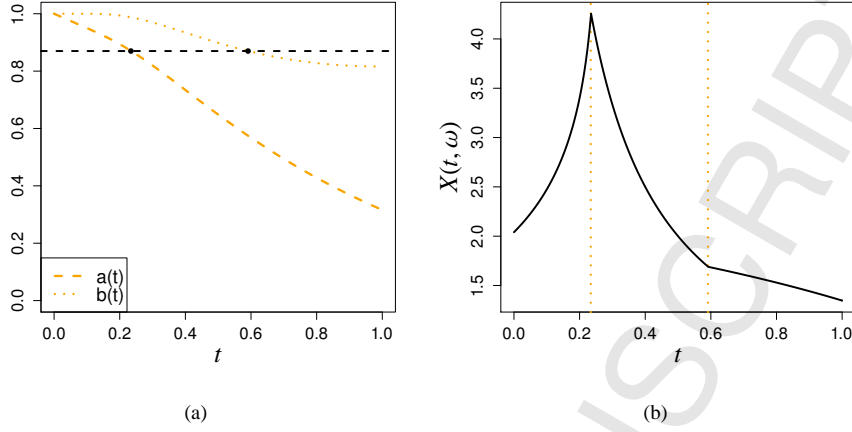


Figure 2: Illustration Example 1. (a) Functions  $a(t)$  and  $b(t)$  dividing the unit square into three regions. These regions determine points in  $[0, 1]$  at which a random function corresponding to  $\omega \in (0, 1)$  (vertical axis) switches branches of the formula for functional value (13). The horizontal line stands for the value  $\omega = 0.86$ . (b) A (completely observed) random function  $X(t, 0.86)$ . The dotted vertical lines represent the points in  $[0, 1]$  at which branches of expression (13) are switched.

Compute first the coordinate-wise expectation function of measure  $P$ ,  $X \sim P$ , corresponding to function  $c$  in (14). Evidently

$$E X(t) = E |X(t)| = \begin{cases} 1 & \text{for } t = 0, \\ \frac{3-t}{2} & \text{for } t \in (0, 1], \end{cases}$$

is a function with a discontinuity at  $t = 0$ .

Sample now an independent sequence  $\{X_n\}_{n=1}^{\infty}$  from  $P$  and define its approximating sequence of functions  $\{\tilde{X}_n\}_{n=1}^{\infty}$  in observation points (10), see Figure 3.

To see that the assertion of part (i) of Theorem 2 is not satisfied for  $P$ , take the point  $t = 0$  at which (8) is violated. Here, obviously, for any  $\omega \in \Omega$ ,

$$\tilde{X}_n(0) = \tilde{X}_n\left(\frac{1}{n}\right) = X_n\left(\frac{1}{n}\right),$$

implying that the assertion of part (i) of Theorem 2 is no longer true since

$$E \tilde{X}_n(0) = E |\tilde{X}_n(0)| = \frac{3n-1}{2n} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{3}{2} > 1 = E |X(0)| = E X(0).$$

Moreover, condition (9) from Theorem 2 is not satisfied for  $P$  as well, since

$$E X^2(t) = \begin{cases} 2 & \text{for } t = 0, \\ 2 - t + \frac{1}{2t} & \text{for } t \in (0, 1]. \end{cases}$$

Thus, we can explore whether the conclusions of parts (ii) and (iii) can be possibly recovered.

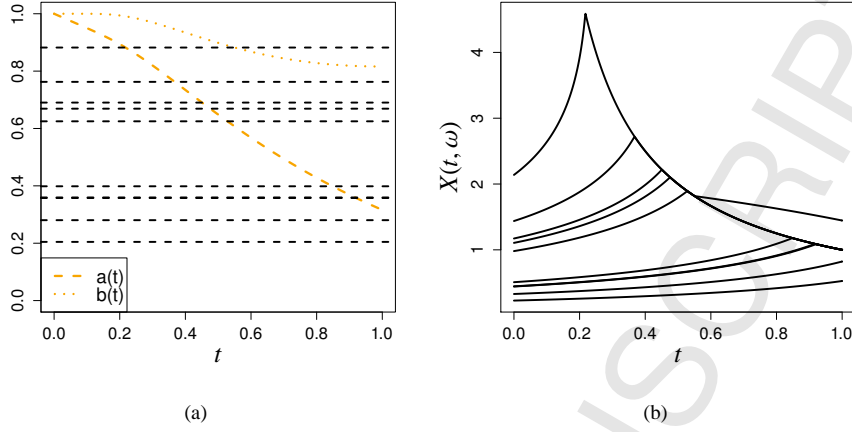


Figure 3: Illustration Example 1. (a) A random sample of elements of  $\Omega$  (horizontal lines) of size 10. (b) The corresponding functional random sample (of completely observed curves) from  $P$ . Here, the central decreasing function to which most of the sample functions stick to in some interval is  $x(t) = 1/t$ , see (13).

To this end, compute for  $n \in \mathbb{N}$  the probability

$$\begin{aligned} \mathbb{P}(\tilde{X}_n(0) \geq n) &= \mathbb{P}\left(X_n\left(\frac{1}{n}\right) \geq n\right) \\ &= c\left(\frac{1}{n}\right) \mathbb{P}(\delta_n \geq n) + \left(1 - c\left(\frac{1}{n}\right)\right) \mathbb{P}(Z \geq n) \geq c\left(\frac{1}{n}\right) \end{aligned} \quad (15)$$

where the second equality follows from (12), and the definitions of the Dirac measure (11) and the random variable  $Z$ , yielding

$$\sum_{n=1}^{\infty} \mathbb{P}(\tilde{X}_n(0) \geq n) \geq \sum_{n=1}^{\infty} \frac{1}{2n} = \infty.$$

Having this, and applying the Borel-Cantelli's lemma shows that the average

$$\frac{1}{n} \sum_{i=1}^n \tilde{X}_i(0) = \frac{1}{n} \sum_{i=1}^n |\tilde{X}_i(0)|$$

is not convergent in  $\mathbb{R}$  as  $n \rightarrow \infty$  for almost all values  $\omega \in \Omega$ , providing an example of a distribution where parts (ii) and (iii) of Theorem 2 do not hold true.

