ORIGINAL PAPER

Level sets of depth measures in abstract spaces

A. Cholaquidis[1] · **R. Fraiman[1]** · **L. Moreno[2]**

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 prove that, on a separable and complete metric space, the level sets of the empirical lens depth based on an iid sample, converge in the Painlevé–Kuratowski sense, to its population counterpart. We also prove that, restricted to compact sets, the empirical level sets and their boundaries are consistent estimators, in Hausdorff distance, of their population counterparts, and analyse 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 (Tukey 1975), the simplicial depth (Liu 1990, 1992), and the multivariate L^1 depth.

✉ A. Cholaquidis
acholaquidis@cmat.edu.uy

R. Fraiman
rfraiman@cmat.edu.uy

L. Moreno
mrleo@iesta.edu.uy

[1] Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

[2] Departamento de Métodos Cuantitativos, FCEA, Instituto de Estadística, Universidad de la República, Montevideo, Uruguay

Other well-known depths are the convex hull peeling depth (Barnett 1976), the Oja depth (Oja 1983), and the spherical depth (Elmore et al. 2006), among others. They have been extended to functional spaces, see (Fraiman and Muniz 2001; López-Pintado and Romo 2009; Claeskens et al. 2014; Cuevas and Fraiman 2009), to Riemannian manifolds, see (Fraiman et al. 2019), and also to general metric spaces, see (Cholaquidis et al. 2020). Several different applications of depths notions have been proposed, in particular for classification problems, by means of the depth-depth method (Vencálek 2017), or to functional data, see (Mosler and Mozharovskyi 2017). Most classical notions of depth, introduced initially on $\mathbb{R}^d$, cannot be directly 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 dimension.

This is not the case of the lens depth, introduced in (Liu and Modarres 2011), 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.

The estimation of the level sets of a depth measure is of current interest, see (Brunel 2019). As indicated in Laketa and Nagy (2021), "a definition of depth-based central regions of the data, which are the regions where the depth exceeds given thresholds, ensues naturally". Level set estimation of depths was initially studied in (Tukey 1975), as a key tool for the visualization and exploration of data. Other significant contributions can be found in (Koshevoy et al. 1997; Zuo and Serfling 2000; Serfling 2002; Dyckerhoff 2016). As it was pointed out in (Liu et al. 1999), "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, (Febrero et al. 2008; Dai and Genton 2019), as well as for supervised classification, see (Ruts and Rousseeuw 1996; Hubert et al. 2017). The boundary of the level sets allows to extend the notion of quantile, see (Agostinelli 2018). As shown in the two real data sets we analyse, the shape of the level sets of the lens depth can help to detect outliers. This is of crucial importance, when dealing with vectorcardiograms, because outliers mean data from children with possibly cardiological problems. More precisely, during the electrical activity of the heart, myocardial cells depolarize and repolarize, producing current flows of different magnitudes and directions. The sum of all these current flows gives rise to a vector indicating the main direction of the flow. Thus, an atrial depolarization vector, three ventricular depolarization vectors and a ventricular repolarization vector are identified. In reality, these vectors do not occur independently of each other, but rather the current, once it begins to flow, originates a vector whose size and magnitude are modified at each moment. Vectorcardiography is the recording of the electrical activity of the heart through the originating vectors, which have the form of a closed loop.

An extension to Riemannian manifolds of the lens depth (called weighted lens depth) was recently introduced in (Cholaquidis et al. 2020) to tackle the supervised classification problem, and a point-wise a.s consistency result is obtained (see Theorem 2 in (Cholaquidis et al. 2020)). Here, we address the more challenging problem of level

set estimation of the lens depth on general metric spaces, while in (Cholaquidis et al. 2020) weighted lens depth is introduced to deal with the density f together with the underlying Riemannian structure to perform supervised classification by means of the depth-depth-G method introduced in Cuesta-Albertos et al. (2017).

In our setting, the lens depth is defined as follows: let X_1, X_2 be two random elements 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 $\mathrm{LD}(x) = \mathbb{P}(x \in A(X_1, X_2))$. Given an iid sample $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ from a distribution P_X, the empirical version of LD is given by the U-statistics of order two,

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

To gain some insight into the shape of $\{\widehat{\mathrm{LD}}_n \ge \lambda\}$, see Fig. 1.

The level set estimation problem is required to define a topology in which the level set estimators will converge. In Molchanov (1998) almost sure consistency is obtained, when the level sets are intersected with compact sets, and the topology inherited from Hausdorff distance is considered. In Cuevas et al. (2006), they are assumed to be compact.

To ask compactness is not too restrictive for data in the classical finite-dimensional Euclidean space. However, this is not the case for infinite-dimensional spaces like in the functional data setting. To overcome this problem, other weaker topologies were considered, for instance, Mosco or Wijsman topology, see Terán (2016).

Following this idea, we prove first that the empirical version of the level sets of the lens depth is almost surely (a.s.) consistent estimators of their population counterpart in the Painlevé–Kuratowski topology.

