

Level sets of depth measures and central dispersion in abstract spaces

Alejandro Cholaquidis ^{*} Ricardo Fraiman [†] Leonardo Moreno [‡]

March 30, 2021

Abstract

The lens depth of a point has been recently extended to general metric spaces, which is not the case for most depths. It is defined as the probability of being included in the intersection of two random balls centred at two random points X and Y , with the same radius $d(X, Y)$. We study the consistency in Hausdorff and measure distance, of the level sets of the empirical lens depth, based on an iid sample on a general metric space. We also prove that the boundary of the empirical level sets are consistent estimators of their population counterparts, and analyze two real-life examples.

1 Introduction

Statistical depth functions have gained importance in the last three decades. The best well-known and most studied among them include the half space depth [1], the simplicial depth [2, 3], and the multivariate L^1 depth.

Other well-known depths are the convex hull peeling depth [4], the Oja depth [5], and the spherical depth [6], among others. They have been extended to functional spaces, see [7, 8, 9, 10], to Riemannian manifolds, see [11], and also to general metric spaces, see [12]. Several different applications of depths notions have been proposed, in particular for classification problems, by means of the depth-depth method [13], or to functional data, see [14]. Most classical notions of depth, introduced initially on \mathbb{R}^d , can not be directly

^{*}acholaquidis@cmat.edu.uy

[†]rfracman@cmat.edu.uy

[‡]mrleo@iesta.edu.uy

extended to general metric spaces or even to functional spaces or manifolds. Some of them are computationally infeasible on high dimensional spaces, such as Liu’s or Tukey’s depth, because the computational complexity is exponential in the number of dimensions. This is not the case of the depth introduced in [15], i.e. the lens depth, whose computational complexity is of order n^2 , and, as we will see, it can be easily extended to general metric spaces. This makes the lens depth particularly suitable for estimating its level sets, by means of the level sets of its empirical version, based on an iid sample, which is one of the main goals of this manuscript. An extension to Riemannian manifolds of the lens depth (called weighed lens depth) was recently introduced in [16] to tackle the supervised classification problem, and a point-wise a.s consistency result is obtained (see Theorem 2). Here we focus on a different problem, i.e: level set estimation of the lens depth on general metric spaces. This require uniform a.s. consistency of the empirical lens depth to the population one, which is developed in Section 3.

Level set estimation of depths was initially studied in [1], as a key tool for the visualization and exploration of data. Other significant contributions can be found in [17, 18, 19, 20]). As it was pointed out in [21], “the shape and size of these levels, as well as the direction and speed at which they expand, provides insight into the dispersion, kurtosis, and asymmetry of the underlying distribution”. It also allows to extend the notion of quantiles to multivariate or functional data, and can be used for outlier detection, see for instance [22, 23], as well as for supervised classification, see [24, 25].

More formally, let X_1, X_2 be two random variables defined on a (rich enough) probability space $(\Omega, \mathcal{A}, \mathbb{P})$, taking values in a complete separable metric space (M, d) endowed with the Borel σ -algebra. Assume that they are independent and identically distributed. In what follows, the distribution of a random element X of M will be denoted by P_X .

Given $x_1, x_2 \in M$ define their “associate lens” by

$$A(x_1, x_2) := B(x_1, d(x_1, x_2)) \cap B(x_2, d(x_1, x_2)),$$

where $B(p, r)$ is the closed ball centred at p with radius $r > 0$. The lens depth of a point $x \in M$ is defined by $\text{LD}(x) = \mathbb{P}(x \in A(X_1, X_2))$. Given an iid sample $\mathcal{X}_n = \{X_1, \dots, X_n\}$ from a distribution P_X , the empirical version of LD is given by the U -statistics of order two,

$$\widehat{\text{LD}}_n(x) = \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{1}_{A(X_{i_1}, X_{i_2})}(x). \quad (1)$$

To gain some insight into the shape of $\{\widehat{\text{LD}}_n \geq \lambda\}$, see Figure 1.

The general approach to obtain consistency results (w.r.t. Hausdorff distance) for the plug-in estimator $\{\widehat{\text{LD}}_n \geq \lambda\}$ to its population counterpart $\{\text{LD} \geq \lambda\}$ is to prove the uniform convergence of $\widehat{\text{LD}}_n$ to LD .

In general metric spaces it will be necessary to restrict the convergence to compact sets. This is proved in Theorem 4, but the results also hold on \mathbb{R}^d under milder conditions, see Theorem 5.

This paper is organized as follows. Section 2 introduces the notation, some previous definitions, and states two important results given in Theorems 1 and 2 that will be used to prove our main results. The almost surely (a.s.) uniform consistency of $\widehat{\text{LD}}_n$ is stated in Section 3, while its asymptotic distribution is formulated in Section 4. The a.s. consistency in Hausdorff and measure distance, for the empirical level sets $\{\widehat{\text{LD}}_n \geq \lambda\}$, as well as the a.s. consistency of its boundary, is stated in Section 5. Lastly, in Section 6 we tackle the study of two interesting real datasets, the vectocardiogram dataset (see subsection 6.1) and the influenza dataset (see subsection 6.2). All proofs are given in the Appendix.

