On the uniform consistency of the zoonoid depth

Ignacio Cascos[a,*], Miguel López-Díaz[b]

[a]Department of Statistics, Universidad Carlos III de Madrid,
UC3M-BS Institue of Financial Big Data
Av. Universidad 30, E-28911 Leganés (Madrid), Spain
[b]Departamento de Estadística e I.O. y D.M., Universidad de Oviedo,
c/ Calvo Sotelo s/n, E-33007 Oviedo, Spain

Abstract

Under some mild conditions on probability distribution P, if $\lim_n P_n = P$ weakly then the sequence of zonoid depth functions with respect to P_n converges uniformly to the zonoid depth function with respect to P.

Keywords: Uniform consistency, Zonoid depth
AMS subject classifications: 62H12, 60D05, 62G20

1. Introduction

The *depth* of a point in $\mathbb{R}^d$ with respect to a probability distribution in the same Euclidean space quantifies the degree of centrality of the point with respect to the distribution. A (possibly non-unique) point of maximal depth is, in some sense, central with regard to the given distribution, while depth decreases along rays emanating from that center. The lack of a natural order in the multivariate Euclidean space together with the possibility of introducing a center-outward ordering of data points based on their depths have provided data depth notions with quite some attention from the multivariate statistics community.

☆Research partially supported by grants MTM2011-22993 and ECO2011-25706 of the Spanish Ministry of Science and Innovation and grant FC-15-GRUPIN14-101 of the Principado de Asturias.

*Ph. (+34) 91 624 8750, Fax: (+34) 91 624 9177
Email addresses: ignacio.cascos@uc3m.es (Ignacio Cascos), mld@uniovi.es (Miguel López-Díaz)

Among the numerous notions of data depth introduced in the statistical literature in the last decades, see [1, 8, 11, 15] for particular notions of data depth and some of their applications, the *zonoid depth* [2, 4] occupies a prominent position, right after the best known data depths, which are Tukey's halfspace depth [13, 14] and Liu's simplicial depth [6].

The *zonoid depth* with regard to the empirical probability of a sample of size n assigns depth 1 to the average of all points from the data cloud, and depth k/n to all points that can be obtained as either averages of k points from the data cloud or as a convex combination of a set of such averages. For a general population distribution, the zonoid depth is commonly introduced in terms of its level sets, which are known as *zonoid trimmed regions*. These zonoid trimmed regions can be as well obtained as a transformation of the so-called *lift zonoid* [5] of a distribution P, which is a convex body in $\mathbb{R}^{d+1}$ that characterizes probability distributions with finite first moment.

The uniform consistency of the empirical depth proves to be relevant when establishing a multivariate order with regard to a data cloud. The reason is that the interest is not only to estimate the depth of a point, but the whole depth function, see Remark A.3 [15] for a brief discussion. Among other particular applications, it has been used in Theorem 6.1 [9] to prove the consistency of a sample quality index built to compare two distributions, and in Proposition 3.1 [7] to support some distribution-free control charts based on ranks induced by a notion of data depth.

The strong uniform consistency of the empirical halfspace depth was established in pp. 1816–1817 [3], while the one of the simplicial depth under the assumption of absolute continuity of the population distribution in Theorem 5 [6]. As for other depth functions, the uniform consistency of the empirical Mahalanobis depth holds for distributions with bounded second absolute moment, and the one of the Majority depth for elliptical distributions, see Remark 2.2 [9]. The pointwise strong consistency of the zonoid depth was established in Theorem 7.1 (iii) [4]. In the current note, we prove the strong uniform consistency of the zonoid depth.

The paper is organized as follows: Section 2 is devoted to some preliminaries about convex geometry and zonoid depth, while our main result is presented in Section 3.

2. Preliminaries

A *convex body* is a compact convex set in the d-dimensional Euclidean space with nonempty interior. The unit sphere in $\mathbb{R}^d$ is represented by S^{d-1}, while $\langle \cdot, \cdot \rangle$ stands for the scalar product.