The general approach we use to obtain a.s. consistency results for the plug-in estimator $\{\widehat{\mathrm{LD}}_n \ge \lambda\}$ to its population counterpart $\{\mathrm{LD} \ge \lambda\}$ is to prove the uniform convergence of $\widehat{\mathrm{LD}}_n$ to LD on the whole space for separable and complete metric spaces. This has interest in itself and extends the results in Liu and Modarres (2011) for the finite-dimensional Euclidean space.

Next, we prove that the empirical version of the level sets of the lens depth converges a.s. in Hausdorff distance, when they are met with any compact set.

This paper is organized as follows. Section 2 introduces the notation, some previous definitions, and the notion of Painlevé–Kuratowski convergence. In Sect. 3, we state the a.s. convergence of level sets in such topology. The a.s. uniform consistency of $\widehat{\mathrm{LD}}_n$ is stated in Sect. 4. Section 5 addresses the problem of a.s. convergence of level
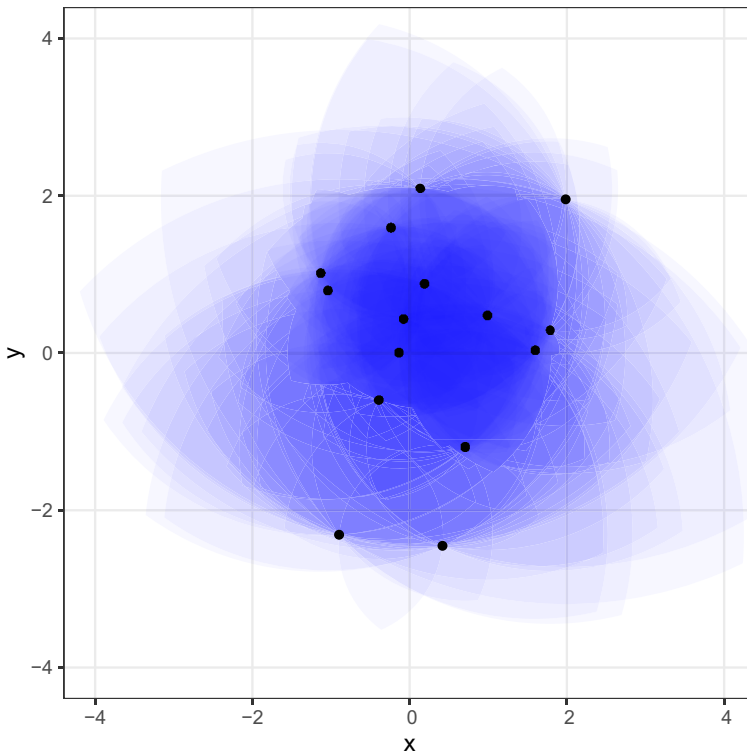
Fig. 1 Level sets of the lens depth based on a sample of 15 iid random vectors, distributed as a bivariate standard normal distribution. The intensity of blue represents the depth

sets in the Hausdorff distance, as well as the a.s. convergence in measure, together with the a.s. consistency of their boundaries. All proofs are given in Appendix.

Lastly, in Sect. 6 we tackle the study of two interesting real data sets, the vectorcardiogram data set (see Sect. 6.1), and the influenza data set (see Sect. 6.2).

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 and complete, 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 ∂A. Given $x \in A^c$, we denote $d(x, A) = \inf_{a \in A} d(x, a)$ and $\text{diam}(A) = \sup_{x, y \in A} d(x, y)$. Given two non-empty closed sets A, B, the Hausdorff distance between them is defined as

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

where $d_H(\emptyset, B) = +\infty$ with $B \neq \emptyset$ and $d_H(\emptyset, \emptyset) = 0$.

Given a finite Borel measure ν and two measurable sets A and B, the distance in measure between them is defined as $d_\nu(A, B) = \nu(A\backslash B) + \nu(C\backslash B)$.

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

2.1 Painlevé–Kuratowski convergence

We will prove first the convergence of the estimated level sets to its population counterpart in the Painlevé–Kuratowski sense. Given a sequence of sets $A_n \subset M$, the Painlevé–Kuratowski limit inferior of A_n as $n \rightarrow \infty$ is

$$\underset{n\rightarrow\infty}{\text{Li}} \, A_n = \left\{x \in M : \underset{n\rightarrow\infty}{\limsup} \, d(x, A_n) = 0\right\}.$$

The Painlevé–Kuratowski limit superior of A_n as $n \rightarrow \infty$ is

$$\underset{n\rightarrow\infty}{\text{Ls}} \, A_n = \left\{x \in M : \underset{n\rightarrow\infty}{\liminf} \, d(x, A_n) = 0\right\}.$$

In general $\underset{n\rightarrow\infty}{\text{Li}} \, A_n \subset \underset{n\rightarrow\infty}{\text{Ls}} \, A_n$. When they agree, the common set is called the Painlevé–Kuratowski limit. It is immediate to prove that if $d_H(A_n, A) \rightarrow 0$, then $A_n \rightarrow A$ in the Painlevé–Kuratowski sense. The converse implication is not true in general, unless M is totally bounded, see Corollary 5.1.11 and page 147 in Beer (1993), since we are assuming that M is complete.