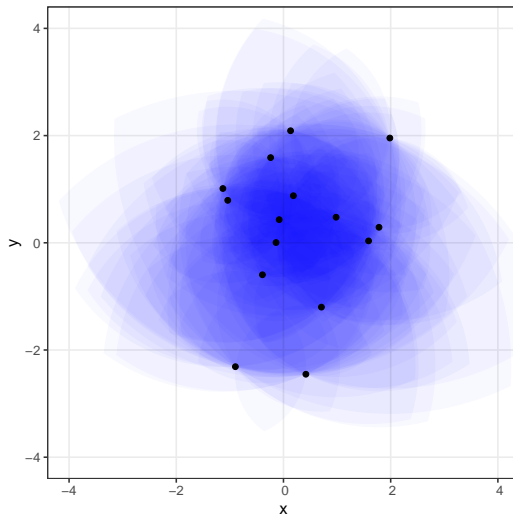


Figure 1: Level sets of the lens depth based on a sample of 15 iid random vectors, distributed as a bivariate standard normal distribution. The intensity of blue represents the depth.

2 Preliminaries

In this section we will introduce the notation and necessary definitions used throughout this paper. Given a metric space (M, d) , which will be assumed to be separable, complete and locally connected, we denote by $B(x, r)$ the closed ball centred at x with radius $r > 0$. The boundary of a set $A \subset M$ is denoted by ∂M . Given $x \in A^c$, $d(x, A) = \inf_{a \in A} d(x, a)$ and $\text{diam}(A) = \sup_{x, y \in A} d(x, y)$. Given two closed sets A, B , the Hausdorff distance between them is defined as

$$d_H(A, B) := \max \left\{ \max_{a \in A} d(a, B), \max_{b \in B} d(b, A) \right\} \quad (2)$$

Given a function $f : M \rightarrow \mathbb{R}$ and $\lambda \in \mathbb{R}$, we denote by $\{f \geq \lambda\}$ the λ -level set $\{x \in M : f(x) \geq \lambda\}$.

For the estimation of level sets in general metric spaces, two main results will play a key role. The first one is Theorem 2.1 in [26]. We will make use of the following slightly restricted version.

Theorem 1 (Molchanov, (1998)). *Let $f_n, f : M \rightarrow \mathbb{R}$ be continuous functions. Assume that for each compact set K_0 , $\sup_{x \in K_0} |f_n(x) - f(x)| \rightarrow 0$. Assume that for all $\lambda \in [c_1, c_2]$, $\{f \geq \lambda\} \subset \overline{\{f > \lambda\}}$. Then*

$$\sup_{c_1 \leq \lambda \leq c_2} d_H(\{f_n \geq \lambda\} \cap K_0, \{f \geq \lambda\} \cap K_0) \rightarrow 0.$$

The convergence of the boundaries of the level sets is in general more involved. To prove that, we will use the following result.

Theorem 2 (Cuevas, Gozález-Manteiga, Rodríguez-Casal, (2006)). *Given a continuous function $f : M \rightarrow \mathbb{R}$, let (Ω, \mathcal{A}, P) be a probability space and $f_n = f_n(\omega, \cdot)$, with $\omega \in \Omega$, a sequence of random functions, $f_n : M \rightarrow \mathbb{R}$, $n = 1, 2, \dots$. Assume that for each n , f_n is continuous with probability one. Assume that the following assumptions are fulfilled.*

h₁) M is locally connected.

h₂) For all $x \in \partial\{f \geq \lambda\}$, there exist sequences $u_n, l_n \rightarrow x$ such that $f(u_n) > \lambda$ and $f(l_n) < \lambda$.

f₁) $\partial\{f \geq \lambda\} \neq \emptyset$. Moreover, there exists a $\lambda^- < \lambda$ such that the set $\{f \geq \lambda^-\}$ is compact.

If $\sup_{x \in M} |f(x) - f_n(x)| \rightarrow 0$ a.s., then

$$d_H(\partial\{f \geq \lambda\}, \partial\{f_n \geq \lambda\}) \rightarrow 0, \quad \text{a.s., as } n \rightarrow \infty.$$

3 Uniform consistency of \widehat{LD}_n

As mentioned above, the key points to prove the consistency in Hausdorff distance of the level sets of \widehat{LD}_n are Theorems 2 and 1. Then we have to prove that \widehat{LD}_n converges uniformly to LD a.s., which is the main goal of this section.

To obtain the a.s. uniform convergence of \widehat{LD}_n , we will use the following version of Theorem 1 in [27].

Theorem 3 (Billingsley and Topsøe, (1967)). *Let $\mathcal{B}(M \times M)$ be the class of all real valued, bounded, measurable functions defined on the metric space $(M \times M, \rho)$, where $\rho(z, y) = \max\{d(z_1, y_1), d(z_2, y_2)\}$. Suppose $\mathcal{F} \subset \mathcal{B}(M \times M)$ is a subclass of functions. Then*

$$\sup_{f \in \mathcal{F}} \left| \int f dP_n - \int f dP \right| \rightarrow 0,$$

for every sequence P_n that converges weakly to P if, and only if,

$$\sup\{|f(z) - f(t)| : f \in \mathcal{F}, z = (z_1, z_2), t = (t_1, t_2) \in M \times M\} < \infty,$$

and for all $\epsilon > 0$,

$$\lim_{\delta \rightarrow 0} \sup_{f \in \mathcal{F}} P \left[\{y = (y_1, y_2) : \omega_f\{B_\rho(y, \delta)\} \geq \epsilon\} \right] = 0, \quad (3)$$

where $\omega_f(A) = \sup\{|f(z) - f(t)| : z, t \in A\}$ and $B_\rho(y, \delta)$ is the open ball in the metric space $(M \times M, \rho)$ of radius $\delta > 0$.

