WEIGHTED HALFSPACE DEPTH

Daniel Hlubinka, Lukáš Kotík and Ondřej Vencálek

Generalised halfspace depth function is proposed. Basic properties of this depth function including the strong consistency are studied. We show, on several examples that our depth function may be considered to be more appropriate for nonsymetric distributions or for mixtures of distributions.

1. INTRODUCTION

Multivariate data became quite common in statistics and its applications in last fifty years. The classical statistical theory for multivariate random vectors is based on the assumption of normality (or a mixture of normal distributions) for which the inference is well developed. This requirement is, however, too strong; the nonparametric approach is desirable and has been studied intensively in last thirty years.

The median or generally the quantiles are very popular in the statistical inference and data analysis of univariate random variables. For the multivariate random vectors, however, the lack of natural ranking means that there is no direct generalisation of the univariate median to the vector case – unlike for the mean. The problem of multivariate quantiles became still very popular in mathematical statistics and recently there have appeared many competitive approaches to that problem. One of the most popular is the *data depth*. Data depth is a tool for ordering the data according to some measure of "centrality", called *the depth*. There are many well known depth functions which have been intensively studied in last two decades. The classical data depth is the *halfspace depth* introduced in [8]. Other popular depth is the *simplicial depth*, [3]. Zuo and Serfling [10] and Mizera [6] study depth from a general point of view. The recent development in the data depth may be found in [4].

In this paper we discuss an alternative definition of the *data depth*; in particular, we generalise the concept of *halfspace depth*. The halfspace depth and its fundamental properties are briefly recalled in Section 2. Some features of halfspace depth may be considered undesirable in different situations, as we illustrate throughout the paper. Therefore, the generalisation of the depth function is introduced in Section 3 and some basic properties of this data depth are discussed in Section 4. In

Section 5, we prove the strong consistency of our depth function. In Section 6, three illustrative examples with discussion are provided.

2. HALFSPACE DEPTH

One of the most popular depth functions is the *halfspace depth*, defined by Tukey [8]. Donoho and Gasko [2] have studied its breakdown properties. The computational aspect may be found, e.g., in [7], Matoušek [5] has proposed fast algorithm for computing the deepest point (point with maximal halfspace depth) of random sample. See also [11] for broad discussion on features of data depth and, in particular, of the halfspace depth.

Definition 2.1. Let P be a probability measure on $\mathbb{R}^p$. The halfspace depth of a point $\boldsymbol{x}$ is defined as

$$\mathrm{HD}(\boldsymbol{x}) = \inf_{\boldsymbol{u}, \|\boldsymbol{u}\|=1} \mathrm{P}\big(\{\boldsymbol{y} : \boldsymbol{u}^T(\boldsymbol{y} - \boldsymbol{x}) \geq 0\}\big).$$

In other words, the halfspace depth of $\boldsymbol{x}$ is the infimum of probability of all closed halfspaces whose border includes $\boldsymbol{x}$. The halfspace depth is well defined for all $\boldsymbol{x} \in \mathbb{R}^p$. The empirical (sample version) halfspace depth $\mathrm{HD}_n(\boldsymbol{x})$ defined on a random sample $X_1, \ldots, X_n$ of the distribution P is defined as a halfspace depth for the empirical probability measure P_n.

This definition is very intuitive and easily interpretable. Moreover, there are many nice properties of the halfspace depth which made this depth popular and widely used. Let us recall some of them.

1. Depth is *affine invariant* function.

2. In many situations (e.g. for absolutely continuous distributions) there is a unique point with highest depth, the *deepest point*.

3. Considering a ray starting at the deepest point, the depth of points along the ray is *nonincreasing as the distance from the deepest point increases.*

4. The depth function HD is vanishing at infinity.

5. The set $\{\boldsymbol{x} : \mathrm{HD}(\boldsymbol{x}) \geq d\}$ of points whose depth is higher than a given value d is *convex* for any d (convexity of central regions, quasi-concavity of depth function).

6. The empirical halfspace depth $\mathrm{HD}_n(\boldsymbol{x})$ *converges almost surely* to $\mathrm{HD}(\boldsymbol{x})$ as $n \to \infty$ for all $\boldsymbol{x} \in \mathbb{R}^p$ (strong consistency).

The properties 1 – 4 are called key properties in [10]. Zuo and Serfling [10] also consider broad classes of depth function and study the possession of these key properties.

Some of these properties may be not desirable for general distributions. For example, when the underlying distribution is not symmetric, then a natural unique candidate for the deepest point needs not to exist. If the level sets of density function

f (i. e., sets defined as $\{x : f(x) \geq a\}$ for all $a > 0$) does not form convex or star-shaped sets, then the properties 3 and 5 may be superfluous.

The convexity of central regions may be considered as a disadvantage of halfspace depth (and of other depth functions) when it is applied to considerably non-convex datasets. Therefore constructions of more general central regions were proposed. Beside the well known level sets of the probability density function, DasGupta et al. [1] considered a general family of star-shaped sets. "Best" shape of central regions is proposed and it is then inflated (deflated) in order to obtain the central region of given probability. The idea behind this approach is substantially different from the halfspace or simplicial depths.

But even for absolutely continuous distributions with convex support, like the bivariate exponential distribution or the bivariate $[0, 1]^2$ uniform distribution, some disadvantages of the halfspace depth may be disclosed; see Section 6. It is the main motivation for us to propose a larger class of depth functions derived from the halfspace depth.

3. WEIGHTED HALFSPACE DEPTH

In this section we propose depth function derived from the halfspace depth function which, in contrary to the halfspace depth, allows the central regions to be more general than convex. The main idea is to use *weights* (weighted probability) in the halfspace rather than the probability of halfspace.

More precisely, let us denote by $\boldsymbol{x}$ the point for which the depth is computed and by $\mathbb{H} \subset \mathbb{R}^p$ the halfspace of interest. Each point $\boldsymbol{y} \in \mathbb{H}$ is assigned a weight $w(\boldsymbol{y})$ which depends on a position of $\boldsymbol{y}$ with respect to $\boldsymbol{x}$ and then the weighted probability $p_{\mathbb{H}} = \int_{\mathbb{H}} w(\boldsymbol{Y}) \, \mathrm{dP}$ of the halfspace $\mathbb{H}$ is computed. The same weights are used to the opposite halfspace $\mathbb{R}^p \setminus \mathbb{H}$ and $p_{\mathbb{R}^p \setminus \mathbb{H}}$ is calculated. The ratio of these two values is used for definition of the weighted depth (in contrary to the halfspace depth where the opposite halfspace need not to be considered). Let us formulate the formal definitions.

Notation of weight functions. In what follows we denote by $w_+ : \mathbb{R}^p \to [0, \infty)$ any measurable weight function which is bounded and such that

$$w_+(\boldsymbol{x}) = w_+(x_1, \ldots, x_p) = 0 \ \ \text{if} \ \ x_p < 0,$$

and denote its "counterweight function" as

$$w_-(\boldsymbol{x}) = w_-(x_1, \ldots, x_p) = w_+(x_1, \ldots, -x_p).$$

Definition 3.1. (Depth function) Let $\boldsymbol{X}$ be a random vector and P its probability distribution. The (population) weighted depth of a point $\boldsymbol{x}$ is defined as

$$\mathrm{D}(\boldsymbol{x}) := \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{\mathrm{E}_{\mathrm{P}} \, w_+\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big)}{\mathrm{E}_{\mathrm{P}} \, w_-\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big)}, \tag{1}$$

where w_+ is the weight function, $\mathbb{O}_p$ denotes the space of all orthogonal $p \times p$ matrices, and the term $0/0$ is defined to be 1.

Notation remark: Sometimes it is useful to emphasize the underlying distribution or the random vector in the depth function. We adopt the notation $D_P(\boldsymbol{x}) = D_{\boldsymbol{X}}(\boldsymbol{x}) = D(\boldsymbol{x})$ where P is the underlying probability measure, and $\boldsymbol{X}$ is a random vector with distribution P.

Remark 3.2. In Definition 3.1 the orthogonal transformations are used to allow full generality of the weight function. For smaller class of symmetric weight functions, i. e., if

$$w_+(x_1, \ldots, x_k, \ldots, x_p) = w_+(x_1, \ldots, -x_k, \ldots, x_p), \quad k = 1, \ldots, p-1$$

holds, it is possible to consider only rotations instead of all orthogonal transformations. In particular, the role of the orthogonal transformation is the same as the role of rotations (directions $\boldsymbol{u}$) of the halfspace in Definition 2.1. In other words, instead of rotating the weight function w_+ the random vector $\boldsymbol{X}$ is orthogonally transformed ("rotated to a direction").

Theorem 3.3. For any p-dimensional random vector $\boldsymbol{X}$ and any $\boldsymbol{x} \in \mathbb{R}^p$ it holds $D_{\boldsymbol{X}}(\boldsymbol{x}) \leq 1$.

P r o o f. It is not difficult to see that $w_-(\boldsymbol{X}) = w_+(\mathbf{I}_-\boldsymbol{X})$ and $w_+(\boldsymbol{X}) = w_-(\mathbf{I}_-\boldsymbol{X})$, where $\mathbf{I}_- = \mathrm{diag}_p(1, 1, \ldots, -1)$ is a $p \times p$ diagonal orthogonal matrix. Since $\{\mathbf{I}_-\mathbf{A} : \mathbf{A} \in \mathbb{O}_p\} = \mathbb{O}_p$ it follows

$$D(\boldsymbol{x}) = \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{Ew_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}))}{Ew_-(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}))} = \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{Ew_-(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}))}{Ew_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}))} \tag{2}$$

and since clearly

$$\min\left\{ \frac{E\,w_+(\boldsymbol{Y})}{E\,w_-(\boldsymbol{Y})}, \frac{E\,w_-(\boldsymbol{Y})}{E\,w_+(\boldsymbol{Y})} \right\} \leq 1$$

the proof is completed. $\qquad\qquad\square$

The connection between the depth function of Definition 3.1 and the halfspace depth function need not be clear at this moment. In the following discussion it is shown that the depth function D is *essentially* a generalisation of the halfspace depth.

Definition 3.4. (Depth function II) Define a depth function

$$\widetilde{D}(\boldsymbol{x}) := \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{Ew_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}))}{Ew_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})) + Ew_-(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}))}, \tag{3}$$

for a weight function w_+; the ratio $0/(0+0)$ is now defined as $1/2$.