3 Painlevé–Kuratowski convergence of level set estimators

We will prove that the limit in the Painlevé–Kuratowski sense of the sequence of sets $\{\widehat{LD}_n \geq \lambda\}$, is the set $\{LD \geq \lambda\}$. This will follow from the uniform convergence of the empirical version of the lens depth given in the next section and the fact that $LD(x)$ is a continuous function of x, which follows from Proposition 3 in (Cholaquidis et al. 2020).

Theorem 1 *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 $\delta > 0$. Assume also that λ is such that if $LD(x) = \lambda$, there exist $u_n \rightarrow x$ such that $LD(u_n) > \lambda$. Then, with probability one, for n large enough,*

$$\{LD \geq \lambda\} = \underset{n\rightarrow\infty}{Li} \{\widehat{LD}_n \geq \lambda\} = \underset{n\rightarrow\infty}{Ls} \{\widehat{LD}_n \geq \lambda\}.$$

Remark 1 Observe that if $P_X(\partial B(x, \delta)) \neq 0$, lens depth may be pathological in infinite-dimensional spaces. Indeed, suppose that the distribution P_X is a discrete distribution supported on an orthogonal basis of L^2 on the unit sphere. Then, all points in the support of the distribution have lens depth one. Regarding the hypothesis on the λ-level set, the following proposition states that the set of λ for which this condition does not holds is at most countable.

Proposition 1 *Let (M, d) be a separable metric space and $f : M \to \mathbb{R}$. Let us denote*

$$\mathcal{A} = \{\lambda : f^{-1}(\lambda) \text{ contains at least a local maximum}\}.$$

Then, $\mathcal{A}$ is countable.

4 Uniform consistency of $\widehat{LD}_n$

A key argument to prove the consistency of level sets in the Painlevé–Kuratowski topology (as well as in the Hausdorff distance) is to prove that $\widehat{LD}_n$ converges uniformly to LD a.s., which is the main goal of this section.

For that purpose, we will use the following version of Theorem 1 in (Billingsley and Topsøe 1967).

Theorem 2 (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,*

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

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

$$\sup\{|f(z) - f(t)| : f \in \mathcal{F}, z = (z_1, z_2), t = (t_1, t_2) \in M \times M\} < \infty, \quad (3)$$

and for all $\epsilon > 0$,

$$\lim_{\delta \to 0} \sup_{f \in \mathcal{F}} P\left[\{y = (y_1, y_2) : \omega_f\{B_\rho(y, \delta)\} \geq \epsilon\} \right] = 0, \quad (4)$$

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 $\delta > 0$.

Theorem 3 *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 $\delta > 0$. Then,*

$$\sup_x |\widehat{LD}_n(x) - LD(x)| \to 0 \quad a.s., \text{ as } n \to \infty.$$

The previous result is proven for $M = \mathbb{R}^d$ in (Liu and Modarres 2011), endowed with the Euclidean norm.

5 Level set estimation in Hausdorff metric

To obtain the consistency in Hausdorff distance of the level set estimator, Theorem 2.1 in (Molchanov 1998) will play a key role, together with the uniform convergence of the previous section. We will make use of the following slightly restricted version.

Theorem 4 (Molchanov (1998)) *Let f_n, $f : M \to \mathbb{R}$ be continuous functions. Assume that for each compact set K_0, $\sup_{x \in K_0} |f_n(x) - f(x)| \to 0$. Assume that for all $\lambda \in [c_1, c_2]$, $\{f \geq \lambda\} \subset \overline{\{f > \lambda\}}$. Then,*

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

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 (Molchanov 1998) together with Theorem 3, and the fact that LD is a continuous function, we get

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

for any compact set K. If $\nu(\{\mathrm{LD} = \lambda\}) = 0$, it follows easily from Theorem 3 that for all compact K,

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

being d_ν any finite Borel measure.

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

Theorem 5 (Cuevas et al. (2006)) *Given a continuous function $f : M \to \mathbb{R}$, let $f_n = f_n(\omega, \cdot)$, with $\omega \in \Omega$, a sequence of functions $f_n : M \to \mathbb{R}$, $n = 1, 2, \ldots,$. Assume that for each n, f_n is continuous with probability one. Assume that the following assumptions are fulfilled.*

- (h_1) *M is locally connected.*
- (h_2) *For all $x \in \partial\{f \geq \lambda\}$, there exist sequences $u_n, l_n \to x$ such that $f(u_n) > \lambda$ and $f(l_n) < \lambda$.*
- (f_1) *$\partial\{f \geq \lambda\} \neq \emptyset$. Moreover, there exists $\lambda^- < \lambda$ such that the set $\{f \geq \lambda^-\}$ is compact.*

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

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

The following theorem states that $\partial\{\widehat{\mathrm{LD}}_n \geq \lambda\} \cap K$ is a consistent estimator of $\partial\{\mathrm{LD} \geq \lambda\} \cap K$. The proof follows the same lines used to prove Theorem 5. However, we cannot apply that theorem directly because $\widehat{\mathrm{LD}}_n(x)$ is not a continuous function, since the range of $\widehat{\mathrm{LD}}_n(x)$ is contained in the set $\{k\binom{n}{2}^{-1} : k = 0, \ldots, \binom{n}{2}\}$.