To prove that (3) is fulfilled, we will use the following lemma.

Lemma 1. *Let $x \in M$ and $y = (y_1, y_2) \in M \times M$ be such that $d(x, \partial A(y_1, y_2)) > 3\delta$. Then $\omega_{f_x}\{B_\rho(y, \delta)\} = 0$, where $f_x(t_1, t_2) = \mathbb{I}_{A(t_1, t_2)}(x)$ with $t_1, t_2 \in M$.*

Theorem 4. *Let (M, d) be a complete separable metric space and P_X a Borel measure on M . Assume that $P_X(\partial B(x, \delta)) = 0$ for all $x \in M$ and $\delta > 0$. For any compact set $K \subset M$,*

$$\sup_{x \in K} |\widehat{LD}_n(x) - LD(x)| \rightarrow 0 \quad \text{a.s., as } n \rightarrow \infty.$$

As a direct consequence of the previous results, if M is a compact manifold, we have that $\sup_{x \in M} |\widehat{LD}_n(x) - LD(x)| \rightarrow 0$. For \mathbb{R}^d , uniform convergence can be obtained, as stated in the following theorem.

Theorem 5. *Assume that $P_X(\partial B(x, \delta)) = 0$ for all $x \in \mathbb{R}^d$ and $\delta > 0$. Then*

$$\sup_{x \in \mathbb{R}^d} |\widehat{LD}_n(x) - LD(x)| \rightarrow 0 \quad \text{a.s., as } n \rightarrow \infty.$$

4 Asymptotic distribution

To obtain the asymptotic law of $\widehat{\text{LD}}_n$ we will use Proposition 10 of [28], which is a very general result for empirical processes, applied to U statistics. This requires proving that the family of sets $\{A(x, y), x, y \in M\}$ has finite Vapnik–Chervonenkis (VC) dimension (see [29]). To prove the result, we introduce the *Vapnik-ball* condition.

Definition 1. *A metric space (M, d) fulfills Vapnik-ball condition if the family $\{F_p\}_{p \in M}$ where $F_p = \{(x, y) \in M \times M : d(p, x) < d(x, y)\}$ has finite VC dimension varying p .*

Let us introduce some notation and general ideas of empirical processes, following [30]. To each $x \in M$ we associate the function $f_x : M^2 \rightarrow \mathbb{R}$ defined by $f_x(a, b) = \mathbb{I}_{A(a, b)}(x)$. The family of all these functions is denoted by \mathcal{F} . Then we can define $\widehat{\text{LD}}_n$, a function of the random variable X and the point x , by

$$\widehat{\text{LD}}_n(P_X, f_x) = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} f_x(X_i, X_j),$$

where X_1, \dots, X_n is an iid sample from X with distribution P_X . Denote the metric space of all real valued bounded functions defined on \mathcal{F} , endowed with the supremum norm, by $l^\infty(\mathcal{F})$. We can consider $\widehat{\text{LD}}_n$ as a random variable with values in $l^\infty(\mathcal{F})$ given by $\omega \in \Omega \rightarrow \widehat{\text{LD}}_n : \mathcal{F} \rightarrow \mathbb{R}$. We define $\text{LD} \in l^\infty(\mathcal{F})$ by $\text{LD}(f_x) = \mathbb{P}(f_x(X_1, X_2))$, with X_1, X_2 being independent copies of X . We want to derive the limit law of the random element of $l^\infty(\mathcal{F})$ given by $G_n = \sqrt{n}(\widehat{\text{LD}}_n - \text{LD})$. The limit law of G_n is denoted by $2GP_X$, where GP_X is a Brownian bridge associated to P_X , which is a random element in $l^\infty(\mathcal{F})$, and is defined through its finite dimensional distributions. More precisely, for all k , $(GP_X(f_{x_1}), \dots, GP_X(f_{x_k}))$ is a zero-mean Gaussian process whose $k \times k$ covariance matrix has (i, j) th element

$$E(GP_X(f_{x_i})GP_X(f_{x_j})) = P_2(f_{x_i}f_{x_j}) - P_2(f_{x_i})P_2(f_{x_j}), \quad (4)$$

where

$$P_2(f_x) := \int_{M^2} f_x(x_1, x_2) dP_X(x_1) \times dP_X(x_2).$$

The convergence of G_n to $2GP_X$ is in law, which in this case is equivalent to the convergence in distribution of the random vector $(G_n(f_{x_1}), \dots, G_n(f_{x_k}))$ to $N(0, \Sigma)$, where the $k \times k$ matrix Σ has (i, j) th element given by (4).

Theorem 6. (*Limit Law*) Let M be a compact manifold fulfilling the Vapnik-ball condition. Following the previous notation, the stochastic processes G_n converge in law to $2GP_X$, as $n \rightarrow \infty$.