Given a convex body $K \subset \mathbb{R}^d$, its *support function*, $h_K : S^{d-1} \mapsto \mathbb{R}$ and, subject to $0 \in K$, its *radius-vector function* $\rho_K : S^{d-1} \mapsto \mathbb{R}$ (see [10]) are respectively given by

$$h_K(u) = \sup\{\langle x, u \rangle : \, x \in K\}\,,$$

$$\rho_K(u) = \sup\{t \geq 0 : \, tu \in K\}\,.$$

The *Hausdorff distance* between two convex bodies $K_1, K_2 \subset \mathbb{R}^d$ is

$$d_H(K_1, K_2) = \sup_{u \in S^{d-1}} |h_{K_1}(u) - h_{K_2}(u)|\,.$$

The standard notion of convergence for convex bodies is the Hausdorff one. We say that the sequence of d-dimensional convex bodies $\{K_n\}_n$ converges to the convex body K in the Hausdorff distance if $\lim_n d_H(K_n, K) = 0$.

All probability measures P considered hereafter are defined on the general d-dimensional Euclidean space equipped with the Borel σ-algebra and are assumed to have finite first moment, that is, $\int_{\mathbb{R}^d} \|x\| \mathrm{d}P(x) < \infty$.

The *lift zonoid* of P is a convex body in $\mathbb{R}^{d+1}$ containing the origin of coordinates and given by

$$\mathrm{Z}(P) = \left\{ \left(\int g(y)\mathrm{d}P(y), \int y g(y)\mathrm{d}P(y) \right), \text{ s.t. } g : \mathbb{R}^d \mapsto [0,1] \text{ measurable} \right\}.$$

The *zonoid depth* of $x \in \mathbb{R}^d$ with respect to P is

$$\mathrm{ZD}(x; P) = \sup\{\alpha \in (0,1] : \, x \in \alpha^{-1}\mathrm{proj}_\alpha(\mathrm{Z}(P))\}\,,$$

where $\mathrm{proj}_\alpha(\mathrm{Z}(P))$ is the projection of the intersection of $\mathrm{Z}(P)$ with the hyperplane $\{(\alpha, x) : \, x \in \mathbb{R}^d\}$ to the last d coordinates. After multiplication by α^{-1} this set is commonly referred to as *zonoid trimmed region of level* α *of* P and denoted by $\mathrm{ZD}^\alpha(P) = \alpha^{-1}\mathrm{proj}_\alpha(\mathrm{Z}(P))$. The family of convex bodies $\{\mathrm{ZD}^\alpha(P)\}_{\alpha \in (0,1]}$ is decreasing on α, while $\mathrm{ZD}^0(P)$ is defined as the closed and convex set $\mathrm{cl}(\cup_{\alpha > 0}\mathrm{ZD}^\alpha(P))$.

If a sequence of probability measures $\{P_n\}_n$ converges weakly to P, then

- $\lim_n \mathrm{Z}(P_n) = \mathrm{Z}(P)$ in the Hausdorff sense (Theorem 3.3 [5]);
- $\lim_n \mathrm{ZD}^\alpha(P_n) = \mathrm{ZD}^\alpha(P)$ in the Hausdorff sense for any $\alpha \in (0,1]$ (Theorem 5.2 (i) [4]);
- if x lies in the interior of the convex hull of the support of P, then $\lim_n \mathrm{ZD}(x; P_n) = \mathrm{ZD}(x; P)$ (Theorem 7.1 (iii) [4]).

3. Main result

The zonoid depth of $x \in \mathbb{R}^d$ with respect to P is the supremum of all $0 < \alpha \leq 1$ such that $(\alpha, \alpha x)$ lies in the lift zonoid of P, and it is thus possible to relate the zonoid depth with the radius-vector function of the lift zonoid. For simplicity, for $x \in \mathbb{R}^d$, we will hereafter write $\boldsymbol{x} = (1, x) \in \mathbb{R}^{d+1}$ and $u(\boldsymbol{x}) = \boldsymbol{x}/\|\boldsymbol{x}\| \in S^d$.

Lemma 3.1. *For any $x \in \mathbb{R}^d$, we have*

$$\mathrm{ZD}(x; P) = \rho_{\mathrm{Z}(P)}(u(\boldsymbol{x}))\|\boldsymbol{x}\|^{-1} \leq \rho_{\mathrm{Z}(P)}(u(\boldsymbol{x}))\,. \tag{1}$$

Proof. For any $x \in \mathbb{R}^d$, we have $\alpha x \in \mathrm{proj}_\alpha(\mathrm{Z}(P))$ as long as $\alpha\boldsymbol{x} = (\alpha, \alpha x) \in \mathrm{Z}(P)$. Finally and since the radius-vector function is only defined on S^{d-1}, we normalize $\boldsymbol{x}$ in order to obtain

$$\mathrm{ZD}(x; P) = \sup\{\alpha \in (0,1] : (\alpha, \alpha x) \in \mathrm{Z}(P)\} = \rho_{\mathrm{Z}(P)}(u(\boldsymbol{x}))\|\boldsymbol{x}\|^{-1}\,.$$

The inequality in (1) follows from $\|\boldsymbol{x}\| \geq 1$ for all $x \in \mathbb{R}^d$. □

Figure 1 illustrates graphically Equation (1) in Lemma 3.1.