Theorem 6 *Let $\lambda > 0$ be such that $\{LD \geq \lambda\} \neq \emptyset$. Under the assumptions of Theorem 3, together with hypotheses $h1$ and $h2$ of Theorem 5, we have that*

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

for all compact sets K.

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

In what follows, we analyse two interesting real data sets: the vectorcardiogram (VCG) and the influenza data sets.

6.1 The vectorcardiogram data set

Until the half of twentieth century, electrocardiograms were the non-invasive methodology used to detect heart diseases. In Frank (1956), an alternative and complementary methodology is introduced, called vectorcardiography (VCG). In many cases, electrocardiogram (ECG) records are not sufficient and complementary studies are necessary for the detection of cardiac failure, such as myocardial infarction, see Hafshejani et al. (2021). In this respect, the VCG can be of great help, but its interpretation is more difficult than an ECG in general. As mentioned in Pastore et al. (2019) "from the year 2000 until the present decade an increasing number of publications related to electrovectorcardiography has been observed. In this context, and mainly regarding the definition of arrhythmogenic substrates, it was observed that the association of the ECG and the vectorcardiogram (VCG) methods could provide much more information about the cardiac electrical phenomena, thus increasing its employment and allowing us to differentiate potentially fatal cases from benign ones". Therefore, an automatic procedure that detects possible anomalous VCGs may be useful for the medical community.

The vectorcardiography 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 Fig. 2. From the location in the body of the electrodes (I, E, C, A, M, H, F) (see Fig. 2), it is possible to obtain electric vectorial signals of the heart in an orthogonal system (x, y, z), called Frank's derivatives

$$\begin{cases} x = 0.610A + 0.171C - 0.781I \\ y = 0.655A + 0.345C - 1.000I \\ z = 0.133A + 0.736M - 0.264I - 0.374E - 0.231C. \end{cases}$$

Our aim is to detect outliers on a real data set of vectorcardiograms of children, corresponding to children with (possibly) cardiological problems, by means of the shape and size of the level sets of lens depth. We consider a real-life data set that consists of 98 vectorcardiograms from children with ages varying between 2 and 19,
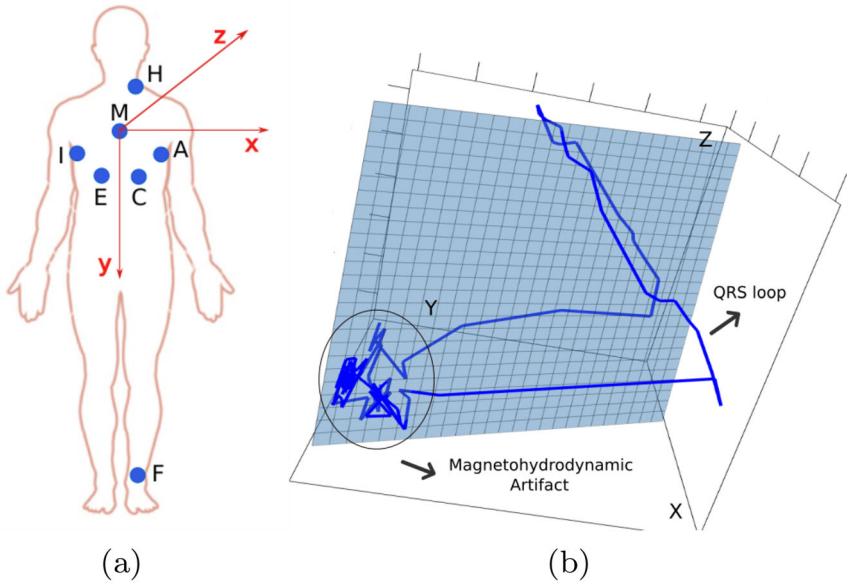
Fig. 2 **a** Location of the electrodes to obtain Frank's derivatives, in an orthogonal coordinate system (x, y, z). **b** QRS curve loop. The plane in blue is the best two-dimensional approximation of the curve by least squares

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 (Hatcher 2002)).

In (Downs and Liebman 1969), there is associated with each curve an element of $SO(3, 2)$ that represents some of the information of the curve, see also (Downs 1972), where they describe the phenomena as "... given the matrix $X = (X_1, X_2) \in SO(3, 2)$, the unitary vector $X_1 \in \mathbb{R}^3$ is the direction from the vertex to the apex of the so-called QRS loop, which is the most important loop among the three virtually planar loops contained in the vectorcardiogram (lies in the plane of the loop). And the vector unitary $X_2 \in \mathbb{R}^3$ is orthogonal to X_1 and to the plane of the loop, where the sense of X_2 is determined by the direction of motion of the moving point generating the loop".

This sample has been previously analysed in the literature, see, for instance, (Chikuse 2012; Chakraborty and Vemuri 2019; Pal et al. 2019).

Figure 3 represents each matrix in $SO(3, 2)$ by 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.

