Journal of Multivariate Analysis 140 (2015) 259–282

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Consistency of non-integrated depths for functional data

Irène Gijbels [a,b,*], Stanislav Nagy [a,b,c]

[a] KU Leuven, Department of Mathematics, Statistics Section, Celestijnenlaan 200b - box 2400, 3001 Leuven, Belgium
[b] KU Leuven, Leuven Statistics Research Centre (LStat), Celestijnenlaan 200b - box 5307, 3001 Leuven, Belgium
[c] Charles University in Prague, Department of Probability and Math. Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic

ARTICLE INFO

Article history:
Received 5 July 2014
Available online 27 May 2015

AMS 2000 subject classifications:
62G20
62G05

Keywords:
Band depth
Functional data
Glivenko–Cantelli theorem
Half-region depth
Infimal depth
Uniform consistency

ABSTRACT

In the analysis of functional data, the concept of data depth is of importance. Strong consistency of a sample version of a data depth is among the basic statistical properties that need to hold. In this paper we discuss consistency properties of three popular types of functional depth: the band depth, the half-region depth and the infimal depth. The latter is a special case of the recently introduced general class of Φ-depths. All three considered depth functions are of a non-integrated type. Counterexamples illustrate some problems with consistency results for these data depths. The main contribution of this paper consists of providing sufficient conditions for consistency of these non-integrated data depths to hold.

1. Introduction: depth for functional data

A statistical depth function, or shortly depth, is a statistical concept aiming to generalize the notion of quantiles to a multidimensional setup, where no natural ordering arises. Given a probability distribution on a general sample space, the depth assigns to a point in the sample space a non-negative number characterizing how "central" the point is with respect to the distribution. By identifying the central part of the distribution as the locus of points having highest depth, it induces a center-outward linear ordering on the sample space. This ordering may afterwards serve as a tool for the introduction of non-parametric methods into the framework of general data sets.

Various fields of application of data depth include multivariate, functional, or even general Banach-valued data. While there is a vast amount of literature concerning depth functions designed for multivariate (finite-dimensional) data (see e.g. [31] for a review), the choice of a depth suitable for functional data is limited to a small number of very similar concepts.

The first notion of depth for functions was introduced by Fraiman and Muniz [10], having a form of an integrated type of depth. Integrated depth functionals compute "average" depths of certain low-dimensional projections of a functional variable. For instance, the original integrated depth of Fraiman and Muniz [10] is computed as an integral of the univariate depth of functional values with respect to the corresponding marginal distribution over the domain of functions.

Other representatives of integrated depths are the integrated dual depths introduced by Cuevas and Fraiman [5] or the recently proposed depth of Claeskens et al. [4], the latter designed for vector-valued functional data.

A distinct, yet perhaps even more compelling approach has been advanced by López-Pintado and Romo [20] by the study of band depths. See also [17,19]. These band depths have been generalized to the setup of vector-valued functions [13,22]

* Corresponding author at: KU Leuven, Department of Mathematics, Statistics Section, Celestijnenlaan 200b - box 2400, 3001 Leuven, Belgium.
E-mail addresses: irene.gijbels@wis.kuleuven.be (I. Gijbels), stanislav.nagy@wis.kuleuven.be (S. Nagy).

http://dx.doi.org/10.1016/j.jmva.2015.05.012
0047-259X/

and sparsely observed functional data [23]. Band depths can be looked upon as infinite-dimensional modifications of known and well-studied multivariate depths, such as the simplicial depth of Liu [16], by using a purely geometrical notion of the graph of a function.

Recently a data depth closely related to band depth, namely the half-region depth, was introduced by López-Pintado and Romo [21]. Furthermore, Mosler and Polyakova [25] introduced a whole family of Φ-depths. All three depth functionals (band depth, half-region depth and Φ-depth) are intrinsically similar and in many aspects they can be studied simultaneously. For brevity purpose we will refer to these three depth functionals as non-integrated depths.

An essential property that a statistical depth function needs to satisfy is the uniform consistency of its sample version. Without this at hand, it is not possible to show the consistency of derived functionals such as a generalized median (the point with the highest depth value), depth contours [12], or depth-based L-statistics (see e.g. [10] for generalized trimmed means, or Fraiman and Meloche [9] for a general survey). All these derived concepts are of central importance in statistical depth methodology.

