The Random Tukey Depth

J.A. Cuesta-Albertos and A. Nieto-Reyes *
Departamento de Matemáticas, Estadística y Computación
Universidad de Cantabria, Spain

April 18, 2008

Abstract

The computation of the Tukey depth, also called halfspace depth, is very demanding, even in low dimensional spaces, because it requires that all possible one-dimensional projections be considered. A random depth which approximates the Tukey depth is proposed. It only takes into account a finite number of one-dimensional projections which are chosen at random. Thus, this random depth requires a reasonable computation time even in high dimensional spaces. Moreover, it is easily extended to cover the functional framework. Some simulations indicating how many projections should be considered depending on the kind of problem, sample size and dimension of the sample space among others are presented. It is noteworthy that the random depth, based on a very low number of projections, obtains results very similar to those obtained with the Tukey depth.

Key words and phrases: Random Tukey depth, one-dimensional projections, multidimensional data, functional data, homogeneity test, supervised classification.

1 Introduction

Here we analyze a conceptually simple and easy to compute multidimensional depth that can be applied to functional problems and that provides results comparable to those obtained with more involved depths. Depths are intended to order a given set in the following way: if a datum is moved toward the center of the data cloud, then its depth increases and if the datum is moved toward the outside, then its depth decreases. More generally, given a probability distribution P defined in a multidimensional (or even infinite-dimensional) space $\mathcal{X}$, a depth tries to order the points in $\mathcal{X}$ from the "center (of P)" to the "outer (of P)". Obviously, this problem includes data sets if we consider P as the empirical distribution associated to the data set at hand. Thus, in what follows, we will always refer to the depth associated to a probability distribution P.

*Corresponding author: Alicia Nieto Reyes, Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Avda. los Castros s.n., 39005 SANTANDER, Spain. E-mail: alicia.nieto@unican.es. Phone: (+34) 942 201369. Fax: (+34) 942 201402.

In the one-dimensional case, it is reasonable to order the points using the order induced by the function

$$x \to D_1(x, P) := \min\{P(-\infty, x], P[x, \infty)\}. \tag{1}$$

Thus, the points are ordered following the decreasing order of the absolute values of the difference between their percentiles and 50, and the deepest points are the medians of P.

Several multidimensional depths have been proposed (see, for instance, the recent book [7]) but here we are mainly interested in the *Tukey (or halfspace) depth* (see [12]). If $x \in \mathbb{R}^p$, then the Tukey depth of x with respect to P, $D_T(x, P)$, is the minimal probability which can be attained in the closed halfspaces containing x. According to [13], this depth behaves very well in comparison with various competitors. An equivalent definition of $D_T(x, P)$ is the following: given $v \in \mathbb{R}^p$, let Π_v be the projection of $\mathbb{R}^p$ on the one dimensional subspace generated by v. Thus, $P \circ \Pi_v^{-1}$ is the marginal of P on this subspace, and it is obvious that

$$D_T(x, P) = \inf\{D_1(\Pi_v(x), P \circ \Pi_v^{-1}) : v \in \mathbb{R}^p\},\ x \in \mathbb{R}^p. \tag{2}$$

I.e., $D_T(x, P)$ is the infimum of all possible one-dimensional depths of the one-dimensional projections of x, where those depths are computed with respect to the corresponding (one-dimensional) marginals of P.

Some other depths based on the consideration of all possible one-dimensional projections have been proposed (see, for instance, [14]). We consider that what follows could be applied to all of them, but we have chosen the Tukey depth to test it specifically. Perhaps the most important drawback of the Tukey depth is the required computational time. This time is more or less reasonable if $p = 2$, but it becomes prohibitive even for $p = 8$ [11, pag. 54]. To reduce the time, in [15, pag. 2234] it is proposed that their values be approximated using randomly selected projections. On the other hand, in [2], a random depth is defined. In this last paper, given a point x, the authors propose that a finite number of vectors $v_1, ..., v_k$ be chosen at random, and the depth of x then be taken as the mean of the values $D_1(\Pi_{v_i}(x), P \circ \Pi_{v_i}^{-1})$, $i = 1, ..., k$.

Our approach follows more closely the suggestion in [15]: we simply replace the infimum in (2) by a minimum over a finite number of randomly chosen projections, obtaining a random approximation to the Tukey depth. Moreover, in Section 2.1 (Theorem 2.3) we show that, with the exception of the invariance under non-singular linear transformations, this approximation satisfies the definition of depth given in [13] (but with the convergence being in probability). Section 2.1 closes with Theorem 2.6, which proves the consistency of the random Tukey depth. The proofs of these theorems are included in the Appendix. One of the main advantages of the random Tukey depth is that it can be extended to infinite-dimensional functional spaces. This possibility is studied in Section 2.2.

The difficulty of the random Tukey depth as an approximation to the Tukey depth is to find the number of random projections required to obtain a good approximation. This question is addressed in Section 3. This number could depend on the kind of application of the depth in which we are interested as well as on the dimension of the underlying space and the size of the random sample we are using. However, the simulations carried

out in Section 3 suggest that a maximum of 250 randomly chosen projections are enough to satisfy a wide range of cases. Section 3 ends with a comparison of the time needed to compute the random Tukey depth and that required for the Mahalanobis depth. The computations have been carried out with MatLab. Computational codes are available from the authors upon request.

2 The random Tukey depth

In this section we define the random Tukey depth, demonstrate that, with the exception of the invariance under non-singular linear transformations, it satisfies the definition of statistical depth given in [13], show its consistency and analyze the possibility of extending it to cover infinite dimensional spaces.

2.1 Finite dimensional spaces

In this section $\mathcal{P}$ denotes the class of distributions on the Borel sets of $\mathbb{R}^p$ and P_X the distribution of a general random vector X. In addition, the symbols $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ respectively denote the usual norm and scalar product in $\mathbb{R}^p$. Now, let us formally define the random Tukey depth.

Definition 2.1 *Let $P \in \mathcal{P}$. Let $\nu \in \mathcal{P}$ absolutely continuous, and let $v_1, ..., v_k$ be independent and identically distributed random vectors with distribution ν. The random Tukey depth of $x \in \mathbb{R}^p$ with respect to P based on k random vectors chosen with ν is*

$$D_{T,k,\nu}(x, P) = \min\{D_1(\Pi_{v_i}(x), P \circ \Pi_{v_i}^{-1}) : i = 1, ..., k\},\ x \in \mathbb{R}^p.$$

In order to simplify the notation, and as long as this not create more confusion, we will delete the subscript ν in the notation and simply write $D_{T,k}$.