We say that a probability distribution P on $\mathbb{R}^d$ satisfies condition (C) if

(C) the probability of any hyperplane that is tangent to the boundary of the convex hull of its support is zero, or equivalently, $P(\partial H) = 0$ for every halfspace H with $P(H) = 1$.

Condition (C) is, e.g. satisfied by all probability distributions that asses probability zero to any hyperplane, or by the smaller family of absolutely continuous distributions.

Lemma 3.2. *If P satisfies condition (C), then the next two statements hold true:*

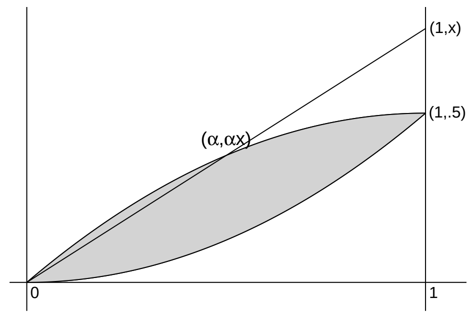

Figure 1: The shaded region is the lift zonoid of a uniform distribution on the unit interval. The values of α and x considered in the graph are $\alpha = 1/2$ and $x = 3/4$. The zonoid depth evaluated at x is the supremum of all α such that $(\alpha, \alpha x)$ is a point of the lift zonoid lying in the line segment with extremes at the origin of coordinates and $(1, x)$.

1. *the boundary of its lift zonoid* $\mathrm{Z}(P)$ *does not contain any line segment with the origin in it,*
2. *for every* $0 < \alpha < \beta \leq 1$ *there exists* $\varepsilon > 0$ *such that* $h_{\mathrm{ZD}^\alpha(P)}(u) - h_{\mathrm{ZD}^\beta(P)}(u) \geq \varepsilon$ *for all* $u \in S^{d-1}$.

Proof. In order to prove 1., assume that the boundary of $\mathrm{Z}(P)$ contains a line segment with the origin in it, then for some $0 < \alpha \leq 1$ and some $x \in \mathbb{R}^d$, we have $(\beta, \beta x) \in \partial\mathrm{Z}(P)$ for all $0 \leq \beta \leq \alpha$. Consequently $x \in \partial\mathrm{ZD}^\beta(P)$ for all $0 < \beta \leq \alpha$ and does also lie in the boundary of the union of the increasing sequence of convex bodies $\cup_{\gamma>0}\mathrm{ZD}^\gamma(P)$, so $x \in \partial\mathrm{ZD}^0(P)$. Theorem 5.5 in [4] establishes that $\mathrm{ZD}^0(P)$ is the intersection of all closed halfspaces whose probability is 1, so there must exist a closed halfspace H with x on its boundary such that $P(H) = 1$. Note that $\mathrm{ZD}(x; P) \geq \alpha$ implies that the probability of the closure of the complementary halfspace to H must be strictly positive. Therefore the probability of the intersection of the two halfspaces is strictly positive, that is, $P(\partial H) > 0$, which contradicts (C).

2. follows from 1. also by contradiction. Assume there exist $0 < \alpha < \beta \leq 1$ such that for every $\varepsilon > 0$ there exists $u \in S^{d-1}$ such that $h_{\mathrm{ZD}^\alpha(P)}(u) - h_{\mathrm{ZD}^\beta(P)}(u) < \varepsilon$. After the continuity of the support function $h_{\mathrm{ZD}^\alpha(P)}(u) =$

$h_{\mathrm{ZD}^\beta(P)}(u)$ for some $u \in S^{d-1}$ and since $\mathrm{ZD}^\beta(P) \subset \mathrm{ZD}^\alpha(P)$ there must exist some point $x \in \mathbb{R}^d$ lying in the boundary of both zonoid trimmed regions. Consequently we obtain $\alpha\boldsymbol{x}, \beta\boldsymbol{x} \in \partial \mathrm{Z}(P)$ which violates 1., so the assumption cannot be valid. □

Statement 2. in Lemma 3.2 means that, under (C), the family of depth-trimmed regions is *strictly* decreasing on α, so for any $0 < \alpha < \beta \leq 1$ there exists an ε*-envelope* of $\mathrm{ZD}^\beta(P)$ that is contained in $\mathrm{ZD}^\alpha(P)$.