6.2 Influenza data set

A problem of great importance, and of current interest in the scientific community, is the early detection of new variants of a virus in a given community. One method to address this problem is to analyse the temporal dynamics of the variability of the virus

Fig. 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
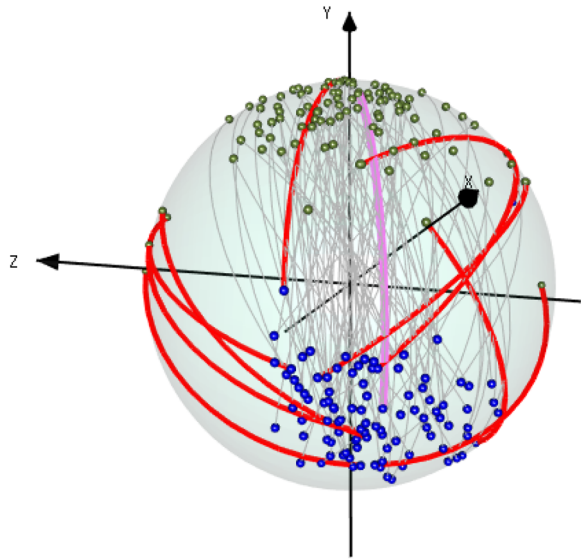
genetic chain. Roughly speaking, an increase in gene-chain diversity over a period of time may indicate diversification of the virus into new variants.

This data set consists of longitudinal data of the influenza virus, belonging to the family Orthomyxoviridae. An important problem analysed in (Cholaquidis et al. 2020) is to model the genomic evolution of the virus, see (Smith et al. 2004). 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 H3N2 variant of the virus, in particular, the subtype haemagglutinin (HA), which produced the SARS pandemic in 2002. This variant is known to have a variability in its genetic arrangement over time, see (Altman 2006; Monod et al. 2018).

The data set can be found in GI-SAID EpiFlu^{TM} database,[1] providing 1089 genomic sequences of H3N1 from 1993 to 2017 in New York, aligned using MUSCULE, see (Edgar 2004). We used reduced trees of 5 leaves, as in (Monod et al. 2018), to capture the structure of the data. This set of trees can be endowed with a distance, see (Billera et al. 2001). 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, (Barden et al. 2018; Brown and Owen 2020; Willis 2019) and references therein.
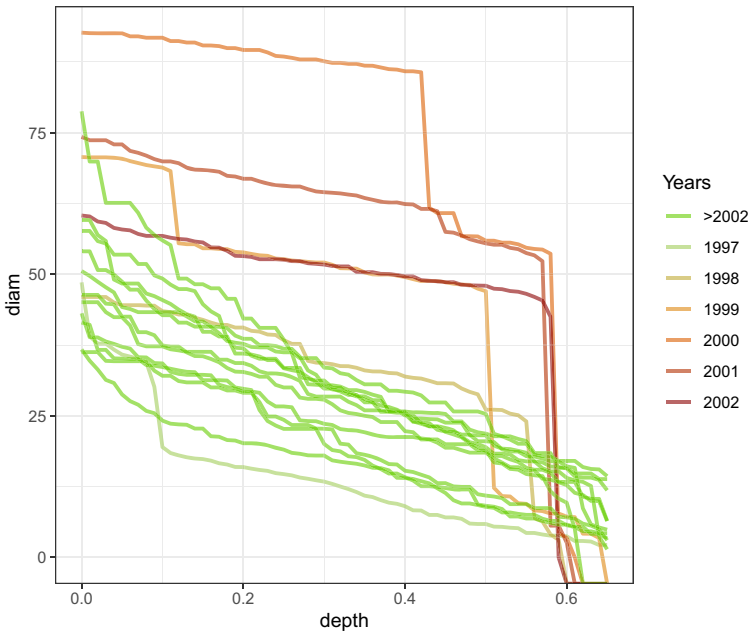
---

[1] www.gisaid.org.

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

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 Fig. 4. As can be seen, there is a larger dispersion in the years prior to the pandemic of 2002 compared to later years. Early detection of this increased variability may have an impact on health policy decision-making, with the aim of preventing the proliferation of the virus variants.

Appendix

Proof of Theorem 1 Let us prove that $\{LD \geq \lambda\} \subset \underset{n\to\infty}{\mathrm{Li}} \{\widehat{LD}_n \geq \lambda\}$. Take a such that $LD(a) = \lambda + \epsilon$ with $\epsilon > 0$. By the law of large numbers for U-statistics, with probability one, for all n large enough $|\widehat{LD}_n(a) - LD(a)| < \epsilon/2$. Then with probability one for n large enough $\widehat{LD}_n(a) > \lambda$, therefore $d(a, \{\widehat{LD}_n \geq \lambda\}) = 0$, which entails that $a \in \underset{n\to\infty}{\mathrm{Li}} \{\widehat{LD}_n \geq \lambda\}$. Let $\delta > 0$ arbitrarily small, if $LD(a) = \lambda$, there exists $u_k \to a$ such that $LD(u_k) > \lambda$ and $d(a, u_k) < \delta$. With probability one, for n large enough, $\widehat{LD}_n(u_k) > \lambda$ and then $d(a, \{\widehat{LD}_n \geq \lambda\}) \leq d(a, u_k) < \delta$ for

all n large enough, from where it follows that $a \in \underset{n\to\infty}{\text{Li}} \{\widehat{\text{LD}}_n \geq \lambda\}$. This proves that $\{\text{LD} \geq \lambda\} \subset \underset{n\to\infty}{\text{Li}} \{\widehat{\text{LD}}_n \geq \lambda\}$.