**Example 2.** Building further on Example 1 we now show that even if the condition of coordinate-wise expectation continuity (8) is satisfied, but the boundedness of higher moment condition (9) is not, the arithmetic average of functional values of approximating functions need not converge to the value of the expectation almost surely.

Consider again the setup as in Example 1 with the only difference that the function  $c$  determining the mixing proportions in the definition of  $P_t$  is now taken to be

$$c(t) = \begin{cases} 0 & \text{for } t = 0, \\ -\frac{t}{2 \log(\frac{t}{2})} & \text{for } t \in (0, 1]. \end{cases}$$

This function is, again, continuous on  $[0, 1]$ , maps into  $[0, 1]$ , and  $c(0) = 0$ . Therefore, we are

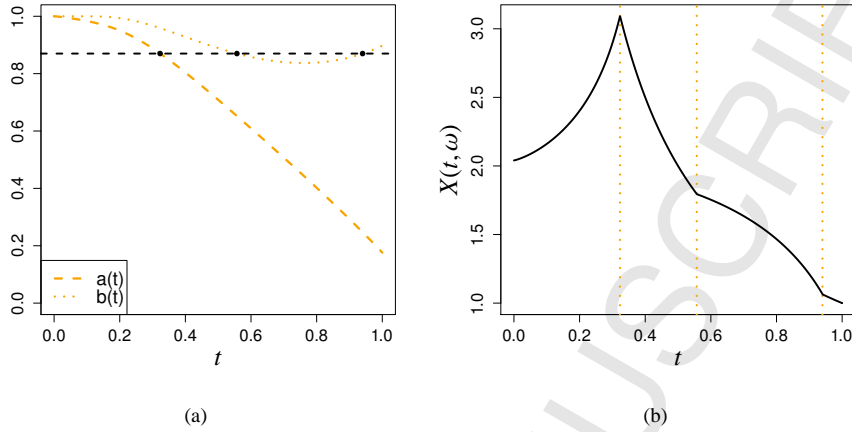


Figure 4: Illustration Example 2. (a) The functions  $a(t)$  and  $b(t)$  dividing the unit square into three regions. These regions determine points in  $[0, 1]$  at which a random function corresponding to  $\omega \in (0, 1)$  (vertical axis) switches branches of the formula for functional value (13). The horizontal line stands for the value  $\omega = 0.86$ . (b) A (completely observed) random function  $X(t, 0.86)$ , with the dotted vertical lines representing the points in  $[0, 1]$  at which branches of expression (13) are switched.

allowed to define  $P \in \mathcal{P}(C([0, 1]))$  and a random function  $X$  by (13) just as in Example 1. In Figure 4 we plot the functions  $a(t)$  and  $b(t)$  together with  $X(t, \omega)$  for  $\omega = 0.86$ . For  $m = 1$  its coordinate-wise expectation is a continuous function, that is (8) holds true, since

$$E X(t) = E |X(t)| = \begin{cases} 1 & \text{for } t = 0, \\ 1 + \frac{t-1}{2 \log(\frac{1}{2})} & \text{for } t \in (0, 1]. \end{cases}$$

On the other hand, for the second coordinate-wise moment we see that

$$E X^2(t) = \begin{cases} 1 & \text{for } t = 0, \\ 1 + \left(t - \frac{1}{t}\right) \frac{1}{2 \log(\frac{1}{2})} & \text{for } t \in (0, 1], \end{cases}$$

and (9) is not satisfied for  $P$  because of its behaviour in the neighbourhood of  $t = 0$ .

To show that the strong law of large numbers does not hold true at  $t = 0$  as written in Theorem 2, notice that by (15) for  $n \in \mathbb{N}$  (and  $\tilde{X}_n$  as in Example 1)

$$\sum_{n=1}^{\infty} P(\tilde{X}_n(0) \geq n) \geq \sum_{n=1}^{\infty} \frac{1}{2n \log(2n)} = \infty,$$

concluding that the condition of boundedness (9) is important, and cannot in general be replaced by the weaker condition (8).

The following example illustrates the problems we may encounter if random observation points, dependent on the random sample functions, are allowed.

**Example 3.** For  $K = 1$ , consider two sequences of points in  $[0, 1]$

$$\begin{aligned} t_j &= 1 - 2^{-j} & \text{for } j = 0, 1, \dots, \\ s_j &= 1 - 3 \times 2^{-(j+1)} & \text{for } j = 1, 2, \dots, \end{aligned}$$

and a sequence of continuous, piecewise linear functions

$$\text{for } j \in \mathbb{N}, \quad x_j(t) = \begin{cases} j \frac{t-t_{j-1}}{s_j-t_{j-1}} & \text{for } t \in (t_{j-1}, s_j), \\ j \frac{t-t_j}{s_j-t_j} & \text{for } t \in [s_j, t_j), \\ 0 & \text{otherwise.} \end{cases}$$

The first five functions of the sequence are depicted in Figure 5.

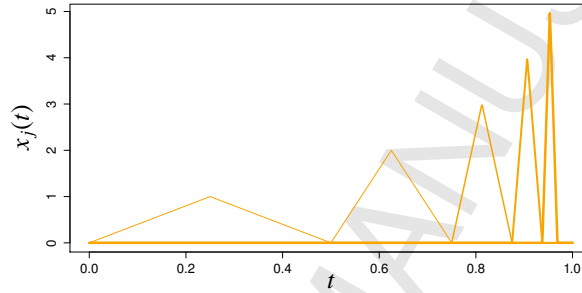


Figure 5: Illustration Example 3. First five functions of the sequence  $\{x_j\}_{j=1}^{\infty} \subset C([0, 1])$  on which the random function  $X$  is supported.

Define a random function  $X \in C([0, 1])$  as

$$\mathbb{P}(X \equiv x_j) = \frac{6}{(\pi j)^2} \quad \text{for } j \in \mathbb{N},$$

and denote its distribution by  $\mathbb{P} \in \mathcal{P}(C([0, 1]))$ . Then it is easy to verify that

$$\begin{aligned} \text{for } j \in \mathbb{N}, \quad \mathbb{E} X(t) &= \begin{cases} \frac{6}{\pi^2 j} \frac{t-t_{j-1}}{s_j-t_{j-1}} & \text{for } t \in [t_{j-1}, s_j), \\ \frac{6}{\pi^2 j} \frac{t-t_j}{s_j-t_j} & \text{for } t \in [s_j, t_j), \\ \mathbb{E} X(1) = 0, \end{cases} \end{aligned}$$

and the continuity of the coordinate-wise expectation (8) for  $m = 1$  holds true for  $\mathbb{P}$ .

