UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 06-30
Statistics and Econometrics Series 12
May, 2006

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

ON THE CONCEPT OF DEPTH FOR FUNCTIONAL DATA

Sara López-Pintado and Juan Romo*

Abstract

The statistical analysis of functional data is a growing need in many research areas. We propose a new depth notion for functional observations based on the graphic representation of the curves. Given a collection of functions, it allows to establish the centrality of a function and provides a natural center-outward ordering of the sample curves. Robust statistics such as the median function or a trimmed mean function can be defined from this depth definition. Its finite-dimensional version provides a new depth for multivariate data that is computationally very fast and turns out to be convenient to study high-dimensional observations. The natural properties are established for the new depth and the uniform consistency of the sample depth is proved. Simulation results show that the trimmed mean presents a better behavior than the mean for contaminated models. Several real data sets are considered to illustrate this new concept of depth. Finally, we use this new depth to generalize to functions the Wilcoxon rank sum test. It allows to decide whether two groups of curves come from the same population. This functional rank test is applied to girls and boys growth curves concluding that they present different growth patterns.

Key Words: Functional data; Data depth; Rank test for functions.

*López-Pintado, Rutgers University, e-mail: saral@stat.rutgers.edu; Romo, Departamento de Estadística, Universidad Carlos III de Madrid, e-mail: juan.romo@uc3m.es.

On the Concept of Depth for Functional Data

Sara López-Pintado and Juan Romo*

Abstract

The statistical analysis of functional data is a growing need in many research areas. We propose a new depth notion for functional observations based on the graphic representation of the curves. Given a collection of functions, it allows to establish the "centrality" of a function and provides a natural center-outward ordering of the sample curves. Robust statistics such as the median function or a trimmed mean function can be defined from this depth definition. Its finite-dimensional version provides a new depth for multivariate data that is computationally very fast and turns out to be convenient to study high-dimensional observations. The natural properties are established for the new depth and the uniform consistency of the sample depth is proved. Simulation results show that the trimmed mean presents a better behavior than the mean for contaminated models. Several real data sets are considered to illustrate this new concept of depth. Finally, we use this new depth to generalize to functions the Wilcoxon rank sum test. It allows to decide whether two groups of curves come from the same population. This functional rank test is applied to girls and boys growth curves concluding that they present different growth patterns.

KEY WORDS: Functional Data; Data Depth; Rank Test for Functions.

*Department of Statistics, Universidad Carlos III de Madrid, 28903, Getafe, Madrid, Spain (e-mail: saral@stat.rutgers.edu; juan.romo@uc3m.es). This research was supported in part by Spanish Ministry of Education and Science grants BEC2000-0167 and BEC2005-03424.

1. INTRODUCTION

The data output sophistication in different research fields requires to advance in the statistical analysis of complex data. In functional data analysis, each observation is a real function $x_i(t)$, $i = 1, ..., n$, $t \in I$, where I is an interval in $\mathbb{R}$. There are several reasons that make necessary the study of functional data. In many research areas (medicine, biology, economics, engineering), the data generating process is naturally a stochastic function. Moreover, many problems are better approached if the data are considered as functions. For instance, if each curve is observed at different points, a multivariate analysis would not be valid, and it is therefore necessary to smooth the data and treat them as continuous functions defined in a common interval.

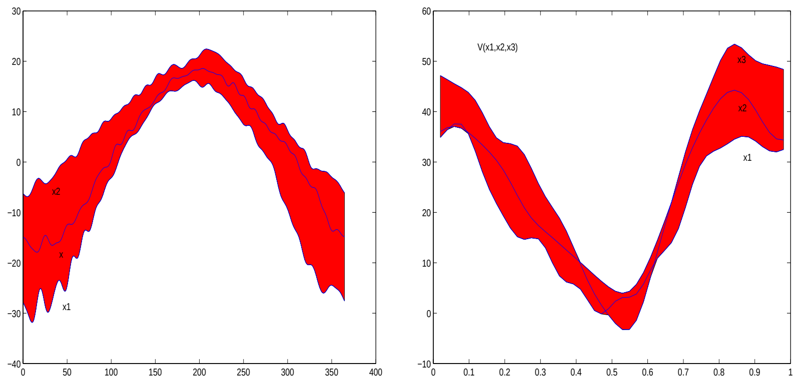
Multivariate techniques such as principal components, analysis of variance and regression methods have already been extended to a functional context (see Ramsay and Silverman, 2005). A fundamental task in functional data analysis is to provide an ordering within a sample of curves that allows the definition of order statistics such as ranks and L-statistics. A natural tool to analyze these functional data aspects is the idea of statistical depth. It has been introduced to measure the "centrality" or the "outlyingness"of an observation with respect to a given data set or an underlying distribution. In this paper, we propose a new definition of depth for functional observations. It permits to order a sample of curves from the center-outward and to extend robust statistics to a functional context. For example, a median function is a curve with the highest depth.

The notion of depth was first considered for multivariate data to generalize order statistics, ranks and medians to higher dimensions. Given a distribution of probability F in $\mathbb{R}^d$, a statistical depth assigns to each point x a real non-negative bounded value $D(x, F)$. Some depth definitions for multivariate data have been proposed by Mahalanobis (1936), Tukey (1975), Oja (1983), Liu (1990), Singh (1991), Fraiman and Meloche (1999), Vardi and Zhang (2001) and Zuo (2003). Liu (1990) and Zuo and Serfling (2000) introduced and studied four key properties a depth should verify: affine invariance, maximality at center, monotonicity and vanishing at infinity. Data depth can be widely applied. For example, Liu and Singh (1993) presented a nonparametric multivariate rank test using a quality depth index and

Liu (1995) proposed control charts for multivariate processes based on depth. Also, Liu, Parelius and Singh (1999) offered depth based tools for multivariate analysis; for instance, they defined trimmed regions, central regions and contours, and constructed a scale curve to visualize sample dispersion. In addition, Rousseeuw and Hubert (1999) introduced the idea of regression depth and Li and Liu (2004) designed a graphic tool and a test to check if two multivariate samples come from the same population.

The main goal of this paper is to propose a new notion of depth for functional data. It is based on the graphic representation of the functions and makes use of the bands defined by their graphs on the plane. Its finite-dimensional version is an alternative definition of depth for multivariate data, verifying essentially all the properties studied in Zuo and Serfling (2000). In addition, it has the advantage of being computationally very fast, which makes it adequate for analyzing high-dimensional data. Some asymptotic results, such as the uniform convergence of the sample depth and deepest point, are established. Most of these properties are extended to functions. With this new definition we can also generalize the concepts of multivariate L-estimates (in particular, trimmed means) to a functional context, where robust methods are possibly more useful than in multivariate problems, because there are more ways for outliers to affect functional statistics. For instance, a curve could be an "outlier" without having any unusually large value. Depth is particularly useful for identifying this kind of outliers; here, besides magnitude, shape is also relevant. Fraiman and Muniz (2001) introduced and studied a previous concept of depth for functional observations based on the integrals of univariate depths.

The paper is organized as follows. In section 2, we present the new definition of band depth for functional data. Section 3 explores its finite-dimensional version. In section 4, the functional version properties are analyzed. A generalized band depth, more convenient for irregular functions, is defined in section 5. Section 6 contains some simulations illustrating the robustness of estimates based on the proposed depth. In section 7, real data examples are discussed and used to check the band depth performance. A rank test for functions is introduced in section 8 and applied to decide whether two groups of real curves come from the same population. Finally, in section 9, we outline the main conclusions of this paper. The proofs are included in the Appendix.



(a) (b)

Figure 1: *(a) Band defined by two curves x_1, x_2 and a third curve x belonging to the band and (b) band determined by three curves x_1, x_2 and x_3.*

2. A BAND DEPTH FOR FUNCTIONAL DATA

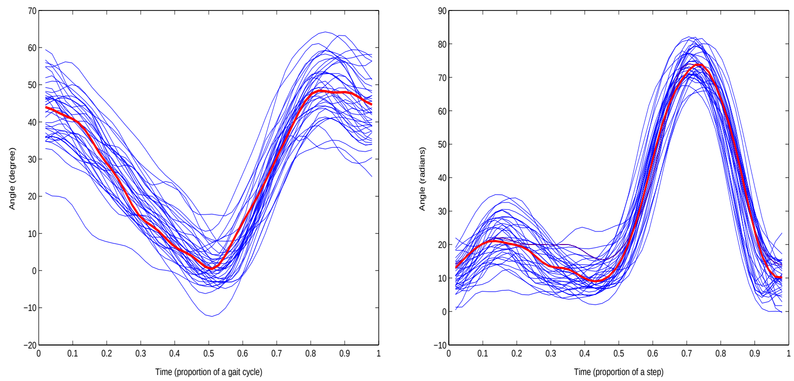
Our proposal follows a graph based approach. We recall definitions about function graphs that will be used throughout the paper. Let $x_1(t), ..., x_n(t)$ be a collection of real functions. Although the following ideas can be given for more general observations, we will restrict the exposition to functions in the space $C(I)$ of real continuous functions on the compact interval I. The graph of a function x is the subset of the plane $G(x) = \{(t, x(t)) : t \in I\}$. The band in $\mathbb{R}^2$ delimited by the curves $x_{i_1}, ..., x_{i_k}$ is

$$\begin{aligned} B(x_{i_1}, x_{i_2}, ..., x_{i_k}) &= \{(t,y):\ t \in I,\ \min_{r=1,...,k} x_{i_r}(t) \leq y \leq \max_{r=1,...,k} x_{i_r}(t)\} = \\ &= \{(t,y):\ t \in I,\ y = \alpha_t \min_{r=1,...,k} x_{i_r}(t) + (1-\alpha_t) \max_{r=1,...,k} x_{i_r}(t), \alpha_t \in [0,1]\}. \end{aligned}$$

Figure 1a presents the band $B(x_1, x_2)$ given by two curves; the graph of the function x is included in the band. Figure 1b shows the band given by three curves $B(x_1, x_2, x_3)$. For any function x in $x_1, ..., x_n$, the quantity

$$S_n^{(j}(x) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < i_2 < ... < i_j \leq n} I\{G(x) \subset B(x_{i_1}, x_{i_2}, ..., x_{i_j})\},\ j \geq 2,$$

expresses the proportion of bands $B(x_{i_1}, x_{i_2}, ..., x_{i_j})$ determined by j different curves $x_{i_1}, x_{i_2}, ..., x_{i_j}$ containing the graph of x. ($I\{A\}$ is one if A is true and zero otherwise).