The second inclusion in Theorem 1 holds by definition; let us prove the third one. Let $a \in \underset{n\to\infty}{\text{Ls}} \{\widehat{\text{LD}}_n \geq \lambda\}$, then there exists $z_n \to a$ such that $\widehat{\text{LD}}_n(z_n) \geq \lambda$. Let $\mathcal{S} = \{\{z_n\} \cup a\}$, then from Theorem 3,

$$\lim_{n\to\infty} \sup_{y\in\mathcal{S}} |\widehat{\text{LD}}_n(y) - \text{LD}(y)| = 0 \quad a.s.$$

Next we bound,

$$|\text{LD}(a) - \widehat{\text{LD}}(z_n)| \leq |\text{LD}(a) - \text{LD}(z_n)| + |\text{LD}(z_n) - \widehat{\text{LD}}(z_n)|$$
$$\leq |\text{LD}(a) - \text{LD}(z_n)| + \sup_{y\in\mathcal{S}} |\widehat{\text{LD}}_n(y) - \text{LD}(y)|.$$

Lastly, $\text{LD}(a) \geq \lambda$ follows from the continuity of LD [see Proposition 3 in (Cholaquidis et al. 2020)]. □

Proof of Proposition 1 Assume by contradiction that $\mathcal{A}$ is uncountable. Then, there exists an uncountable set $I \subset \mathbb{R}$ such that for all $\lambda \in I$, there exists $y_\lambda \in M$ such that y_λ is a local maximum of f and $f(y_\lambda) = \lambda$. Since y_λ is a local maximum, there exists $q = q(y_\lambda) \in \mathbb{Q}$ such that $f(x) \leq f(y_\lambda)$ for all $x \in B(y_\lambda, q)$. Since $\{y_\lambda\}_{\lambda\in I}$ is uncountable, there exists at least one $q^* \in \mathbb{Q}$ and an uncountable subset $J \subset I$, such that $f(x) \leq f(y_\lambda)$ for all $x \in B(y_\lambda, q^*)$ and for all $\lambda \in J$. Let us consider $\{y_\lambda\}_{\lambda\in J}$ the local maximum for which this q^* works. Since M is separable, there exists $y^* \in M$ and a subsequence of distinct y_{λ_n}, associated with distinct λ_n, such that $y_{\lambda_n} \to y^*$. Let n_0 such that for all $n, n' > n_0 \, d(y_{\lambda_{n'}}, y_{\lambda_n}) < q/4$. Then, $y_{\lambda_{n'}} \in B(y_{\lambda_n}, q)$, from where it follows t $\lambda_{n'} \leq \lambda_n$. Also, $y_{\lambda_n} \in B(y_{\lambda_{n'}}, q)$, and then, $\lambda_n \leq \lambda_{n'}$. Lastly, $\lambda_n = \lambda_{n'}$, which contradict that all the λ_n are distinct. □

To prove that (4) is fulfilled, we will use the following lemma.

Lemma 1 *Let* $\delta > 0$, $x \in M$ *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* $t_1, t_2 \in M$.

Proof 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

$$\{x \in A(y_1, y_2) : d(x, \partial A(y_1, y_2)) > 2\delta\} \subset A(z_1, z_2) \tag{5}$$

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 $d(x, y_2) \le d(z_1, z_2)$, then $x \in A(z_1, z_2)$, which implies (5). Next, let us prove that

$$A(z_1, z_2) \subset \{x : d(x, A(y_1, y_2)) \le 3\delta\} \tag{6}$$

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

$$|\mathbb{I}_{A(z_1,z_2)}(x) - \mathbb{I}_{A(t_1,t_2)}(x)| = 0. \tag{7}$$

Observe that if $x \in A(y_1, y_2)$ and $d(x, \partial A(y_1, y_2)) > 3\delta$, then by (5), $x \in A(z_1, z_2)$, so $\mathbb{I}_{A(z_1,z_2)}(x) = 0$ and $x \in A(t_1, t_2)$. Hence, $\mathbb{I}_{A(t_1,t_2)}(x) = 0$ which then shows that (7) holds. Proceeding in the same way if $x \notin A(y_1, y_2)$ and $d(x, \partial A) > 3\delta$, by (6) $x \notin A(z_1, z_2)$, 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 (7) holds. $\qquad\square$

Proof of Theorem 3 We will apply Billingsley's theorem to the set of functions

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

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 $i \ne j$ 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,

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

Clearly, $\sup\{|f(z) - f(t)| : f \in \mathcal{F}, z, t \in M \times M\} = 2$. So we have to prove that (4) holds. By Lemma 1, to prove (4), it is enough to prove that

$$\lim_{\delta \to 0} \sup_x P_X \times P_X\{(y_1, y_2) : d(x, \partial A(y_1, y_2)) \le 3\delta\} = 0. \tag{8}$$