Consider a random triangular array of observation points  $\{T_{j,n}\}_{j=1}^n$  (i.e.  $\{m_n\}_{n=1}^{\infty} = \{n\}_{n=1}^{\infty}$  is fixed and non-random) defined so that **(D)** holds true, and for the last observation point in each row of the array

$$T_{n,n} = T_{n,n}(\omega) = \max \{s_j, t_n\} \text{ if } X_n = X_n(\omega) \equiv x_j,$$

where  $X_1, X_2, \dots$  is a sequence of independent functions distributed as  $X$ . Condition **(D)** is indeed easy to be met, as  $T_{n,n} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 1$ .

Then, by the construction of the approximating functions  $\widetilde{X}_n$ ,  $n \in \mathbb{N}$ , we can write for the right endpoint of the domain ( $t = 1$ )

$$\widetilde{X}_n(1) = \widetilde{X}_n(T_{n,n}) = \begin{cases} 0 & \text{if } X_n \equiv x_j \text{ for } j \leq n, \\ j & \text{if } X_n \equiv x_j \text{ for } j > n, \end{cases}$$

or

$$\widetilde{X}_n(1) = \begin{cases} 0 & \text{with probability } \sum_{j=1}^n \frac{6}{(\pi j)^2}, \\ j & \text{with probability } \frac{6}{(\pi j)^2} \text{ for } j > n. \end{cases}$$

Consequently,

$$\mathbb{E} \widetilde{X}_n(1) = \sum_{j=n+1}^{\infty} \frac{6}{\pi^2 j} = \infty,$$

and the assertion (i) of Theorem 2 is not valid for  $t = 1$  and  $m = 1$ .

#### 4. Application to Functional Data Depth

In the recent literature a great deal of attention has focused on depth functions applicable to complex data sets, in particular to functional data. These are pursued as infinite-dimensional generalizations of better studied statistical depth functions for finite-dimensional multivariate data; see Zuo and Serfling (2000) for a comprehensive theoretical review and Liu et al. (1999) for applications.

In general, data depth is a robust device applicable in nonparametric statistics to complex data. To a given point in the sample space and a probability distribution on this space, it assigns a non-negative number that can be interpreted as a measure of “centrality” of this point with respect to the probability distribution.

Among the multitude of depth proposals for  $\mathbb{R}^K$ -valued data that can be found in the literature, two often used notions of depths are the halfspace depth and the simplicial depth. As we will utilize these depths later in examples, we recall how they are defined.

Given a point  $x \in \mathbb{R}^K$  and a probability  $P \in \mathcal{P}(\mathbb{R}^K)$ , the halfspace depth of  $x$  with respect to  $P$  introduced by Tukey (1975) is defined as

$$hD(x; P) = \inf_{H \in \mathcal{H}(x)} P(H), \quad (16)$$

where  $\mathcal{H}(x)$  denotes the collection of closed halfspaces in  $\mathbb{R}^K$  containing  $x$ . Similarly, the simplicial depth of  $x$  with respect to  $P$  (Liu, 1990) is defined as

$$sD(x; P) = P(x \in S(X_1, \dots, X_{K+1})). \quad (17)$$

Here,  $X_1, \dots, X_{K+1}$  is a random sample from  $P$  and  $S(\cdot)$  stands for a simplex (closed convex hull) in  $\mathbb{R}^K$  whose vertices are formed by the arguments of the mapping.

As far as functional data are concerned, since the pioneering proposal of depth function tailored especially for real-valued functions of Fraiman and Muniz (2001), a vast quantity of papers introducing new depths for functions has been emerging. To name a few of these: the band depth and modified band depth of López-Pintado and Romo (2009), the halfregion depth and modified halfregion depth of López-Pintado and Romo (2011), the  $\Phi$ -depth of Mosler and

Polyakova (2012) and the multivariate functional depth of Claeskens et al. (2014). The realm of depth functionals on  $C^K([0, 1])$  can be rather accurately classified into two groups of depths sharing intrinsic similarities. Assume that  $x \in C^K([0, 1])$  and  $P \in \mathcal{P}(C^K([0, 1]))$  is the distribution with respect to which we want to compute the depth of  $x$ . Then it can be computed by

- **Integrated depths** having a general form of an integral (see Nagy et al., 2014), in the simplest case

$$D(x; P) = \int_0^1 D_K(x(t); P_t) dt, \quad (18)$$

where  $D_K$  is a generic depth on  $\mathbb{R}^K$  ((16) or (17), for example), or

- **Infimal depths** relying more on the graphical representation of the functions, taking the form of an infimum, see Mosler and Polyakova (2012) or Gijbels and Nagy (2014). An elementary representative of infimal depths is the basic infimal depth (Gijbels and Nagy, 2014, Section 5) defined as

$$ID(x; P) = \inf_{t \in [0, 1]} D_1(x(t); P_t), \quad (19)$$

where  $D_1$  is the univariate depth<sup>1</sup> assigning to  $u \in \mathbb{R}$  and  $Q \in \mathcal{P}(\mathbb{R})$  with a distribution function  $F_Q$  the value

$$D_1(u; Q) = 1 - |0.5 - F_Q(u)|.$$

From the aforementioned instances, the depth of Fraiman and Muniz (2001), the modified versions of both band depth and halfregion depth, and the multivariate functional depth, all fall logically into the group of integrated depths, whereas the original versions of band depth, halfregion depth and  $\Phi$ -depths are infimal type depth functionals.

The majority of theoretical results on depths for functional data available in the literature concerns the consistency in the complete design of observation of functional values. The consistency results for the integrated depth of Fraiman and Muniz (2001), band depths, halfregion depths, basic infimal depth as an instance of  $\Phi$ -depths (Gijbels and Nagy, 2014), and general integrated depths (Nagy et al., 2014) are all for complete design of random functions. The only exception, as far as we know, is a recent contribution of Claeskens et al. (2014), where a simplified scenario of the (general) discrete design as described in Section 2 is pursued for multivariate functional depth. See Nagy et al. (2014, Appendix B.2) for further discussion.

