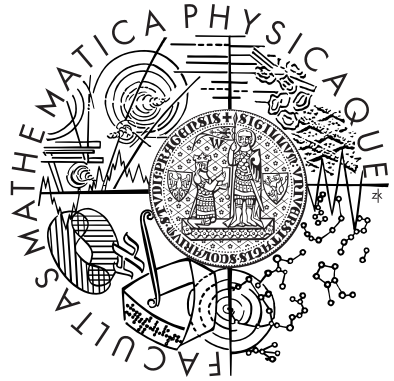
# Weighted Halfspace Depths and Their Properties

Lukáš Kotík




A Dissertation
Presented to the Faculty of Mathematics and Physics
of Charles University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Probability and Mathematical Statistics
Adviser: Doc. RNDr. Daniel Hlubinka, PhD.

Prague 2014

# Abstract

**Název práce:** Vážené poloprostorové hloubky a jejich vlastnosti

**Autor:** Lukáš Kotík

**Katedra / Ústav:** Katedra pravděpodobnosti a matematické statistiky MFF UK

**Vedoucí doktorské práce:** Doc. RNDr. Daniel Hlubinka, PhD.

**Abstrakt:** Statistické hloubkové funkce se staly populárním nástrojem při statistickém neparametrickém zpracování mnohorozměrných dat. Nejznámější hloubkovou funkcí je tzv. poloprostorová hloubka, která má mnoho žádoucích vlastností. Některé její vlastnosti však často vedou k zavádějícím výsledkům, obzvláště v případě jiných než elipticky souměrných rozdělení. Práce zavádí 2 nové třídy hloubkových funkcí. Obě zobecňují poloprostorovou hloubku, zachovávají si některé její vlastnosti a v případě jiných než elipticky souměrných, multimodálních a směsových rozdělení mohou vést k lepším výsledkům a více respektují geometrickou strukturu dat. Definice je založena na použití váženého (polo)prostoru namísto indikátoru samotného poloprostoru. Speciální volbou vah, především v práci zavedených kuželosečkových vah, dostaneme link mezi lokálním pohledem na data, tzv. jádrovými odhady hustoty a mezi globálním pohledem na data v podobě poloprostorové hloubky. Míru lokalizace určuje tvar váhové funkce. V práci jsou odvozeny vlastnosti zavedených hloubkových funkcí, včetně stejnoměrné silné konzistence. Limitní rozdělení je rovněž diskutováno a také jsou zmíněna další témata (regresní hloubka, funkcionální hloubka), která mají spojitost s hloubkou dat a navrhované hloubkové funkce zde mohou přinést určitá vylepšení.

**Klíčová slova:** hloubka dat, mnohorozměrná data, neparametrické metody, uspořádání, asymptotika

**Title:** Weighted Halfspace Depths and Their Properties

**Author:** Lukáš Kotík

**Department / Institute:** Department of Probability and Mathematical Statistics, MFF UK

**Supervisor of the doctoral thesis:** Doc. RNDr. Daniel Hlubinka, PhD.

**Abstract:** Statistical depth functions became well known nonparametric tool of multivariate data analyses. The most known depth functions include the halfspace depth. Although the halfspace depth has many desirable properties, some of its properties may lead to biased and misleading results especially when data are not elliptically symmetric. The thesis introduces 2 new classes of the depth functions. Both classes generalize the halfspace depth. They keep some of its properties and since they more respect the geometric structure of data they usually lead to better results when we deal with non-elliptically symmetric, multimodal or mixed distributions. The idea presented in the thesis is based on replacing the indicator of a halfspace by more general weight function. This provides us with a continuum, especially if conic-section weight functions are used, between a local view of data (e.g. kernel density estimate) and a global view of data as is e.g. provided by the halfspace depth. The rate of localization is determined by the choice of the weight functions and theirs parameters. Properties including the uniform strong consistency of the proposed depth functions are proved in the thesis. Limit distribution is also discussed together with some other data depth related topics (regression depth, functional data depth) where the application of the proposed depth functions can bring some improvements.

**Keywords:** data depth, multivariate data, nonparametric methods, ordering, asymptotics

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.
I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In ..... date ..... signature of the author

# Acknowledgements

I would like to thank my parents who were very supportive of me during my life and study.

I also would like to express my gratitude to ÚTIA and especially SÚRO for offering me a part-time job during my PhD study.

Finally, I would like to thank to my adviser whose thoughts and ideas were very inspiring.

To my parents.

# Contents

# List of Figures

# Preface

Thesis is organized as follows. First, assumptions and notations that are used in the whole thesis are introduced. First chapter shows introduction to the data depth. General definition of the depth function and definitions together with some properties and graphical illustration of different depth functions are shown. Chapter 2 defines new depth functions and shows their properties. Chapter 3 proves the uniform strong consistency of the proposed depth functions and discusses the limit distribution. Chapter 4 defines a special class of the proposed depth function that brings continuum between the *halfspace depth* and the *kernel density estimates*. Chapter 5 includes the other topics that are related to data depth, e.g. directional quantiles, regression depth and functional data depth. Last chapter includes brief discussion. In the Appendix we can find some technical theorems and lemmas used in the proofs. The thesis was written in LaTeX using modified version of Princeton University Thesis class (puthesis.cls). Majority of computations were done in R software ([R Core Team, 2014]). Some procedures were programmed in C, C++ and FORTRAN 77, 95.

## Notation and Basic Assumptions

In the thesis we use the following notations and suppose the following assumptions.

(i) Vectors are written in bold italic math font, matrices are in bold math font, i.e. $\boldsymbol{X}$ denote a vector, $\mathbf{A}$ denote a matrix.

(ii) We suppose a probability space $(\Omega, \mathcal{A}, \mathsf{P})$ and random vectors with values in $\mathbb{R}^p$, where $p \in \mathbb{N}$. A distribution (inducted probability measure on $\mathbb{R}^p$) of a random vector $\boldsymbol{X}$ is denoted by $P$ or $P_{\boldsymbol{X}}$.

(iii) With a *support* $\mathrm{sp}(P)$ of probability measure $P$ we mean the smallest closed set having probability 1, i.e.

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

where $\mathcal{F}$ denotes class of all closed subsets of $\mathbb{R}^p$.

(iv) A *random sample* of sample size $n$ from $P$ is denoted by $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. In this thesis we suppose that $P$ is an *absolutely continuous* (with respect to

Lebesgue measure) probability distribution. Note that, the majority of the results in this thesis holds also for discrete distributions. Some theorems are formulated and proved for general distribution $P$. This way we obtain results for the *empirical version* of proposed depth functions.

(v) The random sample versions of mathematical operators is denoted by '$n$' subscript, i.e. the empirical probability measure we denote by $P_{\mathrm{n}}$ and its corresponding expected value by $\mathsf{E}_n$. We sometimes work with a random vector $\boldsymbol{X}$ with distribution $P_{\mathrm{n}}$ (given a realization of a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$) instead of the random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ itself.

(vi) For some functions (depth functions in particular) when needed we add an additional parameter behind the semicolon. This parameter denotes the distribution. For instance

$$\mathrm{HD}(\boldsymbol{x}; P) \quad \text{or} \quad \mathrm{HD}(\boldsymbol{x}; \boldsymbol{X})$$

denotes the halfspace depth of a point $\boldsymbol{x}$ with respect to distribution $P$ of a random vector $\boldsymbol{X}$. Another example is that the following equality also holds

$$\mathrm{HD}(\boldsymbol{x}; P_{\mathrm{n}}) = \mathrm{HD}_n(\boldsymbol{x}).$$

Further when we work with a depth function we sometimes omit the distribution parameter (stated behind semicolon), i.e. for instance instead of $\mathrm{D}(\boldsymbol{\theta}; P)$ we use only $\mathrm{D}(\boldsymbol{\theta})$ especially in cases where there is no need to specify a distribution.

(vii) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ denotes an inner product of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ and $\|\boldsymbol{u}\|$ denotes a norm of a vector $\boldsymbol{u}$. In whole article we suppose *Euclidean scalar product* and *Euclidean norm* but majority of definitions and theorems can also be generalized in a straightforward way to any *Hilbert space.*

(viii) An *indicator function* is denoted by $\mathbb{1}\{\}$, i.e.

$$\mathbb{1}\{\text{condition}\} = \begin{cases} 1, & \text{condition holds,} \\ 0, & \text{otherwise.} \end{cases}$$

(ix) $M^c$ stands for the complement of a set $M$, i.e. $M^c = \mathbb{R}^p \setminus M$ for any set $M$.

(x) $\overline{M}$, $\mathsf{int}(M)$ and $\partial M$ denotes the *closure, interior* and *border* of a set $M$.

(xi) $\measuredangle(\boldsymbol{u}, \boldsymbol{v})$ stands for the angle between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, such that $\measuredangle(\boldsymbol{u}, \boldsymbol{v}) \in [0, \pi]$.

(xii) Vector and matrix transpose is denoted by "$^T$" superscript.

(xiii) *Orthogonal* matrix $\mathbf{A}$ is matrix such that

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I},$$

where $\mathbf{I}$ denotes identity matrix. Dimension of this matrix is clear from the context.

(xiv) The unit sphere in $\mathbb{R}^p$ is denoted by $\mathcal{S}^p$, i.e.

$$\mathcal{S}^p = \{\boldsymbol{u} \in \mathbb{R}^p : \ \|\boldsymbol{u}\| = 1\}.$$

(xv) In the text we use serif font to emphasize computer programs, software packages etc.

(xvi) We use quotation marks for notions that are not formally defined.

# Chapter 1

# Introduction to Data Depth

## 1.1 Introduction

The data depth is a nonparametric tool for statistical analysis of multivariate data. Roughly speaking, depth is a function which measures for each point in a $p$-dimensional Euclid space its *centrality* (or *outlyingness*) with respect to a given *probability distribution*. Formally the data depth is a function

$$\mathrm{D}: \ \mathbb{R}^p \times \mathcal{P} \longrightarrow [0, +\infty),$$

where $\mathcal{P}$ denotes a set of probability distributions on $\mathbb{R}^p$. The first argument represents the points to which it assigns depth value according to a probability distribution. In general we do not suppose any other properties. The most popular and known depth function is the *halfspace depth* introduced in [Tukey, 1974]. The halfspace depth of a point $\boldsymbol{x}$ is defined as infimum of probabilities of all closed halfspaces containing $\boldsymbol{x}$.

