Journal of Multivariate Analysis 157 (2017) 53–69

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A weighted localization of halfspace depth and its properties

Lukáš Kotík [a], Daniel Hlubinka [b,*]

[a] National Radiation Protection Institute, Bartoškova 28, 140 00 Praha 4, Czech Republic

[b] Charles University, Faculty of Mathematics and Physics, Department of Probability and Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic

ARTICLE INFO

Article history:
Received 13 February 2015
Available online 4 March 2017

AMS subject classifications:
62E10
62H05
62G15
60E05

Keywords:
Data depth
Local depth
Nonconvex depth contours
Strong uniform consistency

ABSTRACT

Statistical depth functions are well-known nonparametric tools for analysing multivariate data. Halfspace depth is most frequently used, and while it has many desirable properties, it is dependent on global characteristics of the underlying distribution. In some circumstances, however, it may be desirable to take into account more local and intrinsic characteristics of the data. To this end, we introduce weighted halfspace depths in which the indicator function of closed halfspace is replaced by a more general weight function. Our approach, which calls in part on functions associated with conic sections, encompasses as special cases the notions of sample halfspace depth and kernel density estimation. We give several illustrations and prove the strong uniform consistency of weighted halfspace depth incorporating mild conditions on the weight function.

1. Introduction

Data depth, or simply depth, is a nonparametric tool for the analysis of multivariate data. Over the last two decades, depth has been intensively studied and its usefulness has been illustrated in many different contexts. For an overview of data depth and its desirable properties, see, e.g., [37,41]. A review of depth functions, their properties, uses, and computational aspects thereof can be found in [25]. For applications of depth in statistical analysis, see, e.g., [7,8,13,15,18,19,21,22,27,29,40].

One of the most commonly used notions of data depth is Tukey's halfspace depth [39]. This concept and its properties are thoroughly studied in [37]. In particular, it was shown in [20] that the halfspace depth may, under certain conditions, characterize the underlying distribution. Within the class of unimodal elliptically symmetric distributions, for example, there is a one-to-one relationship between the depth of a point and the probability density at that point. In contrast, it was proved in [12] that if the underlying distribution is not ℓ_2-symmetric, a discrepancy exists between the halfspace depth contours and the density contours.

This paper is concerned with situations where the halfspace depth and the probability density disagree. This situation is fairly common and occurs, e.g., when the underlying probability distribution has non-convex or disconnected level sets. Such circumstances typically have adverse effects on the properties of the halfspace depth (notably on its affine invariance and quasi-concavity) and its use, e.g., in classification problems, outlier detection, and nonparametric testing. They call for modifications to depth-based procedures or to the notion of depth itself, which needs to be "localized" in order to maintain the good performance of the standard statistical techniques; see, e.g., [4,11,23,31,32].

* Corresponding author.
E-mail addresses: lukaskotik@gmail.com (L. Kotík), hlubinka@karlin.mff.cuni.cz (D. Hlubinka).

http://dx.doi.org/10.1016/j.jmva.2017.02.008

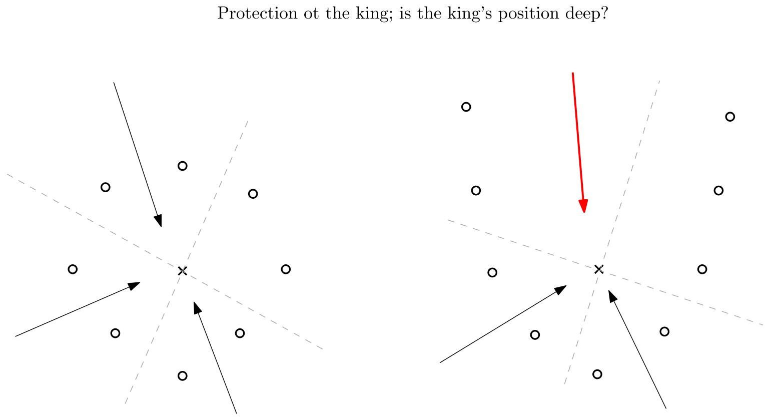

Fig. 1. Both crosses have the same halfspace depth. However, the right cross is not evenly surrounded by circles, hence its defense against an "invasion" from the north is smaller than from other directions. The left cross is well protected from all directions and so its position can be considered deeper.