The depth functions D and $\widetilde{D}$ are equivalent in the sense of the multivariate ordering:

Theorem 3.5. For any weight function w_+ and for all $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^p$ the equivalence

$$\mathrm{D}(\boldsymbol{x}_1) \le \mathrm{D}(\boldsymbol{x}_2) \iff \widetilde{\mathrm{D}}(\boldsymbol{x}_1) \le \widetilde{\mathrm{D}}(\boldsymbol{x}_2) \tag{4}$$

holds. Moreover,

$$\widetilde{\mathrm{D}}(\boldsymbol{x}) \le \frac{1}{2}, \tag{5}$$

and

$$\mathrm{D}(\boldsymbol{x}) = \frac{\widetilde{\mathrm{D}}(\boldsymbol{x})}{1 - \widetilde{\mathrm{D}}(\boldsymbol{x})}. \tag{6}$$

P r o o f. Following similar argument as in proof of Theorem 3.3 it holds

$$\widetilde{\mathrm{D}}(\boldsymbol{x}) = \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{\mathrm{E}_\mathrm{P}\, w_+\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big)}{\mathrm{E}_\mathrm{P}\, w_+\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big) + \mathrm{E}_\mathrm{P}\, w_-\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big)}$$

$$= \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{\mathrm{E}_\mathrm{P}\, w_-\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big)}{\mathrm{E}_\mathrm{P}\, w_+\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big) + \mathrm{E}_\mathrm{P}\, w_-\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big)}.$$

The inequality (5) follows from the obvious fact that

$$\min\left\{ \frac{\mathrm{E}\, w_+(\boldsymbol{Y})}{\mathrm{E}\, w_+(\boldsymbol{Y}) + \mathrm{E}\, w_-(\boldsymbol{Y})}, \frac{\mathrm{E}\, w_-(\boldsymbol{Y})}{\mathrm{E}\, w_+(\boldsymbol{Y}) + \mathrm{E}\, w_-(\boldsymbol{Y})} \right\} \le 1/2.$$

Denote for fixed orthogonal matrix $\mathbf{A}$

$$v_+ = \mathrm{E} w_+\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big) \text{ and } v_- = \mathrm{E} w_-\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big).$$

If $v_- > 0$ then

$$\frac{v_+}{v_-} = \frac{v_+}{v_- + v_+} \left(\frac{v_-}{v_- + v_+} \right)^{-1} = \frac{v_+}{v_- + v_+} \left(1 - \frac{v_+}{v_- + v_+} \right)^{-1}. \tag{7}$$

If $v_- = 0$ and $v_+ > 0$ then v_- and v_+ in (7) may be interchanged (see arguments for (5) and (2)).

If both $v_- = v_+ = 0$ then the 0/0 ratios are defined as

$$\frac{v_+}{v_-} = 1, \quad \frac{v_+}{v_- + v_+} = \frac{1}{2} \Rightarrow \frac{v_+}{v_-} = \frac{v_+}{v_- + v_+} \left(1 - \frac{v_+}{v_- + v_+} \right)^{-1}.$$

Equation(6) now follows.

Since the function $x \mapsto x/(1-x)$ is increasing in x for $x \in [0, 1/2]$, the equivalence (4) follows. $\qquad\square$

Remark 3.6. The previous theorem shows that our definition is in some sense a direct generalisation of the halfspace depth if the underlying distribution is absolutely continuous. Indeed, the halfspace depth $\mathrm{HD}(\boldsymbol{x})$ is equal to $\widetilde{\mathrm{D}}(\boldsymbol{x})$ for $w_+(\boldsymbol{y}) \equiv 1$ (the denominator is 1 for any absolutely continuous distribution).

In the case of non-continuous distribution, it holds $\mathrm{HD}(\boldsymbol{x}) \geq \widetilde{\mathrm{D}}(\boldsymbol{x})$ for all $\boldsymbol{x}$ and the inequality may be strict at some points. Indeed, consider $p \in (0, 1)$ and a bivariate distribution given by $(1 - p)\mathrm{Unif}_{[0,1]^2} + p\delta_{(1,1)}$, i.e., the mixture of the uniform distribution on $[0, 1]^2$ and a point mass at $(1, 1)$. Then, obviously, $\mathrm{HD}(1, 1) = p > p/(1 + p) = \widetilde{\mathrm{D}}(1, 1)$.

Obviously, the empirical measure P_n is used for the definition of the sample weighted depth. In what follows we shall call $\mathrm{D}(\boldsymbol{x})$ simply *the depth of $\boldsymbol{x}$* unless we need to distinguish more depth functions.

Remark 3.7. Usual choice of weight function is *spherically symmetric* about x_p-axis. It means that there exists function $h : [0, +\infty) \times \mathbb{R} \to \mathbb{R}$ such that

$$w_+(x_1, \ldots, x_p) = h(x_1^2 + \ldots + x_{p-1}^2, x_p).$$

Namely, it holds $w_+(\boldsymbol{x}) = w_+(x_1, \ldots, x_{p-1}, x_p) = w_+(-x_1, \ldots, -x_{p-1}, x_p) = w_-(-\boldsymbol{x})$ in this case.

Example 3.8. The *cylinder weight function* is for a chosen $h > 0$ defined as

$$w_+(x_1, \ldots, x_p) = \begin{cases} 1 & \text{if } \sum_{i=1}^{p-1} x_i^2 < h^2, \; x_p > 0, \\ 0 & \text{elsewhere.} \end{cases} \tag{8}$$

In particular, for $\mathbb{R}^2$ the definition has the following meaning. Given fixed point $\boldsymbol{x}$ and a direction $\boldsymbol{s}$ (unit vector in $\mathbb{R}^2$), we consider a line $l = \boldsymbol{x} + t\boldsymbol{s}$, $t \in \mathbb{R}$ for which a band with width $2h$

$$\mathbb{B}(\boldsymbol{x}, \boldsymbol{s}) = \{\boldsymbol{y} \in \mathbb{R}^2 : \; d(\boldsymbol{y}, l) < h\}$$

is defined (d denotes the Euclidean distance). The band $\mathbb{B}(\boldsymbol{x}, \boldsymbol{s})$ is divided by a segment orthogonal to $\boldsymbol{s}$ and containing $\boldsymbol{x}$ into two half-bands $\mathbb{B}_+(\boldsymbol{x}, \boldsymbol{s})$ and $\mathbb{B}_-(\boldsymbol{x}, \boldsymbol{s})$. Denoting $p_+(\boldsymbol{x}, \boldsymbol{s})$ and $p_-(\boldsymbol{x}, \boldsymbol{s})$ the probabilities of $\mathbb{B}_+(\boldsymbol{x}, \boldsymbol{s})$ and $\mathbb{B}_-(\boldsymbol{x}, \boldsymbol{s})$ respectively, the (*band*) weighted depth becomes

$$\mathrm{D}(\boldsymbol{x}) = \inf_{\|\boldsymbol{s}\|=1} \frac{p_+(\boldsymbol{x}, \boldsymbol{s})}{p_-(\boldsymbol{x}, \boldsymbol{s})}.$$

The sample version is calculated from the number of observations in $\mathbb{B}_+(\boldsymbol{x}, \boldsymbol{s})$ and $\mathbb{B}_-(\boldsymbol{x}, \boldsymbol{s})$.

Example 3.9. The *cone weight function* is defined for an angle $\alpha \in (0, \pi/2]$ as

$$w_+(x_1, \ldots, x_p) = \begin{cases} 1 & \text{if } \angle\big((x_1, \ldots, x_p), (0, \ldots, 0, x_p)\big) \leq \alpha \\ 0 & \text{elsewhere,} \end{cases}$$

where $\angle(\boldsymbol{x}, \boldsymbol{y})$ denotes the angle between two vectors. Clearly, for a continuous distribution and $\alpha = \pi/2$ it holds $\mathrm{HD} = \mathrm{D}$.

In some sense the cone weight function is a *modification* of cylinder weight function. To see that, it is sufficient to use an appropriate function $h(x_p)$ instead of a constant h in definition (8).

See Figure 2 for an example of cone weight function and cylinder weight function in $\mathbb{R}^2$.

In both example above, the probability of a halfspace was replaced by the probability of a subset of halfspace. Not only this means a direct generalisation of halfspace depth (replacing halfspace by another subset of $\mathbb{R}^p$), but even more flexibility is allowed by weighted depth.

Example 3.10. The *normal weight function* is defined as

$$
w_+(x_1, \ldots, x_p) = \begin{cases} \phi_{\mathbf{\Sigma}}(x_1, \ldots, x_{p-1}) & \text{if } x_p > 0 \\ 0 & \text{elsewhere,} \end{cases}
$$

where $\phi_{\mathbf{\Sigma}}$ is the density of $p-1$ dimensional normal distribution with zero mean and covariance matrix $\mathbf{\Sigma}$. It is, however, also possible to generalise the weight function in the way that the matrix $\mathbf{\Sigma}$ may be a function of x_p.

4. BASIC PROPERTIES OF WEIGHTED DEPTH

Let us summarise some facts about the depth function D.

Theorem 4.1. The depth function defined by (1) is translation invariant.

P r o o f. It follows directly from the definition that

$$
\mathrm{D}_{\boldsymbol{X}+\boldsymbol{a}}(\boldsymbol{x} + \boldsymbol{a}) = \mathrm{D}_{\boldsymbol{X}}(\boldsymbol{x}).
$$

$\square$

Theorem 4.2. The depth function defined by (1) is rotation invariant.

P r o o f. Every rotation of a vector $\boldsymbol{x} \in \mathbb{R}^p$ may be written as $\mathbf{B}\boldsymbol{x}$, where $\mathbf{B} \in \mathbb{O}_p$ is some orthogonal $p \times p$ matrix. Hence,

$$
\mathrm{D}_{\mathbf{B}\boldsymbol{X}}(\mathbf{B}\boldsymbol{x}) = \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{\mathrm{E}_{\mathrm{P}}\, w_+\big(\mathbf{A}(\mathbf{B}\boldsymbol{X} - \mathbf{B}\boldsymbol{x})\big)}{\mathrm{E}_{\mathrm{P}}\, w_-\big(\mathbf{A}(\mathbf{B}\boldsymbol{X} - \mathbf{B}\boldsymbol{x})\big)} = \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{\mathrm{E}_{\mathrm{P}}\, w_+\big(\mathbf{A}\mathbf{B}(\boldsymbol{X} - \boldsymbol{x})\big)}{\mathrm{E}_{\mathrm{P}}\, w_-\big(\mathbf{A}\mathbf{B}(\boldsymbol{X} - \boldsymbol{x})\big)}
$$

$$
= \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{\mathrm{E}_{\mathrm{P}}\, w_+\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big)}{\mathrm{E}_{\mathrm{P}}\, w_-\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big)} = \mathrm{D}_{\boldsymbol{X}}(\boldsymbol{x})
$$

since $\{\mathbf{A}\mathbf{B} : \mathbf{A} \in \mathbb{O}_p\} = \mathbb{O}_p$ as follows from the orthogonality of $\mathbf{B}$. $\square$

Recall that a *support* sp(P) of probability measure P is the smallest closed set with probability 1, i. e.

$$
\mathrm{sp}(\mathrm{P}) = \bigcap \{ F \in \mathcal{F} : \mathrm{P}(F) = 1 \},
$$

where $\mathcal{F}$ denotes class of all closed subsets. The *closed convex support* csp(P) of probability measure P is defined as closed convex hull of the support sp(P).