Over past two decades the data depth methodology has been intensively studied and it has proved its usefulness in many applications such as *rank tests, nonparametric robust estimation of location and scale parameters* and *data classification*, see e.g. [Liu & Singh, 2006], [Zuo & Serfling, 2000b], [Fraiman et al., 1999] or [Lange et al., 2014, Li et al., 2012].

The depth should be rather global than local characteristics of the position of the point; this is the main difference from the probability density which characterizes local "position" of the point with respect to the distribution. Data depth is mostly used to define *centre-outward ordering* for multivariate data, i.e., the depth induces linear (semi)ordering in situations where no natural ordering exists. Hence it is often used as a nonparametric and usually robust tool of measure of *spread*, *shape* and *symmetry* of data, and as an *outlier* detection tool. The overview of depth and its desirable properties may be found in [Zuo & Serfling, 2000a] or in [Serfling, 2006].

There is a lot of approaches to the data depth using various depth functions. See [Liu et al., 2003] where some of the most popular, and their properties, applications and computational aspects, are presented. In this thesis we propose a modification of the halfspace depth which improves its performance in some ap-

plications, especially in cases when data does not have strictly convex level-sets of probability density function.

The *halfspace depth* has many good properties, see e.g. [Serfling, 2006]. Even some fine properties hold for the halfspace depth, e.g., it was proved in [Kong & Zuo, 2010, Struyf & Rousseeuw, 1999] that the halfspace depth may characterize the underlying distributions. On the other hand a direct application of the halfspace depth to high-dimensional data leads to some problems and surprising results, see [Dutta et al., 2011]. It should be mentioned here, that for the unimodal elliptically symmetric absolutely continuous distributions the halfspace depth and the probability density functions are equivalent in the sense that there exists a bijection between the depth of a point and the probability density at that point. In this case majority of depth functions serves as a robust alternative to "classical" parametric multivariate methods based on assumption of normality or on assumption of elliptical symmetry.

In the thesis we focus on different situations when the halfspace depth (and also majority of other depth functions) and the probability density disagree. This is, for example, typical for probability distributions with *nonconvex* or *disconnected level-sets* of the probability density function. In this case majority of popular depth functions used nowadays tends to produce contours that are "more convex" that expected and that are spread to the direction when only few observations or low probability mass occurs. Hence it can fail to give us proper information about the data and it may have negative influence to the applications of the depth, e.g., in classification problems, but also for outliers detection, tests of unimodality and other.

See Figure 1.1 for a simple illustration of an application of depth functions and also as an illustration of problem stated in previous paragraphs. It shows *bivariate boxplots* for $Exp(1) \times Exp(1)$ distribution with 5 outlying points - group of observations that are in the neighbourhood of point with coordinates $(4, 4)^T$. The data depth is a tool that can, and is, used for generalization of boxplot to higher dimensions. Having a bivariate boxplot, we immediately see shape and spread of the data. The inner set consists of 50% of observations. Usually, points outside the outer set (called *fence*) are suspected of being outliers. The left panel shows so called *bagplot*, that can be obtained in R using function bagplot from aplpack library. In this case the *halfspace depth* was used as a tool that identifies the inner and outer set. No outlying observations are detected. Note that also no false identification of non-outlying observations happened. The right panel shows another generalization of boxplot. It is based on a classical assumption of *elliptical symmetry* - the covariance matrix is (robustly) estimated and then it is used for construction of a bivariate boxplot. It can be also viewed as depth based approach since *Mahalanobis distance* and thus *Mahalanobis depth* is used. This boxplot was computed using R function bivbox that can be downloaded from http://www.stat.sc.edu/~hitchcock/chapter2_R_examples.txt. This method detects all 5 outlying observations. But further it falsely marks non-outlying observations as outliers. Even worse the fence is spread to the direction where no observations occur. It will have negative impact on statistical inference based on this method. Note that, even for almost perfect depth function, there

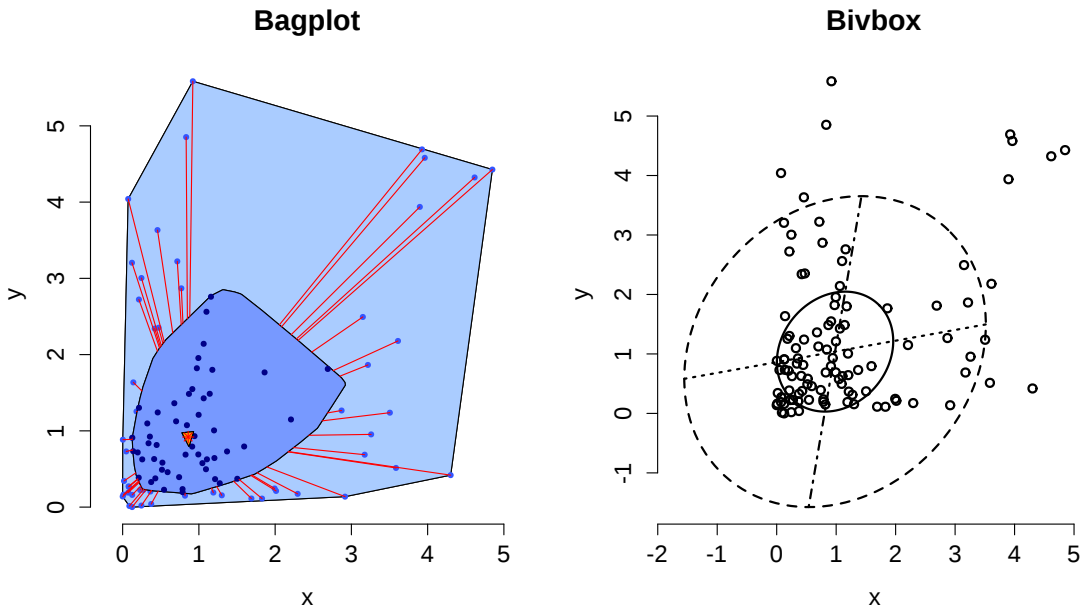



Figure 1.1: Bivariate boxplots for bivariate exponential distribution with contamination of few outlying points. Left panel: bagplot based on the *halfspace depth.* Right panel: bivbox based on assumption of elliptical symmetry, i.e. the *Mahalanobis depth* is used.

exist situations when an outlying point attains high value of depth (or equivalently has low rank) - e.g. see Figure 1.10 – 1.18 to see how various depth functions behave while applied on different datasets. The data depth proposed in the thesis deals better with problems mentioned in these paragraphs but usually in additional cost of need for more observations.

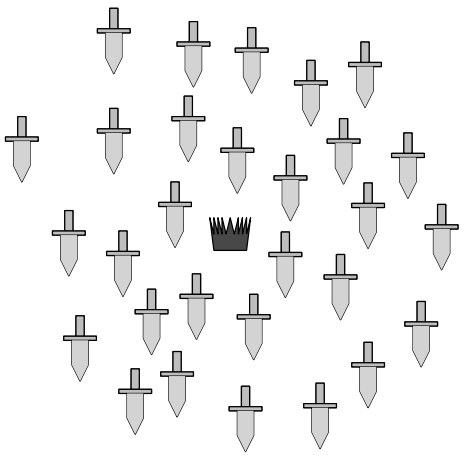


Figure 1.2: Protect the King! - The Fellowship of the Ring. King is well protected - his position is in the so called deepest point.

Here is another motivation for our modification of the halfspace depth (Regina Liu's interpretation of data depth from ICORS 2011 conference):

> *"Suppose we have a king who we want to protect with guards. We place the guards around the king. The depth function should measure how well is the king protected. Intuitively, well protected king is well surrounded by the guards. It means that he is covered against attack from every direction. Clearly, the defence of the king is as weak as its weakest spot – e.g. a direction where few guards are standing or where the guards leave uncovered view on the king. More in the center the king is placed more protection he has."*

Figure 1.2 illustrates a well protected king. In the words of statistical data depth the king is placed in the *deepest point* if we consider the positions of the guards as a bivariate dataset.

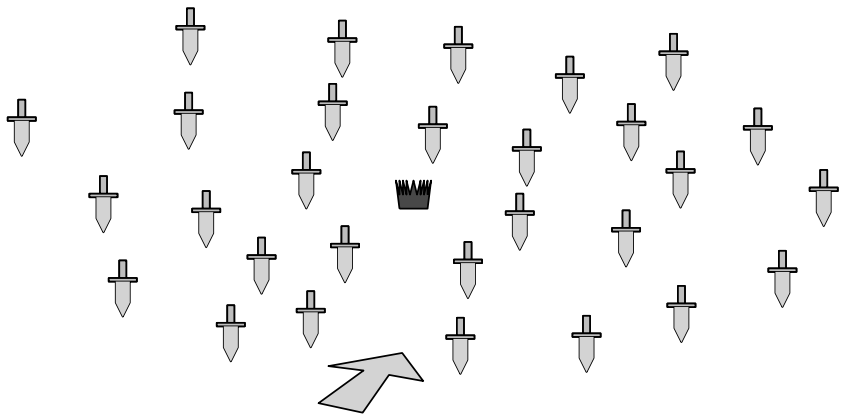


Figure 1.3: Protect the King! - The Return of the King. Affine invariant depth functions say that the king is still well protected.

The main idea of our modification of the halfspace depth may be – in this interpretation – summarized as follows: the halfspace depth is counting the guards in given halfspace regardless how far to the left or to the right they are. In contrary to this we propose to count the distant guards only partially (assigning some weight to them) or not to count them at all. In this sense we *localize* the halfspace depth and it will be shown in the thesis that, using the depth proposed here, one may balance the characterization of the point between the *global* (depth of the point) and *local* (probability density at the point) point of view. Note that in a similar way it is possible to localize other depths as well. However, in the thesis we focus on the halfspace depth only.

Consider the situation in Figure 1.3. Only difference to Figure 1.2 is that the distance between the guards linearly increases in one direction. Nowadays a big attention is given to affine invariant depth functions. Authorities in the field of data depth say that it is one of a key properties. Such depth functions say that the king is still well protected - the king position remains to be in deepest point. There lie the same number of guards in the halfspace indicated by arrow as in Figure 1.2 but there exists an uncovered area heading from outside directly towards the king. In this "protect the king" example there may be better to sacrifices the affine invariance to still keep the king in the well protected position.