(a) (b)

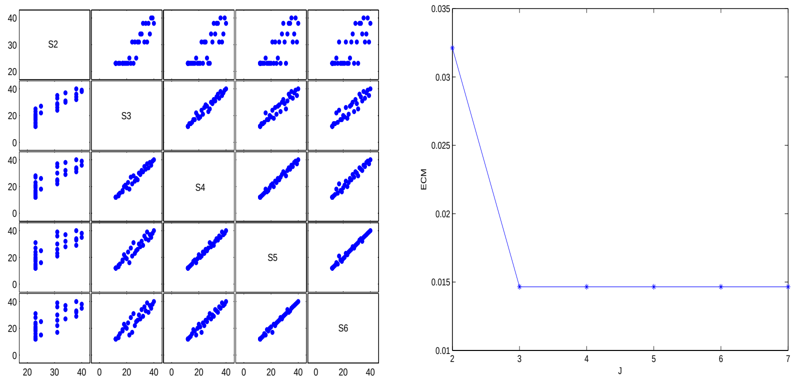
Figure 2: *Angle in the sagittal plane formed by (a) the hip and (b) the knee, as 39 children go through a gait cycle. The deepest curves based on $S_{n,2}$ within each sample are represented in red.*

Definition 1. For functions $x_1, ..., x_n$, the *band depth* of any of these curves x is

$$S_{n,J}(x) = \sum_{j=2}^{J} S_n^{(j}(x), J \geq 2. \tag{1}$$

If $X_1, X_2, ..., X_n$ are independent copies of the stochastic process X generating the observations $x_1, ..., x_n$, the corresponding population versions are $S^{(j}(x) = P\{G(x) \subset B(X_{i_1}, X_{i_2}, ..., X_{i_j})\}$ and $S_J(x) = \sum_{j=2}^{J} S^{(j}(x) = \sum_{j=2}^{J} P\{G(x) \subset B(X_{i_1}, X_{i_2}, ..., X_{i_j})\}$. A sample median function $\widehat{m}_{n,J}$ is a curve from the sample with highest depth value, $\widehat{m}_{n,J} = \underset{x \in \{x_1,...,x_n\}}{\arg\max}\, S_{n,J}(x)$, and a population median is a function m_J in $C(I)$ maximizing $S_J(\cdot)$. If they are not unique, the median will be the average of the curves maximizing depth.

The functions in Figure 2 provide the angle in the sagittal plane formed by the hip (left panel) and by the knee (right panel) as thirty nine children go through a gait cycle (see Ramsay and Silverman, 2005). The deepest curves (or median functions) for $S_{n,2}$ appear in red (they are also the deepest ones for $2 \leq J \leq 7$). The notion of functional depth allows to order the data curves from the center-outward and, consequently, order based statistics can be defined. Thus, L-statistics will be generalized to the functional setting. After extensive simulation, this definition seems to be very stable with respect to the selection of the tuning parameter J. Figure 3a is the multiple scatter plot of the ranks using S_J for different values



(a) (b)

Figure 3: *(a) Multiple scatterplot of the ranks of forty curves simulated from a gaussian process with S_J, $J = 2, ..., 6$ and (b) mean integrated error for the deepest function.*

of J. Forty sample paths in $C[0,1]$ have been simulated from a gaussian process with mean $m(t) = 2t$ and covariance function $\gamma(s,t) = \exp\left\{-|t-s|^2\right\}$. Their ranks were computed for S_2, S_3, ..., S_6. For J larger than 3, the points in the scatterplot fit the line $y = x$ and the ranks do not present dependence on J. To explore the mean integrated error, we have also simulated forty curves from a gaussian process in $C[0,1]$ with mean $f(t) = 4t$ and covariance function $\gamma(s,t) = \exp\left\{-|t-s|^2\right\}$. Figure 3b shows the mean integrated error $EI_{\widehat{m}}(J) = \int_0^1 (\widehat{m}_{n,J}(t) - f(t))^2\, dt$ for different values of J, where $\widehat{m}_{n,J}(t)$ is the deepest observation for $S_{n,J}$. The mean integrated error is calculated as $EI_{\widehat{m}}(J) = \frac{1}{I}\sum_{k=1}^{I}\left[\widehat{m}_{n,J}(k/I) - f(k/I)\right]^2$ at $I = 30$ points equally spaced in $[0,1]$. It is minimized for $J = 3$ and remains constant for $J \geq 3$: this indicates that the deepest point is the same for $J = 3$, 4, 5, 6 and 7.

3. FINITE-DIMENSIONAL VERSION

The finite-dimensional version of the functional band depth provides also a depth for multivariate data. Parallel coordinates (see, e.g., Inselberg, 1981, 1985, Inselberg *et al.*, 1987 and Wegman, 1990) provide a convenient way of visualizing a set of points in $\mathbb{R}^d$. The orthogonal axes in cartesian coordinates become parallel and equally spaced in parallel coordinates; thus, points with dimension larger than three can be easily represented. Observations in $\mathbb{R}^d$ can be seen as real functions defined on the set of indexes $\{1, 2, ..., d\}$ and expressed as

$x = (x(1), x(2), ..., x(d))$. Given points $x_1, x_2, ..., x_n$ in $\mathbb{R}^d$, the corresponding band in parallel coordinates $B(x_1, x_2, ..., x_n) = \left\{(k, y) : \ k \in \{1, 2, ..., d\}, \ \min_{i=1,...,n} x_i(k) \le y \le \max_{i=1,...,n} x_i(k)\right\}$ becomes a d-dimensional interval with sides parallel to the axes and delimited by the minimum and the maximum of the coordinates of $x_1, x_2, ..., x_n$,

$$R(x_1, x_2, ..., x_n) = \left\{x \in \mathbb{R}^d : \ \min_{i=1,...,n} x_i(k) \le x(k) \le \max_{i=1,...,n} x_i(k)\right\}.$$

Figures 4a and 4b present the band delimited by three points in the plane in parallel and cartesian coordinates, respectively. For any point x in $x_1, ..., x_n$,

$$S_n^{(j}(x) = \binom{n}{j}^{-1} \sum_{1 \le i_1 < i_2 < ... < i_j \le n} I\{x \in R(x_{i_1}, x_{i_2}, ..., x_{i_j})\}, \ j \ge 2,$$

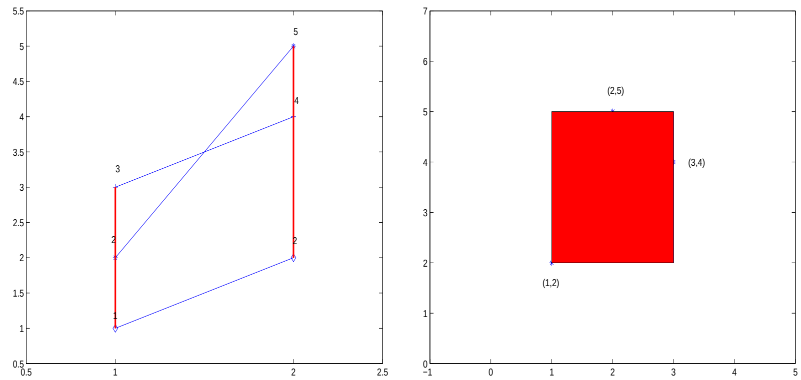
is the proportion of sets $R(x_{i_1}, x_{i_2}, ..., x_{i_j})$ defined by j different points $x_{i_1}, x_{i_2}, ..., x_{i_j}$ containing x. Hence, for points $x_1, ..., x_n$, the band depth of any of these points x is

$$S_{n,J}(x) = \sum_{j=2}^{J} S_n^{(j}(x), \ J \ge 2.$$

If P is a probability distribution in $\mathbb{R}^d$ and X_1, X_2,..., X_n is a random sample from P, the band depth for any point x in $\mathbb{R}^d$ with respect to P is

$$S_J(x) = S_J(x, P) = \sum_{j=2}^{J} P\left(x \in R(X_{i_1}, X_{i_2}, ..., X_{i_j})\right), \ J \ge 2.$$

A sample median $\widehat{m}_{n,J}$ is a sample point with highest depth, $\widehat{m}_{n,J} = \arg\max_{x \in \{x_1,...,x_n\}} S_{n,J}(x)$, and the population median m_J is a point in $\mathbb{R}^d$ maximizing S_J. If there is more than one point with maximum depth, the median is defined as their mean. Figure 5a shows in red the four deepest points for S_3. The deepest observation is marked with a green circle. Also, rectangles determined by two and three points, respectively, can be seen in the same figure. The scatterplot in Figure 5b compares the ranks for thirty points in $\mathbb{R}^2$ using Mahalanobis, Tukey, Liu and S_3 depths. The deepest observation is usually the same for different depths and the orderings induced in the set of points are very similar. Contrary to most of previous definitions of depth, the band depth is not computationally intensive. This makes it convenient to analyze very high-dimensional data.



(a) (b)

Figure 4: *(a) Points* $(1,2)$*,* $(2,5)$ *and* $(3,4)$ *and the corresponding band in parallel coordinates and (b) the same three points and band in cartesian coordinates.*

3.1 Band Depth Properties

The deepest point for the band depth in one dimension is the usual univariate median; moreover, the order induced by the band depth in the real line does not depend on J. Liu (1990) established four natural properties a notion of depth should verify. Zuo and Serfling (2000) analyzed them in a very general framework. The band depth S_J is a *Type A depth function* in their context (just write

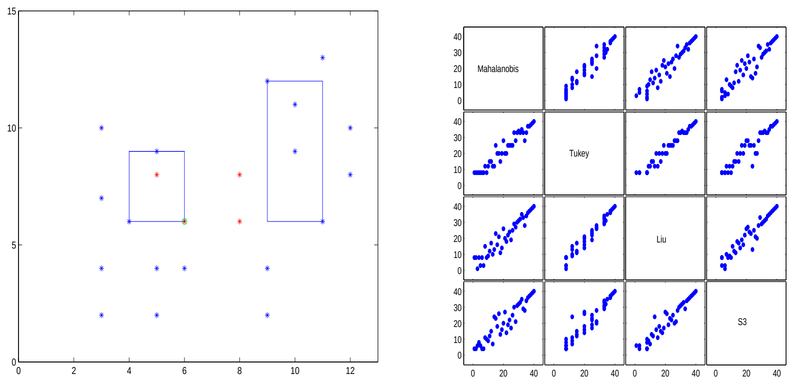
$$h(x, x_1, x_2, ..., x_J) = h_2(x, x_1, x_2) + h_3(x, x_1, x_2, x_3) + ... + h_J(x, x_1, x_2, ..., x_J),$$

with $h_j(x, x_1, x_2, ..., x_j) = I\{x \in R(x_1, x_2, ..., x_j)\}$). Our first theorem gives the structural properties of the band depth: monotonicity, maximality at center, vanishing at infinity and continuity.

Theorem 1. Let P be a probability distribution in $\mathbb{R}^d$. Then:

(i) If P is absolutely continuous and its marginals P_i, $i = 1, 2, ..., d$ are symmetric with respect to the origin then $S_J(\alpha x)$ is a monotone nonincreasing function in $\alpha \geq 0$, for all $x \in \mathbb{R}^d$.