Obviously, $D_{T,k}(x, P)$ is a random variable. It may seem somewhat paradoxical to take a random quantity to measure the depth of a point, which is inherently not-random. We have taken this approach for two reasons. Firstly, Theorem 4.1 in [1] shows that if P and Q are probability distributions on $\mathbb{R}^p$, such that the absolute moments $m_k := \int \|x\|^k dP(x)$ are finite and satisfy $\sum_{k\geq 1} m_k^{-1/k} = \infty$, ν is an absolutely continuous distribution on $\mathbb{R}^p$ and

$$\nu\{v \in \mathbb{R}^p : P \circ \Pi_v^{-1} = Q \circ \Pi_v^{-1}\} > 0,$$

then $P = Q$. In other words, if we have two different distributions and we randomly choose a marginal of them, those marginals are almost surely different. According to this result, one randomly chosen projection is enough to distinguish between two p-dimensional distributions. Since the depths determine one-dimensional distributions, a depth computed on just one random projection allows us to distinguish between two distributions.

Secondly, if the support of ν is $\mathbb{R}^p$, and, for every k, $\{v_1, ..., v_k\} \subset \{v_1, ..., v_{k+1}\}$, then

$$D_{T,k}(x, P) \geq D_{T,k+1}(x, P) \to D_T(x, P), \quad \text{a.s.} \tag{3}$$

Therefore, if we choose k large enough, the effect of the randomness in $D_{T,k}$ will be negligible. Of course, it is of interest to find how large k must be; values of k that are too large would make this definition useless. We will analyze that point in Section 3.

Here, we will show that, for every k, $D_{T,k}$ a.s. satisfies the last three properties of the definition of statistical depth. This definition consists of four key properties desirable for depths. These properties were first conceived and used by Liu in [6] to justify the simplicial depth as a functions of data depth. Subsequently they have been adapted in [13] as the key properties required for any general depth function. They are affine invariance, maximality at center, monotonicity relative to deepest point and vanishing at infinity. In [13], it is shown that the Tukey depth satisfies this definition. Concerning the maximality at center, this states that, having a distribution with a unique center of symmetry (with respect to some notion of symmetry), the depth function should attain the maximum at this center.

Definition 2.2 *The bounded and nonnegative mapping $D(\cdot,\cdot) : \mathbb{R}^p \times \mathcal{P} \longrightarrow \mathbb{R}$ is called a* statistical depth function *if it satisfies the following properties:*

1. *$D(Ax + b, P_{AX+b}) = D(x, P_X)$ holds for any $\mathbb{R}^p$-valued random vector X, any $p \times p$ nonsingular matrix A and any $b \in \mathbb{R}^p$.*
2. *$D(\theta, P) = \sup_{x\in\mathbb{R}^p} D(x, P)$ holds for any $P \in \mathcal{P}$ having center θ.*
3. *For any $P \in \mathcal{P}$ having deepest point θ, $D(x, P) \leq D(\theta + \alpha(x - \theta), P)$ holds for $\alpha \in [0, 1]$.*
4. *$D(x, P) \to 0$ as $\|x\| \to \infty$, for each $P \in \mathcal{P}$.*

With regard to point 2, various notions of symmetry are possible, among them, central, angular and halfspace symmetry. As central symmetry implies angular, which implies halfspace, we will identify the center with the point of halfspace symmetry.

Theorem 2.3 *The random Tukey depth is a bounded and non-negative mapping which satisfies $D_{T,k,\nu}(x+b, P_{X+b}) = D_{T,k,\nu}(x, P_X)$, for any $\mathbb{R}^p$-valued random vector X and any $b \in \mathbb{R}^p$, as well as items 2 and 3 in Definition 2.2.*

Moreover, let $P \in \mathcal{P}$ and $k > 0$. If $\|x\| \to \infty$ with $x \in \mathbb{R}^p$, then $D_{T,k}(x, P)$ converges to zero in probability.

Remark 2.4 It is obvious $D_{T,k,A\nu}(Ax, P_{AX}) = D_{T,k,\nu}(x, P_X)$ for any $\mathbb{R}^p$-valued random vector X and any $p \times p$ nonsingular matrix A but it is not difficult to find examples such that $D_{T,k,\nu}(Ax, P_{AX}) \neq D_{T,k,\nu}(x, P_X)$.

Remark 2.5 The randomness involved in the definition of the Random Tukey depth, only affects to item 4 in Theorem 2.3. The problem is that it would be possible for all the k vectors to be included in the same hyperplane. In this case, it is obvious that property 4 is not satisfied, for example, for some sequence of points orthogonal to this hyperplane, if $D_{T,k}(0, P) > 0$.

Another desirable property for depths is that its sample version converges to the population counterpart. More generally, almost surely, $\sup_x |D(x, P_n) - D(x, P)| \to 0$ where P_n denotes the empirical distribution (i.e. if $x_1, ..., x_n$ is a random sample, $P_n[A] = \#(A \cap \{x_1, ..., x_n\})/n$). This property is satisfied by the Tukey depth (see [13]) and Theorem 2.6 shows that the random Tukey depth also enjoys this property.

Theorem 2.6 *Let $\nu \in \mathcal{P}$ and $v_1, ..., v_k$ be independent and identically distributed random vectors with distribution ν. Let $P \in \mathcal{P}$ and let $\{P_n\}$ be a sequence of empirical distributions computed on a random sample taken from P which is independent of the vectors $v_1, ..., v_k$.*

Then, conditionally on $v_1, ..., v_k$, we have

$$\sup_{x \in \mathbb{R}^p} |D_{T,k}(x, P_n) - D_{T,k}(x, P)| \to 0, \text{ almost surely [P].}$$

Remark 2.7 In Theorem 2.6, the almost surely convergence is with respect to the empirical samples taken from P and the random vectors employed in the computation of the depths are chosen independently of these samples. In fact, this result holds for every fixed vector in $\mathbb{R}^p$, whether randomly chosen with the distribution ν or not, the only condition being that it is independent of the random sample taken from P.

2.2 Infinite dimensional spaces

An interesting possibility of the random Tukey depth is that it can be straightforwardly extended to functional spaces. The only requirement of the main result in [1] is that the sample space be a separable Hilbert space. Thus, in this section we will assume that we are considering a distribution P defined on this kind of space. To fix ideas, we will handle the space, $I\!H$, of square-integrable functions in a given interval which, after rescaling, we can assume to be $[0, 1]$. Thus, $I\!H = L^2[0, 1]$ and given $f, g \in I\!H$ we have $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$ and $\|f\| = \langle f, f\rangle^{1/2}$.