Theorem 4.3. Consider the weight function w_+ such that $w_+(\boldsymbol{x}) > 0$ if $x_1^2 + \cdots + x_{p-1}^2 < k$ and $w_+(\boldsymbol{x}) = 0$ elsewhere (k may be infinite). Then $D_P(\boldsymbol{x}) = 0$ for any $\boldsymbol{x} \notin \mathrm{csp}(P)$.

P r o o f. Note that under the assumptions on w_+ there for all $\boldsymbol{x} \in \mathbb{R}^p$ exists an orthogonal matrix $\mathbf{A}_{\boldsymbol{x}}$ such that $\mathrm{E}w_+\big(\mathbf{A}_{\boldsymbol{x}}(\boldsymbol{X} - \boldsymbol{x})\big) > 0$. It is clear that $D_P(\boldsymbol{x}) > 0$ implies that for all orthogonal matrices $\mathbf{A}$ it holds

$$\mathrm{E}w_+\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big) > 0 \Rightarrow \mathrm{E}w_-\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big) = \mathrm{E}w_+\big(\mathbf{I}_-\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big) > 0. \quad (9)$$

Consider $\boldsymbol{x} \notin \mathrm{csp}(P)$ such that $D_P(\boldsymbol{x}) > 0$. It follows from (9) that $\boldsymbol{x}$ is "surrounded" by points of $\mathrm{sp}(P)$ and therefore $\boldsymbol{x}$ is in the closed convex support of P. $\qquad \square$

Example 4.4. On the other hand, a point $\boldsymbol{x} \in \mathrm{int}\big(\mathrm{csp}(P)\big)$ (here $\mathrm{int}(M)$ denotes the interior of a set) need not to be of positive depth. This is a difference from the halfspace depth, since

$$\boldsymbol{x} \in \mathrm{int}\big(\mathrm{csp}(P)\big) \Rightarrow \mathrm{HD}(\boldsymbol{x}) > 0.$$

Indeed, consider uniform distribution on a set

$$S = \{(x, y) : x > 0, \ 1 < x^2 + y^2 < 2\},$$

and a point $\boldsymbol{a} = (x_0, y_0) = (1/2, 0)$. Consider the depth function based on the band weight function of Example 3.8, where $r^2 < 3/4$. Indeed, for the direction $\boldsymbol{s} = (-1, 0)$ it is clear that $p_+(\boldsymbol{a}, \boldsymbol{s})/p_-(\boldsymbol{a}, \boldsymbol{s}) = 0$ (we follow the notation of Example 3.8) and hence $D\big((1/2, 0)\big) = 0$.

There is a particular interest in the so called *deepest point*, i.e., the point $\widetilde{\boldsymbol{x}}$ for which

$$D(\widetilde{\boldsymbol{x}}) = \max_{\boldsymbol{x}} D(\boldsymbol{x}).$$

Definition (1) in general does not give a unique deepest point even in a situation of an absolutely continuous distribution with connected support.

Example 4.5. Let us consider the uniform distribution on a set

$$S = \{(x_1, x_2)^T : \ 0 \le x_1 \le 10, \ 0 \le x_2 \le 1\} \cup \{(x_1, x_2) : \ 0 \le x_1 \le 1, \ 0 \le x_2 \le 10\}.$$

Let us consider the band weight function (8) with a small h, say $h = 1/20$, and the corresponding weighted depth function. From the shape of the support S it follows that the only unique deepest point may lie on a line $x_1 = x_2$ only. It can be seen that for any point $\boldsymbol{x}$ on the line $x_1 = x_2$ it holds $D(\boldsymbol{x}) \le 1/9$.

Consider the point $\boldsymbol{z} = (5, 1/2)$. After some calculations we get $D(\boldsymbol{z}) > 1/9 \ge D(\boldsymbol{x})$ for any $\boldsymbol{x} = (x_1, x_1)^T$. Indeed, the lower estimate for $D(\boldsymbol{z})$ may be obtained considering a line l connecting $\boldsymbol{z}$ and the point $(0, 10)$ together with a band b of the width $2h$ around l and, on the other hand considering a line l' connecting $\boldsymbol{z}$ and
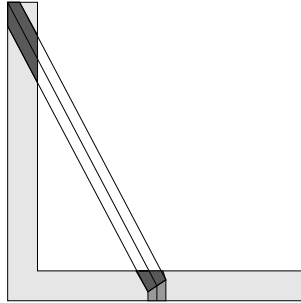
Figure 1: The deepest point need not to be unique, see Example 4.5.

the point $(5, 0)$ with the same band around. See Figure 1 for a visualisation of this example.

In this example there is no natural central point although the distribution is symmetric about the line $x_2 = x_1$. There are two deepest points (symmetric about the line of symmetry). The central regions are symmetric about the $x_1 = x_2$ axis as well.

Remark 4.6. In general the function D does not fulfil two of the key properties (2, 3). The depth need not decrease along a ray from the deepest point (even if the deepest point is unique). And the sets

$$\{\boldsymbol{x} : \mathrm{D}(\boldsymbol{x}) \geq d\}, \quad d \in [0, 1] \tag{10}$$

need not be convex and may be sometimes disconnected. This fact depends on the underlying distribution; however, in some situations these properties are desirable.

In Example 4.5 there isn't any "natural" deepest point. On the other hand, if there is an intuitive deepest point, like the point of central symmetry, we would like to prove that it is the deepest point for the weighted depth function. Indeed it is the case for a suitable weight function.

Before we prove a symmetry of depth for a symmetric distribution, we recall two notions of symmetry for random variable. We denote by $\mathcal{B}(\mathbb{R}^p)$ class of all Borel sets on $\mathbb{R}^p$ and by $\|\cdot\|$ the usual Euclidean norm.

Definition 4.7. Distribution of random vector $\boldsymbol{X} \in \mathbb{R}^p$ is called *centrally symmetric* if there exists a point $\boldsymbol{s} \in \mathbb{R}^p$ such that

$$\mathrm{P}[(\boldsymbol{X} - \boldsymbol{s}) \in B] = \mathrm{P}[-(\boldsymbol{X} - \boldsymbol{s}) \in B] \qquad \forall\, B \in \mathcal{B}(\mathbb{R}^p).$$

We shortly say that $\boldsymbol{X}$ is *centrally symmetric about* $\boldsymbol{s}$.

Definition 4.8. Distribution of random vector $\boldsymbol{X} \in \mathbb{R}^p$ is called *angular symmetric* if there exists a point $\boldsymbol{s} \in \mathbb{R}^p$ such that

$$\mathrm{P}\left[\frac{\boldsymbol{X} - \boldsymbol{s}}{\|\boldsymbol{X} - \boldsymbol{s}\|} \in B\right] = \mathrm{P}\left[-\frac{\boldsymbol{X} - \boldsymbol{s}}{\|\boldsymbol{X} - \boldsymbol{s}\|} \in B\right] \qquad \forall\, B \in \mathcal{B}(\mathbb{R}^p).$$

We shortly say that $\boldsymbol{X}$ is *angular symmetric about* $\boldsymbol{s}$.

Theorem 4.9. Let w_+ be symmetric about x_p-axis, i.e., $w_+(x_1, \ldots, x_{p-1}, x_p) = w_+(-x_1, \ldots, -x_{p-1}, x_p)$ and suppose that the distribution of $\boldsymbol{X}$ is centrally symmetric about point $\boldsymbol{\theta}$. Then

$$\mathrm{D}(\boldsymbol{x}) \leq \mathrm{D}(\boldsymbol{\theta}) = 1, \quad \forall \boldsymbol{x} \in \mathbb{R}^p.$$

Proof. It can be assumed that $\boldsymbol{\theta} = \boldsymbol{0}$ without loss of generality (translation invariance of D). Since w_+ is symmetric about x_p-axis and $w_-(\boldsymbol{x}) = w_+(\mathbf{I}_-\boldsymbol{x})$ it holds

$$w_+(\boldsymbol{x}) = w_-(-\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{R}^p.$$

It follows that

$$\mathrm{E}\, w_-(\mathbf{A}\boldsymbol{X}) = \mathrm{E}\, w_+(-\mathbf{A}\boldsymbol{X}) = \mathrm{E}\, w_+(\mathbf{A}\boldsymbol{X})$$

for $\boldsymbol{X}$ centrally symmetric about $\boldsymbol{0}$ and arbitrary matrix $\mathbf{A} \in \mathbb{O}_p$. Thus $\mathrm{D}(\boldsymbol{0}) = 1$. The fact that $\mathrm{D}(\boldsymbol{x}) \leq 1$, $\forall \boldsymbol{x}$ completes the proof. $\square$

This result may be extended to angular symmetric distributions.

Theorem 4.10. Let w_+ be symmetric about x_p-axis and suppose that the distribution of $\boldsymbol{X}$ is angular symmetric about point $\boldsymbol{\theta}$. If w_+ is such that

$$w_+(k\boldsymbol{x}) = w_+(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{R}^p,\ k \geq 0 \tag{11}$$

then

$$\mathrm{D}(\boldsymbol{x}) \leq \mathrm{D}(\boldsymbol{\theta}) = 1, \quad \forall \boldsymbol{x} \in \mathbb{R}^p.$$

Proof. It is an analogue to the proof of Theorem 4.9. Let $\boldsymbol{\theta} = \boldsymbol{0}$ without loss of generality. Under the assumption (11) it holds

$$\mathrm{E}\, w_-(\mathbf{A}\boldsymbol{X}) = \mathrm{E}\, w_+(-\mathbf{A}\boldsymbol{X}) = \mathrm{E}\, w_+(-\mathbf{A}\boldsymbol{X}/\|\mathbf{A}\boldsymbol{X}\|)$$
$$= \mathrm{E}\, w_+(\mathbf{A}\boldsymbol{X}/\|\mathbf{A}\boldsymbol{X}\|) = \mathrm{E}\, w_+(\mathbf{A}\boldsymbol{X})$$

$\forall \mathbf{A} \in \mathbb{O}_p$, hence

$$\mathrm{D}(\boldsymbol{0}) = \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{\mathrm{E}\, w_+(\mathbf{A}\boldsymbol{X})}{\mathrm{E}\, w_-(\mathbf{A}\boldsymbol{X})} = 1.$$

$\square$

Remark 4.11. In Theorem 4.10 it is sufficient to define the weight function w_+ on the unit halfsphere

$$S_{p,+} = \{\boldsymbol{x} : \|\boldsymbol{x}\| = 1,\ x_p \geq 0\}$$

and use $w_+(\boldsymbol{x}) = w_+(\boldsymbol{x}/\|\boldsymbol{x}\|)$ to ensure (11). Obviously the cylinder (band) depth does not satisfy the assumption of Theorem 4.10. On the other hand the assumption of the theorem is satisfied by the cone weight function defined in Example 3.9.