To utilize the weak convergence result from Section 2 in the framework of depth for functions, one seemingly strong assumption needs to be made for a general depth  $D$  on the space  $C^K([0, 1])$ .

**Definition.** We say that a depth functional  $D$  is weakly continuous at  $P \in \mathcal{P}(C^K([0, 1]))$ , if for any sequence  $\{P_\nu\}_{\nu=1}^\infty \subset \mathcal{P}(C^K([0, 1]))$ ,  $P_\nu \xrightarrow[\nu \rightarrow \infty]{w} P$  it holds true that

$$\sup_{x \in C^K([0, 1])} |D(x; P) - D(x; P_\nu)| \xrightarrow[\nu \rightarrow \infty]{} 0.$$

<sup>1</sup>This one-dimensional depth was introduced in functional data analysis by Fraiman and Muniz (2001). It is easy to see that  $D_1$  and the halfspace depth (16) for  $K = 1$  coincide if  $F_Q$  is continuous. The advantage of using  $D_1$  instead of  $hD$  here is mere notational convenience.



To obtain relevant information about the population version of depth  $D$  with respect to  $P$ , one usually uses the sample depth  $D$  with respect to  $P_n$  defined in (1), the empirical measure based on completely observed functional data. If, however, these complete functions are not available, one has to resort to their approximations, such as functions  $\tilde{X}_n$ , and replace the measures  $P_n$  with  $\tilde{P}_n$  defined in (5). The following theorem states that, if the functional depth  $D$  is weakly continuous at  $P$ , this procedure can be justified, and the sample version of  $D$  based on the reconstructed curves provides a good approximation of the population version of  $D$ . It is a direct consequence of Theorem 1.

**Theorem 3.** *Let  $D$  be weakly continuous at  $P$  and let the assumptions of Theorem 1 be satisfied. Then*

$$\sup_{x \in C^K([0,1])} \left| D(x; P) - D(x; \tilde{P}_n) \right| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0.$$

Although the weak continuity condition may seem to be rather strict, it is obeyed by most of the depths for functions renown from the literature. Namely, if for  $P$

$$P_t(L) = 0 \text{ for all hyperplanes } L \subset \mathbb{R}^K \text{ and } t \in [0, 1], \quad (20)$$

ensured by the absolute continuity of  $P_t$  for all  $t \in [0, 1]$ , is true, then the conditions of Theorem 3 are satisfied for the following types of depth for functional data:

- for any  $K \in \mathbb{N}$ 
  - integrated depth (18) based on halfspace depth  $hD$ ,
  - integrated depth (18) based on simplicial depth  $sD$ ,
- for  $K = 1$ 
  - depth for functional data of Fraiman and Muniz (2001),
  - modified band depth of arbitrary order  $J = 2, 3, \dots$  (López-Pintado and Romo, 2009, Section 5),
  - modified halfregion depth (López-Pintado and Romo, 2011, Section 5),
  - basic infimal depth (19) (see also Mosler and Polyakova (2012)).

As all these depths for functions, apart from the basic infimal depth, are members of a general class of integrated depths, the assertions for these follow from the results of Nagy et al. (2014, Section 4.6) and the results of Appendix A therein. As for the basic infimal depth for functions, Gijbels and Nagy (2014, Theorem 6) provide the affirmation.

If, however, condition (20) is not satisfied for the marginals of  $P$ , the assertion of Theorem 3 can no longer be recovered in full extent. To see this, we consider the following elementary example.

**Example 4.** For  $K = 1$ , assume that  $P \in \mathcal{P}(C([0, 1]))$  is defined as the Dirac measure of a single function

$$f(t) = t - t^2 \text{ for } t \in [0, 1].$$

Then each one-dimensional marginal distribution  $P_t$  is a Dirac measure of a singleton  $f(t)$  in  $\mathbb{R}$ . Thus, (20) is not satisfied for  $P$ , and for both  $hD$  and  $sD$

$$hD(u; P_t) = \begin{cases} 0.5 & \text{for } u = f(t), \\ 0 & \text{otherwise,} \end{cases} \quad sD(u; P_t) = \begin{cases} 1 & \text{for } u = f(t), \\ 0 & \text{otherwise.} \end{cases}$$

For the integrated depth of a general function  $x \in C([0, 1])$  based either on  $hD$  or  $sD$  we have

$$\int_0^1 hD(x(t); P_t) dt = 0.5\lambda(\{t: x(t) = f(t)\}), \quad \int_0^1 sD(x(t); P_t) dt = \lambda(\{t: x(t) = f(t)\}),$$

where  $\lambda$  stands for the Lebesgue measure on  $[0, 1]$ , meaning for the choice  $x \equiv f$

$$\int_0^1 hD(f(t); P_t) dt = 0.5, \quad \int_0^1 sD(f(t); P_t) dt = 1. \quad (21)$$

On the other hand, for any triangular array  $\{T_{j,n}\}_{j=1}^{m_n}$ ,  $n \in \mathbb{N}$ , of points at which random functions from  $P$  are observed, and any random sample  $X_1, \dots, X_n$  from  $P$  we have

$$\tilde{X}_i(t) \begin{cases} = f(t) & \text{for } t = T_{j,i}, j = 1, \dots, m_i, \\ \neq f(t) & \text{otherwise,} \end{cases}$$

yielding for the sample integrated depth based on discretely observed curves  $\tilde{X}_1, \dots, \tilde{X}_n$

$$\int_0^1 hD(f(t); \tilde{P}_{n,t}) dt = 0, \quad \int_0^1 sD(f(t); \tilde{P}_{n,t}) dt = 0, \quad \text{for any } n \in \mathbb{N} \text{ a.s.} \quad (22)$$

See Figure 6 for an illustration associated with this example.

Comparing (21) and (22) it is seen that in this case the sample integrated depth based on both  $hD$  and  $sD$  is not a consistent estimator of its population counterpart.