The random Tukey depth is not a statistical depth in the functional case. The proofs for the invariance under translation and items 2 and 3 in Definition 2.2 (with obvious modifications such as replacing matrices with linear operators) are the same as in Theorem 2.3. However, the following example shows that item 4 fails in this case even for statistical convergences.

Example 2.8 Let $\{\delta_n\}_n \subset \mathbb{R}^+$ with $\lim_n \delta_n = 0$. Let $x_n \in I\!H$ such that $x_n(t) = 1/\delta_n$ if $t \in [0, \delta_n)$ and be zero if otherwise. Obviously, $\|x_n\| = \delta_n^{-1/2}$ and, then $\lim_n \|x_n\| = \infty$. Let us take ν equal to the distribution of the standard Brownian motion and let $P = \nu$. Obviously Theorem 4.1 in [1] works with this distribution. If X is a random element with distribution ν, then $\langle x_n, X\rangle$ converges to zero in probability because

$$\begin{aligned} E|\langle x_n, X\rangle| &= E\left|\int_0^{\delta_n} X(t)\delta_n^{-1}dt\right| \\ &\leq \int_0^{\delta_n} E|X(t)|\delta_n^{-1}dt = \int_0^{\delta_n} (2t/\pi)^{1/2}\delta_n^{-1}dt \leq (2\delta_n/\pi)^{1/2}, \end{aligned} \tag{4}$$

where the last equality holds because the distribution of $X(t)$ is $N(0,t)$. Thus, if $v_1, ..., v_k \in \mathbb{H}$ are randomly chosen with distribution ν, we have

$$\lim_n D_1(< v_i, x_n >, P \circ \Pi_{v_i}^{-1}) = D_1(0, P \circ \Pi_{v_i}^{-1}) = \max_x D_1(x, P \circ \Pi_{v_i}^{-1}) = 2^{-1},$$

because $P \circ \Pi_{v_i}^{-1}$ is a centered Gaussian distribution.

Thus, in this setting the following results hold. Their proofs appear in the Appendix.

Theorem 2.9 *The random Tukey depth is a bounded and non-negative mapping which satisfies $D_{T,k,\nu}(x + b, P_{X+b}) = D_{T,k,\nu}(x, P_X)$, for any $X, b \in \mathbb{H}$, as well as items 2 and 3 in Definition 2.2.*

Theorem 2.10 *Let $v_1, ..., v_k \in \mathbb{H}$. Let P be a probability distribution on $\mathbb{H}$, and let $\{P_n\}$ be a sequence of empirical distributions computed on a random sample taken from P which is independent of the vectors $v_1, ..., v_k$.*

Then, conditionally on $v_1, ..., v_k$, we have

$$\sup_{x \in \mathbb{R}^p} |D_{T,k}(x, P_n) - D_{T,k}(x, P)| \to 0, \text{ almost surely [P].}$$

With regard to the number of random directions to take, we follow the same procedure as in the finite dimensional case, i.e., the choice of method depends on the kind of problem at hand, for instance with bootstrap or cross-validation. However, in this setting we have an additional problem. In the finite dimensional case, it seems reasonable to choose the random directions using the uniform distribution on the sphere because of its invariance properties. Regrettably, in infinite dimensional spaces, there is no distribution with such good properties, making the selection of the random directions more arduous. One interesting possibility is to choose the distribution according on the problem. This way a problem-specific procedure would be designed to select a distribution with some optimality properties. This work is currently in progress. Some related results on the selection of referential measures in functional spaces appear in [4].

3 How many random projections? Testing homogeneity

Obviously, Theorem 4.1 in [1] also holds if ν is a probability distribution absolutely continuous with respect to the surface measure on the unit sphere in $\mathbb{R}^p$. In this section, we fix ν to be the uniform distribution on the unit sphere to analyze the question of the selection of k. Our proposal is to make this selection depending on the problem we have at hand; for instance with bootstrap (as in Section 3.1) or with cross-validation in other settings. However, it is good to first think about the range in which to look for this value.

The obvious way to do this is to make some comparisons between D_T and $D_{T,k}$ for several dimensions, sample sizes and distributions; however, the long computational times required to obtain D_T make those comparisons impractical. Instead of doing these comparisons, we have chosen situations in which the deepness of the points are clearly defined

and can easily be computed with a different depth. A comparison of the results is carried out in Section 3.1.

If P is an elliptical distribution with centralization parameter μ and dispersion matrix Σ, then P is centrally symmetric around μ. It seems that every reasonable depth should consider those points at the same Mahalanobis distance of μ to have the same depth, and that differences in depth should correspond with differences in Mahalanobis distance of μ. In this situation, every depth should be a monotone function of the Mahalanobis depth [9], where, given $x \in \mathbb{R}^p$, this depth is

$$D_M(x, P) := \frac{1}{1 + (x - \mu)^t \Sigma^{-1} (x - \mu)}. \tag{5}$$

Therefore, we can have an idea about the right k in $D_{T,k}$ as follows: if P is elliptical, $D_T(\cdot, P)$, is a monotone function of $D_M(\cdot, P)$. Thus, from (3), the larger the k, the larger the resemblance between $D_{T,k}(\cdot, P)$ and a monotone function of $D_M(\cdot, P)$. However, there should exist a value k_0 from which this resemblance starts to stabilize. This is the value for k we are looking for.

Given that depths only try to rank points according to their closeness to the center of P, it is logical to measure the resemblance between $D_{T,k}(\cdot, P)$ and $D_M(\cdot, P)$ looking only at the ranks of the points. This is equivalent to using the Spearman correlation coefficient, ρ. The resemblance that we handle here is

$$r_{k,P} := \rho \left(D_{T,k}(X, P), D_M(X, P) \right), \tag{6}$$

where X is a random variable with distribution P. If P is an elliptical distribution, then the function $k \to r_{k,P}$ is strictly increasing. We try to identify the point k_0 from which the increments become negligible.

However, in practice, we will not have a distribution P, but a random sample $x_1, ..., x_n$ taken from P. This leads us to replace P in (6) by the empirical distribution P_n. To illustrate the behavior of the function $k \to r_{k,P_n}$, we have represented it for different distributions, sample sizes and dimensions in Figure 1. In this figure, the first column corresponds to centered Gaussian distributions having covariance matrices with ones on the diagonal and 0.9 in all positions off-diagonal. The remaining columns in Figure 1 represent, from left to right, standard Gaussian distributions, distributions with independent double exponential marginals and distributions with independent Cauchy marginals.