Example 4.12. Let $\boldsymbol{X}$ be a two dimensional random vector with normal distribution $\mathsf{N}_2(\boldsymbol{0}, \boldsymbol{I}_2)$. Suppose we have a band weight function

$$w_+(x_1, x_2) = \begin{cases} 1, & \text{if } -h < x_1 < h, \ x_2 > 0 \\ 0, & \text{otherwise} \end{cases}$$

for given $h > 0$. We use the same notation as in Example 3.8, hence

$$\mathrm{D}(\boldsymbol{x}) = \inf_{\|\boldsymbol{s}\|=1} \frac{p_+(\boldsymbol{x}, \boldsymbol{s})}{p_-(\boldsymbol{x}, \boldsymbol{s})}. \tag{12}$$

First we show that for an arbitrary point $\boldsymbol{x}$ it holds

$$\mathrm{D}(\boldsymbol{x}) = \inf_{\|\boldsymbol{s}\|=1} = \min\left\{ \frac{p_+(\boldsymbol{x}, \boldsymbol{s}_0)}{p_-(\boldsymbol{x}, \boldsymbol{s}_0)}, \frac{p_-(\boldsymbol{x}, \boldsymbol{s}_0)}{p_+(\boldsymbol{x}, \boldsymbol{s}_0)} \right\}$$

for $\boldsymbol{s}_0$ such that $\boldsymbol{0} \in \{\boldsymbol{x} + t\boldsymbol{s}_0, \ t \in \mathbb{R}\}$. Without loss of generality we can assume that $\boldsymbol{x} = (0, x_2)^T$ (the distribution is symmetric about $\boldsymbol{0}$ and also about any line containing $\boldsymbol{0}$). For such a point $\boldsymbol{x}$ let $\boldsymbol{s} = (0, 1)^T$. One has

$$p_+(\boldsymbol{x}, (0,1)^T) = \mathrm{P}(X_2 > x_2, \ -h < X_1 < h) = (1 - \Phi(x_2))\mathrm{P}(-h < X_1 < h),$$
$$p_-(\boldsymbol{x}, (0,1)^T) = \mathrm{P}(X_2 < x_2, \ -h < X_1 < h) = \Phi(x_2)\mathrm{P}(-h < X_1 < h),$$

where Φ is the distribution function of $\mathsf{N}(0,1)$. For any other direction $\boldsymbol{u} \neq \boldsymbol{s}$ and bands $\mathbb{B}(\boldsymbol{x}, \boldsymbol{u})$ there exists uniquely determined rotation $\boldsymbol{A} \in \mathbb{O}_2$ such that $\boldsymbol{A}\boldsymbol{u} = (0, 1)^T$ and $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{X}' \sim \mathsf{N}_2(\boldsymbol{0}, \boldsymbol{I}_2)$. For $\boldsymbol{x} = (0, x_2)^T$ it holds $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'$ where $x_2 > x_2'$. It is easy to show that

$$\begin{aligned} p_+(\boldsymbol{x}, \boldsymbol{u}) &= p_+(\boldsymbol{x}', (0,1)^T) = \mathrm{P}(X_2' \geq x_2')\mathrm{P}(x_1' - h < X_1' < x_1' + h) \\ &= (1 - \Phi(x_2'))\mathrm{P}(x_1' - h < X_1' < x_1' + h), \\ p_-(\boldsymbol{x}, \boldsymbol{u}) &= \Phi(x_2')\mathrm{P}(x_1' - h < X_1' < x_1' + h). \end{aligned}$$

Since $\Phi(x_2) > \Phi(x_2')$ it follows

$$\frac{p_+(\boldsymbol{x}, \boldsymbol{u})}{p_-(\boldsymbol{x}, \boldsymbol{u})} = \frac{1 - \Phi(x_2')}{\Phi(x_2')} > \frac{1 - \Phi(x_2)}{\Phi(x_2)} = \frac{p_+(\boldsymbol{x}, (0,1)^T)}{p_-(\boldsymbol{x}, (0,1)^T)}.$$

Hence

$$\mathrm{D}(\boldsymbol{x}) = \frac{1 - \Phi(x_2)}{\Phi(x_2)}.$$

Since both the depth function and the distribution are invariant with respect to rotation, it follows that for any $\boldsymbol{y} \in \mathbb{R}^2$

$$\mathrm{D}(\boldsymbol{y}) = \mathrm{D}\big((0, \|\boldsymbol{y}\|)^T\big) = \frac{1 - \Phi(\|\boldsymbol{y}\|)}{\Phi(\|\boldsymbol{y}\|)}.$$

The depth does not depend on the value of h and it is equal to the halfspace depth.

5. CONSISTENCY OF THE DEPTH FUNCTION

We shall prove in this section a strong pointwise consistency of the depth function under relatively mild conditions on the weight function. Note that the consistency of the halfspace depth is a direct corollary to our result (see Remark 3.6).

In what follows we consider an *absolutely continuous* Borel probability measure P on $\mathbb{R}^p$. Let us denote

$$\angle(\boldsymbol{u}, \boldsymbol{v}) \text{ the angle of vectors } \boldsymbol{u} \text{ and } \boldsymbol{v},$$

and

$$\mathbb{A}_\varphi \subset \mathbb{O}_p \text{ set of all rotation matrices } \mathbf{A} \text{ such that } \angle(\boldsymbol{u}, \mathbf{A}\boldsymbol{u}) \leq \varphi \text{ for all } \boldsymbol{u} \in \mathbb{R}^p.$$

Note that $\mathbb{A}_0 = \{\mathbf{I}_p\}$. Finally, let us denote by $\mathbf{N}_{\boldsymbol{s}}$ any matrix representing an *orthogonal rotation* such that $\mathbf{N}_{\boldsymbol{s}}\boldsymbol{s} = (\mathbf{0}^T, 1)^T$, $\mathbf{N}_{(\mathbf{0}^T,1)^T} := \mathbf{I}_p$. Such a matrix need not to be defined uniquely, however, for any two different $\mathbf{N}_{\boldsymbol{s}}^1$, $\mathbf{N}_{\boldsymbol{s}}^2$ it holds

$$\angle(\mathbf{N}_{\boldsymbol{s}}^1\boldsymbol{u}, \mathbf{N}_{\boldsymbol{s}}^1\boldsymbol{v}) = \angle(\mathbf{N}_{\boldsymbol{s}}^2\boldsymbol{u}, \mathbf{N}_{\boldsymbol{s}}^2\boldsymbol{v}) = \angle(\boldsymbol{u}, \boldsymbol{v}) \quad \text{for all} \quad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^p.$$

Definition 5.1. (Regularity of weight function) We say that weight function w_+ satisfies regularity conditions if

(A) $w_+(x_1, \ldots, x_{p-1}, x_p)$ is spherically symmetric about x_p-axis, i. e. w_+ is a function of $(x_1^2 + \cdots + x_{p-1}^2, x_p)$. In other words, w_+ is a function of the distance from x_p axis and values on x_p axis.

(B) w_+ is measurable and bounded.

(C) For arbitrary point $\boldsymbol{x}$ it holds that

$$\lim_{\varphi \to 0+} \sup_{\mathbf{A} \in \mathbb{A}_\varphi} \left\{ w_+\big(\mathbf{A}\mathbf{N}_{\boldsymbol{s}}(\boldsymbol{X} - \boldsymbol{x})\big) \right\} = w_+\big(\mathbf{N}_{\boldsymbol{s}}(\boldsymbol{X} - \boldsymbol{x})\big) \text{ P-a.s.}$$

$$\lim_{\varphi \to 0+} \inf_{\mathbf{A} \in \mathbb{A}_\varphi} \left\{ w_+\big(\mathbf{A}\mathbf{N}_{\boldsymbol{s}}(\boldsymbol{X} - \boldsymbol{x})\big) \right\} = w_+\big(\mathbf{N}_{\boldsymbol{s}}(\boldsymbol{X} - \boldsymbol{x})\big) \text{ P-a.s.}$$

for every direction $\boldsymbol{s}$, $\|\boldsymbol{s}\| = 1$. In other words, the sup, resp. inf function over all orthogonal rotations is P-a.s. continuous from right in 0 with respect to a rotation angle.

Let us first denote two important subsets of points. Define

$$\mathcal{H}_1 = \{\boldsymbol{x} : \inf_{\mathbf{A} \in \mathbb{O}_p} \mathrm{E}\, w_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})) > 0\},$$

$$\mathcal{H}_2 = \{\boldsymbol{x} : \exists \delta > 0 \,\forall \varepsilon > 0 \,\exists \mathbf{A}_\varepsilon \in \mathbb{O}_p : \mathrm{E}\, w_+(\mathbf{A}_\varepsilon(\boldsymbol{X} - \boldsymbol{x})) < \varepsilon$$
$$\text{and } \mathrm{E}\, w_-(\mathbf{A}_\varepsilon(\boldsymbol{X} - \boldsymbol{x})) > \delta\}.$$

Remark 5.2. It is easy to see that the set $\mathcal{H}_1$ contains the interior of support $\mathrm{sp}(\mathrm{P})$, i. e. points whose open neighbourhood is contained in the support of P. In the case of absolutely continuous distribution $\mathrm{P}(\mathcal{H}_1) = 1$. On the other hand the set $\mathcal{H}_2$ represents points with zero depth and, in particular, for the *complement* of support $\complement\mathrm{sp}(\mathrm{P})$ it holds $\complement\mathrm{sp}(\mathrm{P}) \subset \mathcal{H}_2$ under very weak conditions on the weight function w_+. It is easy to see that if $\boldsymbol{x} \in \mathcal{H}_1$ then $\mathrm{D}_\mathrm{P}(\boldsymbol{x}) > 0$ and if $\boldsymbol{x} \in \mathcal{H}_2$ then $\mathrm{D}_\mathrm{P}(\boldsymbol{x}) = 0$.

Theorem 5.3. Let P_n be an empirical measure defined by a random sample $\boldsymbol{X}_1, \ldots,$ $\boldsymbol{X}_n$ from distribution P. Let the weight function w_+ satisfies the regularity conditions of Definition 5.1. Then for any $\boldsymbol{x} \in \mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$ it holds

$$\mathrm{D}_{\mathrm{P}_n}(\boldsymbol{x}) \to \mathrm{D}_\mathrm{P}(\boldsymbol{x}) \quad \text{P-almost surely.} \tag{13}$$

P r o o f. For our purposes we will use standard conventions from measure theory for the extended real line $[-\infty, +\infty]$, e.g. $0.(\pm\infty) = 0$, $+\infty + \infty = +\infty$, etc. and we define logarithm in zero: $\log 0 := \lim_{x \to 0+} \log x = -\infty$.