5 Level set estimation

Given positive numbers $c_1 < c_2$, assume that for all $\lambda \in [c_1, c_2]$, we have $\{f \geq \lambda\} \subset \overline{\{f > \lambda\}}$. From Theorem 2.1 in [26] together with Theorem 4, and the fact that LD is a continuous function, we get

$$\sup_{c_1 \leq \lambda \leq c_2} d_H(\{\widehat{LD}_n \geq \lambda\} \cap K, \{LD \geq \lambda\} \cap K) \rightarrow 0 \quad a.s., \quad \text{as } n \rightarrow \infty$$

for any compact set K . If $\nu(\{LD = \lambda\}) = 0$, it follows easily from Theorem 4 that for all compact K ,

$$d_\nu(\{\widehat{LD}_n \geq \lambda\} \cap K, \{LD \geq \lambda\} \cap K) \rightarrow 0 \quad a.s., \quad \text{as } n \rightarrow \infty.$$

The following theorem states that $\partial\{\widehat{LD}_n \geq \lambda\} \cap K$ is a consistent estimator of $\partial\{LD \geq \lambda\} \cap K$. The proof follows the same lines used to prove Theorem 1 in [31]. However, we can not apply that theorem directly because $\widehat{LD}_n(x)$ is not a continuous function, since the range of $\widehat{LD}_n(x)$ is contained in the set $\{k \binom{n}{2}^{-1} : k = 0, \dots, \binom{n}{2}\}$.

Theorem 7. Let $\lambda > 0$ be such that $\{LD \geq \lambda\} \neq \emptyset$. Under the assumptions of Theorem 4, together with hypotheses h1 and h2 of Theorem 2, we have that

$$\lim_{n \rightarrow \infty} d_H(\partial\{\widehat{LD}_n \geq \lambda\} \cap K, \partial\{LD \geq \lambda\} \cap K) = 0 \quad a.s.,$$

for all compact sets K .

For the metric space $(\mathbb{R}^d, \|\cdot\|)$ endowed with the Euclidean norm we have the following trivial corollary,

Corollary 1. Assume that hypothesis h2 of Theorem 2 is fulfilled. Assume also that $\{LD \geq \lambda^-\}$ is non-empty and compact, for some $\lambda^- < \lambda$. Then

$$\lim_{n \rightarrow \infty} d_H(\partial\{\widehat{LD}_n \geq \lambda\}, \partial\{LD \geq \lambda\}) = 0 \quad a.s. \quad (5)$$

6 An application to the study of two sets of real data

In what follows we analyze two interesting real data sets: the vectocardiogram and the influenza data sets.

6.1 The vectocardiogram dataset

We consider a real life dataset, where the data belong to the Stiefel manifold $SO(3, 2)$ of all orthonormal 2-frames in \mathbb{R}^3 considered as 3×2 orthogonal matrices (see [32]). The dataset consists of 98 vectocardiograms from children with ages varying between 2 and 19. Vectocardiography is a method that produces a three dimensional curve which comprises the records of the magnitude and direction of the electrical forces generated by the heart over time. These curves are called QRS loops, see Figure 2. In [33] there is associated to each curve an element of $SO(3, 2)$ that represents some of the information of the curve, see also [34].

This sample has been previously analysed in the literature, see for instance [35, 36, 37]. A very important problem is outlier detection, corresponding to children with (possibly) cardiological problems.

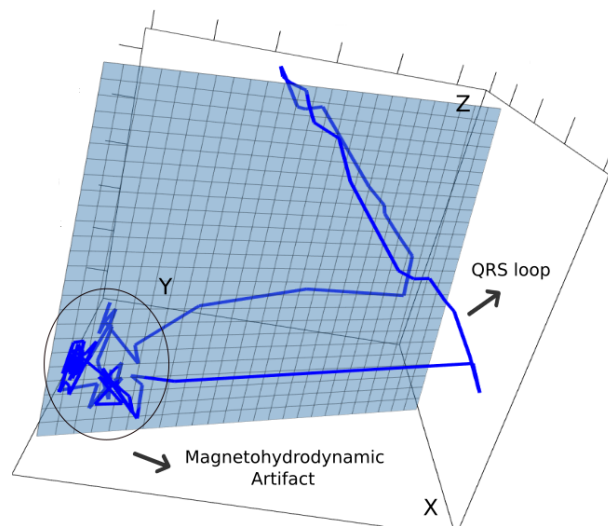


Figure 2: QRS curve loop. The plane in blue is the best two dimensional approximation of the curve by least squares.

Figure 3 represents each matrix in $SO(3, 2)$ as two points in S^2 , one for each column. They are joined by an arc in S^2 . The arc joining the deepest pair

of observations (w.r.t. lens depth) is represented in violet, while the outliers (for a level $\lambda = 0.10$) are represented as red arcs.

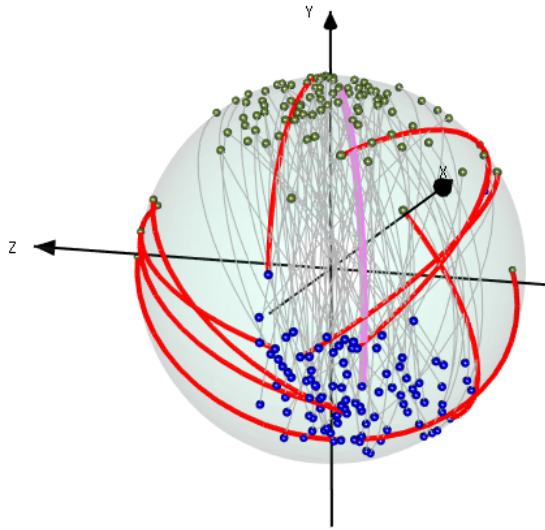


Figure 3: Data visualization of the data from $SO(3, 2)$. The violet arc represents the deepest observation according to lens depth. The outliers represented by red arcs correspond to the data outside the $\lambda = 0.10$ level set of the depth.