(ii) Under the conditions in *(i)*, if the density f is positive in a neighborhood of the center of symmetry m then $S_J(\cdot)$ is uniquely maximized at m.



(a) (b)

Figure 5: *(a) Four deepest points from a sample in $\mathbb{R}^2$ are shown in red and the deepest one in green. Two rectangles determined by two and three points are also represented. (b) Multiple scatterplot diagram for the ranks of thirty points using Mahalanobis, Tukey, Liu and S_3 depths.*

(iii) $\sup_{\|x\|_\infty \geq M} S_J(x) \longrightarrow 0$, as $M \to \infty$ and $\sup_{\|x\|_\infty \geq M} S_{n,J}(x) \xrightarrow{a.s} 0$, as $M \to \infty$.

(iv) $S_J(\cdot)$ is upper- semicontinuous. If the marginal distributions of P are absolutely continuous then $S_J(\cdot)$ is continuous.

The monotonicity in *(i)* is not verified if the underlying distribution is not absolutely continuous: let $d = 1$, $P(X = 0) = 1/5$, $P(X = \pm 1) = 1/5$, and $P(X = \pm 2) = 1/5; X$ is symmetric with respect to 0, $S_2(1/2; P) = 12/25$ and $S_2(1; P) = 15/25$.

Several interesting properties follow from the fact that the band depth is a U-statistic.

Proposition 1. $S_J(x)$ can be expressed as a U-statistic of order J.

The symmetry of P is inherited by the sample distribution of the deepest point. Recall that a variable X with distribution P is antipodally symmetric with respect to c if $(X - c)$ and $-(X - c)$ have the same distribution (see, e.g., Liu, Parelius and Singh, 1999).

Proposition 2. If P is antipodally symmetric then the distribution of $\widehat{m}_{n,J}$ is also antipodally symmetric with respect to the population center of symmetry.

3.2 Asymptotic Results

Theorem 2 provides the consistency properties of the band depth. Consider $S_n^{(j}(x) = \binom{n}{j}^{-1} \sum_{1 \le i_i < \ldots < i_j \le n} I\left\{x \in R(X_{i_1}, ..., X_{i_j})\right\}$, where X_i are independent and identically distributed random variables taking values in $\mathbb{R}^d$ with distribution P and $R(X_1, ..., X_j)$ is the closed interval in $\mathbb{R}^d$ defined by the points $X_1, ..., X_j$. Following Arcones and Giné (1993), $S_n^{(j}(x)$ is a U-process of order j indexed by the class of functions $\mathcal{F} = \left\{R_x : \ x \in \mathbb{R}^d\right\}$, where $R_x = \{I\left\{(x_1, ..., x_j) : \ x \in R(x_1, ..., x_j)\right\}\}$. Then $S_n^{(j}(x) = U_j^n(R_x)$ and its population version is $S^{(j}(x) = P^j\left(x \in R(X_1, ..., X_j)\right)$.

Theorem 2. Let P be a probability distribution in $\mathbb{R}^d$. Then:

(i) $\sup_{x \in \mathbb{R}^d} |S_{n,J}(x) - S_J(x)| \xrightarrow{a.s.} 0$, as $n \to \infty$.

(ii) If $S_J(\cdot)$ is maximized at m_J and $m_{n,J}$ is a sequence of random variables such that $S_{n,J}(m_{n,J}) = \sup_{x \in \mathbb{R}^d} S_{n,J}(x)$ then $m_{n,J} \xrightarrow{a.s.} m_J$, as $n \to \infty$.

(iii) (Consistency of the sample deepest point) If the density f is different from zero in a neighborhood of m_J and $S_J(\cdot)$ is uniquely maximized at m_J then $\widehat{m}_{n,J} \xrightarrow{a.s.} m_J$, as $n \to \infty$.

A particular case of finite-dimensional observations are longitudinal data defined on discrete instants of time $t = 1, 2, ..., T$. All the definitions and properties presented in this section can be applied to this type of data.

4. PROPERTIES OF THE FUNCTIONAL BAND DEPTH

Next we analyze the band depth properties for functions. Let X be a process in $C(I)$ with a tight distribution P, i.e., $P(\|X\|_\infty \ge M) \to 0$, as $M \to \infty$. The first result expresses the compatibility of the functional depth with the increasing finite-dimensional depth versions. Let $\mathbb{Q} \cap I = \{q_1, q_2, ..., q_n, ...\}$ be the sequence of rational numbers in I. Let C_m be the set of vectors in $\mathbb{R}^m$ obtained evaluating a function x at the first m rational numbers of the interval I. S_J denotes both the functional depth and the corresponding finite-dimensional version. The next result provides the convergence of $S_J(x_m)$ to $S_J(x)$ as m tends to infinity, where $S_J(x_m)$ is the depth of x_m in C_m.

Proposition 3. If $x_m = (x(q_1), x(q_2), ..., x(q_m))$ is the function x evaluated at the first m rational numbers in I then $S_J(x_m)$ converges to $S_J(x)$, as m tends to infinity.

The next Theorem gives the basic properties of the functional depth S_J.

Theorem 3.

(i) Let $T(x) = ax + b$, where x, a and b are continuous functions in I, with $a(t) \neq 0$ for each $t \in I$. Then $S_J(x, P) = S_J(ax + b, P_{aX+b})$.

(ii) $\sup_{\|x\|_\infty \geq M} S_J(x) \longrightarrow 0$ and $\sup_{\|x\|_\infty \geq M} S_{n,J}(x) \xrightarrow{a.s} 0$, as $M \to \infty$.

(iii) S_J is an upper-semicontinuous function. Moreover, if the probability distribution P on $C(I)$ has absolutely continuous marginal distributions, then the band depth S_J is a continuous functional on $C(I)$.

Recall that a random variable X on $C(I)$ is symmetric (with respect to the zero function) if X and $-X$ have the same distribution.

Proposition 4. If the population random variable X on $C(I)$ is symmetric then the distribution of $m_{n,J}$ is also symmetric.

The band depth is uniformly consistent on compact sets of functions.

Theorem 4. Let P be a probability distribution on $C(I)$ with absolutely continuous marginal distributions. Then:

(i) $S_{n,J}(\cdot)$ is uniformly consistent on any equicontinuous set E: $\sup_{x \in E} |S_{n,J}(x) - S_J(x)| \overset{a.s.}{\to} 0$, as $n \to \infty$.

(ii) If $S_J(\cdot)$ is uniquely maximized at $m \in E$ and m_n is a sequence of functions in E with $S_{n,J}(m_n) = \sup_{x \in \mathbb{E}} S_{n,J}(x)$ then $m_n \xrightarrow{a.s.} m$, as $n \to \infty$.

For instance, the set $Lip_{\alpha,A}(I) = \{x : I \to \mathbb{R},\ |x(t_1) - x(t_2)| \leq A\,|t_1 - t_2|^\alpha,\ t_1,\ t_2 \in I\}$ is equicontinuous and verifies the condition in Theorem 4; hence, $S_{n,J}$ converges uniformly to S_J over $Lip_{\alpha,A}(I)$. The usual Lipschitzian functions are a particular case of $Lip_{\alpha,A}(I)$ (with $\alpha = 1$) and thus the band depth is uniformly consistent on the Lipschitz functions.

5. A GENERALIZED BAND DEPTH

Instead of considering the indicator function, a more flexible definition can be introduced by measuring the set where the function is inside the corresponding band. For any of the functions x in $x_1, ..., x_n$, let

$$A_j(x) \equiv A(x; x_{i_1}, x_{i_2}, ..., x_{i_j}) \equiv \left\{ t \in I : \min_{r=i_1,...,i_j} x_r(t) \leq x(t) \leq \max_{r=i_1,...,i_j} x_r(t) \right\}, \ j \geq 2, \quad (2)$$

be the set in the interval I where the function x is in the band determined by the observations $x_{i_1}, x_{i_2}, ..., x_{i_j}$. If λ is the Lebesgue measure on I, $\lambda_r(A_j(x)) = \frac{\lambda(A_j(x))}{\lambda(I)}$ gives the "proportion of time" that x is in the band. Now,

$$GS_n^{(j}(x) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < i_2 < ... < i_j \leq n} \lambda_r(A(x; x_{i_1}, x_{i_2}, ..., x_{i_j})), \ j \geq 2, \quad (3)$$

is a generalized version of $S_n^{(j}(x)$: if x is always inside the band, the value $\lambda_r(A_j(x))$ is one and this extends the previous definition.

Definition 2. For functions $x_1, ..., x_n$, the generalized band depth of any of these curves x is

$$GS_{n,J}(x) = \sum_{j=2}^{J} GS_n^{(j}(x), \ J \geq 2. \quad (4)$$

If $X_1, X_2, ..., X_n$ are independent copies of the process X giving the observations $x_1, ..., x_n$, the population version is $GS^{(j}(x) = E\ \lambda_r(A(x; X_{i_1}, X_{i_2}, ..., X_{i_j}))$, $j \geq 2$, and $GS_J(x) = \sum_{j=2}^{J} GS^{(j}(x)$, $J \geq 2$. In the finite-dimensional case, $GS_n^{(j}(x)$ is the proportion of coordinates of x inside the interval established by j different points from the sample:

$$GS_n^{(j}(x) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < ... < i_j \leq n} \frac{1}{d} \sum_{k=1}^{d} I\left\{ \min\left\{x_{i_1}(k), ..., x_{i_j}(k)\right\} \leq x(k) \leq \max\left\{x_{i_1}(k), ..., x_{i_j}(k)\right\}\right\}.$$

In general, the order induced in a sample when J increases is stable. To illustrate it we have simulated forty curves from a gaussian process defined on $[0, 1]$ with zero mean and covariance function $\gamma(s,t) = \exp\left\{-|t-s|^2\right\}$. Figure 6 gives the multiple scatterplot of the curves ranks for the generalized band depth GS_J with different values of J. The ranks present a good fit to the line $y = x$, and so they are essentially the same for different values

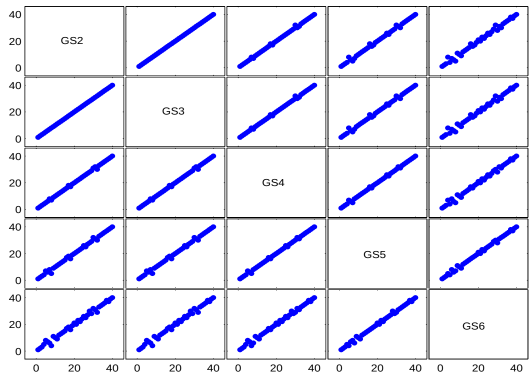


Figure 6: *Multiple scatterplot diagram for ranks of forty curves simulated from a gaussian process based on* GS_J *with* $J = 2, 3, 4, 5, 6$.

of J. Therefore, throughout the paper we will consider the generalized band depth with $J = 2$ and it will be denoted as GS. For $J = 2$,

$$\begin{aligned} GS_n(x) &= \frac{1}{d}\sum_{k=1}^{d}\binom{n}{2}^{-1}\sum_{1\le i_1<i_2\le n} I\left\{\min\left\{x_{i_1}(k), x_{i_2}(k)\right\} \le x(k) \le \max\left\{x_{i_1}(k), x_{i_2}(k)\right\}\right\} \\ &= \frac{1}{d}\sum_{k=1}^{d} SD_{{}_nF_k}(x(k)), \end{aligned}$$

where $SD_{F_{n,k}}(x(k))$ is the univariate simplicial depth of $x(k)$. Hence, $GS_n(x)$ is the average of the univariate simplicial depths for each coordinate $x(k)$. From the simplicial depth properties (see Liu, 1990) it is straightforward to check that the finite-dimensional version of the generalized band depth verifies all the properties established in Theorem 1, except *(iii)*. The band depth is more dependent on the curves shape and more restrictive than the generalized version, which implies the presence of many ties. The generalized band depth relies more on the magnitude or size of the curves than on their shape. Another relevant difference between them is their behavior for curves leaving the center of the sample only for a short interval, i.e., remaining in the interior of the sample almost all the time, but taking extreme values in short subintervals: the generalized band depth can still be large for them but the band depth will always take small values on these curves.