Dimensions and sample sizes vary in rows. We consider, from top to bottom, sample sizes $n = 25, 100$ for $\mathbb{R}^2$, $n = 50, 100$ for $\mathbb{R}^8$ and $n = 100, 500$ for $\mathbb{R}^{50}$. The case $n = 100$ (second, fourth and fifth rows) can be used to see how the dimension affects the function for a fixed sample size. The last row is different. In this row we take advantage of the fact that we know the exact covariance matrix of the theoretical distribution; in row number seven Mahalanobis depth is computed with the exact value of Σ. In this case, we have taken $n = 500$ in $\mathbb{R}^{50}$.

A last comment is related to the computation of the location center and the dispersion matrix (except in the last row) of P_n, to be used in D_M. Those parameters should depend

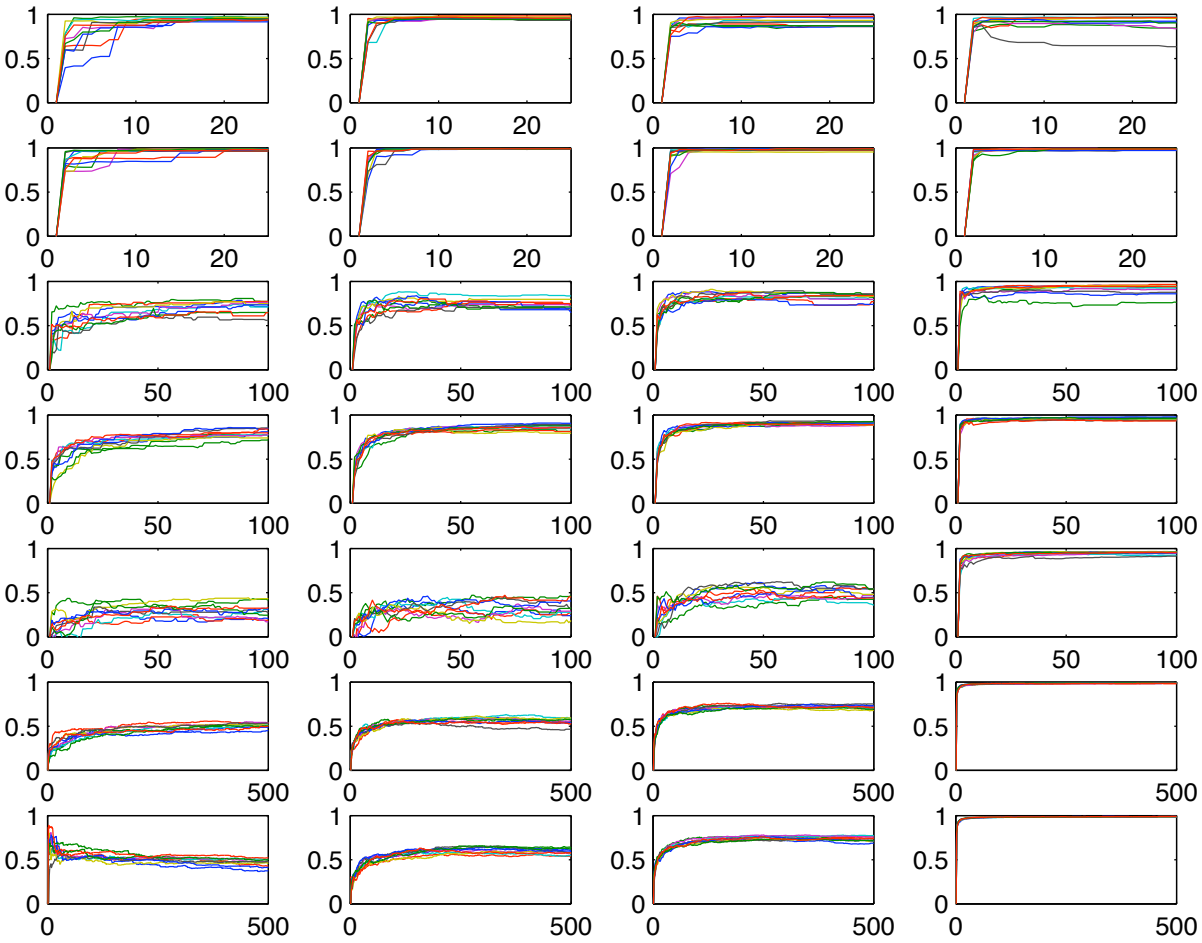

Figure 1: Representation of the function $k \to r_{k,P_n}$ defined in (6) for several dimensions, sample sizes and distributions. The underlying models are described in the text.

on the distribution which generated the sample. Hence, the covariance matrix is an appropriate parameter in the Gaussian and exponential case. However, it is not adequate for the Cauchy distribution, where we have identified Σ with the robust covariance matrix proposed in [10, page 206]. Furthermore, we have replaced μ by the sample mean in the Gaussian case and by the coordinate-wise median in the exponential and Cauchy settings.

In the graphs k varies in set $\{1, ..., 25\}$ in the first and second rows, in $\{1, ..., 100\}$ in the third, fourth and fifth rows, and in $\{1, ..., 500\}$ in the last two rows. Moreover, there are no obvious differences between using the theoretical covariance matrices or their estimation nor between using the case of independent marginals or dependent ones. We have checked more cases (not shown here) with similar results, among which we have analyzed some intermediate dimensions, other sample sizes, and dispersion matrices with 0.5 in all off-diagonal elements for the Gaussian, exponential and Cauchy distributions. It seems that the graphs stabilize for $k \leq 10$ if $p = 2$, $k \leq 60$ if $p = 8$ and $k \leq 250$ if

$p = 50$. These values are suitable values for computations and, of course, are well below those usually used to compute the Tukey depth.

As shown only in the graph in the lower left corner of Figure 1, since P_n does not follow the model exactly, the function r_{k,P_n} is not necessarily increasing and in fact may sometimes, after an initial increase, start to decrease. We believe this occurs because, in spite of $D_T(x, P_n) \neq D_M(x, P_n)$, they are similar. Furthermore, as k increases, $D_{T,k}(x, P_n)$ approaches $D_T(x, P_n)$. Thus, while $D_{T,k}(x, P_n)$ is not too close to $D_T(x, P_n)$, increments in k mean more similarity between $D_{T,k}(x, P_n)$ and $D_M(x, P_n)$. However, from a certain point on, getting closer to $D_T(x, P_n)$ implies to moving away from $D_M(x, P_n)$.

3.1 Testing homogeneity

Our goal in this subsection is to show how the random Tukey depth, with values for k of the order suggested by Figure 1, provides results which are similar to those obtained in practice with the Tukey depth. To this end, we are going to reproduce the simulation study carried out in [8], where the authors apply depth measures to test differences in homogeneity between two 2-dimensional distributions. In only considering the distributions in dimension $p = 2$, our goal is to compare the results of the random Tukey depth with those of the Tukey depth, which can only be computed relatively quickly when $p = 2$.