The first step is to show that the class of functions $\mathcal{W} := \{\boldsymbol{y} \mapsto w_+(\mathbf{A}(\boldsymbol{y} - \boldsymbol{x})) : \mathbf{A} \in \mathbb{O}_p\}$ satisfies the *Uniform law of large numbers*. It means to prove that

$$\sup_{\mathbf{A} \in \mathbb{O}_p} \left| \frac{1}{n} \sum_{i=1}^{n} w_+(\mathbf{A}(\boldsymbol{X}_i - \boldsymbol{x})) - \mathrm{E}_\mathrm{P}\, w_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})) \right| \longrightarrow 0 \quad \text{P-a.s.} \tag{14}$$

To this end it is sufficient to prove that

$$H_{1,B}(\varepsilon, \mathcal{W}, \mathrm{P}) < +\infty, \quad \text{for all } \varepsilon > 0,$$

where $H_{1,B}(\varepsilon, \mathcal{W}, \mathrm{P})$ denotes entropy with ε-bracketing for $L_1(\mathrm{P})$-metric see [9, Lemma 3.1].

For a fixed vector $\boldsymbol{s}$ and a given angle φ we define functions

$$W_{\boldsymbol{s},\varphi}^U(\boldsymbol{z}) = \sup\{w_+(\mathbf{A}\mathbf{N}_{\boldsymbol{s}}(\boldsymbol{z} - \boldsymbol{x})) : \mathbf{A} \in \mathbb{A}_\varphi\},$$
$$W_{\boldsymbol{s},\varphi}^L(\boldsymbol{z}) = \inf\{w_+(\mathbf{A}\mathbf{N}_{\boldsymbol{s}}(\boldsymbol{z} - \boldsymbol{x})) : \mathbf{A} \in \mathbb{A}_\varphi\}.$$

Since $\mathbb{A}_0 = \{\mathbf{I}_p\}$ it holds $W_{\boldsymbol{s},0}^L(\boldsymbol{z}) = W_{\boldsymbol{s},0}^U(\boldsymbol{z}) = w_+(\mathbf{N}_{\boldsymbol{s}}(\boldsymbol{z} - \boldsymbol{x}))$. Further the inequality

$$W_{\boldsymbol{s},\varphi}^L(\boldsymbol{z}) \le w_+(\mathbf{N}_{\boldsymbol{a}}(\boldsymbol{z} - \boldsymbol{x})) \le W_{\boldsymbol{s},\varphi}^U(\boldsymbol{z}) \tag{15}$$

holds for arbitrary $\boldsymbol{z}$ and direction $\boldsymbol{a}$ such that $\angle(\boldsymbol{a}, \boldsymbol{s}) \le \varphi$.

For arbitrary direction $\boldsymbol{s}$ we define function

$$G_{\boldsymbol{s}}(\varphi) = \mathrm{E}_\mathrm{P} W_{\boldsymbol{s},\varphi}^U(\boldsymbol{X}).$$

This definition is correct, because for a measurable function w_+, the function $W_{\boldsymbol{s},\varphi}^U(\boldsymbol{z})$ is (universally) measurable; see Lemma 5.5 and its proof.

We will show that $G_{\boldsymbol{s}}$ is continuous from right in 0. Since w_+ is bounded, one has that $G_{\boldsymbol{s}}(\varphi) < +\infty$ for all $\varphi \in [0, \pi]$. Measurability and integrability together with condition (C) directly imply continuity from right of $G_{\boldsymbol{s}}$ in 0 using Lebesgue's dominated convergence theorem.

It follows that for all $\varepsilon > 0$ there exists φ_0 such that for all $\varphi \in [0, \varphi_0)$ holds

$$\begin{aligned}
\varepsilon > |G_{\boldsymbol{s}}(\varphi) - G_{\boldsymbol{s}}(0)| &= \left| \mathrm{E}_\mathrm{P}\left[W_{\boldsymbol{s},\varphi}^U(\boldsymbol{X}) - w_+(\mathbf{N}_{\boldsymbol{s}}(\boldsymbol{X} - \boldsymbol{x})) \right] \right| \\
&= \mathrm{E}_\mathrm{P} \left| W_{\boldsymbol{s},\varphi}^U(\boldsymbol{X}) - w_+(\mathbf{N}_{\boldsymbol{s}}(\boldsymbol{X} - \boldsymbol{x})) \right|.
\end{aligned}$$

Since inequality (15) holds, the last equation is correct. An analogous inequality holds for $W_{\boldsymbol{s},\varphi}^L$.

Hence, for arbitrary s, $\|s\| = 1$, and for every $\varepsilon > 0$ there exists $\varphi_s > 0$ such that

$$\mathrm{E_P}\big|W^U_{s,\varphi_s}(\boldsymbol{X}) - W^L_{s,\varphi_s}(\boldsymbol{X})\big| < \varepsilon. \tag{16}$$

Now, for arbitrary $\varepsilon > 0$, we construct ε-bracketing for $\mathcal{W}$. Let's consider the metric space $(\mathbb{S}_p, \rho)$, where $\mathbb{S}_p = \{s : \ \|s\| = 1\}$ and ρ is the Euclidean distance metric. Space $(\mathbb{S}_p, \rho)$ is closed and bounded, hence it is compact. For arbitrary $s \in \mathcal{S}$ an angle φ_s which satisfies (16) may be found. Denote by $\mathcal{C}(s, \varphi_s)$ a set of all $\boldsymbol{u} \in \mathbb{S}_p$ such that $\angle(\boldsymbol{u}, s) < \varphi_s$. $\mathcal{C}(s, \varphi_s)$ are open sets in the metric space $(\mathbb{S}_p, \rho)$ and form an open cover of $\mathbb{S}_p$. Since $\mathbb{S}_p$ is compact it follows that for any open cover there exists a finite subcover. In other words there exists a finite subset U of $\mathbb{S}_p$ such that

$$\mathbb{S}_p = \bigcup_{\boldsymbol{u} \in U} \mathcal{C}(\boldsymbol{u}, \varphi_{\boldsymbol{u}}).$$

Every function from $\mathcal{W}$ is determined by a direction $s \in \mathbb{S}_p$ in the sense that for an arbitrary function $v \in \mathcal{W}$ there exists $s \in \mathbb{S}_p$ such that $v(\boldsymbol{y}) = w_+(\mathbf{N}_s(\boldsymbol{y} - \boldsymbol{x}))$ and obviously there exists $\boldsymbol{u} \in U$ such that $s \in \mathcal{C}(\boldsymbol{u}, \varphi_{\boldsymbol{u}})$. Hence $W^U_{\boldsymbol{u},\varphi_{\boldsymbol{u}}}$ and $W^L_{\boldsymbol{u},\varphi_{\boldsymbol{u}}}$ are the corresponding bracketing functions which satisfy (15) and (16).

Finally, we obtain

$$H_{1,B}(\varepsilon, \mathcal{W}, \mathrm{P}) \leq \mathrm{card}(U) < +\infty.$$

and thus (14) holds.

Now we can come up to the proof of consistency of depth $\mathrm{D}_{\mathrm{P}_n}(\boldsymbol{x})$. It is a consequence of (14). Let us use the notation

$$\widehat{\mathrm{D}}_{\mathrm{P}}(\boldsymbol{x}, \mathbf{A}) = \frac{\mathrm{E_P}\, w_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}))}{\mathrm{E_P}\, w_-(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}))},$$

where the term $0/0$ is defined again as 1.

First the case $\boldsymbol{x} \in \mathcal{H}_1$ is treated. It holds

$$0 < \mathrm{D}_{\mathrm{P}}(\boldsymbol{x}) \leq \widehat{\mathrm{D}}_{\mathrm{P}}(\boldsymbol{x}, \mathbf{A}) \leq 1/\mathrm{D}_{\mathrm{P}}(\boldsymbol{x}) < +\infty, \quad \forall \mathbf{A} \in \mathbb{O}_p.$$

It follows from Lemma 5.7 below that

$$\begin{aligned}
|\log \mathrm{D}_{\mathrm{P}_n}(\boldsymbol{x}) - \log \mathrm{D}_{\mathrm{P}}(\boldsymbol{x})| &= |\inf_{\mathbf{A} \in \mathbb{O}_p} \log \widehat{\mathrm{D}}_{\mathrm{P}_n}(\boldsymbol{x}, \mathbf{A}) - \inf_{\mathbf{A} \in \mathbb{O}_p} \log \widehat{\mathrm{D}}_{\mathrm{P}}(\boldsymbol{x}, \mathbf{A})| \\
&\leq \sup_{\mathbf{A} \in \mathbb{O}_p} |\log \widehat{\mathrm{D}}_{\mathrm{P}_n}(\boldsymbol{x}, \mathbf{A}) - \log \widehat{\mathrm{D}}_{\mathrm{P}}(\boldsymbol{x}, \mathbf{A})| \\
&\leq \sup_{\mathbf{A} \in \mathbb{O}_p} \left(\Big| \log \frac{1}{n} \sum_{i=1}^n w_+(\mathbf{A}(\boldsymbol{X}_i - \boldsymbol{x})) - \log \mathrm{E_P}\, w_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})) \Big| \right. \\
&\qquad \left. + \Big| \log \frac{1}{n} \sum_{i=1}^n w_-(\mathbf{A}(\boldsymbol{X}_i - \boldsymbol{x})) - \log \mathrm{E_P}\, w_-(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})) \Big| \right) \\
&\leq 2 \sup_{\mathbf{A} \in \mathbb{O}_p} \Big| \log \frac{1}{n} \sum_{i=1}^n w_+(\mathbf{A}(\boldsymbol{X}_i - \boldsymbol{x})) - \log \mathrm{E_P}\, w_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})) \Big|
\end{aligned} \tag{17}$$

almost surely.