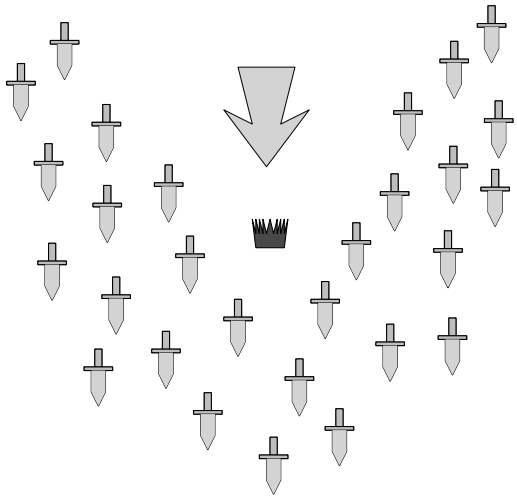


Figure 1.4: Protect the King! - The Two Towers. Non-convexly supported distribution of guards "around" the king. Majority of nowadays used depth functions says that the king is still well protected.

Figure 1.4 is an example of how non-convexly supported distribution of guards may still provide the good defence for the king. It is obvious that the arrow indicates the weakest direction. But there is still a lot of guards in the halfspace towards that arrow. The halfspace depth, similarly as other depth functions, says that the king has the best protection available. In this case the localized versions of depth will help to keep the king in better protected position. The depth proposed in the thesis suggests to use a depth function which we places the king in the spot where the least protected direction is better protected than the least protected direction from any other spot.

As noted in the beginning of this chapter the depth functions have many applications. Besides the robust rank test, outlier detection tool, robust estimates of scale and location it can be used to evaluate the difference in two distributions. Figure 1.5 shows so called *depth–depth plot.* It is used as a tool to visualize spread and location differences between two distributions. Here the *kernel weighted conic depth* (see Section 4.2, in particular Definition 13 and Example 15) was used to plot the depth of all observations with respect to $Exp(1) \times Exp(1)$ distribution (first class of observations) against the depth with respect to $Exp(1/2) \times Exp(1/2)$ distribution (the second class of observations). The layout of the observations indicates difference in scale and also in location. In recent developments of data depth methodology this tool has been used for classification. In the *depth–depth plot* a dividing curve is fitted and for a new observation the depth is computed with respect to both distributions (datasets from both classes). The class of the observation is assigned according to on which side of the dividing curve the observation lies. In some cases, the localized depth proposed in this thesis brings some improvements. Some of these results have been published in [Hlubinka & Vencalek, 2013].

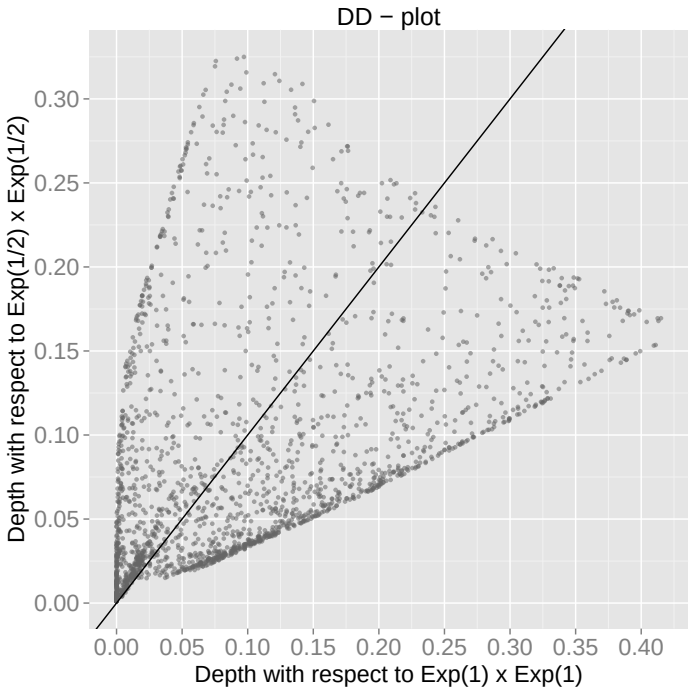



Figure 1.5: *Depth–depth plot* for $Exp(1) \times Exp(1)$ and $Exp(1/2) \times Exp(1/2)$ distributions. As a depth function the *generalized halfspace depth* was used (see Chapter 2) with the *kernel weighted conic* function with *Gaussian kernel*, eccentricity $e = 1.5$ (hyperbolic depth) and smoothing parameter $\sigma = 0.4$ were used. For more details see Section 4.2.

Following sections introduce a general definition of the statistical depth function. Some properties emerging from the general definition may lead to intriguing properties discussed in this section. Then a short overview of depth functions is shown. Only *halfspace depth* is discussed more thoroughly because it is a starting point for the depth functions proposed by us. Our goal was to find a depth function that generalizes the halfpace depth, keeps some of it properties and that works well with non-convexly supported data or bimodal data.

Chapter 2 shows definitions of the proposed depth functions and shows their properties. The affine invariant modification of proposed depths is defined. Chapter 3 proves the uniform strong consistency. Also the influence function is derived and the limit distribution is discussed.

In Chapter 4 we define the class of depth functions that provides us with a continuum between the halfspace depth and the kernel density estimate. Properties of this class are proved and also one possible characterization / interpretation of one member from this class is shown. This characterization can be used in econometric optimization.

Chapter 5 is not directly about data depth functions. It deals with topics that are directly connected to the data depth or where the data depth, the proposed depth in particular, can be used. In the first section the directional depth and directional quantiles are defined. In fact, the directional quantiles were one of our

main motivation to work on the proposed depth. The directional approach need a central point. This center should be well balanced - it should be well surrounded by observations. First we used the *Tukey median* (the deepest point with respect to the halfspace depth) to estimate the center. But for some data, especially for "banana" shaped data, this point fails to provide us with good results. There was a need to find a depth which provides us with the deepest point that more respect the shape of data. It seems that the proposed depth leads to a choice of a very natural center for directional quantiles. Other sections deal with a regression depth and a functional data depth. In both fields our proposed depth can be applied and it may bring some improvements.

The last chapter discusses the results and possible future development of proposed data depth.

## 1.2 General Definition of Data Depth

In the late nineties and the years that follows, statistical depth functions became increasingly used. Main reason is an increase of computational power - till nineties no big attention was put on the data depth concepts since their ideas were usually impossible to realize due weak computation power. Nowadays it is easy to obtain computer that can handle heavy computations that are usually needed to calculate a depth of one or more points in multivariate data space. So the data depth concepts have become little bit more widely spread. Many new definitions, concepts and application occurs in a great variety. Some statistical depth authorities (e.g. see [Zuo & Serfling, 2000a, Zuo & Serfling, 2000c]) noted that an exact definition of the statistical depth function is needed.

Since the depth provides us with *center-outward* ordering and the center or the deepest point plays relatively important role in the data depth, we should first note what *center* is. Not all classes of probability distributions posses naturally given central point - e.g. consider bivariate exponential distribution $Exp(\lambda_1) \times Exp(\lambda_2)$ for some $\lambda_1, \lambda_2 > 0$. But there exists a broad class of distributions where concept of the centres seems to be very natural and intuitive. It is a class of *symmetric* distributions with respect to some notion of symmetry. Let us mention some:

- The distribution is *elliptically symmetric* about a point $\boldsymbol{\xi} \in \mathbb{R}^p$ if its density $f$ can be written in the following form:

$$f(\boldsymbol{x}) = g\big((\boldsymbol{x}-\boldsymbol{\xi})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\xi})\big),$$

  where $g$ is a nonnegative function and $\boldsymbol{\Sigma}^{-1}$ is a definitely positive $p \times p$ matrix.

- The distribution of random vector $\boldsymbol{X}$ is *centrally symmetric* about a point $\boldsymbol{\xi}$ if the distribution of $\boldsymbol{X} - \boldsymbol{\xi}$ is the same as the distribution of $\boldsymbol{\xi} - \boldsymbol{X}$.

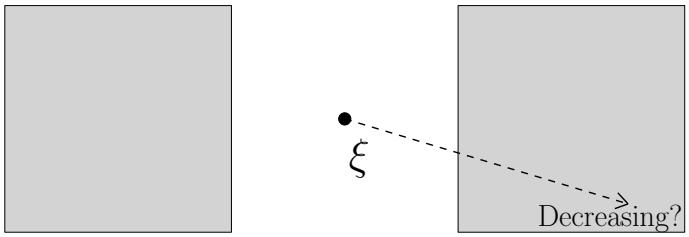



Figure 1.6: Are the key properties always desirable? The deepest point $\boldsymbol{\xi}$ of general depth function lies in the center but also outside the support. Since the depth should decreases along rays starting in $\boldsymbol{\xi}$ the points outside the support have higher depth values than the points well surrounded by data.

- The distribution of random vector $\boldsymbol{X}$ is *angularly symmetric* about a point $\boldsymbol{\xi}$ if the distribution of $(\boldsymbol{X}-\boldsymbol{\xi})/\|\boldsymbol{X}-\boldsymbol{\xi}\|$ is the same as the distribution of $(\boldsymbol{\xi}-\boldsymbol{X})/\|\boldsymbol{X}-\boldsymbol{\xi}\|$.

Now we present a general definition of depth function on $\mathbb{R}^p$ according to [Zuo & Serfling, 2000a] (Def. 2.1). Denote by $\mathcal{P}$ the class of distributions on the Borel sets of $\mathbb{R}^p$. Distribution of a given random vector $\boldsymbol{X}$ is denoted by $P_{\boldsymbol{X}}$.

**Definition 1** (Key properties of depth function)**.** Let the mapping $\mathrm{D}(\cdot;\cdot) : \mathbb{R}^p \times \mathcal{P} \longrightarrow \mathbb{R}$ be bounded, nonnegative, and satisfies:

1. $\mathrm{D}(\mathbf{A}\boldsymbol{x}+\boldsymbol{b}; P_{\mathbf{A}\boldsymbol{X}+\boldsymbol{b}}) = \mathrm{D}(\boldsymbol{x}; P_{\boldsymbol{X}})$ holds for any random vector $\boldsymbol{X} \in \mathbb{R}^p$, and $p \times p$ nonsingular matrix $\mathbf{A}$, and any vector $\boldsymbol{b} \in \mathbb{R}^p$.
2. $\mathrm{D}(\boldsymbol{\xi}; P) = \sup_{\boldsymbol{x}\in\mathbb{R}^p} \mathrm{D}(\boldsymbol{x}; P)$ for any $P \in \mathcal{P}$ having center $\boldsymbol{\xi}$ (with respect to at least one of latter notions of symmetry).
3. For any $P \in \mathcal{P}$ having the deepest point $\boldsymbol{\xi}$, $\mathrm{D}\big(\boldsymbol{\xi}+t(\boldsymbol{x}-\boldsymbol{\xi}); P\big) \geq \mathrm{D}(\boldsymbol{x}; P)$ holds for $t \in [0,1]$.
4. $\mathrm{D}(\boldsymbol{x}; P) \xrightarrow{\|\boldsymbol{x}\|\to+\infty} 0$ for any $P \in \mathcal{P}$.