Let us begin by giving a brief description of the problem and the procedure. Additional details can be found in [8]. Assume that we have two random samples $\{X_1, ..., X_{n_1}\}$ and $\{Y_1, ..., Y_{n_2}\}$ taken from the centered distributions P and Q respectively. Let us assume that those distributions coincide except for a scale factor, i.e., we assume that there exists $r > 0$ such that the r.v.'s $\{rX_1, ..., rX_{n_1}\}$ and $\{Y_1, ..., Y_{n_2}\}$ are identically distributed. The problem consists in testing, at level α, the hypotheses:

$$\begin{aligned} H_0 &: \quad r = 1 \text{ (both scales are the same)} \\ H_a &: \quad r > 1 \text{ (Q has a larger scale).} \end{aligned}$$

Under the alternative, the observations in the second sample should appear in the outside part of the joint sample $\{X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}\}$, and, consequently, should have lower depths than the points in the first sample. Thus, it is possible to test H_0 against H_a by computing the depths of the points $\{Y_1, ..., Y_{n_2}\}$ in the joint sample, replacing them by their ranks and rejecting H_0 if those ranks are small. The Wilcoxon rank-sum test can be used to test when the ranks of the points $\{Y_1, ..., Y_{n_2}\}$ are small. In [8] several possibilities are proposed to break the ties. We have tried all of them, with no relevant differences. Thus, we have chosen random tie-breaking as the only method to be shown here.

To select the number of random projections, we have come up with the following process. According to the graphs in Figure 1, the number of required directions should be less than 10. Just to be on the safe side, we begin by selecting 25 vectors at random, $v_1, ..., v_{25}$. Our aim is to choose a subset $v_1, ..., v_k$, with $k \leq 25$. To do this we have applied the test 100 times to two bootstrap samples taken from $Z := \{X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}\}$ where the second bootstrap sample is modified so that it satisfies the alternative hypothesis for some

r in the grid $R := \{1.1, 1.2, 1.3, 1.5, 1.7, 2, 2.5, 3, 4, 5, 7, 10, 20, 40, 80, ...\}$. The process is as follows:

1. Draw two bootstrap samples, I and J, from Z, respectively with sizes n_1 and n_2.
2. Center I and J in median, separately.
3. Initialize $k = 1$ and $i = 1$.
4. Multiply the vectors in J by the i-th element in R.
5. Compute the random Tukey depth of the points in $I \cup J$ using the vectors $v_1, ..., v_k$.
6. If H_0 is rejected at level α, then keep record of k and finish.

   Else: if $k = 25$, then go back to point (4) with $k = 1$ and $i = i + 1$.

   Else: go back to point (5) with $k = k + 1$.

At the end of the bootstrap step, we have 100 values of k, which correspond to the number of vectors used the first time the homogeneity hypothesis was rejected in the bootstrap world. Equivalently, every $k = 1, ..., 25$ has associated the number of times, $n_k \geq 0$, in which the null hypothesis was rejected using the vectors in the set $\{v_1, ..., v_k\}$. Since we have done 100 bootstrap replications, it is obvious that $n_1 + ... + n_{25} = 100$. The precise subset to be used in the real-world test should be chosen based on the information provided by the probability distribution which gives mass $n_k/100$ to the point $k = 1, 2, \ldots, 25$. We have considered four possibilities: the mean, the median, the 80% percentile or even the maximum (i.e. the 100% percentile) of this distribution. We have repeated the procedure 5,000 times and we have not obtained significant differences in the rejection rates. Despite the low differences, the worst rates were obtained with the median and then with the mean. The rates were very close when employing the 80% percentile and the maximum.

With respect to the centering in median of the bootstrap samples, it should be noted that when there is at most one observation with the same value as the median, there is no problem. However, let us assume that the sample J contains several values (let us say $h > 1$) which coincide with its median, m_J. After centering, those values go to zero, and they are not modified when multiplying by a large value of r. If h is moderate, since those values remain in the inner part of the joint sample, the null hypothesis is never rejected in the bootstrap world. In order to avoid this undesirable behavior we have defined

$$J_l := \sup\{z \in J : z < m_J\} \text{ and } J_u := \inf\{z \in J : z > m_J\},$$

replaced those repeated values which coincide with the median by a random sample with size h taken with the uniform distribution in the interval $((J_l + m_J)/2, (J_u + m_J)/2)$ and centered the modified sample using the median computed in this modified sample which contains no repeated data.

In Table 3.1 we show the rate of rejections under these conditions when we carry out the test at the significance level $\alpha = .05$. The table also includes, between parenthesis,

the rejection rates when the random depth is replaced by the Tukey depth computed using 1,000 directions uniformly scattered on the upper halfspace. The distributions used in the simulations are the 2-dimensional standard Gaussian, and the double exponential and Cauchy with independent marginals. We have considered the values $r = 1, 1.2, 2$, and $n_1 = n_2 = n$ with $n \in \{20, 30, 100\}$, and have done 5,000 simulations for each combination of distribution, sample size and r. As explained, the rejection rates depend on the samples we have each time and not so much on the method used to select the number of random projections. Table 3.1 only contains the rejections obtained with the 80% percentile. Remaining tables are available from the authors upon request.

Table 3.1 *Rate of rejections in 5,000 simulations using the random Tukey depth (between parentheses, the rate with D_T) for the considered distributions, sample sizes and values of r. The dimension is $p = 2$. The significance level is .05.*

Sample size	Scale factor	Distribution					
		Cauchy		Gaussian		D. exponential	
$n = 20$	$r = 1$	.054	(.055)	.048	(.049)	.051	(.057)
	$r = 1.2$	.120	(.125)	.212	(.216)	.170	(.174)
	$r = 2$	.540	(.539)	.938	(.940)	.828	(.824)
$n = 30$	$r = 1$	.053	(.049)	.046	(.052)	.050	(.050)
	$r = 1.2$	.159	(.146)	.297	(.292)	.216	(.223)
	$r = 2$	.716	(.704)	.995	(.995)	.950	(.943)
$n = 100$	$r = 1$	.053	(.049)	.048	(.049)	.048	(.048)
	$r = 1.2$	.279	(.291)	.687	(.699)	.476	(.495)
	$r = 2$	.993	(.991)	1	(1)	1	(1)

The number of projections used varies with the method used to select them, the distribution and the sample size. In particular, they decrease with the sample size. Furthermore, the medians of the number of vectors used vary between 17–24 when selecting k as the maximum in the bootstrap procedure. The medians are 2 or 3 if k is selected with the median and oscillate between 2–5 if the employed method is the mean. When employing the 80% percentile, the medians are between 6–8 for the sample size $n = 20$, between 5 and 6 when $n = 30$ and they are always 3 when $n = 100$. The small differences in the rejection rates suggest that the precise value of k is not significant in terms of application. The small values obtained for all the procedures (except the maximum) reinforce the impression provided by Figure 1 that values for k well below 10 are enough for dimension $p = 2$.