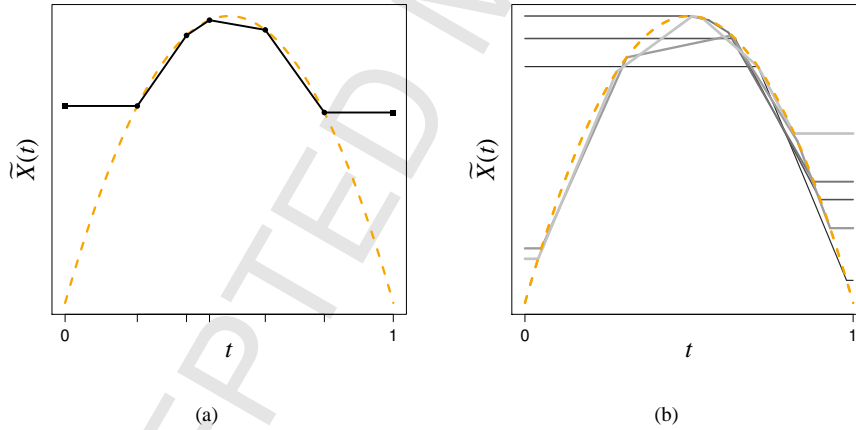


Figure 6: Illustration Example 4. (a) The function  $f$  (in dashed line) at which the measure  $P$  is concentrated with its approximation (in solid line) based on  $m = 5$  randomly distributed discrete observations. (b) A random sample of  $n = 5$  approximating functions following a sampling process from  $P$  with increasing number of observation points.

It may seem that the problem in Example 4 is caused by the upper semicontinuity of univariate depths  $hD$  and  $sD$ ; namely that a single point can be the only one with a positive value of depth. Nevertheless, a similar counterexample can be elaborated also for lower semicontinuous

depths, showing that the essence of Example 4 does not lie in the argument of semicontinuity. See Section S.3 in the Supplementary Material.

In Gijbels and Nagy (2014), a number of counterexamples was constructed to show that neither of the three established infimal (non-integrated) depth functionals, namely band depth, halfregion depth, and basic infimal depth (or more general  $\Phi$ -depth), is universally consistent in the usual, complete design sense. In all these examples measures on  $C([0, 1])$  were constructed so that  $P_t$  was absolutely continuous on  $\mathbb{R}$ , unless  $t = 1$ .

Let us recall, for instance, Example 4 from Gijbels and Nagy (2014) presenting a probability measure  $P \in \mathcal{P}(C([0, 1]))$ ,  $X \sim P$ , with respect to which the basic infimal depth as a representative of  $\Phi$ -depths is not consistent under complete design sampling.

In the remaining part of the paper we expand that example substantially to show that even under discrete design the sample basic infimal depth does not consistently estimate its population counterpart. This resolves a principal question that suggests itself when contemplating all the counterexamples presented in Gijbels and Nagy (2014). Therein, the conclusions of non-consistency claims are drawn upon inspecting examples where the probability measure  $P$  misbehaves only in small sets of neighbourhoods of singletons. It might be of interest to see if this pathological behaviour persists also when moving towards discretely observed functions. Then, any odd local behaviour of functions might have been “smoothed out” of the random sample by considering only approximating functions  $\tilde{X}_n$  that act reasonably predictably, being simple piecewise linear continuous curves. What is shown in Example 5 below is that this is not the case. In summary, all these three non-integrated depths can be revealed to be not universally consistent even for discrete design scenario.

**Example 5.** In Gijbels and Nagy (2014, Example 4), the following probability measure  $P \in \mathcal{P}(C([0, 1]))$  was constructed to show the (complete design) non-consistency of the infimal depth  $ID$  for  $K = 1$ .

Initially, partition the interval  $[0, 1]$  into infinitely many disjoint subintervals by setting

$$t_0 = 0, \quad t_j = \sum_{i=1}^j \frac{1}{2^i} \text{ for } j = 1, 2, \dots \quad (23)$$

For an infinite sequence of independent random variables  $\{V_j\}_{j=0}^{\infty}$ ,  $V_j$  being uniformly distributed on the closed interval  $[0, \frac{1}{2^j}]$ , set  $X(t_j) = V_j$  for  $j = 0, 1, \dots$ , and  $X(1) = 0$ . On each subinterval  $(t_{j-1}, t_j)$  for  $j = 1, 2, \dots$  define  $X$  to be a linear function interpolating the adjacent points  $X(t_{j-1})$  and  $X(t_j)$ . One realization of such a function is depicted in Figure 7(a).

Then almost surely each random function  $X$  is piecewise linear, uniformly bounded and Lipschitz continuous with constant 2. It is also easy to see that for each  $t \in [0, 1)$  the marginal measure  $P_t$  is absolutely continuous with a bounded density supported in the set  $[0, 1 - t]$ .

However, notice that  $P$  satisfies condition (20) for  $t$  on any subset  $[0, 1 - \varepsilon]$  for  $\varepsilon > 0$ , but not on  $[0, 1]$ .

To show that the basic infimal depth is not consistent for  $P$ , consider the fixed deterministic function

$$x(t) = \frac{1-t}{2} \text{ for } t \in [0, 1]. \quad (24)$$

At  $t_j$  function  $x$  takes the value of the expectation (and also median) of the random variable  $V_j$  for each  $j \geq 1$ , and  $x$  is zero at  $t = 1$ .

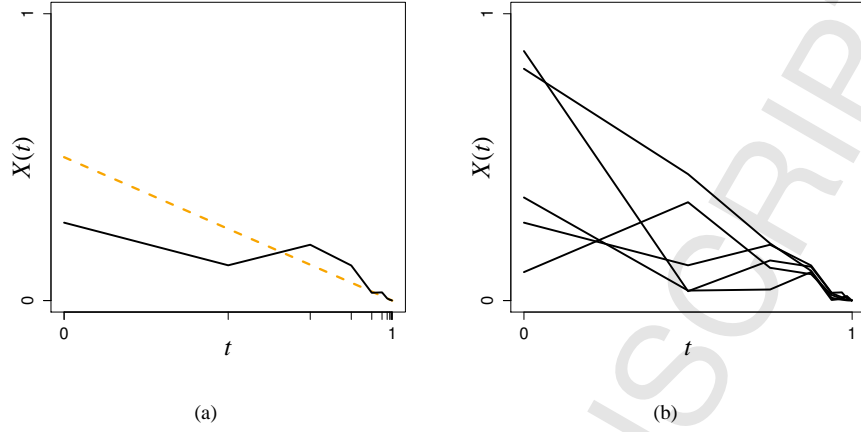


Figure 7: Illustration Example 5. (a) A single (completely observed) random function  $X$  from  $P$  in solid line, and the function  $x$  at which consistency of integrated depth fails to hold in dashed line. The ticks on the horizontal axis represent points  $t_j$  as in (23). (b) A random sample (of size  $n = 5$ ) of (completely observed) functions generated from  $P$ .