Then we say that D satisfies the *key properties.*

The *deepest point* in latter definition is defined as follows.

**Definition 2** (The deepest point)**.** The *deepest point* with respect to a distribution $P$ and a depth function D is the point $\boldsymbol{\theta}$ such that

$$\mathrm{D}(\boldsymbol{\theta}; P) = \max_{\boldsymbol{x}} \mathrm{D}(\boldsymbol{x}; P).$$

If there exists more such points the deepest point is defined as their mean.

A function that satisfies properties of Definition 1 is what some statisticians call the *general depth function.* These "desirable" properties may be too strict in some cases and can also bring some controversies. Let us briefly discuss some:

1. Depth is an affine invariant function. Consider the situations on Fig. 1.2 and Fig. 1.3. Affine invariant depth functions say that king has the same protection in both cases.

2. Depth is maximal at the center of distribution. It is usually a good assumption unless we work with mixtures of distributions or with a distribution where no natural center exists. Consider situation on Fig. 1.6. $\boldsymbol{\xi}$ is the center of symmetry with respect to the all mentioned symmetry notions and it even does not lie inside the support of the distribution (shadow rectangles). Another example can be seen on Fig. 1.4 where no natural center exists.

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.* Again consider the case on Fig. 1.6. Closer to the support of the data we move, the depth decreases. Another example is shown on Fig. 1.4. In this case some of the depth functions (the halfspace depth, the Liu (simplicial) depth, ...) say that the king is almost in the deepest spot. Depth functions satisfying this non-increase property claim that more the king approaches his guards less protection he gets.

4. The depth function is vanishing at infinity. That is indeed the property that every depth should posses in any occasion.

Zuo and Serfling ([Zuo & Serfling, 2000a]) consider broad classes of depth functions and study the possession of these key properties. Properties 1–3 are all connected with the *convexity* of sets of points whose depth is higher than a given value. Formal definitions of such regions is following.

**Definition 3.** Suppose $d > 0$. The *d-central region* of a depth function D and a distribution $P$ is the set

$$\mathcal{C}(d) = \{\boldsymbol{x} : \ \mathrm{D}(\boldsymbol{x}; P) \geq d\}.$$

In [Zuo & Serfling, 2000c] these regions are called *d-trimmed regions.* Usually using a depth function that satisfies key properties 1–3 leads to fact that *central regions* are convex (or near to convex) and it happens even in cases when one can intuitively expects another shape of such regions. Therefore constructions of more general central regions may be needed. Besides the well known *level sets* of the probability density function, [DasGupta et al., 1995] considered a general family of star-shaped sets (for definition see Def. 16). The "best" shape of central regions is proposed and it is then inflated (or deflated) in order to obtain the central region of given probability. The idea behind this approach is substantially different from the depth based approach.

But, as we will see, 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 depth functions (especially the halfspace depth function) satisfying key properties may be disclosed. It is the main motivation for us to propose a larger class of depth functions derived from the halfspace depth.

There also has been particular interest in the deepest point (see Def. 2). This is usually used as a measure of position of the data. The sample deepest point is often used as a robust estimate of the mean value. It plays similar role as median in univariate data analysis. So there is a need for the deepest point whose position respects the geometry of the data. As we have already seen and we will see some of the depth functions (e.g. the halfspace depth, the Liu/simplicial depth) that satisfy the properties of Def. 1 provide us with the deepest point that may lie quite near the border (or even outside the border) of the support of the distribution.

## 1.3 The Halfspace Depth

One of the most popular, the most known and surely the most used depth functions is the halfspace depth, defined by [Tukey, 1974]. [Donoho, 1982, Donoho & Gasko, 1992] have studied its breakdown properties. The computational aspect may be found, for example, in [Rousseeuw & Ruts, 1996] and [Struyf & Rousseeuw, 2000], [Matoušek, 1992] has proposed fast algorithm for computing the deepest point of random sample. See also [Zuo & Serfling, 2000c] for broad discussion on features of data depth and, in particular, of the halfspace depth.

We can calculate the halfspace depth of a point using R package depth (function depth). Besides the halfspace depth the package contains functions for calculating the *Liu* (*simplicial*) and the *Oja depth*.

**Definition 4.** Let $P$ be a probability measure on $\mathbb{R}^p$. The halfspace depth of a point $\boldsymbol{\theta} \in \mathbb{R}^p$ is defined as

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

In other words, the halfspace depth of $\boldsymbol{x}$ is the infimum of probability mass of all closed halfspaces whose border includes $\boldsymbol{\theta}$. Figure 1.7 illustrates this definition. The depth is assigned according to the probability in the "worst direction". In case of elliptically symmetric distributions the minimum probability halfspace is the halfspace whose border is tangent to an elliptic contour. There needs not to exist unique minimum probability halfspace. For instance, consider case of multivariate normal distribution and its mean. All the halfspaces with border going through mean attain probability 1/2.

The halfspace depth is well defined for all $\boldsymbol{\theta} \in \mathbb{R}^p$. The empirical (sample version of) halfspace depth $\mathrm{HD}_n(\boldsymbol{\theta})$ for a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of the distribution $P$ is defined as a halfspace depth for the empirical probability measure $P_n$, i.e.

$$\mathrm{HD}_n(\boldsymbol{\theta}) = \inf_{\|\boldsymbol{u}\|=1} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{\boldsymbol{u}^T\boldsymbol{X}_i \geq \boldsymbol{u}^T\boldsymbol{\theta}\right\}.$$

This definition is illustrated on Fig. 1.8. We simply calculate the number of points lying in each halfspace and as the depth we take proportion of observations lying

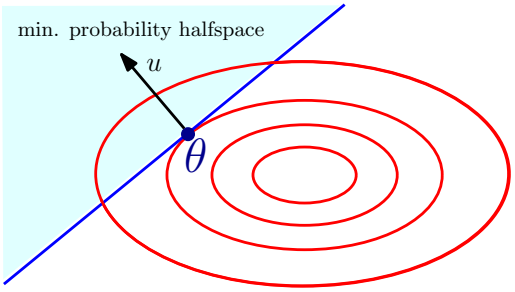



Figure 1.7: The halfspace depth of a point $\boldsymbol{\theta}$ is the probability of minimum halfspace that contains this point. In the case of elliptically symmetric distribution it is the halfspace whose border is tangent to an elliptical contour.

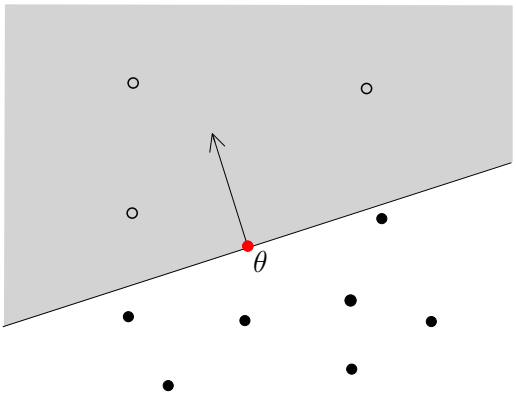



Figure 1.8: The sample halfspace depth of a point $\boldsymbol{\theta}$ is the minimum number of points lying in halfspace that contains point $\boldsymbol{\theta}$ divided by size of the sample. In this case $n = 10$. $\mathrm{HD}_n(\boldsymbol{\theta}) = 3/10$.

in a halfspace with the smallest number of observations. The minimum empirical probability halfspace needs not to be unique.

The definitions of the halfspace depth and its empirical version may be, using *inner product* notation, equivalently rewritten as

$$\begin{aligned} \mathrm{HD}(\boldsymbol{\theta}) &= \inf_{\|\boldsymbol{u}\|=1} \mathsf{E}\, \mathbb{1}\left\{\langle \boldsymbol{X} - \boldsymbol{\theta}, \boldsymbol{u}\rangle \geq 0\right\}, \\ \mathrm{HD}_n(\boldsymbol{\theta}) &= \inf_{\|\boldsymbol{u}\|=1} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{\langle \boldsymbol{X}_i - \boldsymbol{\theta}, \boldsymbol{u}\rangle \geq 0\right\}. \end{aligned} \tag{1.1}$$

All mentioned definitions are very intuitive and easily interpretable. Moreover, there are many properties of the halfspace depth which made this depth popular and widely used. Besides the *key properties* from Definition 1 additional properties can be proved:

1. The *d-central region*, $\{\boldsymbol{x} : \mathrm{HD}(\boldsymbol{x}) \geq d\}$, 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).

2. 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).

3. The deepest point is unique for absolutely continuous distributions.

4. Under mild conditions the distribution of $\sqrt{n}\big(\mathrm{HD}(\boldsymbol{\theta}; P_\mathrm{n}) - \mathrm{HD}(\boldsymbol{\theta}; P)\big)$ converges to normal distribution if $\boldsymbol{\theta}$ is not a center of symmetry (see [Massé, 2004]).

5. The breakdown point of the *halfspace depth median* (the deepest point) is at least $1/(p+1)$.

Figures 1.10, 1.13 and 1.16 show *halfspace depth contours* (borders of central regions from Def. 3). More info about datasets used in this figures can be found in Example 1, 2 and 3. Another example can be seen on Fig. 4.6. We can see that the shape of central regions may sometimes differ from our expectation. Especially in the case of distribution similar to bivariate exponential (see Example 2) one would like to obtain more triangle shaped contours. [Dutta et al., 2011] makes very detailed review of some counterintuitive properties of the halfspace depth. It seems that the halfspace depth is not compatible with nonconvexly supported distributions. In application it can lead to some undesirable consequences (e.g. in multivariate ordering a point outside the data can have same order as a point well surrounded by the data). This is why some researches try to improve this depth functions - to keep its good properties and to make it compatible with noncovexly supported data. One of the possibilities is *localization* - see [Agostinelli & Romanazzi, 2011] and [Paindaveine & Van bever, 2013] or Subsection 1.4.9 for short summary of this concept. Other possibility is proposed in this thesis. It brings another generalization of the halfspace depth that can more reflects the shape of the data.