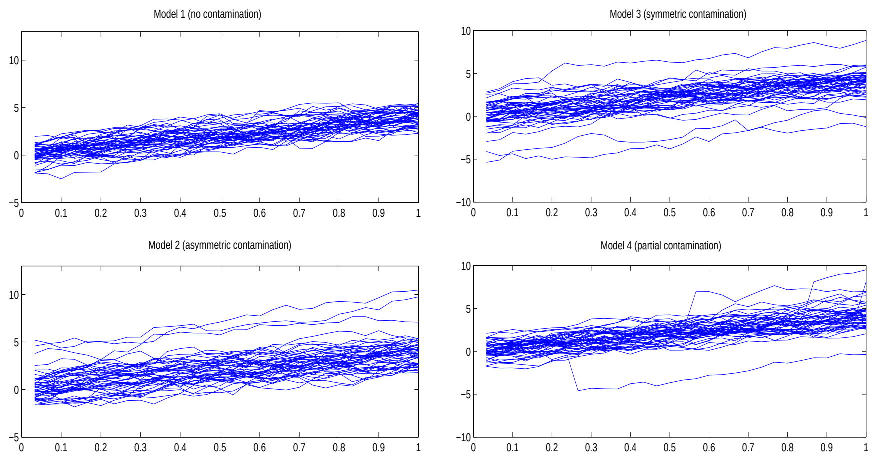


Figure 7: *Curves generated from model 1 (without contamination), model 2 (asymmetric contamination), model 3 (symmetric contamination) and model 4 (partial contamination) with $M = 5$ and $q = 0.1$ in all cases.*

6. SIMULATION RESULTS

We compare now the functional trimmed mean based on $S_{n,3}$ and GS_n with the mean in terms of robustness. We have generated curves from different models: a basic one without contamination and several models with different types of contaminations. They include those previously analyzed by Fraiman and Muniz (2001). Model 1 is the basic one, $X_i(t) = g(t) + e_i(t), 1 \leq i \leq n$, where $e_i(t)$ is a stochastic gaussian process with zero mean and covariance function $\gamma(s,t) = (\frac{1}{2})(\frac{1}{2})^{5|t-s|}$ and $g(t) = 4t$, with $t \in [0,1]$. An asymmetric contamination appears in model 2, $Y_i(t) = X_i(t) + c_i M$, $1 \leq i \leq n$, where c_i is 1 with probability q and 0 with probability $1-q$; M is the contamination size constant. A symmetric contamination is included in model 3, $Y_i(t) = X_i(t) + c_i\sigma_i M$, $1 \leq i \leq n$, where c_i and M are defined in model 2 and σ_i is a sequence of random variables independent of c_i taking values 1 and -1 with probability $1/2$. Model 4 is partially contaminated, $Y_i(t) = X_i(t) + c_i\sigma_i M$, if $t \geq T_i$, $1 \leq i \leq n$, and $Y_i(t) = X_i(t)$, if $t < T_i$, where T_i is a random number generated from a uniform distribution on $[0,1]$. Curves simulated from these four models can be seen in Figure 7.

Model 5 is contaminated by peaks, $Y_i(t) = X_i(t) + c_i\sigma_i M$, if $T_i \leq t \leq T_i + l$, $1 \leq i \leq n$, and $Y_i(t) = X_i(t)$, if $t \notin [T_i, T_i + l]$, where $l = 2/30$ and T_i is a random number from

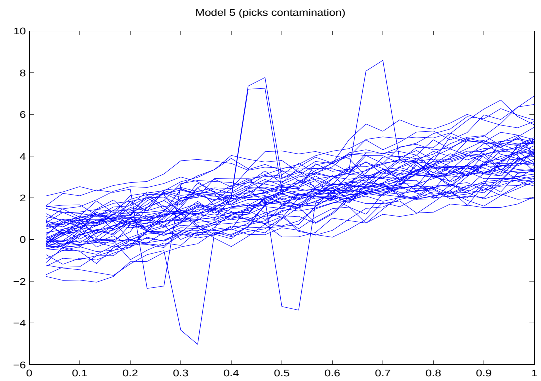


Figure 8: *Fifty curves from model 5 ($M = 5$ and $q = 0.1$).*

a uniform distribution in $[0, 1 - l]$. The contamination only occurs for a short subinterval of length l. Figure 8 shows an example of model 5 curves. Our goal is to analyze the robustness of statistics based on functional depth. We compare the mean and the α-trimmed mean, given by $\widehat{\mu}_n(t) = \frac{\sum_{i=1}^{n} Y_i(t)}{n}$ and $\widehat{m}_n^{\alpha}(t) = \frac{\sum_{i=1}^{n-[n\alpha]} Y_{(i)}(t)}{n-[n\alpha]}$, where $Y_{(1)}, Y_{(2)}, ...Y_{(n)}$ is the sample ordered from the deepest to the least deep curve and $[n\alpha]$ is the integer part of $n\alpha$. For each model, we have considered R replications, generating n curves in each replication and calculating the integrated error evaluated at $T = 30$ equally spaced points in $[0, 1]$. The integrated errors for each replication j are $EI_{\mu}(j) = \frac{1}{T}\sum_{k=1}^{T} \left[\widehat{\mu}_n(k/T) - g\left(k/T\right)\right]^2$ and $EI_S^{\alpha}(j) = \frac{1}{T}\sum_{k=1}^{T} \left[\widehat{m}_n^{\alpha}(k/T) - g\left(k/T\right)\right]^2$, respectively. Table 1 contains the mean integrated error for each model considering $R = 500$ replications with $n = 50$ curves, contamination probability $q = 0.1$, two contamination constants, $M = 5$, $M = 25$ and $\alpha = 0.2$. In the model without contamination the mean behaves always better than the trimmed mean. For models 2, 3 and 4 the best mean integrated error corresponds to GS (and with S_3 the results improve over the mean); however, for model 5 with $M = 25$, the mean integrated error is minimized for S_3. The reason is that contamination in model 5 appears in a small domain interval and the generalized depth is not very robust with respect to this type of contamination; a curve with contaminated values in a short interval of the domain can still be deep using the generalized band depth whereas it will not be deep based on the band depth.

Const.	Estimator	Model 1	Model 2	Model 3	Model 4	Model 5
M=5	Mean	**0.0094** (**0.0067**)	0.3106 (0.2564)	0.0578 (0.0656)	0.0322 (0.0326)	**0.0153** (**0.0086**)
	S_3	0.0126 (0.0096)	0.2519 (0.2584)	0.0267 (0.0290)	0.0314 (0.0313)	0.0179 (0.0113)
	GS	0.0139 (0.0104)	**0.0228** (**0.0359**)	**0.0168** (**0.0112**)	**0.0184** (**0.0142**)	0.0192 (0.0113)
M=25	Mean	**0.0099** (**0.0069**)	2.1158 (2.1485)	0.7634 (1.0839)	0.3296 (0.4674)	0.1442 (0.0769)
	S_3	0.0128 (0.0094)	1.1345 (1.8690)	0.2907 (0.7061)	0.2558 (0.3944)	**0.0876** (**0.0812**)
	GS	0.0129 (0.0110)	**0.0781** (**0.3481**)	**0.0148** (**0.0109**)	**0.0951** (**0.1587**)	1.3762 (0.0662)

Table 1: *Means and standard deviations of integrated errors with* $R = 500$ *replications,* $q = 0.1$ *and* $\alpha = 0.2$.

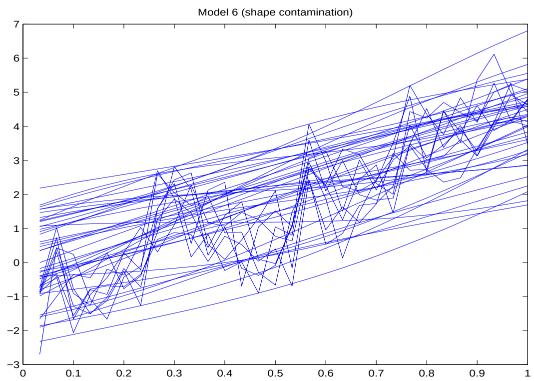


Figure 9: *Curves from model 6 with* $k_2 = c_2 = 1$, $\mu_2 = 0.2$ *and contamination probability* $q = 0.1$.