As discussed in Gijbels and Nagy (2014, Example 4), for the population basic infimal depth we can write<sup>2</sup>

$$ID(x; P) = \inf_{t \in [0,1)} D_1(x(t); P_t) \geq 0.75,$$

but for the sample depth it is possible to show that

$$ID(x; P_n) = \inf_{t \in [0,1)} D_1(x(t); P_{n,t}) = 0.5 \text{ for all } n \text{ a.s.}, \quad (25)$$

and the basic infimal depth is not consistent at  $x$  in the sense of complete design sampling.

On the contrary, using the complete design sampling consistency result for basic infimal depth (Gijbels and Nagy, 2014, Theorem 5), restricting ourselves to any domain  $[0, 1 - \varepsilon]$  for  $\varepsilon > 0$  bounded away from  $t = 1$ , it is possible to guarantee the consistency

$$\sup_{x \in C([0,1))} \left| \inf_{t \in [0,1-\varepsilon]} D_1(x(t); P_t) - \inf_{t \in [0,1-\varepsilon]} D_1(x(t); P_{n,t}) \right| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0.$$

Thus, it is evident that the sequence of points  $\{t_n\}_{n=1}^\infty \subset [0, 1]$  at which the consistency is violated for various sample sizes  $n$  must necessarily tend to 1. Taking this into account, it is possible to argue that the same problem needs not occur when sampling random functions within the discrete design framework. This would be very beneficiary indeed, because in the more likely setup of discretely observed functions the common major difficulty of all non-integrated depths — lack of consistency — would possibly not be an issue. As we show further, however, this is not true, as the same problem persists also when using the discrete design sampling scheme.

<sup>2</sup>Excluding the point of degeneracy  $t = 1$  since for the discrete distribution  $P_1$  the univariate depth  $D_1$  exhibits anomalous behaviour, see also Gijbels and Nagy (2014, Example 4).

To show so, consider first an infinite sequence of positive integers  $\{a_l\}_{l=1}^{\infty}$  defined recursively

$$a_1 = 1, \quad a_l = l \sum_{k=1}^{l-1} a_k \text{ for } l = 2, 3, \dots$$

This rapidly growing sequence has a peculiar asymptotic property that each consequent term is proportionally as large as the whole partial sum of the sequence up to this term. To be precise, it can be written as

$$\frac{a_l}{\sum_{k=1}^l a_k} = \frac{l \sum_{k=1}^{l-1} a_k}{l \sum_{k=1}^{l-1} a_k + \sum_{k=1}^{l-1} a_k} = \frac{l}{l+1} \xrightarrow{l \rightarrow \infty} 1. \quad (26)$$

To get an idea how fast the sequence grows, the first 10 terms of it are shown in the second column of Table 1. In column 3 of the table we list some values of  $\sum_{k=1}^{l-1} a_k$ .

Table 1: First few terms of the sequences  $a_l$ ,  $i_l$  and  $j_l$  from Example 5. The  $l$ th batch of functions consists of  $a_l$  consecutive functions of random sample starting at the index  $i_l + 1$ . The functions from the  $l$ th batch are observed (among others) at  $2^{a_l}$  consecutive points  $t_j$  starting at index  $j = j_l + 1$ .

$l$	$a_l$	$i_l = \sum_{k=1}^{l-1} a_k$	$j_l = \sum_{k=1}^{l-1} 2^{a_k}$
1	1	0	0
2	2	1	2
3	9	3	6
4	48	12	518
5	300	60	$\approx 2.81 \times 10^{14}$
6	2 160	360	$\approx 2.04 \times 10^{90}$
7	17 640	2 520	$\approx 1.68 \times 10^{650}$
8	161 280	20 160	$\approx 5.77 \times 10^{5307}$
9	1 632 960	181 440	$\approx 1.31 \times 10^{48550}$
10	18 144 000	1 814 400	$\approx 1.04 \times 10^{491358}$

Exploiting this sequence, we preset a fixed triangular array of observation points as follows. Assign to the  $n$ th unobserved (complete) random sample function  $X_n \in C([0, 1])$  a unique batch label  $l \in \mathbb{N}$  so that the number  $l$  satisfies the inequality

$$\sum_{k=1}^{l-1} a_k < n \leq \sum_{k=1}^l a_k.$$

The  $l$ th batch then consists of  $a_l$  consecutive functions from the random sample. Similarly, group the terms of the sequence  $\{t_j\}_{j=0}^{\infty} \subset [0, 1]$  from (23) so that in the  $l$ th group,  $l \in \mathbb{N}$ , will be those indices  $j$  for which

$$\sum_{k=1}^{l-1} 2^{a_k} < j \leq \sum_{k=1}^l 2^{a_k}.$$

Suppose that all functions from the  $l$ th batch are observed at the same vector of observation points. This vector always consists of

- $t = 0$  and  $t = 1$ ,

- an equidistant sequence of points  $\{k \frac{t_{j_l}}{l}\}_{k=1}^l$ , where  $j_l = \sum_{k=1}^{l-1} 2^{a_k}$ ,
- and the terms belonging to the  $l$ th group of the sequence  $\{t_j\}_{j=0}^\infty$ .

The values of  $j_l$  for the first batches are given in column 4 of Table 1. For clarity Table 2 gives the part of the triangular array of observation points for the first few functions from the random sample. In Figure 8 plots of functions from the first three batches are provided. It is evident from the construction that this triangular array enjoys the condition of getting denser **(D)**.