Besides the well known halfspace depth there exists a lot of other depth functions. Next section make a short overview of them.

## 1.4 Data Depth - Overview of Some Existing Depth Functions

A lot of data depth related results has appeared in the last two decades. Unfortunately there does not exist a comprehensive overview of existing depth functions and their properties. Although some summary results can be found in [Liu et al., 1999], [Zuo & Serfling, 2000a] and [Liu et al., 2003]. In-depth overview of existing depth functions and their properties could fulfil a whole single thesis. Hence this section shows only brief overview. Instead of a list of depth functions' properties we show graphical results of application of all depth functions mentioned in this section to 3 different datasets. See Figures 1.10 – 1.18 which show 20%, 40%, 60% and 80% sample depth contours. With $\tau \cdot 100\%$ depth contour we mean the border of the central set $\mathcal{C}(d)$ (see Definition 3) such

that

$$P\big(\mathcal{C}(d)\big) = \tau.$$

For $\tau$ depth contours with higher values of $\tau$ we should be aware of so called *course of dimensionality* (see e.g. [Hastie et al., 2009], Chapter 1, Section 2.5). In multidimensional space many points lie on the border of data cloud. For instance, consider uniform distribution on unit sphere in dimension 2. About 25% of observations lie in distance greater than 0.87 from the center. Compare it to dimension 1 where this value is 0.75. A quarter of observations lie in a relatively narrow band near the border of the support. This is why the 80% depth contour is often very rugged.

In the following paragraphs the depth functions are defined for an absolutely continuous probability measure $P$ on $\mathbb{R}^p$ and for a point $\boldsymbol{\theta} \in \mathbb{R}^p$. Functions are listed in random order. Some figures show contours that are expected to be smooth but that are little bit rugged (e.g. in the case of the Liu depth). It is usually consequence of the fact that contours are interpolated from a grid points whose depth is calculated (ggplot with geom_contour from excellent R library ggplot2 is used for interpolation - see [Wickham, 2009]).

### 1.4.1 The Mahalanobis and the Euclidean Depth

The Mahalanobis depth is one of the oldest depth function. It is derived from the well known *Mahalanobis distance* (see [Mahalanobis, 1936]) and it is still often applied in simple classification and clustering problems, psychology, data visualizations etc. It is strongly connected with the *elliptical symmetry* and its use for non-elliptically symmetric distributions can lead to distorted results. Central regions $\mathcal{C}(d)$ are always ellipsoids. The Mahalanobis depth of a point $\boldsymbol{\theta} \in \mathbb{R}^p$ is defined as

$$\mathrm{MHD}(\boldsymbol{\theta}) = \Big(1 + (\boldsymbol{\theta} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu})\Big)^{-1},$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix of distribution $P$. If we replace these characteristics by their sample version we obtain the sample version of the Mahalanobis depth.

If no covariance structure of data is considered we obtain the *Euclidean depth*. Its center regions are spheres. It is defined as

$$\mathrm{ED}(\boldsymbol{\theta}) = \Big(1 + \|\boldsymbol{\theta} - \boldsymbol{\mu}\|\Big)^{-1}.$$

The Mahalanobis depth satisfies the *key properties* (see Definition 1). Together with Euclidean depth it can can be calculated using depth function from R library depthproc. This function is very fast and it also can be set as hyper-threaded. Untill now (july 2014) it has not been finished yet but its beta version can be installed from R-Forge. See Figures 1.11, 1.14 and 1.17 for examples of these depth functions. For alpha vs. gamma intakes (Fig. 1.17) the data were first standardized before depth computations.

### 1.4.2 The Liu (Simplicial) Depth

The Liu depth is one of the most known depth function. It is also called the simplicial depth. It is defined in a relatively intuitive way. First it appeared in [Liu, 1990]. The definition is as follows

$$\mathrm{SD}(\boldsymbol{\theta}) = \mathsf{P}\big(\boldsymbol{\theta} \in S[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{p+1}]\big),$$

where $S[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{p+1}]$ is a closed simplex formed by $p+1$ independent random variables on $\mathbb{R}^p$ with distribution $P$. Hence the more probability mass is distributed around the point the deeper the point lies. This depth function satisfies the *key properties* from Definition 1 (see [Liu, 1990, Zuo & Serfling, 2000a]). The sample version is defined as

$$\mathrm{SD}_n(\boldsymbol{\theta}) = \binom{n}{p+1}^{-1} \sum_{1 \le i_1 < \ldots < i_{p+1} \le n} \mathbb{1}\left\{\boldsymbol{\theta} \in S[\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_{p+1}}]\right\}.$$

The sample deepest point need not to be unique. An improved definition, that leads to unique sample deepest point for discrete distributions, was introduced in [Burr et al., 2003]. This depth function can be calculated using depth function from R library depth. Computation is relatively fast. For dataset from Example 3 the computation does not lead to sensible results. Hence a graphical presentation for this data is not shown. Contours for other 2 datasets can be seen on Figure 1.10 and 1.13. Contours should be always convex. The discrepancy from convexity are most likely due the interpolation of contours from a set of grid points and also because of some numerical instabilities of the Liu depth computations using function depth.

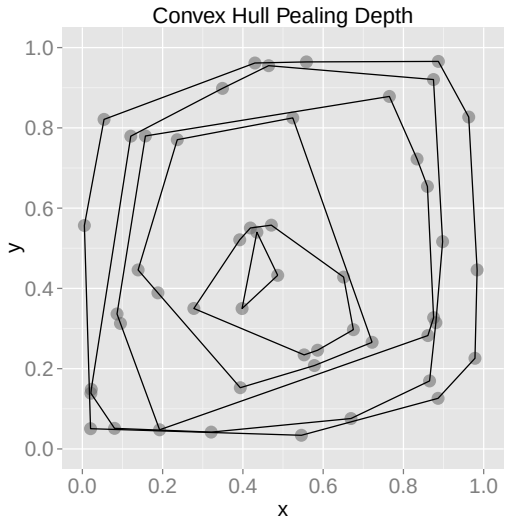



Figure 1.9: The convex hull peeling depth.

### 1.4.3 The Convex Hull Peeling Depth

The convex hull peeling depth was introduced in [Barnett, 1976]. Only population version was defined. It is one of the most intuitive depth functions - we peel a layer by layer of data cloud until we get to the deepest spot. Layers consist of borders of convex hulls of remaining points after previous layers were removed. Depth is assigned according to how deep the layer is. See Fig. 1.9 for an illustration of this principle, where convex hulls are plotted for the uniform distribution on $[0,1] \times [0,1]$. Formal definition of sample version can be found in [Barnett, 1976] or [Porzio & Ragozini, 2007]. Last cited article also shows definition of the population version of the depth - authors call it *convex hull probability depth*. Authors further proved that population version satisfies *key properties*. There are many functions in R that can be used for fast computation of convex hulls of a set of points, e.g. see libraries grDevices (chull function), geometry (convhulln function) and spatstat (function convexhull.xy). Figures 1.10, 1.13 and 1.16 show contours for the convex hull pealing depth.

### 1.4.4 The Oja Depth

[Oja, 1983] defined the *Oja depth* as follows.

$$\mathrm{OD}(\boldsymbol{\theta}) = \Big(1 + \mathsf{E}\,\mathrm{volume}\big(S[\boldsymbol{\theta}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_p]\big)\Big)^{-1},$$

where $S[\boldsymbol{\theta}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_d]$ is, similarly as in Subsection 1.4.2, the closed simplex with vertices $\boldsymbol{\theta}$, and $p$ random observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p$. The sample version is defined in the same way as the sample version of Liu depth. The central regions of the Oja depth tend to be elliptic with no respect to the geometry of the data - see Figures 1.10, 1.13 and 1.16 for examples. For alpha vs. gamma intakes (Fig. 1.16) the data were first standardized before depth computations. This depth has never been widely used. The Oja depth is not affine invariant but an affine invariant modification exists (see [Zuo & Serfling, 2000a]). It can be calculated using depth function from R library depth. Computation lasts longer than computation for other depth functions from this library.

### 1.4.5 The Zonoid Depth

It was introduced in [Koshevoy & Mosler, 1997]. The definition is a little bit complicated and not so intuitive as previous depth functions' definitions. Consider a family $D_\alpha$ of nested sets in $\mathbb{R}^p$. Suppose that expectation $\mathsf{E}\,\boldsymbol{X}$ is finite. The sets are called the *zonoid regions* if $D_0 = \mathbb{R}^p$ and for $\alpha \in (0,1]$

$$D_\alpha = \left\{ \mathsf{E}\,\boldsymbol{X} g(\boldsymbol{X}) : \ g : \mathbb{R}^p \to \left[0, \frac{1}{\alpha}\right], \mathsf{E}\, g(\boldsymbol{X}) = 1 \right\}.$$

The zonoid depth is defined as

$$\mathrm{ZD}(\boldsymbol{\theta}) = \max\{\alpha : \ \boldsymbol{\theta} \in D_\alpha\}.$$

The sample version is obtained if the empirical expectation $\mathsf{E}_n$ is used in the definition of $D_\alpha$. Instead of functions it suffices to use constants $0 \leq \lambda_i \leq 1/\alpha,\ i = 1, \ldots, n$. The deepest point is always the mean $\mathsf{E}\,\boldsymbol{X}$. If the mean is also the center of a symmetry then the zonoid depth satisfies the *key properties* from Definition 1. This depth function is not robust. If the depth is applied to non-symmetric distributions central regions sometimes tend to be stretched to the direction where no big density of observations occurs - see Figures 1.10, 1.13 and 1.16. This depth can be computed using depth.zonoid function from R ddalpha library.