Besides magnitude contamination, we have also considered shape contamination. There is no general definition of outlier for functional observations. A curve could be outlier for different reasons: it can be very distant from the mean (magnitude outlier) or have a pattern different from the other curves, being, e.g., decreasing when the remaining ones are increasing or very irregular in a set of smooth curves (shape outlier). To generate shape outliers, we use the covariance $\gamma(s,t) = k\exp\{-c\,|t-s|^{\mu}\}$, with $s,\ t \in [0,1]$, and k, c, $\mu > 0$ (see Wood and Chan, 1994). Different values of k, c and μ change the shape of the generated functions. For example, increasing μ and k, the curves are smoother; however, increasing c the curves are more irregular. Model 6 is a mixture of $X_i(t) = g(t) + e_{1i}(t)$, $1 \leq i \leq n$, with $g(t) = 4t$ and $e_{i1}(t)$ a gaussian stochastic process with zero mean and covariance function $\gamma_1(s,t) = \exp\left\{-|t-s|^2\right\}$ and $Y_i(t) = g(t) + e_{2i}(t)$, $1 \leq i \leq n$, with $e_{i2}(t)$ a gaussian process with zero mean and covariance function with values k_2, c_2 and μ_2 chosen to generate more irregular curves (for example, $\mu_2 < 2$ or $c_2 > 1$). The contaminated model 6 is $Z_i(t) = (1-\varepsilon)X_i(t) + \varepsilon Y_i(t)$, $1 \leq i \leq n$, where ε is a Bernoulli variable $Be(q)$ and q is a small contamination probability; thus, we contaminate a sample of smooth curves from $X_i(t)$ with curves from $Y_i(t)$ having different covariance function and providing more irregular curves. Figure 9 shows curves simulated from model 6: the contaminated curves behave more irregularly than the remaining functions but they are not far from them in terms of distance. Table 2 contains the simulation results for model 6 with $R = 500$, $n = 50$

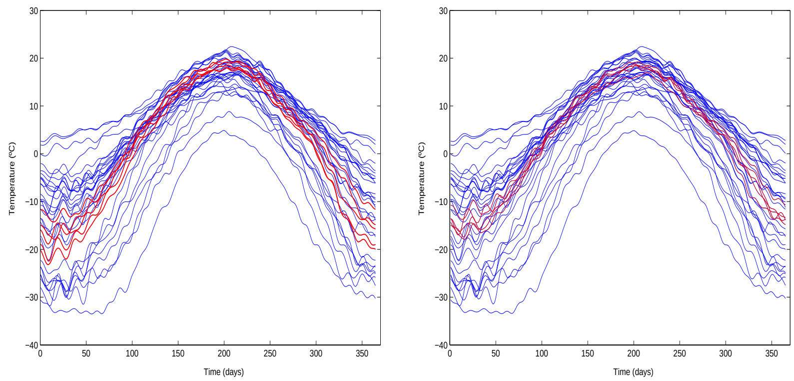
	Estimation	$\mu_2 = 0.2$ $k_2 = 1$	$\mu_2 = 0.1$ $k_2 = 1$	$\mu_2 = 0.1$ $k_2 = 2$
$q = 0.15$	Mean	0.0455 (0.0487)	0.0483 (0.0622)	0.0468 (0.0418)
	S_3	**0.0321** (**0.0455**)	**0.0295** (**0.0351**)	**0.0250** (**0.0233**)
	GS	0.0542 (0.0633)	0.0488 (0.0447)	0.0364 (0.0339)
$q = 0.1$	Mean	0.0318 (0.0334)	0.0360 (0.0320)	0.0355 (0.0297)
	S_3	**0.0293** (**0.0344**)	**0.0282** (**0.0313**)	**0.0272** (**0.0299**)
	GS	0.0345 (0.0508)	0.0414 (0.0369)	0.0380 (0.0381)

Table 2: *Simulation results with $R = 500$ replications, $n = 50$ curves and $\alpha = 0.2$.*

and $\alpha = 0.2$. We have considered two contamination probabilities ($q = 0.15$ and $q = 0.1$) and different values of μ_2, k_2 to modify the covariance of $Y(t)$ ($c_2 = 1$ in all cases). The minimum mean integrated error corresponds always to the trimmed mean based on the band depth S_3. This is due to the contamination type, shape more than magnitude; thus, the band depth performs well in terms of robustness with respect to this form of outliers. However, the generalized band depth is less robust against shape contamination, since most of the contaminated curves values can still be very central in the sample although the curve behavior is different from the remaining functions.

7. REAL DATA EXAMPLES

The first real data set describes daily temperature in different Canadian weather stations (Ramsay and Silverman, 2005). The original data was smoothed using a Fourier basis with sixty five elements. Figure 10a shows the temperature curves; the five deepest functions for the band depth are in red. Figure 10b presents the five deepest curves using the generalized



(a) (b)

Figure 10: *Daily temperature in 35 Canadian weather stations for one year. The five deepest curves (in red) based on (a)* $S_{n,3}$ *and (b)* GS*.*

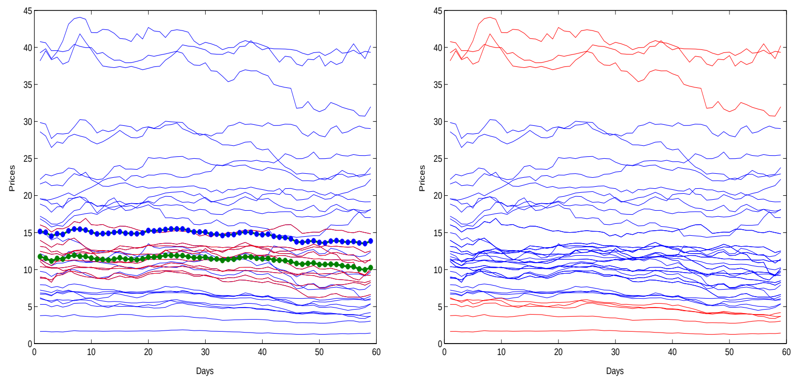
band depth. The second example contains the price curves for the 35 firms in the Spanish IBEX 35 index. The functions include fifty nine daily measurements starting at June 22, 2002. Figure 11a gives in red the 30% deepest elements for S_3. The mean curve is represented with blue asterisks and the trimmed mean with $\alpha = 0.3$ appears in green asterisks. The mean curve is more sensitive to the extreme functions than the trimmed one. In Figure 11b the 25% most extreme (least deep) curves from the sample are in red.

8. A RANK TEST FOR FUNCTIONS

The depth definition for curves allows us to extend the rank test to functional data. Liu and Singh (1993) generalized to multivariate data the univariate Wilcoxon rank test through the order induced by a multivariate depth. Brown and Hettmansperger (1989) and Hettmansperger and Oja (1994) have also proposed different rank tests for multivariate observations. Following Liu and Singh (1993), let

$$R(P_n, x_i) = R(x_i) = \text{proportion of } x_j\text{'s from the sample with } S(x_j) \leq S(x_i). \qquad (5)$$

It takes values between 0 and 1. We rank the observations x_i according to the increasing values of R, assigning them an integer from 1 to n. If there are curves with the same value of R, $R(x_{i_1}) = R(x_{i_2}) = ... = R(x_{i_j})$, with $i_1 < i_2 < ... < i_j$, we consider the rank of $x_{i_{k+1}}$ as the rank of x_{i_k} plus one. We propose a test based on these ranks to decide if



(a) (b)

Figure 11: *Prices included in the IBEX 35 index. (a) The* 30% *deepest curves are in red, and mean and median are marked with blue and green asterisks, respectively. (b) The* 25% *most extreme curves appear in blue.*

two groups of curves come from the same population. Let $x_1, ..., x_n$ be a sample of curves from population P_1 and let $y_1, ..., y_m$ be a sample of curves from population P_2. Assume that there is a third reference sample $Z = \{z_1, z_2, ..., z_{n_0}\}$ from one of the two populations, for example P_1, with n_0 greater than n and m. Let P_{n_0} be the corresponding empirical distribution. Calculate $R(P_{n_0}, x_i)$ = proportion of z_j's with $S(z_j, P_{n_0}) \leq S(x_i, P_{n_0,})$ and $R(P_{n_0}, y_i)$ = proportion of z_j's with $S(z_j, P_{n_0}) \leq S(y_i, P_{n_0})$, that express the position of each x_i and y_i with respect to Z. Order these values, $R(P_{n_0}, X_i)$ and $R(P_{n_0}, Y_i)$, from smallest to highest giving them a rank from 1 to $n + m$. If there are ties, we apply the previous criterion. The proposed statistic to test $H_0 : P_1 = P_2$ is $W = \sum_{j=1}^{m}$ ranks $R(P_{n_0}, y_j)$. The ranks of $R(P_{n_0}, y_j)$ behave under H_0 as m numbers randomly chosen from $\{1, 2, ..., n + m\}$. Hence, the distribution of W is the distribution of $\rho_1 + ... + \rho_m$, where $\rho_1, ..., \rho_m$ is a sample without replacement of $\{1, 2, ..., n + m\}$ (see Liu and Singh, 1993). The null hypothesis is rejected when W is small, because this indicates that $R(P_{n_0}, y_j)$ take on average lower values than $R(P_{n_0}, x_i)$, implying that the observations y_j are less deep with respect to P_{n_0} than x_i. The alternative hypothesis is that on average more than 50% of population P_1 is inner or more central than any observation from P_2, indicating that the distributions are not the same.

We have applied this test to real data representing the relative diameter versus the relative height of two groups of trees (Laricio and Radiata). Due to technical restrictions in the measurements, the relative diameter is defined as the ratio between its value at the corresponding height and the diameter at a fix height (1.3 centimeters). Relative height is height over the total height of the tree. Figure 12 shows the curves corresponding to seventy Laricio trees (left panel) and one hundred and forty Radiata trees (right panel). Since the number of observations per tree is very irregular (from 3 to 25), the data has been smoothed using a spline basis. To apply the rank test, we have considered seventy functions randomly chosen from the Radiata trees group as the reference group to compute the ranks of the remaining curves. The obtained p-values, using both the band depth and the generalized band depth are very close to zero. Therefore, we conclude that there exists significative differences between both groups. The second real data set includes the growth curves for boys and girls (see Ramsay and Silverman, 2005). We have applied the rank test to decide if there are no differences between both groups curves. We consider thirty two curves randomly chosen from the group of girls as the reference group. The remaining twenty two curves constitute the test group together with the thirty growth curves for boys. The obtained p-value with S_3 is 0.00011; hence, we reject the null hypothesis, concluding that there exist significative differences between the growth curves for boys and girls. This difference could be caused by a change either in mean or in dispersion. Graphically, the groups do not seem to be similar (see Figure 13). The shape is different and the heights of boys achieve higher values at the end. The rank test detects these differences. However, if we apply the rank test using the generalized band depth (GS) or Fraiman and Muniz's depth (FM) instead of S_3, we obtain p-values of 0.1199 and 0.1636 respectively, concluding in this case that there is no evidence for rejecting the null hypothesis. The reason is that GS and FM consider only the magnitudes, ignoring the curves shape. Note that the average values of boys and girls heights only differ in the final years (17 to 18 years) and FM and GS do not detect differences between both groups because they occur over a short interval. Since the essential difference between both groups is shape, we have applied the rank test to the curves derivatives (growth speed) (see Figure 14). The rank test p-values using S_3, GS and FM are very close to zero ($1.6 * 10^{-4}$, $4.54 * 10^{-7}$ and $6.01 * 10^{-7}$, respectively) and the null hypothesis is rejected. The instant of growth maximum velocity is different for boys

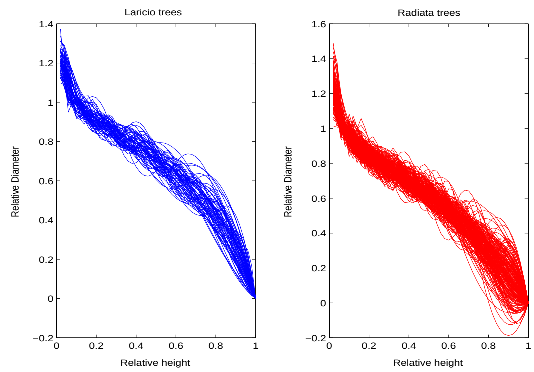


Figure 12: *Relative diameter versus relative height for Laricio trees represented in blue (left panel) and for Radiata trees represented in red (right panel).*

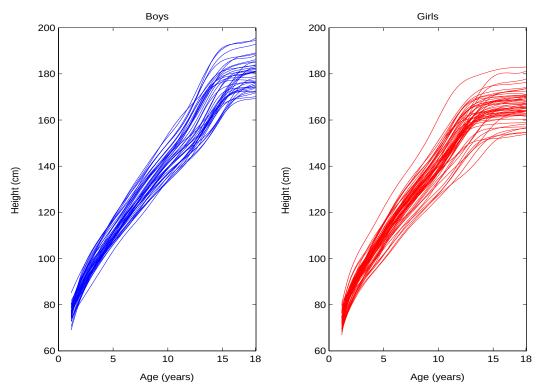


Figure 13: *Growth curves for a sample of thirty boys (left panel) and fifty four girls (right panel).*

and girls (see Figure 14). Girls reach maximum speed at an earlier age than boys. The band depth rank test detects these shape differences for boys and girls both in the original sample of growth curves and in the derivatives set.