In contrast to the depths of integrated type, there are few theoretical results known about the depths of non-integrated type. The research has been focused on computational issues, and the only major theoretical results are the consistency theorems: López-Pintado and Romo [20, Theorem 4] for band depth and López-Pintado and Romo [21, Theorem 3] for half-region depth. Recently, Chakraborty and Chaudhuri [2, Section 3] pointed out some degenerate behavior of the band depth and the half-region depth for certain probability distributions. For Φ-depths of Mosler and Polyakova [25] there are, as far as we know, no results, yet, on treating its theoretical statistical properties.

The present paper concerns the investigation of uniform consistency results for the triple of non-integrated type depth functionals. As such we address a fundamental question for a great portion of depth functionals present in the state-of-the-art literature.

In Section 2 we provide preliminaries and introduce notations.

Band depth, as the first representative of non-integrated depths, is considered in Section 3. Firstly, we point out that band depth is, in general, not a member of the family of general graph depths, contrary to what was stated in [25]. Consequently, a separate treatment of band depths is justified. Secondly, we provide examples to illustrate problems with the available consistency results for band depths. To solve these problems we propose an adjusted band depth for which it is possible to guarantee a uniform consistency result.

In Section 4 we discuss consistency results for the half-region depth. As for band depth, we first show that the half-region depth cannot be seen as a general graph depth (cf. [25]), and it needs to be studied separately. Next, we point out some problems with the available consistency result for the half-region depth. We establish sufficient conditions under which uniform consistency of the half-region depth is guaranteed.

Finally, in Section 5 we establish consistency of the most important representative of Φ-depths recently adduced in the literature, the infimal depth. As a key for proving this consistency, we derive a new functional version of Glivenko–Cantelli's theorem for empirical cumulative distribution function processes. This key result, stated in Theorem 4, is of independent interest. Its proof is provided in the Appendix. Some further discussion on this theorem is given in Section 6.

In this paper we do not discuss the modified versions of band depths [20, Section 5] and half-region depths [21, Section 5], as they naturally belong to the class of integrated depths as described above. A general study on, among others, the consistency of this class of depths can be found in [26]. In that paper general consistency results for integrated depths are shown under minimal assumptions. These results include, as special cases, consistency properties for the modified versions of both band depth and half-region depth.

2. Preliminaries and notations

Let $\mathcal{C}\left([0,1]\right)$ denote the set of continuous real-valued functions defined on $[0,1]$. The space $\mathcal{C}\left([0,1]\right)$ is equipped with the uniform norm

$$\|x\| = \sup_{t\in[0,1]} |x(t)| \quad \text{for } x \in \mathcal{C}\left([0,1]\right), \tag{1}$$

which naturally induces a metric

$$\|x-y\| \quad \text{for } x, y \in \mathcal{C}\left([0,1]\right),$$

and a distance between a function and a set of functions A

$$\mathrm{d}\left(x, A\right) = \inf_{a\in A} \|x-a\| \quad \text{for } x \in \mathcal{C}\left([0,1]\right),\ A \subset \mathcal{C}\left([0,1]\right). \tag{2}$$

Let Ω be the sample space and $\mathcal{F}$ a σ-field on Ω, constituting a measurable space $(\Omega, \mathcal{F})$ on which all random elements in the paper are defined. For an arbitrary measurable space S, $\mathcal{P}\left(S\right)$ is the collection of all possible probability measures on S. Consider a functional random variable X taking on values in $\mathcal{C}\left([0,1]\right)$. For X having probability distribution P, with $\mathrm{P} \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$, we write $X \sim \mathrm{P}$. For $t \in [0,1]$, we use $\mathrm{P}_t \in \mathcal{P}\left(\mathbb{R}\right)$ to denote the univariate marginal distribution of $X(t)$. The (cumulative) marginal distribution function of $X(t)$ is denoted by F_t. For a positive integer n, P_n stands for the empirical measure of a random sample of size n taken from P, $\mathrm{P}_{n,t}$ for the marginal measure of P_n at t and $\mathrm{F}_{n,t}$ for the corresponding empirical distribution function.