6.2 Influenza dataset

This dataset consists of longitudinal data of the influenza virus, belonging to the family Orthomyxoviridae. An important problem analyzed in [12], is to model the genomic evolution of the virus, see [38]. In this paper we focus on another important issue: modelling the temporal variability of the virus by means of the lens depth, and use this to predict and anticipate a possible pandemic. The influenza virus has an RNA genomic which is very common: it produces diseases like yellow fever and hepatitis and annually costs half a million deaths worldwide. It is well known that these viruses

change their genetic pattern over time, which is vital for developing a possible vaccine. We will study the the H3N2 variant of the virus, in particular, the subtype hemagglutini (HA), which produced the SARS pandemic in 2002. This variant is known to have a variability in its genetic arrangement over time, see [39, 40].

The dataset can be found in GI-SAID EpiFluTM database ¹, providing 1089 genomic sequences of H3N1 from 1993 to 2017 in New York, aligned using MUSCULE, see [41]. We used reduced trees of 5 leaves, as in [40], to capture the structure of the data. This set of trees can be endowed with a distance, see [42]. The database used was obtained from the GitHub repository <https://github.com/antheamonod/FluPCA>.

For this kind of data, constructing measures of centrality and variability, as well as confidence regions, is a problem that has been previously addressed in the literature, see for instance [43, 44, 45] and references therein.

For each year we computed the empirical lens depth of the trees, considered on the manifold of phylogenetic trees. We estimated the diameter of the level sets from the sample points that belong to the level sets, see Figure 4. As can be seen, there is a larger dispersion in the years prior to the pandemic of 2002 compared to later years.

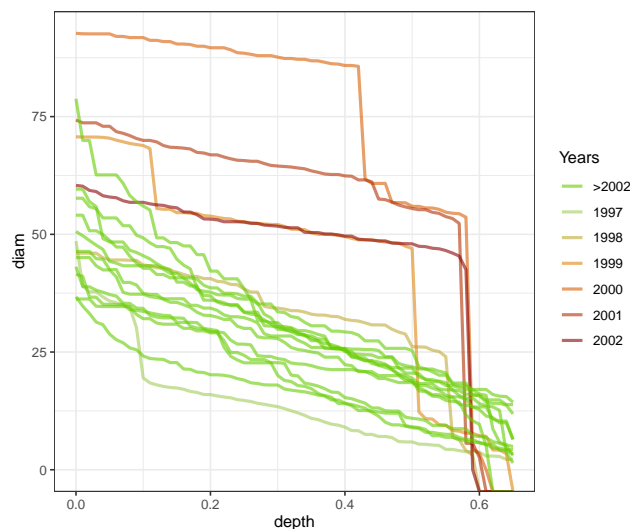


Figure 4: Diameters of the lens depth level sets from the sample of trees for the different years, with respect to λ .

¹www.gisaid.org

Appendix

Proof of Lemma 1

If $z = (z_1, z_2) \in B_\rho(y, \delta)$, then $d(z_1, y_1) < \delta$ and $d(z_2, y_2) < \delta$. Let us prove first that

$$\{x \in A(y_1, y_2) : d(x, \partial A(y_1, y_2)) > 2\delta\} \subset A(z_1, z_2) \quad (6)$$

for all $z = (z_1, z_2) \in B_\rho(y, \delta)$. If $x \in A(y_1, y_2)$ and $d(x, \partial A(y_1, y_2)) > 2\delta$, then $d(x, y_1) \leq d(y_1, y_2) - 2\delta \leq d(y_1, z_1) + d(z_1, z_2) + d(z_2, y_2) - 2\delta \leq d(z_1, z_2)$. In the same way, if $d(x, y_2) \leq d(z_1, z_2)$, then $x \in A(z_1, z_2)$, which implies (6). Next let us prove that

$$A(z_1, z_2) \subset \{x : d(x, A(y_1, y_2)) \leq 3\delta\} \quad (7)$$

For any $x \in A(z_1, z_2)$, $d(x, y_1) \leq d(x, z_1) + d(z_1, y_1) \leq d(z_1, z_2) + \delta \leq d(y_1, y_2) + 3\delta$. In the same way, $d(x, y_2) \leq d(x, z_2) + \delta \leq d(z_1, z_2) + \delta \leq d(y_1, y_2) + 3\delta$. Then $x \in \{x : d(x, A(y_1, y_2)) \leq 3\delta\}$. To prove that $\omega_{f_x}\{B_\rho(y, \delta)\} = 0$, it is enough to prove that for all $(z_1, z_2), (t_1, t_2) \in B_\rho((y_1, y_2), \delta)$

$$|\mathbb{I}_{A(z_1, z_2)}(x) - \mathbb{I}_{A(t_1, t_2)}(x)| = 0. \quad (8)$$

Observe that if $x \in A(y_1, y_2)$ and $d(x, \partial A(y_1, y_2)) > 3\delta$, then by (6), $x \in A(z_1, z_2)$, so $\mathbb{I}_{A(z_1, z_2)}(x) = 0$ and $x \in A(t_1, t_2)$ and hence $\mathbb{I}_{A(t_1, t_2)}(x) = 0$ which then shows that (8) holds. Proceeding in the same way if $x \notin A(y_1, y_2)$ and $d(x, \partial A) > 3\delta$, by (7) $x \notin A(z_1, z_2)$, which implies that $\mathbb{I}_{A(z_1, z_2)}(x) = 0$. Also $x \notin A(t_1, t_2)$, which implies that $\mathbb{I}_{A(t_1, t_2)}(x) = 0$, so again (8) holds. \square