Remark 3.1. Condition (C) actually guarantees the continuity of the zonoid depth as a function of x in $\mathbb{R}^d$. Such continuity was established for the x's inside the convex hull of the support of P in Theorem 7.1 (ii) [4]. Whenever (C) holds, the zonoid depth at the boundary of support is 0, as it is outside the convex hull of the support.

Under condition (C) we prove the uniform convergence of a sequence of zonoid depths of a converging sequence of probabilities. The main argument we use here is, in essence, borrowed from Lemma 3.2 [12].

Theorem 3.1. *If a probability P satisfies condition (C) and the sequence $\{P_n\}_n$ converges weakly to P, we have that*

$$\lim_n \sup_{x \in \mathbb{R}^d} |\mathrm{ZD}(x; P) - \mathrm{ZD}(x; P_n)| = 0\,.$$

Proof. For any $0 < \varepsilon < 1$, we obviously have

$$\begin{aligned}&\sup_{x \in \mathbb{R}^d} |\mathrm{ZD}(x; P) - \mathrm{ZD}(x; P_n)| \\ &= \max\{\sup_{x \in \mathrm{ZD}^{\varepsilon/2}(P)} |\mathrm{ZD}(x; P) - \mathrm{ZD}(x; P_n)|, \sup_{x \notin \mathrm{ZD}^{\varepsilon/2}(P)} |\mathrm{ZD}(x; P) - \mathrm{ZD}(x; P_n)|\}\,.\end{aligned}$$

In first place, we will show that the last supremum is bounded by ε for n large enough. After statement 2. in Lemma 3.2, the inclusion relation $\mathrm{ZD}^\varepsilon(P) \subset \mathrm{ZD}^{\varepsilon/2}(P)$ is strict and the convergence of the zonoid trimmed regions ensures the existence of N_1 such that $\mathrm{ZD}^\varepsilon(P_n) \subset \mathrm{ZD}^{\varepsilon/2}(P)$ for all $n \geq N_1$, so

$$\sup_{x \notin \mathrm{ZD}^{\varepsilon/2}(P)} |\mathrm{ZD}(x; P) - \mathrm{ZD}(x; P_n)| < \varepsilon \quad \text{for } n \geq N_1\,.$$

Let us now concentrate on $\mathrm{ZD}^{\varepsilon/2}(P)$. Notice that $\mathrm{ZD}^{\varepsilon/2}(P) \subset \mathrm{ZD}^{\varepsilon/4}(P_n)$ for all $n \geq N_2$ and some N_2. Assume that there exists a sequence $x_n \in$

$\mathrm{ZD}^{\varepsilon/2}(P)$ with a subsequence satisfying $|\mathrm{ZD}(x_n;P) - \mathrm{ZD}(x_n;P_n)| \geq \varepsilon$ for all n, which after Lemma 3.1 guarantees $|\rho_{\mathrm{Z}(P)}(u(\boldsymbol{x}_n)) - \rho_{\mathrm{Z}(P_n)}(u(\boldsymbol{x}_n))| \geq \varepsilon$. Due to the compactness of $\mathrm{ZD}^{\varepsilon/2}(P)$, such a subsequence must have a converging subsequence, which we will again denote as x_n, converging to some $x \in \mathrm{ZD}^{\varepsilon/2}(P)$. Define

$$y_n = \rho_{\mathrm{Z}(P)}(u(\boldsymbol{x}_n))u(\boldsymbol{x}_n) \in \partial\mathrm{Z}(P) \quad \text{and} \quad z_n = \rho_{Z(P_n)}(u(\boldsymbol{x}_n))u(\boldsymbol{x}_n) \in \partial\mathrm{Z}(P_n)\,.$$