This problem motivates us to propose here the notion of weighted halfspace depth. It will be shown that this concept leads to a characterization of each point in terms of a global component (its data depth) and a local one (its probability density function). Localization of different depth functions was previously considered by Agostinelli and Romanazzi [1], mostly in terms of simplicial depth. These authors' localization of halfspace depth is a special variant of the weighted halfspace depth discussed herein. Another approach to depth localization with applications to classification and to testing symmetry is considered in [11,31,32].

An analogy can be used to give a heuristic interpretation for our modified notion of halfspace depth. Consider a king and guards around him. When the king (represented by a cross) is surrounded more or less evenly by his guards (represented by circles), as in Fig. 1(a), he is uniformly protected against an attack from every direction. The more central the king lies within this configuration of guards, the better protected he is; his position is deep with respect to the guards.

In this analogy, measuring the halfspace depth amounts to counting the guards in each halfspace evenly around the king, regardless of how far they are, "left or right". However, if the guards are positioned unevenly around the king, there are directions in which potential enemies stand a higher chance of reaching the king, as depicted in Fig. 1(b). In this case there is a neighbourhood of the half-line in which there are fewer guards, even though the king still lies deeply in his lands, as measured by the classical halfspace depth.

This weakness can be assessed by evaluating the position of each guard relative to the king and the direction of attack, thereby allowing us to assign a weight to each guard's contribution to the overall defense of the king. In so doing, we gain a better understanding of the king's vulnerability. By analogy, the use of the same concept in a data model will help reduce vulnerability of the analysis to misinterpretation.

In Section 2, we define the weighted halfspace depth and we introduce some elementary weight functions. Basic properties of weighted halfspace depth are given in Section 3. In Section 4, we study the balance between local and global characteristics through the class of conic section weight functions. Finally, conditions guaranteeing the uniform strong consistency of weighted halfspace depth are given in Section 5. The proofs are relegated to Appendix A, and additional discussion concerning weight functions can be found in Appendix B. Finally, computational aspects of the procedure are briefly discussed in Appendix C.

In what follows we assume a fixed probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and random vectors defined on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}^p$, $p \geq 2$. The distribution (i.e., the induced probability measure on $\mathbb{R}^p$) of a random vector X is denoted either P or P_X and the probability measure on $(\Omega, \mathcal{A})$ is denoted $\mathbb{P}$. In this paper we consider only absolutely continuous random vectors. Furthermore, the following notational conventions are used. A random sample of size n from P is denoted $X_1, \ldots, X_n$. A subscript n is added to denote sample analogues of all mathematical operators, e.g., the empirical probability measure is denoted by P_n and its corresponding expected value by E_n. The inner product between two vectors u and v is denoted $\langle u, v\rangle$ and $\|u\|$ denotes the norm of u. We use the Euclidean scalar product and Euclidean norm throughout the paper but most definitions and theorems could be generalized in a straightforward way to any Hilbert space. The indicator function of a set A is denoted $\mathbf{1}(A)$ and A^c stands for the complement of A, i.e., for any set $A \in \mathbb{R}^p$ is $A^c = \mathbb{R}^p \setminus A$. By an orthogonal matrix $\mathbf{A}$ we mean a matrix such that $\mathbf{A}\mathbf{A}^\top = \mathbf{A}^\top\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ denotes an identity matrix. Finally, the unit sphere in $\mathbb{R}^p$ is denoted $\mathbb{S}^{p-1}$, i.e., $\mathbb{S}^{p-1} = \{u \in \mathbb{R}^p : \|u\| = 1\}$.

2. Weighted halfspace depth

First recall the definition of the well-known notion of halfspace depth. Let X be a random variable with probability distribution P on $\mathbb{R}^p$. The halfspace depth of a point $x \in \mathbb{R}^p$ with respect to P is defined as

$$HD(x; P) = \inf_{\|u\|=1} \mathbb{P}(\langle X - x, u\rangle \geq 0) = \inf_{\|u\|=1} P\{y : \langle y - x, u\rangle \geq 0\}.$$