Let $0 < \epsilon < 1$, and let K be a compact set such that $P_X(K^c) < \epsilon$

$$\sup_x P_X \times P_X\{(y_1, y_2) : d(x, \partial A(y_1, y_2)) \le 3\delta\}$$

$$\le \sup_x P_X \times P_X\big\{(y_1, y_2) : y_1, y_2 \in K, d(x, \partial A(y_1, y_2)) \le 3\delta\big\} + 2\epsilon$$

$$\sup_x P_X \times P_X\{(y_1, y_2) : y_1, y_2 \in K, d(x, \partial A(y_1, y_2)) \le 3\delta\}$$

$$\int_K P_X(y_2 : d(y_2, \partial B(y_1, d(y_1, y_2))) \le 3\delta) P_X(dy_1)$$

$$\int_K P_X(y_2 : d(y_2, \partial B(y_1, d(y_1, y_2))) \le 3\delta) P_X(dy_1)$$

$$\leq \sup_{y_1 \in K} P_X(y_2 : d(y_2, \partial B(y_1, d(y_1, y_2))) \leq 3\delta)$$

Since $P_X(\partial B(x, \delta)) = 0$ for all $\delta > 0$ for a fixed y_1,

$$P_X(y_2 : d(y_2, \partial B(y_1, d(y_1, y_2))) \leq \delta) \to 0 \text{ as } \delta \to 0 \tag{9}$$

By the dominated convergence theorem, it follows that for fixed δ, $P_X(y_2 : d(y_2, \partial B(y_1, d(y_1, y_2))) \leq \delta)$ is a continuous function of y_1. Then, the supremum $\sup_{y_1 \in K} P_X(y_2 : d(y_2, \partial B(y_1, d(y_1, y_2))) \leq 3\delta)$ it attained, at some $y^* \in K$ for which $P_X(y_2 : d(y_2, \partial B(y^*, d(y^*, y_2))) \leq 3\delta) \to 0$ as $\delta \to 0$. Then, for all $0 < \epsilon < 1$

$$\limsup_{\delta \to 0} P_X \times P_X\{(y_1, y_2) : d(x, \partial A(y_1, y_2)) \leq 3\delta\} < 2\epsilon,$$

which implies (8), and concludes the proof of Theorem 3 □

Proof of Theorem 6 Let K be any compact set. Observe that $\{LD \geq \lambda^-\} \cap K$ is non-empty and compact for any $0 < \lambda^- < \lambda$. Following the same ideas as those used in the proof of Theorem 1 in (Cuevas et al. 2006), one can derive that for all $\epsilon > 0$

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

The proof of

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

is slightly different from the analogous inclusion in (Cuevas et al. 2006). If we proceed by contradiction, there exists an $x_n \in \{\widehat{LD}_n \geq \lambda\} \cap K$, but

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

Observe that if $x_n \in \partial\{\widehat{LD}_n \geq \lambda\} \cap K$, 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 $d(y_n, x_n) < \epsilon/2^n$, $y_n \in \partial\{\widehat{LD}_n(x) \geq \lambda\}$, y_n in the boundary of only one $\partial A(X_i, X_j)$ and $|\widehat{LD}_n(y_n) - \lambda| < \binom{n}{2}^{-1}$. 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 also have $|\widehat{LD}_n(y_n) - \lambda| < \binom{n}{2}^{-1}$

Since K is compact, there exists an $x \in K$ such that $x_n, y_n \to x$ (by considering a subsequence if necessary). Then, $|LD(x) - \lambda| \leq |LD(x) - LD(y_n)| + |LD(y_n) - \widehat{LD}(y_n)| + |\widehat{LD}(y_n) - \lambda|$. The first terms converge to 0 by the continuity of LD at x. The second term converges to 0 a.s., because the set $K' = \{\bigcup_n\{y_n\}\} \cup \{x\}$ is compact and by Theorem 3, $\sup_{z \in K'} |\widehat{LD}_n(z) - LD(z)| \to 0$. The last term is bounded from above by $\binom{n}{2}^{-1} \to 0$ as $n \to \infty$. Then, $LD(x) = \lambda$, which implies that $d(x_n, \partial\{LD \geq \lambda\} \cap K) \leq d(x_n, x) \to 0$, which contradicts (10). □

References

Agostinelli C (2018) Local half-region depth for functional data. J Multivar Anal 163:67–79

Altman LK (2006) This season's flu virus is resistant to 2 standard drugs. New York Times

Barden D, Le H, Owen M (2018) Limiting behaviour of fréchet means in the space of phylogenetic trees. Ann Instit Stat Math 70(1):99–129

Barnett V (1976) The ordering of multivariate data. J R Stat Soc Ser A 139(3):318–355

Beer G (1993) Topologies on closed and closed convex sets. Math Appl 268

Billera LJ, Holmes SP, Vogtmann K (2001) Geometry of the space of phylogenetic trees. Adv Appl Math 27(4):733–767

Billingsley P, Topsøe F (1967) Uniformity in weak convergence. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 7(1):1–16

Brown DG, Owen M (2020) Mean and variance of phylogenetic trees. Syst Biol 69(1):139–154

Brunel V-E (2019) Concentration of the empirical level sets of Tukey's halfspace depth. Prob Theory Relat Fields 173(3):1165–1196

Chakraborty R, Vemuri B (2019) Statistics on the stiefel manifold: theory and applications. Ann Stat 47(1):415–438