The compactness of $\mathrm{Z}(P)$ and the convergence of $\{\mathrm{Z}(P_n)\}_n$ to $\mathrm{Z}(P)$ in the Hausdorff distance guarantees the existence of a closed ball containing all $\mathrm{Z}(P_n)$ and $\mathrm{Z}(P)$, thus $\rho_{\mathrm{Z}(P_n)}$ and $\rho_{\mathrm{Z}(P)}$ are uniformly bounded. It is therefore possible to obtain converging subsequences of $\{y_n\}_n$ and $\{z_n\}_n$ whose limits are respectively denoted as y and z. Notice that the first component of each y_n is at least $\varepsilon/2$, while the first component of each z_n is at least $\varepsilon/4$, so both of y and z are different from the origin of coordinates. Since $\partial\mathrm{Z}(P)$ is closed, $y \in \partial\mathrm{Z}(P)$, while the convergence of $\{\mathrm{Z}(P_n)\}_n$ to $\mathrm{Z}(P)$ in the Hausdorff distance guarantees that $z \in \partial\mathrm{Z}(P)$. Finally $\|y_n - z_n\| \geq \varepsilon$ implies $\|y - z\| \geq \varepsilon$ while both of them belong to a ray from the origin with direction $u(\boldsymbol{x})$, which contradicts 1. from Lemma 3.2. □

As the next example shows, condition (C) cannot be omitted in Theorem 3.1.

Example 3.1. For any $x \in \mathbb{R}$, let $\delta_{\{x\}}$ stand for the degenerated distribution at x. Clearly the sequence of probabilities $\{\delta_{\{1/n\}}\}_n$ converges weakly to $\delta_{\{0\}}$, but for any n, we have $\sup_x |\mathrm{ZD}(x;\delta_{\{1/n\}}) - \mathrm{ZD}(x;\delta_{\{0\}})| = 1$.

Finally we obtain the strong uniform consistency of the empirical zonoid depth.

Corollary 3.1. *Given P with finite first moment and such that the probability of any hyperplane (that is tangent to the boundary of the convex hull of the support of P) is zero, the empirical zonoid depth is a strongly uniformly consistent estimator of the population zonoid depth.*

Acknowledgements

The authors gratefully appreciate the comments of a referee which led to improvements in the note.

References

[1] Cascos, I., 2010. Data depth: multivariate statistics and geometry. In Kendall, W.S. and Molchanov, I. (Eds.) *New Perspectives in Stochastic Geometry*, Oxford University Press, pp. 398–423.

[2] Dyckerhoff, R., Koshevoy, G., Mosler, K., 1996. Zonoid data depth: Theory and computation. In Prat, A. (Ed.), *COMPSTAT 1996—Proceedings in Computational Statistics*, Physica-Verlag, Heidelberg, pp. 235–240

[3] Donoho, D., Gasko, M., 1992. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. Ann. Statist. 20, 1803–1827.

[4] Koshevoy, G., Mosler, K., 1997. Zonoid trimming for multivariate distributions. Ann. Statist. 25,1998–2017.

[5] Koshevoy, G., Mosler, K., 1998. Lift zonoids, random convex hulls and the variability of random vectors. Bernoulli 4, 377–399.

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

[7] Liu, R., 1995. Control charts for multivariate processes. J. Amer. Statist. Assoc. 90, 1380–1387.

[8] Liu, R., Parelius, J, Singh, K., 1999. Multivariate analysis by data depth: Descriptive statistics, graphics and inference. With discussion and a rejoinder by Liu and Singh. Ann. Statist. 27, 783–858.

[9] Liu, R., Singh, K., 1993. A Quality Index Based on Data Depth and Multivariate Rank Tests. J. Amer. Statist. Assoc. 88, 252–260.

[10] Molchanov, I., 2005. *Theory of Random Sets.* Springer, London.

[11] Mosler, K., 2002. Multivariate dispersion, central regions and depth. The Lift Zonoid Approach *Lecture Notes in Statistics*, vol. 165, Springer, Berlin.

[12] Ritoré, M., Vernadakis, E., 2015. Isoperimetric inequalities in Euclidean convex bodies Trans. Amer. Math. Soc. In press.

[13] Rousseeuw, P.J., Ruts, I., 1999. The depth function of a population distribution. Metrika 49, 213–244.

[14] Tukey, J.W., 1975. Mathematics and the picturing of data, in: James, R.D. (Ed.), Proceedings of the International Congress of Mathematics, Vancouver 1974, vol. 2, pp. 523–531.

[15] Zuo, Y., Serfling, R., 2000. General notions of statistical depth function. Ann. Statist. 28, 461–482.