Proof of Theorem 4

We will apply Billingsley's theorem to the set of functions

$$\mathcal{F} = \{f_x(z, y) = \mathbb{I}_{A(z, y)}(x) : x \in K\},$$

where the sequence P_n of probability measures on $M \times M$ is such that $P_n(X_i, X_j) = (1/2)\binom{n}{2}^{-1}$ if $i \neq j$ and $P_n(X_i, X_j) = 0$ if $i = j$. Let P be the product measure $P_X \times P_X$. In this case

$$\sup_{f \in \mathcal{F}} \left| \int f dP_n - \int f dP \right| = \sup_{x \in K} |\widehat{\text{LD}}_n(x) - \text{LD}(x)|.$$

Clearly, $\sup\{|f(z) - f(t)| : f \in \mathcal{F}, z, t \in M \times M\} = 2$. So we have to prove that (3) holds. By Lemma (1), to prove (3), it is enough to prove that

$$\limsup_{\delta \rightarrow 0} \sup_{x \in K} P_X \times P_X\{(y_1, y_2) : d(x, \partial A(y_1, y_2)) \leq 3\delta\} = 0.$$

By the dominated convergence theorem, since $P_X(\partial B(x, \delta)) = 0$ for all $\delta > 0$ and all $x \in M$, it follows that for fixed δ , $P_X \times P_X\{(y_1, y_2) : d(x, \partial A(y_1, y_2)) \leq 3\delta\}$ is a continuous function of x . Also, for a fixed x , $P_X \times P_X\{(y_1, y_2) : d(x, \partial A(y_1, y_2)) \leq 3\delta\} \rightarrow 0$, again by the dominated convergence theorem.

By Lemma 1,

$$\sup_{x \in K} P_X \times P_X \left[\{y = (y_1, y_2) : \omega_{f_x}\{B_\rho(y, \delta)\} \geq \epsilon\} \right] \leq \sup_{x \in K} P_X \times P_X \{(y_1, y_2) : d(x, \partial A(y_1, y_2)) \leq 3\delta\}$$

which converges to 0 as $\delta \rightarrow 0$. \square

Proof of Theorem 5

Let $\epsilon > 0$ and suppose that R is large enough so that $P_X(\overline{B(0, R)^c}) < \epsilon$, and take $K = \overline{B(0, R)}$. We will apply the previously mentioned theorems of Billingsley and Topsøe to the set of functions

$$\mathcal{F} = \{f_x(z, y) = \mathbb{I}_{A(z, y)}(x) : x \in \mathbb{R}^d\}.$$

We start by splitting into two terms

$$\begin{aligned} \sup_x P_X \times P_X \left[\{y = (y_1, y_2) : \omega_{f_x}\{B_\rho(y, \delta)\} \geq \epsilon\} \right] &\leq \\ \sup_{x \in K} P_X \times P_X \left[\{y = (y_1, y_2) : \omega_{f_x}\{B_\rho(y, \delta)\} \geq \epsilon\} \right] &+ \\ \sup_{x \in K^c} P_X \times P_X \left[\{y = (y_1, y_2) : \omega_{f_x}\{B_\rho(y, \delta)\} \geq \epsilon\} \right] &= I_1 + I_2. \end{aligned}$$

By Lemma 1, $I_1 \leq \sup_{x \in K} P_X \times P_X\{(y_1, y_2) : d(x, \partial A(y_1, y_2)) \leq 3\delta\}$, which converges to 0 as $\delta \rightarrow 0$. Regarding I_2 , we bound

$$\begin{aligned} P_X \times P_X \left[\{y = (y_1, y_2) : \omega_{f_x}\{B_\rho(y, \delta)\} \geq \epsilon\} \right] &\leq \\ P_X \times P_X \left[\{y = (y_1, y_2) : y_1, y_2 \in K \text{ and } \omega_{f_x}\{B_\rho(y, \delta)\} \geq \epsilon\} \right] &+ \\ + P_X \times P_X \left[\{y = (y_1, y_2) : y_1 \in K^c \text{ or } y_2 \in K^c \text{ and } \omega_{f_x}\{B_\rho(y, \delta)\} \geq \epsilon\} \right]. & \end{aligned}$$

The second term is bounded from above by $2P_X(K^c) < 2\epsilon$. Lastly, to tackle the first term, by Lemma 1, it only remains to prove that

$$\lim_{\delta \rightarrow 0} \sup_{x \in K^c} P_X \times P_X \left[\{y = (y_1, y_2) : y_1, y_2 \in K \text{ and } d(x, \partial A(y_1, y_2)) < 3\delta\} \right] = 0.$$

Lastly, let δ_0 be small enough such that $P_X((\partial K)^{3\delta}) < \epsilon$ for all $\delta < \delta_0$. If $y_1, y_2 \in B(0, R - 3\delta)$, then $d(x, A(y_1, y_2)) \geq 3\delta$. Therefore,

$$P_X \times P_X \left[\{y = (y_1, y_2) : y_1, y_2 \in K \text{ and } d(x, \partial A(y_1, y_2)) < 3\delta\} \leq P_X(y_1 \in (\partial K)^{3\delta} \text{ or } y_2 \in (\partial K)^{3\delta}) < 2\epsilon. \quad (9)$$