In [8] previous ideas are also applied to check the homogeneity between K samples, $K > 2$. Let $\{X_{1,1}, ..., X_{1,n_1}\}, ..., \{X_{K,1}, ..., X_{K,n_K}\}$ be random samples obtained, respectively, from the distributions $P_1, ..., P_K$ and let us assume that there exists $r_1, ..., r_{K-1} > 0$ such that the random vectors $r_1X_{1,1}, ..., r_1X_{1,n_1}, ..., r_{K-1}X_{K-1,1}, ..., r_{K-1}X_{K-1,n_{K-1}}, X_{K,1}, ..., X_{K,n_K}$

are identically distributed. We are interested in testing the following hypotheses:

$$H_0: \quad r_i = 1, i = 1, ..., K-1 \text{ (all scales are the same)}$$
$$H_a: \quad \text{there exists } r_i \neq 1 \text{ (scales are different).}$$

If we center each sample separately, join all the observations in a single sample, compute the depths of all the points and transform those depths in ranks, then we can apply the Kruskal-Wallis test [5] to check if there is a lack of homogeneity between the ranks in each sub-sample.

We have carried out a simulation study applying previous procedure to the Tukey depth and to the random Tukey depth in the 2-dimensional case with Gaussian distributions, $K = 3$ and sample sizes $n_1 = n_2 = n_3 = n$, where $n \in \{20, 30\}$. We have carried out 5,000 replications in each case at the significance level $\alpha = .05$. To select k, we have applied bootstrap in a similar way to that used previously. The only difference is that now we have taken three bootstrap samples and that, after centering, we have multiplied just one of them by the values of r in the grid. However, the Tukey depth has been computed similarly to the previous case.

At the end of the bootstrap step, we have tried the same four procedures to select k as in the previous case, also obtaining here similar rejection rates among the four cases. In Table 3.2 we present the rejection rates obtained when the precise value of k is selected with the 80% percentile procedure. The tables with the results of the remaining procedures are available from the authors upon request. We have computed the medians of the selected numbers of vectors used in each of the four procedures. They are 22 or 23 when k is selected with the maximum. They are 2 in every case, when k is selected with the median, 3 or 4 when selected with the mean and 5 (when $n = 20$) or 3 (when $n = 30$) if selected with the 80% percentile.

Table 3.2 *Rate of rejections in 5,000 simulations using $D_{T,k}$ (between parentheses the rate with D_T) to test the homogeneity in three samples of Gaussian distributions with independent, identically distributed marginals and the exposed values of r. The dimension is $p = 2$. The significance level is .05.*

Covariance matrices	Sample sizes			
	$n = 20$		$n = 30$	
$r_1 = r_2 = 1$	.04	(.04)	.04	(.04)
$r_1 = r_2 = 1.2$	.12	(.13)	.18	(.18)
$r_1 = 2$, $r_2 = 1.2$	.85	(.85)	.97	(.98)
$r_1 = r_2 = 2$	.94	(.94)	1	(.99)

The results of both studies in this subsection are quite encouraging because there are no important differences between the rejection rates with both depths despite the comparatively low number of directions used to compute the random Tukey depth.

3.2 Computational time

We end this section by paying some attention to the computational time required to compute the random Tukey depth. As a comparison we have selected the time necessary to compute the Mahalanobis depth, which is one of the quickest depths according to Table 1 in [11]. In Table 3.3 we present the mean time, obtained from 200 simulations, necessary to compute the random Tukey and Mahalanobis depths for all points in a sample with the shown sizes and dimensions. The numbers of random directions used correspond with those obtained in Figure 1.

To make a reliable comparison between the computational times we need to compare the time needed to compute the random Tukey depth of a sample to the time needed to compute the Mahalanobis depth of the same sample. However, we must keep in mind that the first depth to be computed may have an advantage as the RAM memory may be cleaner than when the second depth is computed. In order to avoid this, we have computed the random Tukey depth first 100 times and the Mahalanobis depth first 100 times. The computations have been carried out on a Xserve G5 PowerPC G5 Dual 2.3 GHz computer with 2Gb of RAM memory.

Table 3.3 *Time, in seconds, to compute the random Tukey and the Mahalanobis depths of all points in a sample with size n taken from a standard Gaussian distribution.*

Dimension	Random vectors	Sample size	Random Tukey	Mahalanobis
$p = 2$	$k = 10$	$n = 25$	$4.349 \cdot 10^{-4}$	.0014
		$n = 100$	$6.322 \cdot 10^{-4}$	.0024
$p = 8$	$k = 60$	$n = 50$	.0047	.0017
		$n = 100$	.0105	.0028
$p = 50$	$k = 250$	$n = 100$	.1153	.0047
		$n = 500$	.5596	.0158

It can be observed that the time needed to compute the random Tukey depth is acceptable in every case. Moreoever, it is better than the one needed to compute the Mahalanobis depth for low dimensions like $p = 2$, of the same order for $p = 8$ and worse for dimensions around 50.

4 Discussion

We introduce the random Tukey depth, which can be considered as a random approximation of the Tukey depth. The new depth is interesting because of the little effort required in its computation and because it can be extended to cover Hilbert valued data. This depth satisfies most of the properties of a depth according to the definition in [13]. Moreover, this depth can be consistently estimated from a random sample both in the finite and the infinite dimensional settings.

The interest of the random Tukey depth lies in the fact that by taking only a few one-dimensional projections, it is possible to obtain results similar to those obtained with more involved depths. The number of required projections is surprisingly low indeed. This is shown in the comparisons with the Tukey depth that we have carried out. Those studies do not show relevant differences between the results obtained with the Tukey depth and with the random Tukey depth. Thus, we conclude that, at least under the considered conditions, the random Tukey depth is an alternative which is worth considering because of the little time required to compute it.

Due to the generalization of the main results in [1], which appears in [3], it is possible to extend the results of this paper to more general spaces.

Acknowledgments

The authors wish to thank an anonymous referee and the Associated Editor in charge of this paper whose comments have improved it considerably. They also thank Prof. R. Fraiman who pointed out a mistake in a previous version of this paper.