9. CONCLUSIONS

We have introduced a band depth for functional data based on the graphic representation of the curves. It provides a criterion to order a sample of functions from center-outward and robust statistics for functional observations, such as the median and trimmed mean, can be constructed. Its finite-dimensional version provides a new multivariate depth. It is very

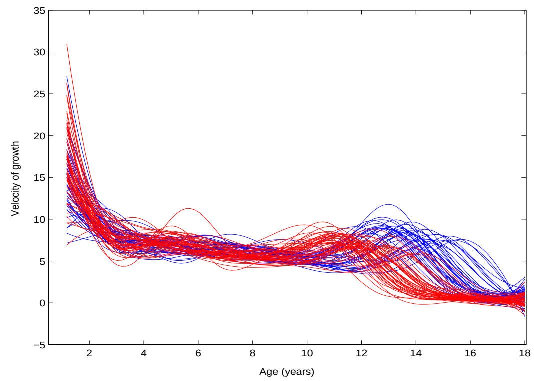


Figure 14: *First derivative of the growth curves for boys (in blue) and girls (in red).*

convenient to deal with high-dimensional data since it is computationally very fast, avoiding the main drawback of other finite-dimensional depth definitions. This new notion verifies natural depth properties. We have also established the uniform consistency of the sample depth in the finite and functional case. We have illustrated the robustness of this new depth with a simulation study and several real examples. As an application, a rank test for functional data is introduced and applied to decide whether two groups of curves come from the same population.

APPENDIX: PROOFS

Proof of Theorem 1: *(i)* The proof follows along the lines of Liu (1990). We check it for $d = 2$ and analyze the events contributing to the difference $S^{(j}(x) - S^{(j}(\alpha x)$, with $\alpha \geq 1$. Consider the sets defined by the arrow from x to αx entering or leaving $R(X_1, ..., X_j)$ with sides parallel to the axes. Given two points a and b, the line containing the segment $\overline{ab}$ divides the plane into two half planes. If the origin does not belong to this line, the half plane including the origin is called the inner side $I(a, b)$. Let $C = \left\{(a, b) : \ \overline{ab} \cap \overline{x, \alpha x} \neq \emptyset\right\}$; hence, C is the set of all segments that intersect the segment $\overline{x, \alpha x}$. Define the sets $r_1 = \left(\min_{i=1,...,j} \{X_i(1)\}, \min_{i=1,...,j} \{X_i(2)\}\right), r_2 = \left(\max_{i=1,...,j} \{X_i(1)\}, \min_{i=1,...,j} \{X_i(2)\}\right)$

$$r_3 = \left(\max_{i=1,...,j} \{X_i(1)\}, \max_{i=1,...,j} \{X_i(2)\}\right), r_4 = \left(\min_{i=1,...,j} \{X_i(1)\}, \max_{i=1,...,j} \{X_i(2)\}\right).$$

We have that $x \in R(X_1, ..., X_j)$ is equivalent to express that x belongs to the rectangle determined by the random points r_1, r_2, r_3, r_4. Consider the disjoint sets

$$\begin{aligned} A_{in}^{r_1 r_2} &= \{(X_1, ..., X_j) \text{ s.t. } \ (r_1, r_2) \in C \text{ and } r_3 \notin I(r_1, r_2)\} \\ A_{in}^{r_2 r_3} &= \{(X_1, ..., X_j) \text{ s.t. } \ (r_2, r_3) \in C \text{ and } r_4 \notin I(r_2, r_3)\} \\ A_{in}^{r_3 r_4} &= \{(X_1, ..., X_j) \text{ s.t. } \ (r_3, r_4) \in C \text{ and } r_1 \notin I(r_3, r_4)\} \\ A_{in}^{r_4 r_1} &= \{(X_1, ..., X_j) \text{ s.t. } \ (r_4, r_1) \in C \text{ and } r_2 \notin I(r_4, r_1)\}. \end{aligned} \tag{6}$$

Denote by A_{in} the events where the vector $\overline{x, \alpha x}$ enters a random rectangle with vertexes (r_1, r_2, r_3, r_4). Therefore, $A_{in} = A_{in}^{r_1 r_2} \cup A_{in}^{r_2 r_3} \cup A_{in}^{r_3 r_4} \cup A_{in}^{r_4 r_1}$. The events A_{out} are those where the arrow $\overline{x, \alpha x}$ leaves a random rectangle, and can be decomposed as a union of disjoint events in a similar way as A_{in}. Let $B_\alpha = \{\alpha x \in R(X_1, ..., X_j)\}$. Since $P(B_1 \setminus B_\alpha) = P(B_1) - P(B_1 \cap B_\alpha)$, $S^{(j}(x) - S^{(j}(\alpha x) = P(B_1 \setminus B_\alpha) - P(B_\alpha \setminus B_1)$. Now, $B_1 \setminus B_\alpha = A_{out} \setminus A_{in}$ and $B_\alpha \setminus B_1 = A_{in} \setminus A_{out}$, $S^{(j}(x) - S^{(j}(\alpha x) = P(A_{out} \setminus A_{in}) - P(A_{in} \setminus A_{out}) = P(A_{out}) - P(A_{out} \cap A_{in}) - [P(A_{in}) - P(A_{in} \cap A_{out})] = P(A_{out}) - P(A_{in})$, and it is sufficient to prove that $P(A_{out}^{r_i r_j}) - P(A_{in}^{r_i r_j})$ is positive for each i, j in (6). We present the proof only for one of them. For example, $P(A_{out}^{r_1 r_2}) - P(A_{in}^{r_1 r_2}) =$

$$= \int_{\{(x_1, ..., x_j):\ (r_1, r_2) \in C\}} \{P(r_3 \in I(r_1, r_2)) - P(r_3 \notin I(r_1, r_2))\}\, dF(x_1)...dF(x_j)$$

and it will be positive when $P(r_3 \in I(r_1, r_2)) \geq 1/2$, since this implies that

$$\{P(r_3 \in I(r_1, r_2)) - P(r_3 \notin I(r_1, r_2))\} \geq 0.$$

If the second coordinate of r_1 is positive, we will denote this event as r_1 *up*, and if it is negative, r_1 *down*. So,

$$P(r_3 \in I(r_1, r_2)) = P(r_3 \in I(r_1, r_2) \mid r_1\ up) \times P(r_1\ up) +$$
$$+ P(r_3 \in I(r_1, r_2) \mid r_1\ down) \times P(r_1\ down) = 0 + P(r_1\ down).$$

Now, by the definition of r_1, and due to the symmetry with respect to the origin,

$$\begin{aligned} P(r_1\ down) &= P\left(\min\{X_1(2), ..., X_j(2)\} < 0)\right) = 1 - P\left(\min\{X_1(2), ..., X_j(2)\} > 0)\right) = \\ &= 1 - [P(X_1(2) > 0) \times ... \times P(X_j(2) > 0)] = 1 - \left(\frac{1}{2}\right)^j > \frac{1}{2}. \end{aligned}$$

Then, for $\alpha \geq 1$, $S^{j)}(x) - S^{j)}(\alpha x) \geq 0$; thus, $S^{j)}(\alpha x)$ is non increasing in α, for $\alpha \geq 0$. Hence, $S_J(\alpha x)$ is also nonincreasing in α, for $\alpha \geq 0$.

(iii) We establish first the inclusion

$$\left\{(X_1, X_2, ..., X_j)\colon\ x \in R(X_1, X_2, ..., X_j)\right\} \subset \bigcup_{r=1}^{j} \left\{(X_1, X_2, ..., X_j) : \|X_r\|_\infty \geq \|x\|_\infty\right\} \qquad (7)$$

by contradiction. If $x \in R\left(X_1, X_2, ..., X_j\right)$ then, for each $k \in \{1, ..., d\}$,

$$\min_{r=1,...,j} \{X_r(k)\} \leq x(k) \leq \max_{r=1,...,j} \{X_r(k)\}. \qquad (8)$$

Assume that $\|X_r\|_\infty < \|x\|_\infty$ for each $r = 1, ..., j$; this implies that, for each r, we have

$$\max_{k \in \{1,...,d\}} |X_r(k)| < \max_{k \in \{1,...,d\}} |x(k)|.$$

Let k^* be the point where the maximum of $x(k)$ is achieved. Then, for all $r = 1, ..., j$, $|X_r(k^*)| < |x(k^*)|$, and this contradicts (8). Therefore,

$$\begin{aligned}
\sup_{\|x\|_\infty \geq M} S_J(x) &\leq \sum_{j=2}^{J} \sup_{\|x\|_\infty \geq M} P\left(x \in R(X_1, X_2, ..., X_j)\right) \\
&\leq \sum_{j=2}^{J} \sup_{\|x\|_\infty \geq M} \sum_{r=1}^{j} P\left(\|X_r\|_\infty \geq \|x\|_\infty\right) \\
&\leq \sum_{j=2}^{J} \sum_{r=1}^{j} \sup_{\|x\|_\infty \geq M} P\left(\|X_r\|_\infty \geq \|x\|_\infty\right)
\end{aligned}$$

and $\sup_{\|x\|_\infty \geq M} S_J(x) \longrightarrow 0$, when $M \to \infty$. To prove that the sample depth converges almost surely to zero we use again the inclusion in (7). $S_n^{(j}(x)$ is bounded by

$$\binom{n}{j}^{-1} \sum_{1 \leq i_1 < ... i_j \leq n} I\left\{\bigcup_{r=1}^{j} \left\{\|X_{i_r}\|_\infty \geq \|x\|_\infty\right\}\right\}, \qquad (9)$$

where $I\left\{\bigcup_{r=1}^{j} \left\{\|X_{i_r}\|_\infty \geq \|x\|_\infty\right\}\right\}$ is 1 if $\|X_{i_r}\|_\infty \geq \|x\|_\infty$ for some r and 0 in any other case. Then,

$$\begin{aligned}
\sup_{\|x\|_\infty \geq M} S_n^{(j}(x) &\leq \sup_{\|x\|_\infty \geq M} \binom{n}{j}^{-1} \sum_{1 \leq i_1 < ... i_j \leq n} I\left\{\bigcup_{r=1}^{j} \left\{\|X_{i_r}\|_\infty \geq \|x\|_\infty\right\}\right\} \\
&\leq \binom{n}{j}^{-1} \sum_{1 \leq i_1 < ... < i_j \leq n} \sum_{r=1}^{j} \sup_{\|x\|_\infty \geq M} I\left\{\|X_{i_r}\|_\infty \geq \|x\|_\infty\right\}.
\end{aligned}$$