□

Proof of Theorem 6

By assumption, we have that \mathcal{F} has a finite VC -dimension. This implies that the assumptions in Proposition 10 in [28] hold for an order 2 U -statistic and the asymptotic distribution is derived from Theorem 4.10 in [30]. □

Proof of Theorem 7

Let K be any compact set. Observe that $\{\text{LD} \geq \lambda^-\} \cap K$ is nonempty and compact for any $0 < \lambda^- < \lambda$. Following the same ideas as those used in the proof of Theorem 1 in [31], one can derive that for all $\epsilon > 0$

$$\partial\{\text{LD} \geq \lambda\} \cap K \subset B(\partial\{\widehat{\text{LD}}_n \geq \lambda\} \cap K, \epsilon).$$

The proof of

$$\partial\{\widehat{\text{LD}}_n \geq \lambda\} \cap K \subset B(\partial\{\text{LD} \geq \lambda\} \cap K, \epsilon)$$

is slightly different from the analogous inclusion in [31]. If we proceed by contradiction, there exists an $x_n \in \{\widehat{\text{LD}}_n \geq \lambda\} \cap K$, but

$$d(x_n, \partial\{\text{LD} \geq \lambda\} \cap K) > \epsilon. \quad (10)$$

Observe that if $x_n \in \partial\{\widehat{\text{LD}}_n \geq \lambda\} \cap K$, then $x_n \in \partial A(X_i, X_j)$ for some $X_i \neq X_j$. If x_n is in the boundary of two or more sets $A(X_i, X_j)$, we can take y_n such that $d(y_n, x_n) < \epsilon/2^n$, $y_n \in \partial\{\widehat{\text{LD}}_n(x) \geq \lambda\}$, y_n in the boundary of only one $\partial A(X_i, X_j)$ and $|\widehat{\text{LD}}_n(y_n) - \lambda| < \binom{n}{2}^{-1}$. If x_n is in the boundary of only one $A(X_i, X_j)$, we choose $x_n = y_n$. In this case we clearly have also have $|\widehat{\text{LD}}_n(y_n) - \lambda| < \binom{n}{2}^{-1}$

Since K is compact, there exists an $x \in K$ such that $x_n, y_n \rightarrow x$ (by considering a subsequence if necessary). Then $|\text{LD}(x) - \lambda| \leq |\text{LD}(x) - \text{LD}(y_n)| +$

$|\text{LD}(y_n) - \widehat{\text{LD}}(y_n)| + |\widehat{\text{LD}}(y_n) - \lambda|$. The first terms converge to 0 by the continuity of LD at x . The second term converges to 0 a.s., because the set $K' = \{\cup_n \{y_n\}\} \cup \{x\}$ is compact and by Theorem 4, $\sup_{z \in K'} |\widehat{\text{LD}}_n(z) - \text{LD}(z)| \rightarrow 0$. The last term is bounded from above by $\binom{n}{2}^{-1} \rightarrow 0$ as $n \rightarrow \infty$. Then $\text{LD}(x) = \lambda$, which implies that $d(x_n, \partial\{\text{LD} \geq \lambda\} \cap K) \leq d(x_n, x) \rightarrow 0$, which contradicts (10). \square

References

- [1] J. W. Tukey, Mathematics and the picturing of data, in: Proceedings of the international congress of mathematicians, Vol. 2, 1975, pp. 523–531.
- [2] R. Y. Liu, On a Notion of Data Depth Based on Random Simplices, *The Annals of Statistics* 18 (1) (1990) 405–414.
- [3] R. Y. Liu, Data depth and multivariate rank tests, L1-statistical analysis and related methods (Y. Dodge, ed.) (1992) 279–294.
- [4] V. Barnett, The ordering of multivariate data, *Journal of the Royal Statistical Society. Series A (General)* 139 (3) (1976) 318–355.
URL <http://www.jstor.org/stable/2344839>
- [5] H. Oja, Descriptive statistics for multivariate distributions, *Statistics & Probability Letters* 1 (6) (1983) 327–332.
- [6] R. T. Elmore, T. P. Hettmansperger, F. Xuan, Spherical data depth and a multivariate median, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 72 (2006) 87.
- [7] R. Fraiman, G. Muniz, Trimmed means for functional data, *Test* 10 (2) (2001) 419–440.
- [8] S. López-Pintado, J. Romo, On the concept of depth for functional data, *Journal of the American Statistical Association* 104 (486) (2009) 718–734.
- [9] G. Claeskens, M. Hubert, L. Slaets, K. Vakili, Multivariate functional halfspace depth, *Journal of the American Statistical Association* 109 (505) (2014) 411–423.
- [10] A. Cuevas, R. Fraiman, On depth measures and dual statistics. a methodology for dealing with general data, *Journal of Multivariate Analysis* 100 (4) (2009) 753–766.