Table 2: The distribution of points at which first 60 random sample functions from Example 5 are observed. With increasing batch number the norm of the partition induced by these observation points vanishes, that is **(D)** holds true.

batch $l$	r.s. functions $n$	observation points $\{T_{j,n}\}_{j=1}^{m_n}$
1	1	$(0, t_1, t_2, 1)^\top$
2	2, 3	$(0, 1 \frac{t_2}{2}, 2 \frac{t_2}{2}, t_3, \dots, t_6, 1)^\top$
3	4, 5, \dots, 12	$(0, 1 \frac{t_6}{3}, 2 \frac{t_6}{3}, 3 \frac{t_6}{3}, t_7, t_8, \dots, t_{518}, 1)^\top$
4	13, 14, \dots, 60	$(0, 1 \frac{t_{518}}{4}, 2 \frac{t_{518}}{4}, 3 \frac{t_{518}}{4}, 4 \frac{t_{518}}{4}, t_{519}, t_{520}, \dots, t_{j_s}, 1)^\top$

Let us now explore the depth of the function  $x$  defined in (24) with respect to the empirical measure  $\tilde{P}_n$ . This is based on the sequence of functions  $\tilde{X}_1, \dots, \tilde{X}_n$  approximating a random sample  $X_1, \dots, X_n$  from  $P$  on a triangular array  $\{T_{j,n}\}_{j=1}^{m_n}$  as described in detail in Section 2. To see that (25) still holds true even when  $P_n$  is replaced by  $\tilde{P}_n$ , we show that for almost all  $\omega \in \Omega$

$$\liminf_{n \rightarrow \infty} \inf_{t \in [0,1]} D_1(x(t); \tilde{P}_{n,t}(\omega)) = 0.5. \quad (27)$$

This is enough to prove the asserted lack of consistency in the discrete design for basic infimal depth, just as it was in the complete design case.

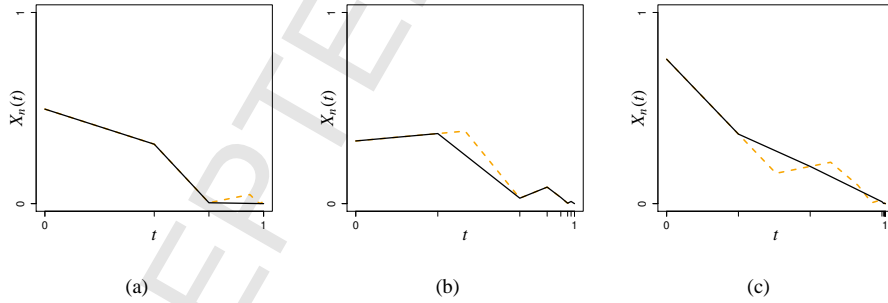


Figure 8: Discrete design sampling from Example 5: a single function  $\tilde{X}_n$  (solid line) and the function it approximates  $X_n$  (dashed line) generated from  $P$  for  $n$  in batches 1 (a), 2 (b) and 3 (b), respectively. The ticks on the horizontal axis represent points of the domain on which the approximations  $\tilde{X}_n$  are based.

To show (27), for  $n, j \in \mathbb{N}$  designate  $A_{n,j} = \{\omega \in \Omega : \tilde{X}_n(t_j) \leq x(t_j)\} \subset \Omega$ . Notice first that by the construction of the array of observation points, if  $n$ th random sample function comes from

the  $l$ th batch, then  $\tilde{X}_n(t_j) = X_n(t_j)$  for all  $j = j_l + 1, \dots, j_{l+1}$ . Therefore, for arbitrary  $n$  and  $X_n$  in the  $l$ th batch, the sequence of random events  $A_{n,j}$  for  $j = j_l + 1, \dots, j_{l+1}$  is independent of each other and  $P(A_{n,j}) = 0.5$ . Since  $\tilde{X}_1, \dots, \tilde{X}_n$  are independent of each other, the latter observation yields that the sequence of random events

$$\{A_{n,j}: l \in \mathbb{N}, X_n \text{ belongs to the } l\text{th batch and } j = j_l + 1, \dots, j_{l+1}\} \subset \mathcal{F}$$

is a sequence of independent events of equal probability 0.5. Evaluate now for fixed  $l \in \mathbb{N}$  the probability in the following expression

$$P\left(\bigcup_{j=j_l+1}^{j_{l+1}} \bigcap_{i: X_i \text{ is in the } l\text{th batch}} A_{i,j}\right) = \sum_{k=1}^{2^{a_l}} (-1)^{k-1} \binom{2^{a_l}}{k} 0.5^{a_l k} = 1 - \left(1 - \frac{1}{2^{a_l}}\right)^{2^{a_l}}.$$

This probability tends to  $1 - e^{-1} > 0$  as  $l \rightarrow \infty$ , hence

$$\sum_{l=1}^{\infty} P\left(\bigcup_{j=j_l+1}^{j_{l+1}} \bigcap_{i: X_i \text{ is in the } l\text{th batch}} A_{i,j}\right) = \infty, \quad (28)$$

and by Borel-Cantelli's lemma — recall the independence of the events — there almost surely exists an infinite sequence in  $l$  so that the random events in (28) occur for each  $l$  in the sequence. In other words, for almost all  $\omega \in \Omega$  in an infinite number of batches one can find a point  $t_j$  (depending on the current batch) close to 1 so that all the functions  $\tilde{X}_n$  for  $X_n$  from the batch are smaller than  $x(t_j)$  at  $t_j$ . Consequently, if  $n$  is the  $l$ th term in the sequence  $\{i_l\}_{l=1}^{\infty}$  (see Table 1) and the event above occurs for  $n$ , there exists a point  $t_{j(l)} \in [0, 1]$  so that at least the last  $a_l$  functions from  $\tilde{X}_1, \dots, \tilde{X}_n$  lie at  $t_{j(l)}$  below  $x(t_{j(l)})$ , yielding

$$\inf_{t \in [0,1]} D_1(x(t); \tilde{P}_{n,t}) \leq D_1(x(t_{j(l)}); \tilde{P}_{n,t_{j(l)}}) = 1 - \left| 0.5 - \frac{1}{n} \sum_{i=1}^n \mathbb{I}[\tilde{X}_n(t_{j(l)}) \leq x(t_{j(l)})] \right|.$$

For the right hand term in the previous display we can use (26) and bound

$$1 \geq \frac{1}{n} \sum_{i=1}^n \mathbb{I}[\tilde{X}_n(t_{j(l)}) \leq x(t_{j(l)})] \geq \frac{a_l}{\sum_{k=1}^l a_k} \xrightarrow{l \rightarrow \infty} 1$$

to see that (27) holds true indeed for almost all  $\omega \in \Omega$ .