Next, we prove that $X_M = \sup_{\|x\|_\infty \geq M} I\left\{\|X_{i_r}\|_\infty \geq \|x\|_\infty\right\}$ converges almost surely to 0 when M tends to infinity (for any r). Let $Y_M = I\left\{\|X_{i_r}\|_\infty \geq M\right\}$; since $0 \leq X_M \leq Y_M$, it is sufficient to prove that $Y_M \xrightarrow{a.s.} 0$ or, equivalently, that $P\left(\sup\limits_{M\geq l} I\left\{\|X_{i_r}\|_\infty \geq M\right\} > \varepsilon\right) \longrightarrow 0$, when $l \to \infty$. Since $\sup\limits_{M\geq l} I\left\{\|X_{i_r}\|_\infty \geq M\right\} \leq I\left\{\|X_{i_r}\|_\infty \geq l\right\}$, $P\left(\sup\limits_{M\geq l} I\left\{\|X_{i_r}\|_\infty \geq M\right\} > \varepsilon\right) \leq P\left(I\left\{\|X_{i_r}\|_\infty \geq l\right\} > \varepsilon\right) = P\left(\|X_{i_r}\|_\infty \geq l\right) \longrightarrow 0$, when $l \longrightarrow \infty$. Therefore, $\sup\limits_{\|x\|_\infty \geq M} S_n^{(j}(x) \xrightarrow{a.s.} 0$, when $M \to \infty$ and $\sup\limits_{\|x\|_\infty \geq M} S_{n,J}(x) \xrightarrow{a.s} 0$, when $M \to \infty$.

(iv) $S^{(j}$ is an upper- semicontinuous function: if $x_n \to x$,

$$\limsup_{n\to\infty} S^{(j}(x_n) = \limsup_{n\to\infty} P(x_n \in R(X_1, ...X_j)) \leq P\left(\limsup_{n\to\infty} \{x_n \in R(X_1, ...X_j)\}\right)$$
$$\leq P(x \in R(X_1, ..., X_j)) = S^{(j}(x).$$

S_J is also upper- semicontinuous:

$$\limsup_{n\to\infty} S_J(x_n) = \limsup_{n\to\infty} \sum_{j=2}^{J} S^{(j}(x_n) \leq \sum_{j=2}^{J} \limsup_{n\to\infty} S^{(j}(x_n) \leq \sum_{j=2}^{J} S^{(j}(x) = S_J(x).$$

To establish that $S_J(x)$ is continuous under distributions P with continuous marginals, it is enough to prove it for $S_n^{j)}(x)$. Assume that x_n converges to x. Since

$$\{x_n \in R(X_1, ..., X_j)\} \subset \{x \in R(X_1, ..., X_j)\} \cup \{x_n \in R(X_1, ..., X_j) \cap x \notin R(X_1, ..., X_j)\},$$

it follows that

$$|P(x_n \in R(X_1, ..., X_j)) - P(x \in R(X_{1,...,}X_j))| \leq P(x_n \in R(X_1, ..., X_j) \cap x \notin R(X_1, ..., X_j)) +$$
$$+ P(x_n \notin R(X_1, ..., X_j) \cap x \in R(X_1, ..., X_j)).$$

Therefore,

$$\begin{aligned} \left|S^{(j}(x_n) - S^{(j}(x)\right| &= |P(x_n \in R(X_1, ..., X_j)) - P(x \in R(X_1, ..., X_j))| \\ &\leq P(x_n \in R(X_1, ..., X_j), x \notin R(X_1, ..., X_j)) + \\ &\quad + P(x_n \notin R(X_1, ..., X_j), x \in R(X_1, ..., X_j)) \leq 2P(A_n), \end{aligned}$$

where $A_n = \bigcup\limits_{i=1}^{j} \bigcup\limits_{k=1}^{d} \{\min\{x(k), x_n(k)\} \leq X_i(k) \leq \max\{x(k), x_n(k)\}\}$ and, since the marginals of P are absolutely continuous, $P(A_n)$ converges to zero when $x_n \to x$.

Proof of Proposition 1: $S^{(j}(x)$ is a U-statistic with kernel $h_j(x_1, x_2, ..., x_j) = I\{x \in R(x_1, x_2, ..., x_j)\}$ (see, e.g., Serfling, 1980). We prove that the sum of U-statistics of orders $\{2, ..., J\}$ is a U-statistic of order J. Let $H_J = \sum_{j=2}^{J} H^{j)}(x_1, ..., x_n)$ with

$$H^{j)}(x_1, ..., x_n) = \binom{n}{j}^{-1} \sum_{1 \le i_1 < ... < i_j \le n} h_j(x_{i_1}, ..., x_{i_j}). \tag{10}$$

We have that

$$H_J = \frac{\frac{\binom{n}{J}}{\binom{n}{2}} \sum_{1 \le i_1 < i_2 \le n} h_2(x_{i_1}, x_{i_2}) + ... + \frac{\binom{n}{J}}{\binom{n}{j}} \sum_{1 \le i_1 < ... < i_j \le n} h_j(x_{i_1}, ..., x_{i_j}) + ... + \sum_{1 \le i_1 < ... < i_J \le n} h_J(x_{i_1}, ..., x_{i_J})}{\binom{n}{J}}, \tag{11}$$

and, in general, the coefficient of the j-th term in H_J is $\frac{\binom{n}{J}}{\binom{n}{j}} = \frac{j!(n-j)!}{J!(n-J)!}$. On the other hand, consider the projection $h^{pj} : \mathbb{R}^J \to \mathbb{R}$ given by $h^{pj}(x_1, ..., x_J) = h_j(x_1, ..., x_j)$.

The symmetrized version of h^{pj} is $h^{*j}(x_1, ..., x_J) = \frac{\sum_* h^{pj}(x_{\sigma(1)}, ..., x_{\sigma(J)})}{J!}$, where the sum is over all permutations of indexes $1, ..., J$. The function h^{*j} is a symmetric function defined on $\mathbb{R}^J$. It holds that

$$\sum_{1 \le i_1 < i_2 < ... < i_J \le n} h^{*j}(x_{i_1}, ..., x_{i_J}) = \frac{\binom{n-j}{J-j} j!(J-j)! \sum_{1 \le i_1 < i_2 < ... < i_j \le n} h_j(x_{i_1}, ..., x_{i_j})}{J!} \tag{12}$$

and

$$\sum_{1 \le i_1 < i_2 < ... < i_j \le n} h_j(x_{i_1}, ..., x_{i_j}) = \frac{J!}{\binom{n-j}{J-j} j!(J-j)!} \sum_{1 \le i_1 < i_2 < ... < i_J \le n} h^{*j}(x_{i_1}, ..., x_{i_J}). \tag{13}$$

Plugging (13) in equation (11), we obtain

$$H_J = \frac{\sum_{1 \le i_1 < ... < i_J \le n} h^{*2}(x_{i_1}, ..., x_{i_J}) + ... + \sum_{1 \le i_1 < ... < i_J \le n} h_J(x_{i_1}, ..., x_{i_J})}{\binom{n}{J}}.$$

Defining the symmetric kernel $h^{def} : \mathbb{R}^J \to \mathbb{R}$ as

$$h^{def}(x_1, ..., x_J) = h^{*2}(x_1, ..., x_J) + ... + h^{*j}(x_1, ..., x_J) + ... + h_J(x_1, ..., x_J),$$

H_J is a U-statistic of order J with kernel h^{def},

$$H_J(x_1, ..., x_n) = \frac{\sum_{1 \le i_1 < i_2 < ... < i_J \le n} h^{def}(x_{i_1}, ..., x_{i_J})}{\binom{n}{J}}.$$

The expected value of H_J is

$$E^J\left(H_J\right) = \frac{\sum\limits_{1\leq i_1<i_2<...<i_J\leq n} E^J \left(h^{def}(X_{i_1},...,X_{i_J})\right)}{\binom{n}{J}} = E^J\left(h^{def}(X_1,...,X_J)\right)$$
$$= E^J(h^{*2}(X_1,...,X_J)) + ... + E^J\left(h^{*j}(X_1,...,X_J)\right) + ... + E^J\left(h_J(X_1,...,X_J)\right)$$

and since the sample is identically distributed the expectation of a general term is

$$E^J\left(h^{*j}(X_1,...,X_J)\right) = E^J\left(h^{pj}(X_1,...,X_J)\right) = E^J\left(h_j(X_1,...,X_j)\right) = E^j\left(h_j(X_1,...,X_j)\right).$$

Then $E^J\left(H_J\right) = E^2(h_2(X_1,X_2)) + ... + E^J(h_J(X_1,...,X_J))$.

Proof of Proposition 2: Assume that P is antipodally symmetric about zero (i.e., X and $-X$ are equally distributed). Let $X_s = \{X_1,...,X_n\}$ be a random sample from P and let $\widehat{m}_{n,J}(X_s)$ be the sample median based on $S_{n,J}$. Due to invariance under $T(X) = AX + b$, where A is a diagonal and invertible matrix, it can be shown that $\widehat{m}_{n,J}(-X_s) = -\widehat{m}_{n,J}(X_s)$, where $-X_s = -X_1,...,-X_n$. Since X and $-X$ are identically distributed, $\widehat{m}_{n,J}(-X_s)$ and $\widehat{m}_{n,J}(X_s)$ are also identically distributed; hence, $\widehat{m}_{n,J}(X_s)$ and $-\widehat{m}_{n,J}(X_s)$ are identically distributed and $\widehat{m}_{n,J}(X)$ is symmetric around zero.