This research has been partially supported by the Spanish Ministerio de Ciencia y Tecnología, grant MTM2005-08519-C02-02 and the Consejería de Educación y Cultura de la Junta de Castilla y León, grant PAPIJCL VA102/06.

Appendix: The proofs

The proofs are identical for finite or infinite dimensional spaces (except, of course, the proof of item 4 in Theorem 2.3, which only works for the finite dimensional case). Then, in this appendix the symbol $\mathcal{X}$ will refer here, indistinctly, to $\mathbb{R}^p$ or $I\!H$. Given a set $B \subset \mathcal{X}$, B^c and ∂B will respectively be their topological complement and boundary. If $x, v \in \mathcal{X}$ and $P \in \mathcal{P}$, we will denote

$$S^P_{x,v} := \begin{cases} \{y : \langle y - x, v\rangle \leq 0\} & \text{if } P\{y : \langle y - x, v\rangle \leq 0\}] \leq P[y : \langle y - x, v\rangle \geq 0] \\ \{y \, \langle y - x, v\rangle \geq 0\} & \text{otherwise} \end{cases}.$$

To simplify, if there is no risk of confusion, the super-index P will be omitted. With this notation, we have that $D_1(\Pi_v(x), P \circ \Pi_v^{-1}) = P(S_{x,v})$.

Proof of Theorems 2.3 and 2.9. Clearly, the random Tukey depth is bounded and nonnegative because it is a minimum of probabilities. To check the remaining properties, let ν be an absolutely continuous distribution on $\mathcal{X}$, $k > 0$ and $P \in \mathcal{P}$.

1. Invariance under translation. It is straightforward due to the linearity of the projections.

2. Maximality at center. First remember that a distribution P is halfspace symmetric about θ if $P[H] \geq 1/2$ for every closed halfspace H containing θ. Assume that $\theta \in \mathcal{X}$ is

the center of P, and that there exists $x \in \mathcal{X}$ satisfying

$$D_{T,k}(x, P) > D_{T,k}(\theta, P). \tag{7}$$

By definition, there exists $v \in \{v_1, \ldots, v_k\}$ such that $D_{T,k}(\theta, P) = D_1(\Pi_v(\theta), P \circ \Pi_v^{-1}) = P(S_{\theta,v}) \geq 1/2$, due to the halfspace symmetry. Thus, from (7) we get

$$P(S_{x,v}) > 1/2. \tag{8}$$

We have three possibilities for the sets $S_{x,v}$ and $S_{\theta,v}$. The first one is that $S_{x,v} \subseteq S_{\theta,v}$. However, then $D_{T,k}(x, P) \leq P(S_{x,v}) \leq P(S_{\theta,v}) = D_{T,k}(\theta, P)$ which contradicts (7). The second one is $S_{\theta,v} \subset S_{x,v}$. From here and $P(S_{\theta,v}) \geq 1/2$, we obtain that $P(S_{x,v}^c \cup \partial S_{x,v}) \leq 1/2$. By (8), $\min(P(S_{x,v}^c \cup \partial S_{x,v}), P(S_{x,v})) = P(S_{x,v}^c \cup \partial S_{x,v})$, which contradicts the definition of $S_{x,v}$. The final possibility is that $S_{\theta,v} \subset S_{x,v}^c$. Thus, $P(S_{x,v}^c) \geq 1/2$. Therefore, by (8) $1 = P(S_{x,v}) + P(S_{x,v}^c) > 1$.

3. Monotonicity relative to deepest point. Let us assume that P has a deepest point θ and there exist $x \in \mathcal{X}$ and $\alpha \in [0, 1]$ with

$$D_{T,k}(x, P) > D_{T,k}(\theta + \alpha(x - \theta), P). \tag{9}$$

Obviously, cases $\alpha = 0$ and $\alpha = 1$ are not possible. Then, $\alpha \in (0, 1)$. Since θ is the deepest point, we have

$$D_{T,k}(\theta, P) \geq D_{T,k}(y, P), \text{ for all } y \in \mathcal{X}. \tag{10}$$

Let $v \in \{v_1, ..., v_k\}$ such that $D_{T,k}(\theta + \alpha(x - \theta), P) = P(S_{\theta+\alpha(x-\theta),v})$. From (9) and (10) it is inferred that

$$P(S_{\theta,v}) > P(S_{\theta+\alpha(x-\theta),v}). \tag{11}$$

Since $\alpha \in (0, 1)$, we have the point $\theta + \alpha(x - \theta)$ laying in the open segment joining the points x and θ. Then, from (10) and (11), following a reasoning similar to the final part of the proof of Statement 2, we have that $S_{x,v} \subset S_{\theta+\alpha(x-\theta),v}$. Thus,

$$D_{T,k}(\theta + \alpha(x - \theta), P) = P(S_{\theta+\alpha(x-\theta),v}) \geq P(S_x, v) \geq D_{T,k}(x, P),$$

which contradicts (9).

4. Vanishing at infinity (only for Theorem 2.3). We will show this property for the case in which ν is not a probability, but the Lebesgue measure. The proof for probabilities follows the same steps until statement (14) below. From this point on, only some additional technicalities are required.

Let $\epsilon > 0$. Since $\lim_{H\to\infty} P\{y : \|y\| \leq H\} = 1$, there exists $H_\epsilon > 0$ such that $P\{y : \|y\| \leq H_\epsilon\} > 1 - \epsilon$. Furthermore, if $v \in \mathbb{R}^p$ then

$$\Pi_v^{-1}[-H_\epsilon\|v\|, H_\epsilon\|v\|] \supset \{y : \|y\| \leq H_\epsilon\}.$$

Thus, $P \circ \Pi_v^{-1}[-H_\epsilon\|v\|, H_\epsilon\|v\|] > 1-\epsilon$, for all $v \in \mathbb{R}^p$. In consequence, if

$$\sup\left(D_1(-H_\epsilon\|v\|, P\circ\Pi_v^{-1}), D_1(H_\epsilon\|v\|, P\circ\Pi_v^{-1})\right) < \epsilon, \text{ for all } v \in \mathbb{R}^p. \tag{12}$$

Let $M > 0$ and let $x \in \mathbb{R}^p$ with $\|x\| \geq M$. Thus

$$\begin{aligned}\nu^k\left\{(v_1,...,v_k) \in (\mathbb{R}^p)^k : D_{T,k}(x,P) < \epsilon\right\} &\geq \nu^k\left\{(v_1,...,v_k) \in (\mathbb{R}^p)^k : D_{T,1}(x,P) < \epsilon\right\} \\ &= \nu\left\{v \in \mathbb{R}^p : D_{T,1}(x,P) < \epsilon\right\}, \end{aligned} \tag{13}$$

where we assume that $D_{T,1}$ is computed using v_1.