Since (14) holds it follows that also

$$\sup_{\mathbf{A} \in \mathbb{O}_p} \left| \log \frac{1}{n} \sum_{i=1}^{n} w_+(\mathbf{A}(\boldsymbol{X}_i - \boldsymbol{x})) - \log \mathrm{E}_\mathrm{P} \, w_+(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})) \right| \longrightarrow 0 \quad \text{P-a.s.}$$

From (17) one has that

$$|\log \mathrm{D}_{\mathrm{P}_n}(\boldsymbol{x}) - \log \mathrm{D}_\mathrm{P}(\boldsymbol{x})| \longrightarrow 0 \quad \text{P-a.s.}$$

So eventually,

$$|\mathrm{D}_{\mathrm{P}_n}(\boldsymbol{x}) - \mathrm{D}_\mathrm{P}(\boldsymbol{x})| \longrightarrow 0 \quad \text{P-a.s.}$$

We shall now consider the case $\mathcal{H}_2$. For $\boldsymbol{x} \in \mathcal{H}_2$ there exists $\delta > 0$ such that for any $\varepsilon > 0$ there exists $\mathbf{A}_\varepsilon$ and for any $\eta > 0$ there exists n_η such that for $n \geq n_\eta$

$$\frac{1}{n} \sum_{i=1}^{n} w_+(\mathbf{A}_\varepsilon(\boldsymbol{X}_i - \boldsymbol{x})) < \mathrm{E}_\mathrm{P} \, w_+(\mathbf{A}_\varepsilon(\boldsymbol{X}_i - \boldsymbol{x})) + \eta < \varepsilon + \eta,$$
$$\frac{1}{n} \sum_{i=1}^{n} w_-(\mathbf{A}_\varepsilon(\boldsymbol{X}_i - \boldsymbol{x})) > \mathrm{E}_\mathrm{P} \, w_-(\mathbf{A}_\varepsilon(\boldsymbol{X}_i - \boldsymbol{x})) - \eta > \delta - \eta,$$

(18)

holds P=a.s. (see the definition of $\mathcal{H}_2$ and (14)). It follows that for $n \geq n_\eta$

$$\begin{aligned}
|\mathrm{D}_{\mathrm{P}_n}(\boldsymbol{x}) - \mathrm{D}_\mathrm{P}(\boldsymbol{x})| &= \left| \inf_{\mathbf{A} \in \mathbb{O}_p} \frac{\frac{1}{n} \sum_{i=1}^{n} w_+(\mathbf{A}(\boldsymbol{X}_i - \boldsymbol{x}))}{\frac{1}{n} \sum_{i=1}^{n} w_-(\mathbf{A}(\boldsymbol{X}_i - \boldsymbol{x}))} - 0 \right| \\
&\leq \frac{\frac{1}{n} \sum_{i=1}^{n} w_+(\mathbf{A}_\varepsilon(\boldsymbol{X}_i - \boldsymbol{x}))}{\frac{1}{n} \sum_{i=1}^{n} w_-(\mathbf{A}_\varepsilon(\boldsymbol{X}_i - \boldsymbol{x}))} \\
&< \frac{\varepsilon + \eta}{\delta - \eta},
\end{aligned}$$

and since ε and η may be chosen arbitrary small the proof is completed.

$\square$

It is clear that the most restrictive regularity condition is (C). In the next theorem a simple sufficient condition for (C) is stated.

Corolary 5.4. Let us have $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ a p-dimensional sample from absolutely continuous probability distribution P and suppose spherically symmetric weight function w_+ about x_p axis (see Remark 3.7). Further assume that w_+ is continuous on some connected set $\mathcal{M} \subseteq \mathbb{R}^{p-1} \times [0, +\infty)$ of positive Lebesgue measure and that w_+ is equal to zero on $\mathbb{R}^p \setminus \mathcal{M}$. Then for any $\boldsymbol{x} \in \mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$ it holds

$$\mathrm{D}_{\mathrm{P}_n}(\boldsymbol{x}) \longrightarrow \mathrm{D}_\mathrm{P}(\boldsymbol{x}) \quad \text{P-a.s.}$$

Proof. We need to check the validity of regularity conditions.

Condition (C) for supremum can be equivalently expressed in the form:

$$\lim_{\varphi \to 0+} \sup_{\mathbf{A} \in \mathbb{A}_\varphi} \left\{ w_+\big(\mathbf{A}\mathbf{N}_s(\boldsymbol{y} - \boldsymbol{x})\big) f(\boldsymbol{y}) \right\} = w_+\big(\mathbf{N}_s(\boldsymbol{y} - \boldsymbol{x})\big) f(\boldsymbol{y})$$

for almost all $\boldsymbol{y}$ and for every direction $\boldsymbol{s}$, $\|\boldsymbol{s}\| = 1$. f denotes density of probability distribution P. In the following we will use this form of condition (C) and for fixed $\boldsymbol{s}$ we will work with shifted and rotated random vector $\mathbf{N}_s(\boldsymbol{X} - \boldsymbol{x})$ instead of random vector $\boldsymbol{X}$. Its density we denote by $f_{\boldsymbol{s}}$.

If $\boldsymbol{y} \notin \mathrm{clo}(\mathcal{M})$ then $\mathbb{R}^p \setminus \mathrm{clo}(\mathcal{M})$ is open set and thus there exists $\varphi_0 > 0$ such that for all $0 \le \varphi < \varphi_0$ it holds that $w_+(\mathbf{B}\boldsymbol{y})f_{\boldsymbol{s}}(\boldsymbol{y}) = 0$, where $\mathbf{B} \in \mathbb{O}_p$ is arbitrary orthogonal rotation about angle φ.

If $\boldsymbol{y} \in \mathrm{int}(\mathcal{M})$ then, since $\mathrm{int}(\mathcal{M})$ is open and w_+ is there continuous, one has that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\mathcal{B}(\boldsymbol{y}, \delta) = \{\boldsymbol{u} : \|\boldsymbol{u} - \boldsymbol{y}\| < \delta\} \subseteq \mathrm{int}(\mathcal{M})$ and inequality $|w_+(\boldsymbol{u}) - w_+(\boldsymbol{y})| < \varepsilon$ holds for every $\boldsymbol{u} \in \mathcal{B}(\boldsymbol{y}, \delta)$. For every such δ there exists angle $\varphi_0 > 0$ such that for arbitrary rotation $\mathbf{B} \in \mathbb{O}_p$ about angle smaller than φ_0 one has $\mathbf{B}\boldsymbol{y} \in \mathcal{B}(\boldsymbol{y}, \delta)$ and thus $|w_+(\mathbf{B}\boldsymbol{y}) - w_+(\boldsymbol{y})| < \varepsilon$. For any angle ξ, $0 \le \xi < \varphi_0$, we define set

$$\mathcal{U}_\xi(\boldsymbol{y}) = \{\boldsymbol{u} : \|\boldsymbol{u}\| = \|\boldsymbol{y}\|, \; \angle(\boldsymbol{y}, \boldsymbol{u}) \le \xi\} \subset \mathcal{B}(\boldsymbol{y}, \delta).$$

$\mathcal{U}_\xi(\boldsymbol{y})$ is compact and w_+ is continuous on this set. Thus

$$\sup_{\mathbf{A} \in \mathbb{A}_\xi} \left\{ w_+(\mathbf{A}\boldsymbol{y}) f_{\boldsymbol{s}}(\boldsymbol{y}) \right\} = f_{\boldsymbol{s}}(\boldsymbol{y}) \max\left\{ w_+(\boldsymbol{u}) : \; \boldsymbol{u} \in \mathcal{U}_\xi(\boldsymbol{y}) \right\}.$$

Therefor for all $\varepsilon > 0$ there exists angle $\varphi_0 > 0$ such that for all ξ, $0 \le \xi < \varphi_0$, inequality

$$\Big| \sup_{\mathbf{A} \in \mathbb{A}_\xi} \{w_+(\mathbf{A}\boldsymbol{y})f_{\boldsymbol{s}}(\boldsymbol{y})\} - w_+(\boldsymbol{y})f_{\boldsymbol{s}}(\boldsymbol{y}) \Big| = f_{\boldsymbol{s}}(\boldsymbol{y}) \Big| \max_{\boldsymbol{u} \in \mathcal{U}_\xi(\boldsymbol{y})} w_+(\boldsymbol{u}) - w_+(\boldsymbol{y}) \Big| < \varepsilon$$

holds for all $\boldsymbol{y} \in \mathbb{R}^p \setminus (\partial\mathcal{M} \cup \mathcal{K})$, where $\mathcal{K} = \{\boldsymbol{y} : \; f_{\boldsymbol{s}}(\boldsymbol{y}) = +\infty\}$. Whence condition (C) holds, because Lebesgue measure of $(\partial\mathcal{M} \cup \mathcal{K})$ is equal to zero.

The regularity of infimum function is proved analogically. □

There is a natural question what can be said about the points outside $\mathcal{H}$ and about the set $\mathcal{H}$ itself. First of all, let us show two counterexamples to the consistency of sample depth (see Figure 2).

We consider a uniform distribution on a "hourglass" set, and a uniform distribution on "four tiles". In both cases the distributions are symmetric around a naturally defined central point x and it is exactly the point x where the problem arises. For any sample size n there exists a.s. an orthogonal transformation A such that $\mathrm{E}_n w_+\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big) = 0$ while $\mathrm{E}_n w_-\big(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\big) > 0$. In both cases the central point x is *the only* point for which the sample depth is not consistent. Both points are also points of discontinuity of the depth function. Indeed, the theoretical depth $\mathrm{D}(\boldsymbol{x}) = 1$ as follows from the symmetry of distribution. On the other hand there exists sequence $\boldsymbol{x}_n \to \boldsymbol{x}$ such that $\mathrm{D}(\boldsymbol{x}_n) = 0$ for all n.
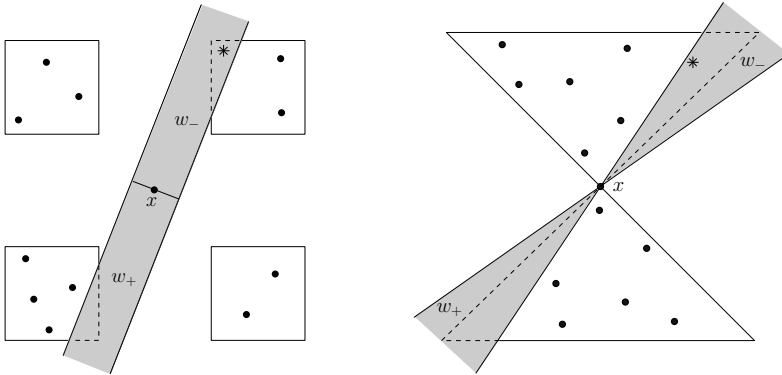
Figure 2: The sample depth need not to be consistent.

The nature of the problem lies in the limit of $0/0$ type. Assume without loosing the generality that the central point $\boldsymbol{x} = \boldsymbol{0}$. In both cases there exists an orthogonal transformation $\mathbf{A}_0$ and a sequence of orthogonal transformations $\mathbf{A}_n$ such that

$$
\begin{aligned}
&\mathrm{E}w_+\big(\mathbf{A}_0\boldsymbol{X}\big) = 0, \quad \mathrm{E}w_-\big(\mathbf{A}_0\boldsymbol{X}\big) = 0 \\
&\mathrm{E}w_+\big(\mathbf{A}_n\boldsymbol{X}\big) > 0, \quad \mathrm{E}w_-\big(\mathbf{A}_n\boldsymbol{X}\big) > 0 \quad \forall n \\
&\mathrm{E}w_+\big(\mathbf{A}_n\boldsymbol{X}\big) \to 0, \quad \mathrm{E}w_-\big(\mathbf{A}_n\boldsymbol{X}\big) \to 0 \quad \text{as } n \to \infty
\end{aligned}
\tag{19}
$$

There exist technical assumptions on the support of probability measure P and on the weight function (beside the regularity conditions of Definition 5.1) such that (19) does not hold for any point $\boldsymbol{x} \in \mathbb{R}^p$. Obviously, the critical points are in the interior of convex support and simultaneously in the complement of interior of support itself.

Therefore, if $\mathrm{sp}(\mathrm{P}) = \mathrm{csp}(\mathrm{P})$ then $\mathcal{H} = \mathbb{R}^p$ and the strong consistency holds for any point. An example may be normal distribution, bivariate exponential distribution, and many others.