Proof of Theorem 2: (i) By triangular inequality,

$$\sup_{x\in R^d}\left|S_{n,J}(x) - S_J(x)\right| \leq \sup_{x\in R^d}\left|\sum_{j=2}^{J}\left(S_n^{(j}(x) - S^{(j}(x)\right)\right| \leq \sum_{j=2}^{J}\sup_{x\in R^d}\left|S_n^{(j}(x) - S^{(j}(x)\right|$$

and it is enough to show that

$$\sup_{x\in R^d}\left|S_n^{(j}(x) - S^{(j}(x)\right| \xrightarrow{a.s.} 0, \quad \text{when} \quad n\to\infty. \tag{14}$$

Since $S_n^{(j}(\cdot)$ is a U-process, we can use the results in Arcones and Giné (1993) establishing that if the family of functions $\mathcal{F}$ is VC with envelope G and $P^m G < \infty$ then $\|U_m^n - P^m\|_{\mathcal{F}} \xrightarrow{a.s.} 0$. In this case $\mathcal{F} = \left\{R_x :\ x\in\mathbb{R}^d\right\}$ is a VC class of functions (see Corollary 6.7 in Arcones and Giné, 1993) and the envelope G is the function identically equal to one in its domain $(\mathbb{R}^d)^j$. The second part follows by adapting Theorem 6.9 in Arcones and Giné (1993) to the band depth S_J. *(iii)* follows as the second part of Theorem 5 in Liu (1990) since $S_J(\cdot)$ verifies that theorem conditions.

Proof of Proposition 3: We have that

$$S^{(j}(x_m) = P(\min_{i=1,\ldots,j}\{X_{m,i}(q_k)\} \leq x_m(q_k) \leq \max_{i=1,\ldots,j}\{X_{m,i}(q_k)\}\,, \, k=1,...,m) =$$
$$= P\left(\bigcap_{k=1}^{m}\{\min_{i=1,\ldots,j}\{X_{m,i}(q_k)\} \leq x_m(q_k) \leq \max_{i=1,\ldots,j}\{X_{m,i}(q_k)\}\}\right).$$

Let $A_m = \bigcap_{k=1}^{m}\{\min_{i=1,\ldots,j}\{X_{m,i}(q_k)\} \leq x_m(q_k) \leq \max_{i=1,\ldots,j}\{X_{m,i}(q_k)\}\}$. Thus $A_{m+1} \subset A_m$ for each m and $\lim_{m\to\infty} P(A_m) = P(\bigcap_{m=1}^{\infty} A_m)$. Hence,

$$\lim_{m\to\infty} S^{(j}(x_m) = P(\bigcap_{k=1}^{\infty}\{\min_{i=1,\ldots,j}\{X_i(q_k)\} \leq x(q_k) \leq \max_{i=1,\ldots,j}\{X_i(q_k)\}\})$$
$$= P\left(\min_{i=1,\ldots,j}\{X_i(t)\} \leq x(t) \leq \max_{i=1,\ldots,j}\{X_i(t)\}\,, \, t \in I\right) = S^{(j}(x),$$

what implies that $\lim_{m\to\infty} S_J(x_m) = S_J(x)$.

Proof of Theorem 3: *(i)* follows from the definition. The proof of *(ii)* is analogous to the finite-dimensional case. *(iii)* S_J is upper-semicontinuous (following the finite-dimensional argument). For S_J continuity, we express it in terms of the rational numbers in I,

$$S^{(j}(x) = P\left(\min_{i=1,\ldots,j}\{X_i(q)\} \leq x(q) \leq \max_{i=1,\ldots,j}\{X_i(q)\}\,, \quad q \in \mathbb{Q} \cap I\right).$$

For any sequence $\{x_n\}$ in $C(I)$ such that $x_n \overset{\|\cdot\|_\infty}{\longrightarrow} x$,

$$\begin{aligned}
\left|S^{(j}(x_n) - S^{(j}(x)\right| &= \left|P\left(G(x_n) \subset V(X_1,...,X_j)\right) - P\left(G(x) \subset R(X_1,...,X_j)\right)\right| \\
&\leq P\left(G(x_n) \subset V(X_1,...,X_j), G(x) \subsetneq V(X_1,...,X_j)\right) + \\
&\quad + P\left(G(x_n) \subsetneq V(X_1,...,X_j), G(x) \subset V(X_1,...,X_j)\right) \\
&\leq 2P(A_n),
\end{aligned}$$

where

$$A_n = \bigcup_{i=1}^{j} \bigcup_{q\in\mathbb{Q}\cap I} \{\min\{x(q), x_n(q)\} \leq X_i(q) \leq \max\{x(q), x_n(q)\}\}.$$

If the marginal distributions of P are absolutely continuous, then, for $x \in C(I)$,

$$P\left(\bigcup_{q\in\mathbb{Q}\cap I}\{X(q) = x(q)\}\right) = 0.$$

Therefore, $P(A_n)$ converges to zero when n tends to infinity, what implies that $S^{(j}$ is continuous and this is inherited by S_J.

Proof of Theorem 4: Without loss of generality, consider $I = [0,1]$. We have that

$$\sup_{x \in E} |S_{n,J}(x) - S_J(x)| \leq \sup_{x \in E,\ \|x\|_\infty < M} |S_{n,J}(x) - S_J(x)| + \sup_{x \in E,\ \|x\|_\infty \geq M} |S_{n,J}(x) - S_J(x)|.$$

Due to *(ii)* in Theorem 3, the second term converges to zero when M tends to infinity. Therefore, it is sufficient to prove that for M large enough, the first term converges to zero when $n \to \infty$. By Ascoli-Arzela's Theorem (see, e.g., Pollard, 1984), the closure $\overline{E}_M$ of $E_M = \{x \in E : \|x\|_\infty < M\}$ is compact and, hence, totally bounded. Then, for every $\varepsilon > 0$, there exists a finite set $\{x_1, x_2, ..., x_{N_\varepsilon}\}$ of functions such that $\overline{E}_M \subseteq \bigcup_{m=1}^{\infty} B(x_m, \varepsilon)$. By Theorem 3. *(iii)*, to establish that $\sup_{x \in \overline{E}_M} |S_{n,J}(x) - S_J(x)| \xrightarrow{a.s.} 0$, when $n \to \infty$, it is enough to prove $\max_{x \in \{x_m\}_{m=1}^{N_\varepsilon}} |S_{n,J}(x) - S_J(x)| \xrightarrow{a.s.} 0$, when $n \to \infty$. But, we have that

$$\begin{aligned} P\left\{\max_{x \in \{x_m\}_{m=1}^{N_\varepsilon}} |S_{n,J}(x) - S_J(x)| \geq \epsilon\right\} &\leq N_\varepsilon \max_{x \in \{x_m\}_{m=1}^{N_\varepsilon}} P\left\{|S_{n,J}(x) - S_J(x)| \geq \epsilon\right\} \\ &\leq N_\varepsilon \max_{x \in \{x_m\}_{m=1}^{N_\varepsilon}} \frac{E\left[|S_{n,J}(x) - S_J(x)|^4\right]}{\varepsilon^4} = O(n^{-2}), \end{aligned}$$

and Borel-Cantelli's Lemma provides the result in *(i)*. To prove *(ii)*, we use that $m_n \xrightarrow{a.s.} m$ if, and only if, $P\left(\sup_{n \geq l} \|m_n - m\|_\infty \geq \varepsilon\right) \underset{l \to \infty}{\longrightarrow} 0$. $(E, \|\cdot\|_\infty)$ is a metric space and $S_J(\cdot)$ is an upper-semicontinuous function in E and verifies that $\sup_{\|x\|_\infty \geq M, x \in E} S_J(x) \underset{M \to \infty}{\longrightarrow} 0$; thus, the proof is analogous to that of Theorem 2 *(ii)*.

REFERENCES

Arcones, M.A. and Giné, E. (1993), Limit theorems for U-processes, *The Annals of Probability*, **21,** 1494-1542.

Brown, B. and Hettmansperger, T. (1989), The affine invariant bivariate version of the sign test, *Journal of the Royal Statistical Society B*, **51**, 117-125.

Fraiman, R. and Meloche, J. (1999), Multivariate L-estimation, *Test,* **8,** 255-317.

Fraiman, R. and Muniz, G. (2001), Trimmed means for functional data, *Test*, **10,** 419-440.

Hettmansperger, T. and Oja, H. (1994), Affine invariant multivariate multisample sign tests, *Journal of the Royal Statistical Society B*, **56**, 235-249.

Inselberg, A. (1981), N-dimensional graphics, Part I-lines and hyperplanes, in IBM LASC, Technical Report, G320-27111.

Inselberg, A. (1985), The plane parallel coordinates, Invited paper, *Visual Computer*, **1**, 69-91.

Inselberg, A., Reif, M. y Chomut, T. (1987), Convexity algorithms in parallel coordinates, *Journal of ACM*, **34**, 765-801.

Li, J. and Liu, R. (2004), New Nonparametric Tests of Multivariate Locations and Scales using Data Depth, *Statistical Science*, **19**, 4, 686-696.

Liu, R. (1990), On a notion of data depth based on random simplices, *The Annals of Statistics*, **18**, 405-414.

Liu, R. (1995), Control charts for multivariate processes, *Journal of the American Statistical Association*, **90,** 1380-1388.

Liu, R., Parelius, J.M. and Singh, K. (1999), Multivariate analysis by data depth: Descriptive statistics, graphics and inference, *Annals of Statistics*, **27**, 783-858.

Liu, R. and Singh, K. (1993), A quality index based on data depth and multivariate rank test, *Journal of the American Statistical Association*, **88**, 257-260.

Mahalanobis, P. C. (1936), On the generalized distance in statistics, *Proceedings of National Academy of Science of India*, **12**, 49-55.

Oja, H. (1983), Descriptive statistics for multivariate distributions, *Statistics and Probability Letters*, **1**, 327-332.

Pollard, D. (1984), *Convergence of stochastic processes.* Springer Verlag, New York.

Ramsay, J.O. and Silverman, B.W. (2005), *Functional Data Analysis.* Second Edition.

Springer Verlag, New York.

Rousseeuw, P. and Hubert, M. (1999), Regression depth (with discussion), *Journal of the American Statistical Association*, **4**, 388-433.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Singh, K. (1991), A notion of majority depth, Unpublished document.

Tukey, J. (1975), Mathematics and picturing data, *Proceedings of the 1975 International Congress of Mathematics*, **2**, 523-531.

Vardi, Y. and Zhang, C.-H.(2001), The multivariate L1-median and associated data depth, *Proceedings of the National Academy of Science USA*, **97**, 1423-1426.

Wegman, E. (1990), Hyperdimensional data analysis using parallel coordinates, *Journal of the American Statistical Association*, **85**, 664-675.

Wood, A.T.A. and Chan, G. (1994), Simulation of stationary gaussian processes in $C[0,1]$, *Journal of Computational and Graphical Statistics*, **3**, 409-432.

Yeh, A. and Singh, K. (1997), Balanced confidence sets based on the Tukey depth, *Journal of the Royal Statistical Society Series B*, **3**, 639-652.

Zuo, Y. (2003), Projection based depth functions and associated medians, *The Annals of Statistics*, **31**, 1460-1490.

Zuo, Y. and Serfling, R. (2000), General notions of statistical depth function, *The Annals of Statistics*, **28**, 461-482.