**Acknowledgement.** The authors thank the Managing Editor, the Editor and two referees for their valuable comments. This research is supported by the IAP research network nr. P7/06 of the Federal Science Policy (Belgium). The first author is a Research Assistant of the Research Foundation - Flanders, and acknowledges support from this foundation. The second author gratefully acknowledges support from the GOA/12/014 — project of the Research Fund KU Leuven. The third author gratefully acknowledges support from GAČR 14-07234S project of Czech Science Foundation.

## References

- Aliprantis, C. D., Border, K. C., 2006. Infinite dimensional analysis. A hitchhiker's guide, 3rd Edition. Springer, Berlin.
- Aneiros, G., Vieu, P., 2014. Variable selection in infinite-dimensional problems. *Statistics & Probability Letters* 94, 12–20.  
URL <http://dx.doi.org/10.1016/j.spl.2014.06.025>
- Araujo, A., Giné, E., 1980. The central limit theorem for real and Banach valued random variables. John Wiley & Sons, New York-Chichester-Brisbane, Wiley Series in Probability and Mathematical Statistics.
- Blumenthal, R. M., Corson, H. H., 1972. On continuous collections of measures. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory. Univ. California Press, Berkeley, Calif., pp. 33–40.
- Bongiorno, E. G., Goia, A., Salinelli, E., Vieu, P. (Eds.), 2014. Contributions in infinite-dimensional statistics and related topics. Società Editrice Esculapio, Bologna.
- Claeskens, G., Hubert, M., Slaets, L., Vakili, K., 2014. Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.* 109 (505), 411–423.  
URL <http://dx.doi.org/10.1080/01621459.2013.856795>
- Cuevas, A., 2014. A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference* 147, 1–23.  
URL <http://dx.doi.org/10.1016/j.jspi.2013.04.002>
- Darling, D. A., 1953. On a class of problems related to the random division of an interval. *Ann. Math. Statistics* 24, 239–253.
- Diestel, J., Uhl, Jr., J. J., 1977. Vector measures. American Mathematical Society, Providence, R.I., with a foreword by B. J. Pettis, *Mathematical Surveys*, No. 15.
- Ferraty, F., Vieu, P., 2006. Nonparametric functional data analysis. Theory and practice. Springer Series in Statistics. Springer, New York.
- Fraiman, R., Muniz, G., 2001. Trimmed means for functional data. *Test* 10 (2), 419–440.  
URL <http://dx.doi.org/10.1007/BF02595706>
- Gijbels, I., Nagy, S., 2014. Consistency of non-integrated depths for functional data, under review.
- Goia, A., Vieu, P., 2014. A partitioned single functional index model. *Computational Statistics*, 1–20.  
URL <http://dx.doi.org/10.1007/s00180-014-0530-1>
- Grenander, U., 1981. Abstract inference. John Wiley & Sons, Inc., New York, Wiley Series in Probability and Mathematical Statistics.
- Horváth, L., Kokoszka, P., 2012. Inference for functional data with applications. Springer Series in Statistics. Springer, New York.  
URL <http://dx.doi.org/10.1007/978-1-4614-3655-3>
- Ledoux, M., Talagrand, M., 2011. Probability in Banach spaces: Isoperimetry and processes. *Classics in Mathematics*. Springer-Verlag, Berlin, reprint of the 1991 edition.
- Liu, R. Y., 1990. On a notion of data depth based on random simplices. *Ann. Statist.* 18 (1), 405–414.  
URL <http://dx.doi.org/10.1214/aos/1176347507>
- Liu, R. Y., Parelius, J. M., Singh, K., 1999. Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.* 27 (3), 783–858.  
URL <http://dx.doi.org/10.1214/aos/1018031260>
- López-Pintado, S., Romo, J., 2009. On the concept of depth for functional data. *J. Amer. Statist. Assoc.* 104 (486), 718–734.  
URL <http://dx.doi.org/10.1198/jasa.2009.0108>
- López-Pintado, S., Romo, J., 2011. A half-region depth for functional data. *Comput. Statist. Data Anal.* 55 (4), 1679–1695.  
URL <http://dx.doi.org/10.1016/j.csda.2010.10.024>
- Mosler, K., Polyakova, Y., 2012. General notions of depth for functional data. arXiv preprint arXiv:1208.1981.
- Nagy, S., 2013. Coordinatewise characteristics of functional data. In: Vojáčková, H. (Ed.), *Proceedings 31th Int. Conf. Mathematical Methods in Economics 2013*, Jihlava, Czech Republic. College of Polytechnics Jihlava, pp. 655–660 (Part II.).
- Nagy, S., Gijbels, I., Omelka, M., Hlubinka, D., 2014. Integrated depth for functional data: Statistical properties and consistency, under review.
- Osius, G., 1989. Some results on convergence of moments and convergence in distributions with applications in statistics. *Mathematik-arbeitspapiere* no. 33, Universität Bremen.
- Ramsay, J. O., Silverman, B. W., 2002. Applied functional data analysis: Methods and case studies. Springer Series in Statistics. Springer-Verlag, New York.  
URL <http://dx.doi.org/10.1007/b98886>



- Ramsay, J. O., Silverman, B. W., 2005. Functional data analysis, Second Edition. Springer Series in Statistics. Springer, New York.
- Tukey, J. W., 1975. Mathematics and the picturing of data. In: Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2. Canad. Math. Congress, Montreal, Que., pp. 523–531.
- Varadarajan, V. S., 1958. On the convergence of sample probability distributions. *Sankhyā* 19, 23–26.
- Winkler, G., 1985. Choquet order and simplices with applications in probabilistic models. Vol. 1145 of Lecture Notes in Mathematics. Springer-Verlag, Berlin.
- Zuo, Y., Serfling, R., 2000. General notions of statistical depth function. *Ann. Statist.* 28 (2), 461–482.  
URL <http://dx.doi.org/10.1214/aos/1016218226>