As we have mentioned above, there are technical conditions on the support of probability measure P and on the weight function w_+ such that the consistency hold for $\boldsymbol{y} \in \mathbb{R}^p$. An example of such sufficient conditions may be

- There exist $r > 0$ and $w > 0$ such that $w_+(\boldsymbol{y}) \geq w$ if $y_1^2 + \cdots + y_{p-1}^2 \leq r$.

- There exists a compact set C such that $\mathrm{csp}(\mathrm{P}) \setminus \mathrm{sp}(\mathrm{P}) \subset C$.

- The interior of support $\mathrm{sp}(\mathrm{P})$ is a connected set.

These conditions are neither necessary conditions, nor the only possible sufficient conditions. In general, the set of points for which the consistency does not hold is, however, small in the sense of probability. Indeed, for any absolutely continuous distribution P it holds

$$
\mathrm{P}\{\boldsymbol{y} : \mathrm{D}_{\mathrm{P}_n}(\boldsymbol{y}) \to \mathrm{D}_{\mathrm{P}}(\boldsymbol{y}), \ \text{a.s.}\} = 1.
$$

The non-consistent points are, as may be clear from the counterexamples, special cases and may be considered as rather "pathological". In particular, consider the "hourglass" distribution together with the band weight function (rather than with the cone weight function) then the consistency of depth holds for the central point $\boldsymbol{x}$ as well as for any other points $\boldsymbol{y} \in \mathbb{R}^2$. Hence, it is a combination of a specific weight function and a specific distribution which causes the trouble at $\boldsymbol{x}$.

The following two technical lemmas are necessary for the proof of consistency.

Lemma 5.5. Let the weight function w_+ satisfy regularity conditions and consider fixed $\boldsymbol{s}, \|\boldsymbol{s}\| = 1$ and $\varphi \in [0, \pi]$. Then the function

$$\boldsymbol{z} \mapsto \sup\{w_+(\mathbf{AN}_{\boldsymbol{s}}(\boldsymbol{z} - \boldsymbol{x})) : \ \mathbf{A} \in \mathbb{A}_\varphi\}$$

is universally measurable.

P r o o f .   The function w_+ may be considered as a function of a distance $(d = \|\boldsymbol{x}\|)$ and the "direction" $\boldsymbol{s} = \boldsymbol{x}/\|\boldsymbol{x}\|$ where $\boldsymbol{s} \in \mathbb{S}_p$, the unit sphere. We use the metric $\rho(\boldsymbol{s}, \boldsymbol{z}) = \angle(\boldsymbol{s}, \boldsymbol{z})$ for $\boldsymbol{s}, \boldsymbol{z} \in \mathbb{S}_p$.

The problem is therefore equivalent to a problem of measurability of a function

$$g(d, s) = \sup\{f(d, z) : \tau(z, s) \leq e\}$$

if $f : [0, +\infty) \times \mathbb{M} \to [0, +\infty)$ is a measurable function, where $(\mathbb{M}, \tau)$ is a separable metric space. Denote $B^a = \{(d, z) : \ f(d, z) > a\}$ and note that B^a is a Borel set for any a due to the measurability of f. Denote $C^a := \{(d, s) : \ g(d, s) > a\}$. It is clear that for any d

$$C^a_d = \mathcal{U}_e(B^a_d),$$

where $M_d = \{s : \ (d, s) \in M\}$ denotes the d-section of a set M and $\mathcal{U}_e(N)$ denotes the e-neighbourhood of a set $N \subset \mathbb{M}$. The set C^a is therefore a projection of a Borel set

$$D^{a,e} = \{(d, s, z) \in [0, +\infty) \times \mathbb{M} \times \mathbb{M} : \ (d, z) \in B^a, \ \tau(s, z) \leq e\}$$

into the first two coordinates.

Since the projection of a Borel set is an analytic and hence a universally measurable set it follows that $g(y, x)$ is universally measurable function. $\square$

Remark 5.6. If a function g is universally measurable then for any finite Borel measure μ on $[0, +\infty) \times \mathbb{R}$ (in particular for any probability measure) there exist a pair of Borel functions g_1, g_2 such that $g_1(y, x) \leq g(y, x) \leq g_2(y, x)$ and $g_2 = g_1$ μ-almost surely. Hence the Lebesgue integral of universally measurable function is well defined.

Lemma 5.7. Consider two bounded functions f, $g : \ M \to \mathbb{R}$. Then

$$\sup\{|f(x) - g(x)| : \ x \in M\} \geq |\inf\{f(x) : \ x \in M\} - \inf\{g(x) : \ x \in M\}|.$$

P r o o f . If $\inf f = \inf g$ then it follows immediately because $\sup |f - g| \geq 0$.

If $\inf f > \inf g$ then there exists $\varepsilon_0 > 0$ such that for all ε, $0 < \varepsilon < \varepsilon_0$, exists $x_g \in M$ which satisfies

$$\inf g \leq g(x_g) < \inf g + \varepsilon < \inf f \leq f(x_g).$$

Therefor

$$\sup |f - g| \geq |f(x_g) - g(x_g)| \geq |\inf f - g(x_g)| > |\inf f - \inf g| - \varepsilon$$

for all ε, $0 < \varepsilon < \varepsilon_0$ and the proof of Lemma is completed. $\square$

6. EXAMPLES

In this section, we first shortly discuss the computational aspects of sample depth computation. Then few examples are given to show the main differences between the halfspace depth and the weighted depth.

Since the weighted halfspace depth is defined for a broad class of weight functions, a general fast algorithm for depth computing doesn't exist. Also, the theoretical depth $\mathrm{D_P}(\boldsymbol{x})$ of point $\boldsymbol{x}$ under a general absolutely continuous distribution P cannot be usually calculated exactly and some numerical approximation is needed. It is caused by the fact that $w_+(\mathbf{A}\boldsymbol{x})$ can attain different values for every transformation $\mathbf{A} \in \mathbb{O}_p$, which means that possibly uncountable number of values must be considered. The symmetric weight functions (see Remark 3.2) allow to use only rotation rather than all orthogonal transformations $\mathbb{O}_p$.

On the other hand, in some special cases the empirical depth may be computed exactly. It is the case when the weight function is piecewise constant. The cone weighted depth, the band weighted depth, the halfspace depth are, in particular, examples of such depths. The set $\{\sum_{i=1}^{n} w_+ \mathbf{A}(\boldsymbol{X}_i - \boldsymbol{x}), \mathbf{A} \in \mathbb{O}_p\}$ is finite for each $\boldsymbol{x}$ in such a case.

Straightforward algorithm is used to compute the sample depth of a given point $\boldsymbol{x}$. It uses a predefined number of vectors in $\mathbb{R}^{p-1} \times [0, +\infty)$ which represent halfspaces in which we compute sample weighted probability. These vectors are normal vectors of hyperplanes which determine appropriate halfspaces. For every such vector we rotate our dataset so that normal vector goes to x_p axis. Then we make, for rotated dataset, two computations of sample weighted probability – for halfspace where $x_p \geq 0$ and for halfspace where $x_p \leq 0$. Finally the depth is set to the smallest value of portions of sample weighted probabilities in $x_p \geq 0$ and $x_p \leq 0$ halfspaces. For sample size n the computation of weighted probability in given halfspace takes $O(n)$ steps. There are $2k$ halfspaces, hence computation of depth of given point takes $O(2kn)$ steps. If one wants to compute the depth of all points in dataset it takes $O(2kn^2)$ steps. Note that for two dimensional dataset setting the choice $2k = 1000$ halfspaces brings very precise answer.

We illustrate some differences between the weighted depth and the halfspace depth. In the following four examples we use the band weight function of Example 3.8, where $h = 0.25$ or $h = 0.5$, respectively, is the "radius" of the band, i. e., the

band width itself is $2h$. Therefore we speak about band weighted depth or simply about the *band depth*. Four bivariate distributions of random vector $\boldsymbol{X} = (X_1, X_2)^T$ are considered:

- Normal $\mathsf{N}_p(\mathbf{0}, \mathbf{I}_p)$

- Uniform on $[0, 1] \times [0, 1]$

- Exponential: $X_1 \sim \mathsf{Exp}(1)$, $X_2 \sim \mathsf{Exp}(1)$, where X_1 and X_2 are independent

- Mixture of two normal distributions, namely

$$\mathsf{N}_2\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 & -0.9\sqrt{3} \\ -0.9\sqrt{3} & 1 \end{pmatrix}\right) \text{ and } \mathsf{N}_2\left(\begin{pmatrix} -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 & 0.8\sqrt{2} \\ 0.8\sqrt{2} & 1 \end{pmatrix}\right)$$

We simulate 2500 points for each particular distribution and we compute sample depth of these points. In next figures the areas of 25%, 50% and 75% of the deepest points (points with the highest depth) are plotted. The rest of points (25% points with the lowest depth) are marked by light grey. A triangle marks the sample deepest point.

First let us consider two cases with natural centre - normal distribution and uniform distribution on the unit square $[0, 1] \times [0, 1]$.
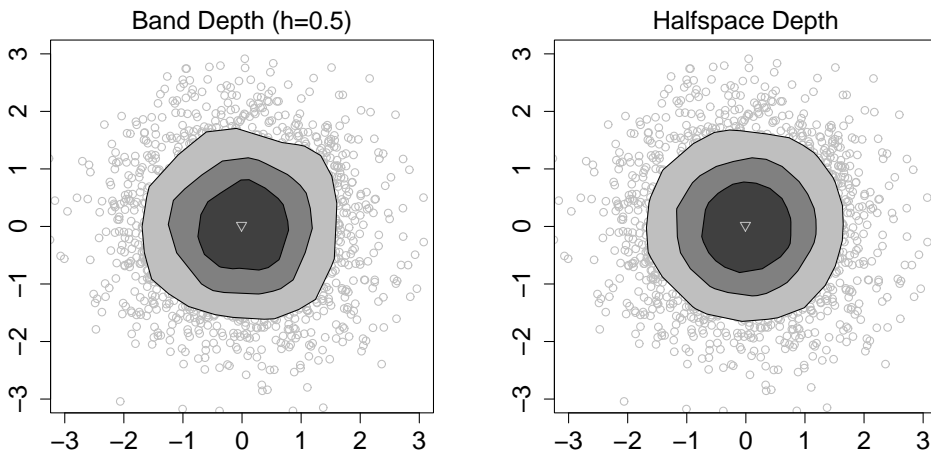
Figure 3: Normal distribution $\mathsf{N}_2(\mathbf{0}, \mathbf{I}_2)$: areas of 25%, 50% and 75% of deepest points.