If $v \in \mathbb{R}^p$ satisfies that $|\langle x, v\rangle| \geq H_\epsilon\|v\|$, then by (12), it happens that $D_1(\Pi_v(x), P\circ\Pi_v^{-1})) < \epsilon$. Therefore, from (13),

$$\begin{aligned}\nu^k\left\{(v_1,...,v_k) \in (\mathbb{R}^p)^k : D_{T,k}(x,P) < \epsilon\right\} &\geq \nu\left\{v \in \mathbb{R}^p : |\langle x, v\rangle| \geq H_\epsilon\|v\|\right\} \\ &\geq \nu\left\{v \in \mathbb{R}^p : \frac{|<x,v>|}{\|x\|\|v\|} \geq \frac{H_\epsilon}{M}\right\} \quad (14) \\ &= \nu\left\{v \in \mathbb{R}^p : \frac{|<e_1,v>|}{\|v\|} \geq \frac{H_\epsilon}{M}\right\}, \end{aligned}$$

where (14) comes from $\|x\| > M$. In the last equality, e_1 denotes the first element in a fixed orthonormal base of $\mathbb{R}^d$ and the equality holds because ν is rotationally invariant. Therefore, from this chain we have

$$\inf_{x:\|x\|\geq M} \nu^k\left\{(v_1,...,v_k) \in (\mathbb{R}^p)^k : D_{T,k}(x,P) < \epsilon\right\} \geq \nu\left\{v \in \mathbb{R}^p : \frac{|<e_1,v>|}{\|v\|} \geq \frac{H_\epsilon}{M}\right\},$$

and the proof ends because, trivially,

$$\lim_{M\to\infty} \nu\left\{v \in \mathbb{R}^p : \frac{|<e_1,v>|}{\|v\|} \geq \frac{H_\epsilon}{M}\right\} = 1.$$

•

Proof of Theorems 2.6 and 2.10. Let $k > 0$ and $v_1, ..., v_k \in \mathcal{X}$, which will remain fixed during the proof. Let P be a probability distribution on $\mathcal{X}$, let $x_1, ..., x_n \in \mathcal{X}$ be a random sample taken from P and let P_n be the associated empirical distribution. Let $v \in \{v_1, ..., v_k\}$. It is obvious that $\Pi_v(x_1), ..., \Pi_v(x_n)$ is a random sample taken from the distribution $P\circ\Pi_v^{-1}$ and that the empirical distribution associated to those projections coincide with $P_n\circ\Pi_v^{-1}$. Moreover, $P\circ\Pi_v^{-1}$ is a distribution on the real line and, then, the Glivenko-Cantelli theorem gives

$$\sup_{y\in\mathbb{R}} \sup\left(\left|P_n\circ\Pi_v^{-1}(-\infty,y] - P\circ\Pi_v^{-1}(-\infty,y]\right|, \left|P_n\circ\Pi_v^{-1}[y,\infty) - P\circ\Pi_v^{-1}[y,\infty)\right|\right) \to 0, \text{a.s.}$$

From here, we obtain

$$\sup_{y\in\mathbb{R}}\left(|D_1(y, P_n\circ\Pi_v^{-1}) - D_1(y, P\circ\Pi_v^{-1})|\right) \to 0, \text{a.s.} \tag{15}$$

Therefore,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}} |D_{T,k}(x,P_n) - D_{T,k}(x,P)| \\
&\qquad = \sup_{x\in\mathcal{X}} |\min_{i=1,\ldots,k} D_1(\Pi_{v_i}(x), P_n\circ\Pi_{v_i}^{-1}) - \min_{i=1,\ldots,k} D_{T,k}(\Pi_{v_i}(x), P\circ\Pi_{v_i}^{-1})| \\
&\qquad \leq \sup_{x\in\mathcal{X}, i=1,\ldots k} |D_1(\Pi_{v_i}(x), P_n\circ\Pi_{v_i}^{-1}) - D_1(\Pi_{v_i}(x), P\circ\Pi_{v_i}^{-1})| \\
&\qquad = \sup_{y\in\mathbb{R}, i=1,\ldots k} |D_1(y, P_n\circ\Pi_{v_i}^{-1}) - D_1(y, P\circ\Pi_{v_i}^{-1})|,
\end{aligned}$$

which converges a.s. to zero because of (15). •

References

[1] Cuesta-Albertos, J., Fraiman, R. and Ransford, T., 2007. A sharp form of the Cramér-Wold theorem. J. Theoret. Probab., 20 201-209.

[2] Cuevas, A., Febrero, M. and Fraiman, R., 2007. Robust estimation and classification for functional data via projection-based depth notions. To appear in Computation. Statist.

[3] Cuevas, A. and Fraiman, R., 2007. On depth measures and dual statistics: A methodology for dealing with general data. Preprint.

[4] Ferraty, F. and Vieu, P. 2006. Nonparametric functional data analysis: theory and practice. Springer Verlag, New York.

[5] Hettmansperger, T.P., 1984. Statistical inference based on ranks. John Wiley & Sons, New York.

[6] Liu, R., 1990. On a notion of data depth based on random simplices. Ann. Statist. 18 405-414.

[7] Liu, R., Serfling, R. and Souvaine, D.L., editors, 2006. Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications. American Mathematical Society, DIMACS Series, Vol. 72.

[8] Liu, R.Y. and Singh, K., 2006. Rank tests for nonparametric description of dispersion. In: R. Liu, R. Serfling and D.L. Souvaine (Ed.), Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, American Mathematical Society, DIMACS Series, Vol. 72, 17-35.

[9] Mahalanobis, P. C., 1936. On the Generalized Distance in Statistics. Proceed. Nat. Academy of India, 12 49-55.

[10] Maronna, R.A., Martin, R. D. and Yohai, V.J., 2006. Robust Statistics. Theory and Methods. John Wiley & Sons, Chichester.

[11] Mosler, K. and Hoberg, R., 2006. Data analysis and classification with the zonoid depth. In: R. Liu, R. Serfling and D.L. Souvaine (Ed.), Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, American Mathematical Society, DIMACS Series, Vol. 72 49-59.

[12] Tukey, J.W., 1975. Mathematics and picturing of data. Proceedings of ICM, Vancouver, 2 523-531.

[13] Zuo, Y. and Serfling, R., 2000. General notions of statistical depth function. Ann. Statist., 28(2) 461-482.

[14] Zuo, Y., 2003. Projection-based depth functions and associated medians. Ann. Statist., 31(5) 1460-1490.

[15] Zuo, Y., 2006. Multidimensional trimming based on projection depth. Ann. Statist., 34(5) 2211-2251.