### 1.4.6 The Projection Depth

Definition of the projection depth is relatively straightforward. *Projection outlyingness* of a point $\boldsymbol{\theta}$ is defined as

$$\mathrm{PO}(\boldsymbol{\theta}) = \sup_{\boldsymbol{u}:\ \|\boldsymbol{u}\|=1} \frac{\left|\boldsymbol{u}^T\boldsymbol{\theta} - \mathrm{Med}(\boldsymbol{u}^T\boldsymbol{X})\right|}{\mathrm{MAD}(\boldsymbol{u}^T\boldsymbol{X})},$$

where $\mathrm{Med}(Y)$ and $\mathrm{MAD}(Y)$ denote univariate median and mean absolute deviation of a random variable $Y$ ($\mathrm{MAD}(Y) = \mathrm{Med}\big|(Y - \mathrm{Med}(Y))\big|$). The projection depth can be now defined as

$$\mathrm{PD}(\boldsymbol{\theta}) = \big(1 + \mathrm{PO}(\boldsymbol{\theta})\big)^{-1}.$$

This depth function satisfies the *key properties* - see [Zuo & Serfling, 2000a]. The sample version is obtained if we replace population statistics with their random sample counterparts. Projection to lines has some (geometric) information loss. As it can be seen in Figures 1.10, 1.13 and 1.16 the projection depth central regions tend to stretch to places where no points occur. In contrary to concluding remarks in [Zuo & Serfling, 2000a] the author of the thesis does not recommend to use this depth function. [Zuo & Serfling, 2000a] claims that this depth has an advantage, as an implementation of the *projection pursuit method*, to powerfully extract information from the data. Projection pursuit method is indeed very powerful method but its power lies more in application of so called ridge functions on the projections - see e.g. [Hastie et al., 2009] for more information. The projection depth can be calculated with depth function from R depthproc library. See Subsection 1.4.1 for more details about this library.

### 1.4.7 The $L^q$ Depth

The $L^q$, $q > 0$, norm can serve as a distance (or outlyingness) measure. The $L^q$ norm we denote by $\|\cdot\|_q$. It holds $\|\boldsymbol{x}\|_q = (|x_1|^q + \ldots + |x_p|^q)^{1/q}$. Using this distance measure the $L^q$ depth function is defined as

$$\mathrm{L^qD}(\boldsymbol{\theta}) = \big(1 + \mathsf{E}\|\boldsymbol{\theta} - \boldsymbol{X}\|_q\big)^{-1}.$$

The depth is not affine invariant (see [Zuo & Serfling, 2000a]) hence it does not satisfy all the properties from Def. 1. The sample version is obtained if we replace $\mathsf{E}$ with $\mathsf{E}_n$. Figures 1.11, 1.14 and 1.17 show contours for this depth if $q = 2$. In this case the deepest point is so called *spatial median*. For alpha vs. gamma intakes (Fig. 1.17) the data were standardized before the depth was calculated. The central regions of the $L^q$ depth have tendency to be in spherical shape (if the scale of all coordinates is the same) even if the "shape" of the data is very different. That's why the author does not recommend to use this popular depth function. The depth can be computed with depth function from R depthproc library. See Subsection 1.4.1 for more details about this library.

### 1.4.8 The Likelihood Depth

This depth is very simple. Suppose that $f$ is the probability density of $P$, then

$$\mathrm{LD}(\boldsymbol{\theta}) = f(\boldsymbol{\theta})$$

is the likelihood depth. This depth was first considered in [Fraiman et al., 1999]. The deepest point is the *modus*. It is clear that this depth does not satisfy the *key properties*. The sample version is obtained if any of density estimates is used, for example the *kernel density estimate*. Central regions $\mathcal{C}(d)$ are called *levelsets* and have the smallest volume possible. Some authorities (e.g. [Zuo & Serfling, 2000a]) object against consider LD as a depth function because it reflects more "local" than "global" property of the data. Figures 1.11, 1.14 and 1.17 show contours for this depth. For alpha vs. gamma intakes the data were first standardized before the depth was computed. Function kde from R library ks was used for computation. Automatic selection of bandwidth was used.

### 1.4.9 Other Depth Functions

Of course we do not make complete survey of all existing depth notions. Besides other we did not mention the *majority depth*, which had not been widely used because no fast algorithm existed. Fast and effective calculation of this depth was recently investigated in [Chen & Morin, 2013], where also definition can be found.

Another not mentioned depth notion is *localized depth*. Since majority of depth functions produce convex central regions even in the case of multimodal distribution some researchers tried to find a connection between local and global view of the data. Paper [Agostinelli & Romanazzi, 2011] proposes simple localization of the halfspace and Liu depth. For former they consider only finite width slabs instead of halfspaces, for latter they consider only simplices with smaller volume than some fixed threshold. Implementation of this approach can be found in R library localdepth. [Paindaveine & Van bever, 2013] shows another concept of localization - they claim that they introduce concept that makes data depth compatible with multimodal or non-convexly supported distributions. Suppose we want to calculate the depth of a point $\boldsymbol{\theta}$. First we need symmetrize the dis-

tribution $P$ about $\boldsymbol{\theta}$, $P_{\text{sym}} = 1/2\, P_{\boldsymbol{X}} + 1/2\, P_{2\boldsymbol{\theta}-\boldsymbol{X}}$ . Then for any depth function satisfying key properties 2. and 3. (see Def. 1) we consider the depth with respect to the conditional symmetrized probability measure to the central regions with a given confidence level. Smaller the confidence level is the higher degree of localization occurs. Formal definition can be seen in the article. As a disadvantage of majority of depth function dealing with non-convexity (including the generalized halfspace depth and the weighted halfspaces ratio depth introduced in this thesis) authors point out that in extreme localization the depth usually converges to a density or to a constant. Their localized depth does not posses this property.

### 1.4.10 Examples

This subsection graphically illustrates properties of majority of the depth functions mentioned in the thesis. We used 3 datasets described in Examples 1 - 3. See also introduction of Section 1.4 for more information. The figures titled *Band Depth, Parabolic Depth, Kernel Parabolic Depth* and *Kernel Hyperbolic Depth* show contours for the depth functions proposed later in the thesis (Chapter 2 and Chapter 4).

**Example 1** (A mixture of 2 Gaussian distributions)**.** This example illustrates how data depth functions deal with bimodal non-convexly shaped data, i.e. with data where is hard to measure centrality. The data (sample size is 1000) consists of 50%:50% mixture of two Gaussian distributions,

$$\mathcal{N}\left(\begin{pmatrix}0\\6\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 9\end{pmatrix}\right), \mathcal{N}\left(\begin{pmatrix}6\\0\end{pmatrix}, \begin{pmatrix}9 & 0\\0 & 1\end{pmatrix}\right).$$

See Fig. 1.10 - 1.12.

**Example 2** (Exponential distribution)**.** It is almost classical data depth example. Fig. 1.13 - 1.15 show sample depths contours for a "bivariate exponential distribution". The dataset consists of 1000 simulated bivariate observations from $Exp(1) \times Exp(1)$ distribution.

**Example 3** (Long lived radionuclides intake vs. gamma intake)**.** This dataset consists of 10503 observations of long lived radionuclides (in Bq) and gamma (in mSv) monthly intakes for Rožná mine miners during 2003 - 2012. Data were collected from personal dosimeters. One would expect linear relationship between these two variables. Unfortunately there exist many unobservable and unknown variables and mechanisms that affect this relation. Consequence is that both variables are not strongly correlated. If the depth function is not affine invariant, the data were first scaled to have unit variance. Then the depth function was applied and finally the result was retransformed back to the original scale. See Figures 1.16 - 1.18.

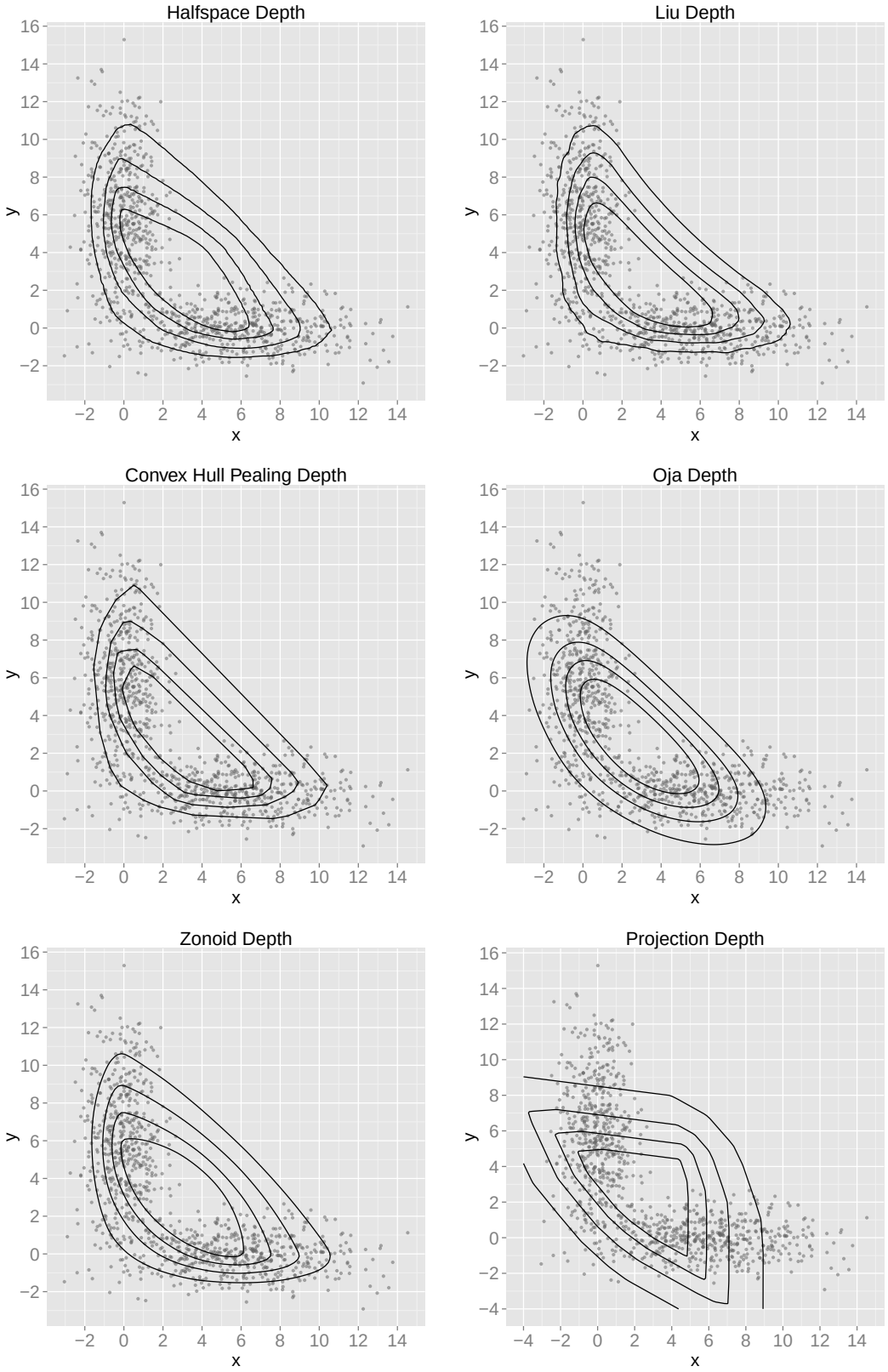



Figure 1.10: Depth functions - contours (20%, 40%, 60% and 80%) for a mixture of 2 Gaussian distributions.