Chikuse Y (2012) Statistics on special manifolds, vol 174. Springer Science & Business Media

Cholaquidis A, Fraiman R, Moreno L (2020) Weighted lens depth: some applications to supervised classification. arXiv preprint arXiv:2011.11140

Claeskens G, Hubert M, Slaets L, Vakili K (2014) Multivariate functional halfspace depth. J Am Stat Assoc 109(505):411–423

Cuesta-Albertos JA, Febrero-Bande M, de la Fuente MO (2017) The dd^g-classifier in the functional setting. Test 26(1):119–142

Cuevas A, Fraiman R (2009) On depth measures and dual statistics. A methodology for dealing with general data. J Multivar Anal 100(4):753–766

Cuevas A, González-Manteiga W, Rodríguez-Casal A (2006) Plug-in estimation of general level sets. Austr N Z J Stat 48(1):7–19

Dai W, Genton MG (2019) Directional outlyingness for multivariate functional data. Comput Stat Data Anal 131:50–65

Downs TD (1972) Orientation statistics. Biometrika 59(3):665–676

Downs TD, Liebman J (1969) Statistical methods for vectorcardiographic directions. IEEE Tran Biomed Eng 1:87–94

Dyckerhoff R (2016) Convergence of depths and depth-trimmed regions. arXiv preprint arXiv:1611.08721

Edgar RC (2004) Muscle: multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res 32(5):1792–1797

Elmore RT, Hettmansperger TP, Xuan F (2006) Spherical data depth and a multivariate median. DIMACS Ser Discr Mathem Theor Comput Sci 72:87

Febrero M, Galeano P, González-Manteiga W (2008) Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. Environ Met Off J Int Environ Soc 19(4):331–345

Fraiman R, Muniz G (2001) Trimmed means for functional data. Test 10(2):419–440

Fraiman R, Gamboa F, Moreno L (2019) Connecting pairwise geodesic spheres by depth: Dcops. J Multivar Anal 169:81–94

Frank E (1956) An accurate, clinically practical system for spatial vectorcardiography. Circulation 13(5):737–749

Hafshejani NJ, Mehridehnavi A, Hajian R, Boudagh S, Behjati M (2021) Identification of myocardial infarction using morphological features of electrocardiogram and vectorcardiogram. IET Signal Process 15(9):674–685

Hatcher A (2002) Algebr Topol. Cambridge University Press

Hubert M, Rousseeuw P, Segaert P (2017) Multivariate and functional classification using depth and distance. Adv Data Anal Classif 11(3):445–466

Koshevoy G, Mosler K et al (1997) Zonoid trimming for multivariate distributions. Ann Stat 25(5):1998–2017

Laketa P, Nagy S (2021) Halfspace depth for general measures: the ray basis theorem and its consequences. Statistical Papers

Liu RY (1992) Data depth and multivariate rank tests. L1-statistics and related methods. In: Dodge Y, ed., pp 279–294. North-Holland, Amsterdam

Liu RY (1990) On a notion of data depth based on random simplices. Ann Stat 18(1):405–414

Liu Z, Modarres R (2011) Lens data depth and median. J Nonparam Stat 23(4):1063–1074

Liu RY, Parelius JM, Singh K et al (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by liu and singh). Ann Stat 27(3):783–858

López-Pintado S, Romo J (2009) On the concept of depth for functional data. J Am Stat Assoc 104(486):718–734

Molchanov IS (1998) A limit theorem for solutions of inequalities. Scand J Stat 25(1):235–242

Monod A, Lin B, Yoshida R, Kang Q (2018) Tropical geometry of phylogenetic tree space: a statistical perspective. arXiv preprint arXiv:1805.12400

Mosler K, Mozharovskyi P (2017) Fast dd-classification of functional data. Stat Papers 58(4):1055–1089

Oja H (1983) Descriptive statistics for multivariate distributions. Stat Probab Lett 1(6):327–332

Pal S, Sengupta S, Mitra R, Banerjee A (2019) A bayesian approach for analyzing data on the stiefel manifold. arXiv preprint arXiv:1907.04303

Pastore CA, Samesima N, Pereira HG, Tobias NMMDO, Madaloso BA, Facin ME (2019) Applicability of the electro-vectorcardiogram in current clinical practice. Arquivos brasileiros de cardiologia 113:87–99

Ruts I, Rousseeuw PJ (1996) Computing depth contours of bivariate point clouds. Comput Stat Data Anal 23(1):153–168

Serfling R (2002) Quantile functions for multivariate analysis: approaches and applications. Stati Neerlandica 56(2):214–232

Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA (2004) Mapping the antigenic and genetic evolution of influenza virus. Science 305(5682):371–376

Terán P (2016) A multivalued strong law of large numbers. J Theor Probab 29(2):349–358

Tukey JW (1975) Mathematics and the picturing of data. Proc Int Congr Math 2:523–531

Vencálek O (2017) Depth-based classification for multivariate data. Austr J Stat 46(3–4):117–128

Willis A (2019) Confidence sets for phylogenetic trees. J Am Stat Assoc 114(525):235–244

Zuo Y, Serfling R (2000) General notions of statistical depth function. Ann Stat 28(2):461–482