In Figure 3 we can see that there is no big difference between the band weighted depth and the halfspace depth for bivariate normal distribution $\mathsf{N}_2(\mathbf{0}, \mathbf{I}_2)$. Both methods find point $(-0.008, 0.019)$ as the sample deepest point, which is the "observation" (sample point) with the smallest distance (in the standard Euclidean metric) from the theoretical centre $(0, 0)$. Areas of the deepest points are similarly large. The only remarkable difference is in the value of sample depth in sample deepest

point (recall that this is the same point for both methods) which is 0.88 for the band depth and 0.94 for the halfspace depth (theoretical depth of the deepest point is equal to one in both cases, see Theorem 4.9; the halfspace depth is here defined as the band depth with infinite bandwidth; see Remark 3.6 and Theorem 3.5). Differences between the sample band weighted depth and the sample halfspace depth for fixed sample size become smaller as h increases. Note that in this case the sample band depth for different values of h approach each other with increasing sample size (see Example 4.12 and Theorem 5.4).
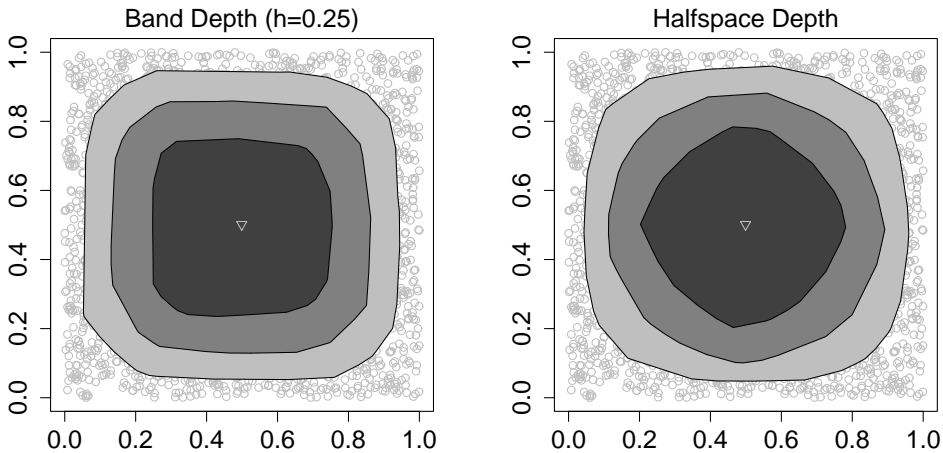
Figure 4: Uniform distribution on $[0,1] \times [0,1]$: areas of 25%, 50% and 75% of deepest points.

In Figure 4 areas of deepest points for uniform distribution on square $[0,1] \times [0,1]$ are displayed. The difference between band weighted depth (with $h = 0.25$, hence the band width is 0.5) and halfspace depth is obvious. The main difference is in the shape of areas of deepest points. Band depth keeps more faithfully the shape of the support i. e. square whereas halfspace depth areas are rather going to be a circle. It is not a surprise that for uniform distribution the areas are similarly large for both methods and there is a common sample deepest point $(0.499, 0.502)$, which is pretty close to theoretical centre $(0.5, 0.5)$. Again the sample depth of the sample deepest point is remarkably smaller for band depth (0.91 for band depth, 0.96 for halfspace depth).

We should note that differences are going to be smaller and smaller as h increases for both normal and uniform distribution. In the case of uniform distribution there is even no difference between theoretical band depth and halfspace depth if h is greater than the diagonal of the square ($h > \sqrt{2}$).

In two previous examples we have considered centrally symmetric distributions. For such a distribution there is naturally defined a unique centre. Now we will consider some distributions that are not symmetric and the notion of *centre* may be questionable.

In Figure 5 big differences between the band weighted depth and the halfspace
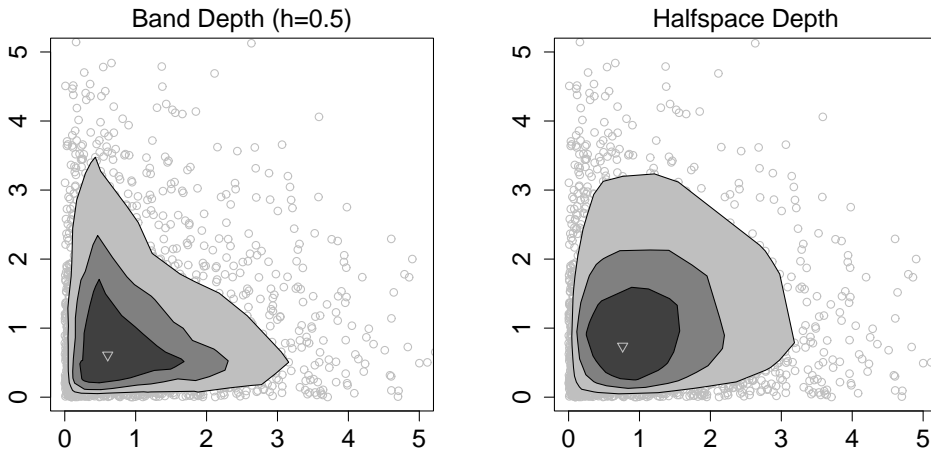
Figure 5: Exponential distribution ($X_1 \sim \mathsf{Exp}(1)$, $X_2 \sim \mathsf{Exp}(1)$, X_1 and X_2 are independent) : areas of 25%, 50% and 75% of deepest points.

depth for exponential distribution can be easily seen. Band depth areas are rather triangular whereas the halfspace depth areas are rather oval. Note that band depth areas correspond better to level sets of the density (level sets of this distribution are rectangular isosceles triangles with vertex in $(0,0)$). Also there is a remarkable difference in position of sample deepest point which is $(0.606, 0.610)$ for band depth (depth $= 0.68$) and $(0.763, 0.739)$ for halfspace depth (depth $= 0.77$). Both are close to line $y = x$, but sample deepest point for band depth is closer to 0. Another difference is that areas for the halfspace depth are about 30%-40% larger than for band depth.

In Figure 6 a mixture of two bivariate normal distributions is plotted and a remarkable difference between the halfspace depth and the band weighted depth is shown. Band weighted depth areas again correspond more faithfully to level sets of density. The shape of the areas for band depth give evidence that the distribution is mixture of two other distributions. Areas for halfspace depth are about 25% larger than for band depth. The difference in position of sample deepest point is not surprising. In such a situation (we have two distinct natural centres) estimator of the deepest point for band depth may be quite unstable, because in such cases there need not exist unique deepest point. For band depth the sample deepest point is $(0.099, 0.538)$ (depth $= 0.60$), for halfspace depth it is $(-0.534, 0.958)$ (depth $= 0.70$). Both these points are quite close to an abscissa that connects theoretical centres of normal distributions (these centres are marked by light circle).

There is an interesting question about the choice of bandwidth h. In general the smaller the bandwidth we use the more "local" behaviour of depth we obtain and the more shattered (due to the discontinuity of weight function) are the sample depth contours. It is desirable to take into account the variability of data and the tradeoff between the "local" and the "global" features of the data which should be emphasised.
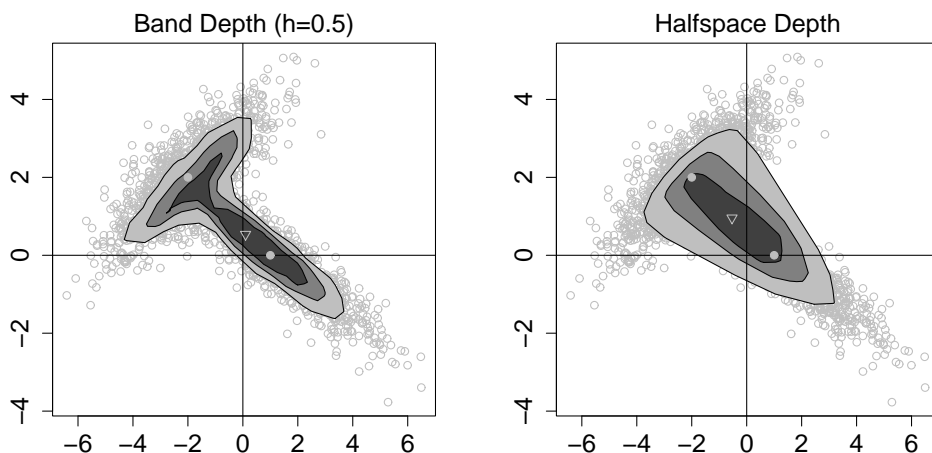
Figure 6: Mixture of two bivariate normal distributions: areas of 25%, 50% and 75% of deepest points.

Concluding this section we should note that

- Main differences between the band and the halfspace depth are in the shape of areas of deepest points.

- For considered nonsymmetric distributions the areas for the halfspace depth were remarkably larger than for the band depth.

- For symmetric distribution both depths localise the centre of symmetry quite well, for nonsymmetric distributions there are differences in localisation of the deepest point (which may not be unique for the band depth).

- Sample depth of the sample deepest point is usually higher for the halfspace depth.

ACKNOWLEDGEMENT

The author are indebted to anonymous referees for their careful reading and comments which led to substantial improvement of the original manuscript.

(Received June 4, 2008)

REFERENCES

[1] A. DasGupta, J. K. Ghosh, and M. M. Zen: A new general method for constructing confidence sets in arbitrary dimensions with applications. Ann. Statist. *23* (1995), 1408–1432.

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

[3] R. Y. Liu: On a notion of data depth based on random simplices. Ann. Statist. *18* (1990), 405–414.

[4] R. Y. Liu, R. Serfling, and D. L. Souvaine (eds.): DIMACS; Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications. American Mathematical Society, Providence RI 2006.

[5] J. Matoušek: Computing the center of planar point sets. In: DIMACS; Discrete and Computational Geometry (J. E. Goodman, R. Pollack, and W. Steiger, eds.). American Mathematical Society, 1992.

[6] I. Mizera: On depth and deep points: a calculus. Ann. Statist. *30* (2002), 1681–1736.

[7] P. Rousseeuw and I. Ruts: Algorithm AS307: bivariate location depth. J. Roy. Statist. Soc.-C *45* (1996), 516–526.

[8] J. Tukey: Mathematics and picturing data. In: Proc. 1975 International Congress of Mathematics, Vol. *2* (1975), pp. 523–531.

[9] S. van de Geer: Empirical Processes and M-Estimates. Cambridge, 2000.

[10] Y. Zuo and R. Serfling: General notion of statistical depth function. Ann. Statist. *28* (2000), 461–482.

[11] Y. Zuo and R. Serfling: Structural properties and convergence results for contours of sample statistical depth functions. Ann. Statist. *28* (2000), 483–499.

Daniel Hlubinka, Katedra pravděpodobnosti a matematické statistiky, Matematicko-fyzikální fakulta Univerzity Karlovy v Praze, Sokolovská 83, 186 75 Praha 8. Czech Republic.
    e-mail: hlubinka@karlin.mff.cuni.cz

Ondřej Vencálek, Katedra pravděpodobnosti a matematické statistiky, Matematicko-fyzikální fakulta Univerzity Karlovy v Praze, Sokolovská 83, 186 75 Praha 8. Czech Republic.
    e-mail: vencalek@karlin.mff.cuni.cz

Lukáš Kotík, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
    e-mail: kotik@utia.cas.cz