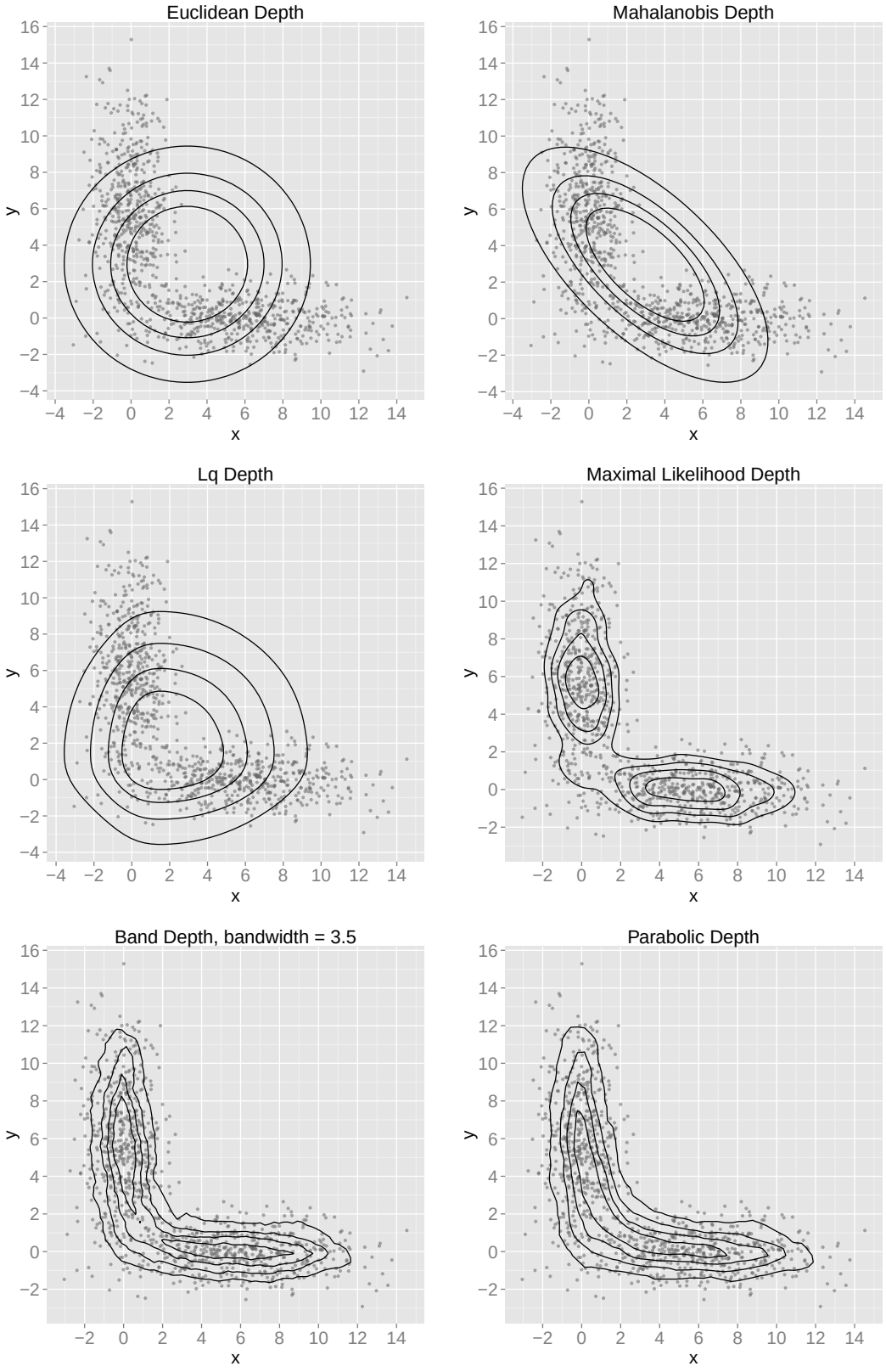



Figure 1.11: Depth functions - contours (20%, 40%, 60% and 80%) for a mixture of 2 Gaussian distributions.

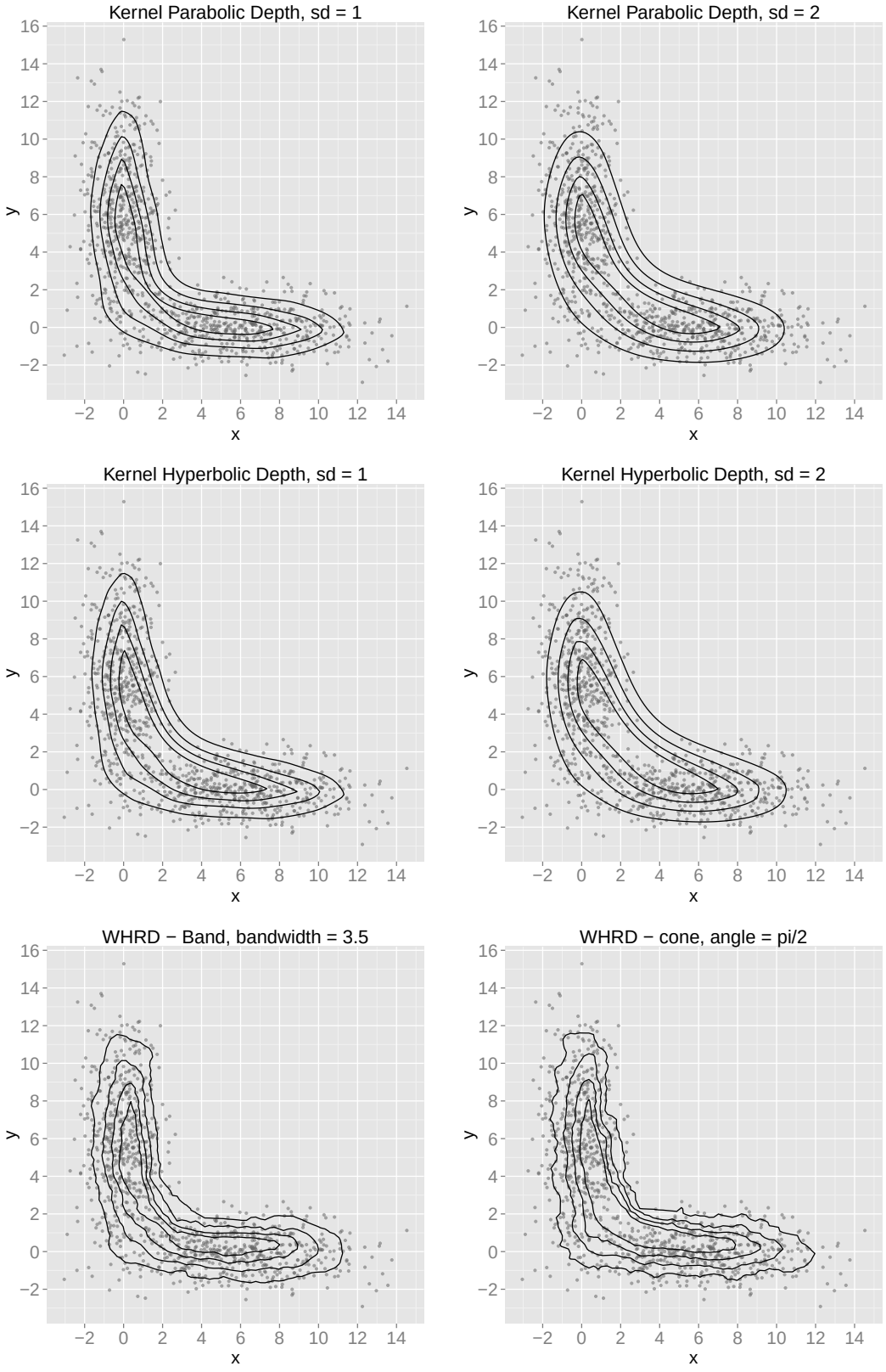



Figure 1.12: Depth functions - contours (20%, 40%, 60% and 80%) for a mixture of 2 Gaussian distributions.