- [11] R. Fraiman, F. Gamboa, L. Moreno, Connecting pairwise geodesic spheres by depth: Dcops, *Journal of Multivariate Analysis* 169 (2019) 81–94.
- [12] R. Fraiman, F. Gamboa, L. Moreno, Weighted lens depth: Some applications to supervised classification, *arXiv* (2020).
- [13] O. Vencálek, Depth-based classification for multivariate data, *Austrian Journal of Statistics* 46 (3-4) (2017) 117–128.
- [14] K. Mosler, P. Mozharovskyi, Fast dd-classification of functional data, *Statistical Papers* 58 (4) (2017) 1055–1089.
- [15] Z. Liu, R. Modarres, Lens data depth and median, *Journal of Nonparametric Statistics* 23 (4) (2011) 1063–1074.
- [16] A. Cholaquidis, R. Fraiman, L. Moreno, Level set and density estimation on manifolds, *arXiv* (2020).
- [17] G. Koshevoy, K. Mosler, et al., Zonoid trimming for multivariate distributions, *Annals of Statistics* 25 (5) (1997) 1998–2017.
- [18] Y. Zuo, R. Serfling, General notions of statistical depth function, *The Annals of Statistics* 28 (2) (2000) 461–482.
- [19] R. Serfling, Quantile functions for multivariate analysis: approaches and applications, *Statistica Neerlandica* 56 (2) (2002) 214–232.
- [20] R. Dyckerhoff, Convergence of depths and depth-trimmed regions, *arXiv preprint arXiv:1611.08721* (2016).
- [21] R. Y. Liu, J. M. Parelius, K. Singh, et al., Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by liu and singh), *The annals of statistics* 27 (3) (1999) 783–858.
- [22] M. Febrero, P. Galeano, W. González-Manteiga, Outlier detection in functional data by depth measures, with application to identify abnormal nox levels, *Environmetrics: The official journal of the International Environmetrics Society* 19 (4) (2008) 331–345.
- [23] W. Dai, M. G. Genton, Directional outlyingness for multivariate functional data, *Computational Statistics and Data Analysis* 131 (2019) 50 – 65.

- [24] I. Ruts, P. J. Rousseeuw, Computing depth contours of bivariate point clouds, *Computational Statistics and Data Analysis* 23 (1) (1996) 153 – 168.
- [25] M. Hubert, P. Rousseeuw, P. Segaert, Multivariate and functional classification using depth and distance, *Advances in Data Analysis and Classification* 11 (3) (2017) 445–466.
- [26] I. S. Molchanov, A limit theorem for solutions of inequalities, *Scandinavian Journal of Statistics* 25 (1) (1998) 235–242.
- [27] P. Billingsley, F. Topsøe, Uniformity in weak convergence, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 7 (1) (1967) 1–16.
- [28] E. Giné, Empirical processes and applications: an overview, *Bernoulli* 2 (1) (1996) 1–28.
- [29] L. Devroye, L. Györfi, G. Lugosi, A probabilistic theory of pattern recognition, Vol. 31, Springer Science & Business Media, 2013.
- [30] M. A. Arcones, E. Gine, Limit theorems for u -processes, *The Annals of Probability* 21 (3) (1993) 1494–1542.
URL <http://www.jstor.org/stable/2244585>
- [31] A. Cuevas, W. González-Manteiga, A. Rodríguez-Casal, Plug-in estimation of general level sets, *Australian & New Zealand Journal of Statistics* 48 (1) (2006) 7–19.
- [32] A. Hatcher, *Algebraic topology*, 2005.
- [33] T. D. Downs, J. Liebman, Statistical methods for vectorcardiographic directions, *IEEE Transactions on Biomedical Engineering* 1 (1969) 87–94.
- [34] T. D. Downs, Orientation statistics, *Biometrika* 59 (3) (1972) 665–676.
- [35] Y. Chikuse, *Statistics on special manifolds*, Vol. 174, Springer Science & Business Media, 2012.
- [36] R. Chakraborty, B. C. Vemuri, et al., Statistics on the stiefel manifold: theory and applications, *The Annals of Statistics* 47 (1) (2019) 415–438.
- [37] S. Pal, S. Sengupta, R. Mitra, A. Banerjee, A bayesian approach for analyzing data on the stiefel manifold, arXiv preprint arXiv:1907.04303 (2019).
- [38] D. J. Smith, A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. Osterhaus, R. A. Fouchier, Mapping the antigenic

- and genetic evolution of influenza virus, *science* 305 (5682) (2004) 371–376.
- [39] L. K. Altman, This season’s flu virus is resistant to 2 standard drugs, *New York Times* (2006).
 - [40] A. Monod, B. Lin, R. Yoshida, Q. Kang, Tropical geometry of phylogenetic tree space: A statistical perspective, arXiv preprint arXiv:1805.12400 (2018).
 - [41] R. C. Edgar, Muscle: multiple sequence alignment with high accuracy and high throughput, *Nucleic acids research* 32 (5) (2004) 1792–1797.
 - [42] L. J. Billera, S. P. Holmes, K. Vogtmann, Geometry of the space of phylogenetic trees, *Advances in Applied Mathematics* 27 (4) (2001) 733–767.
 - [43] D. Barden, H. Le, M. Owen, Limiting behaviour of fréchet means in the space of phylogenetic trees, *Annals of the Institute of Statistical Mathematics* 70 (1) (2018) 99–129.
 - [44] D. G. Brown, M. Owen, Mean and variance of phylogenetic trees, *Systematic Biology* 69 (1) (2020) 139–154.
 - [45] A. Willis, Confidence sets for phylogenetic trees, *Journal of the American Statistical Association* 114 (525) (2019) 235–244.