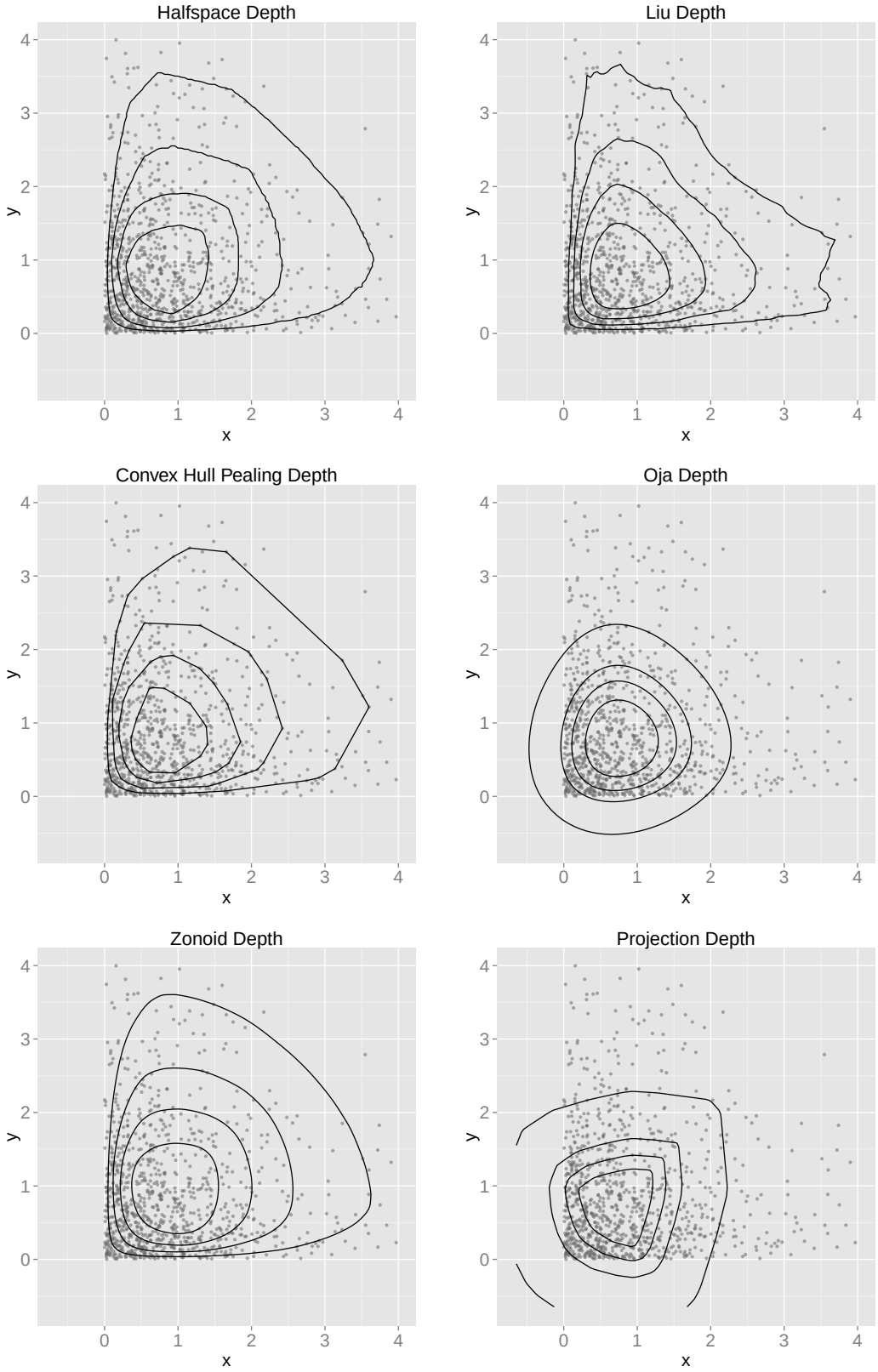



Figure 1.13: Depth functions - contours (20%, 40%, 60% and 80%) of exponential distribution.

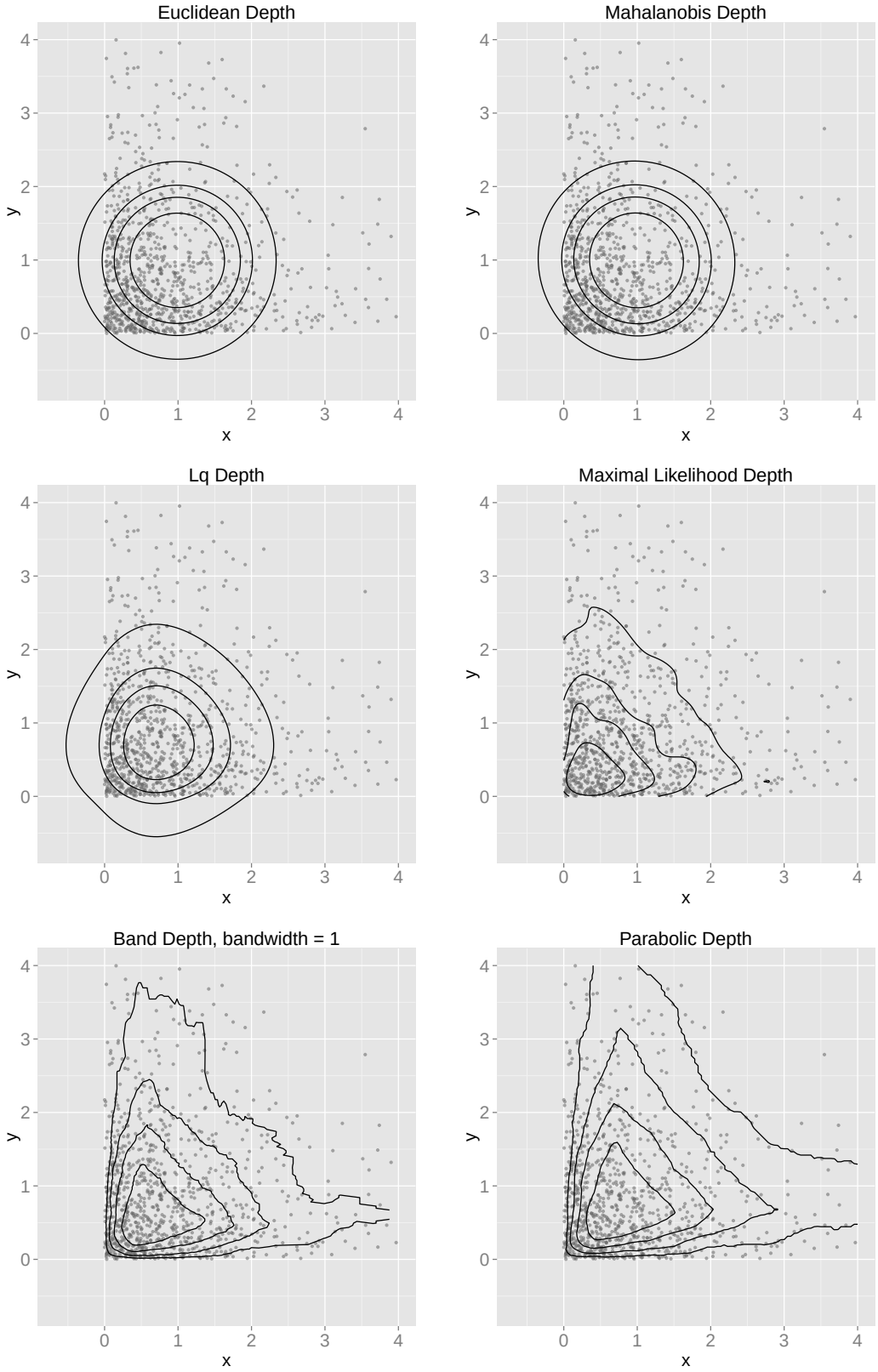



Figure 1.14: Depth functions - contours (20%, 40%, 60% and 80%) of exponential distribution.

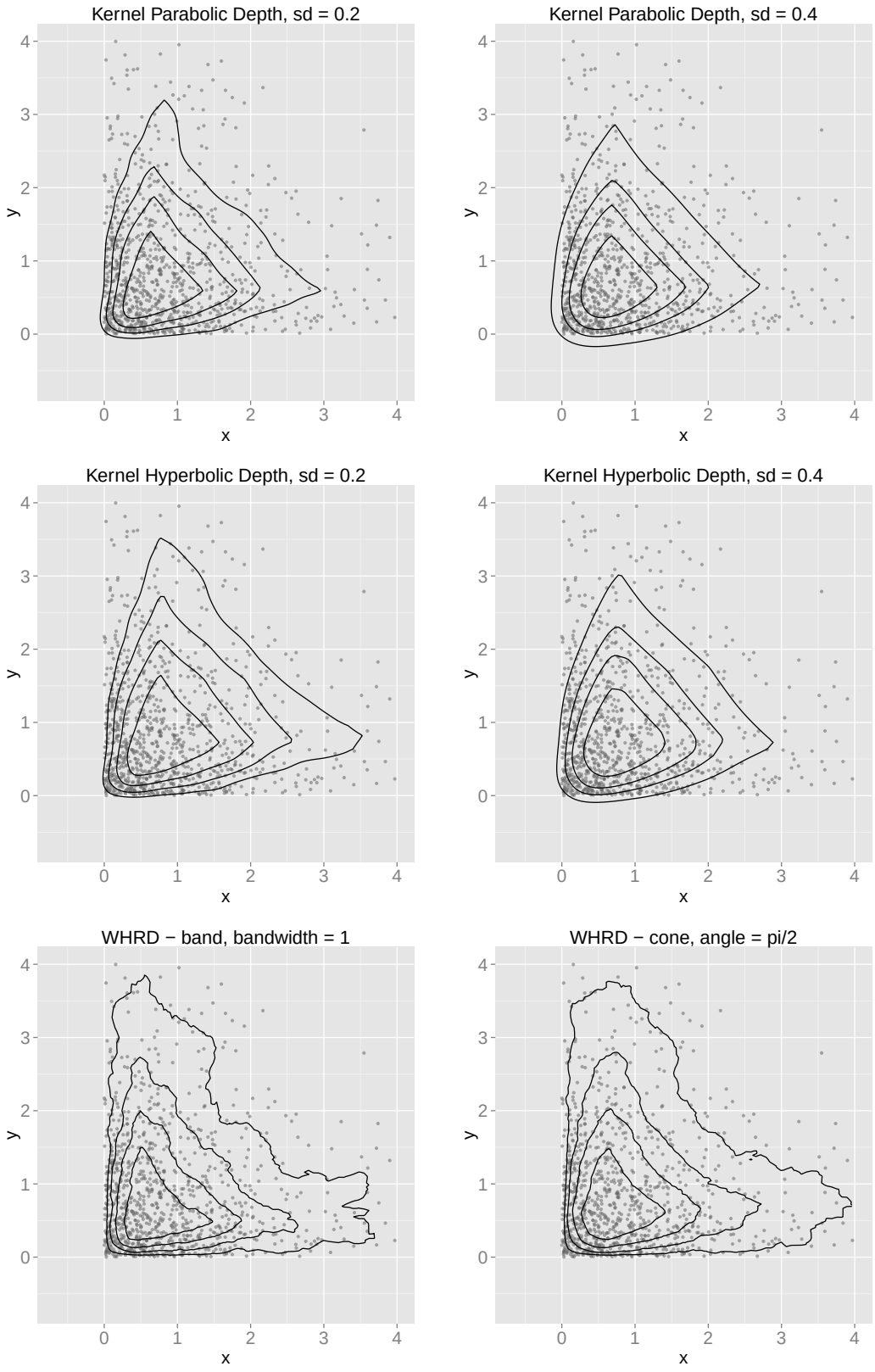



Figure 1.15: Depth functions - contours (20%, 40%, 60% and 80%) of exponential distribution.

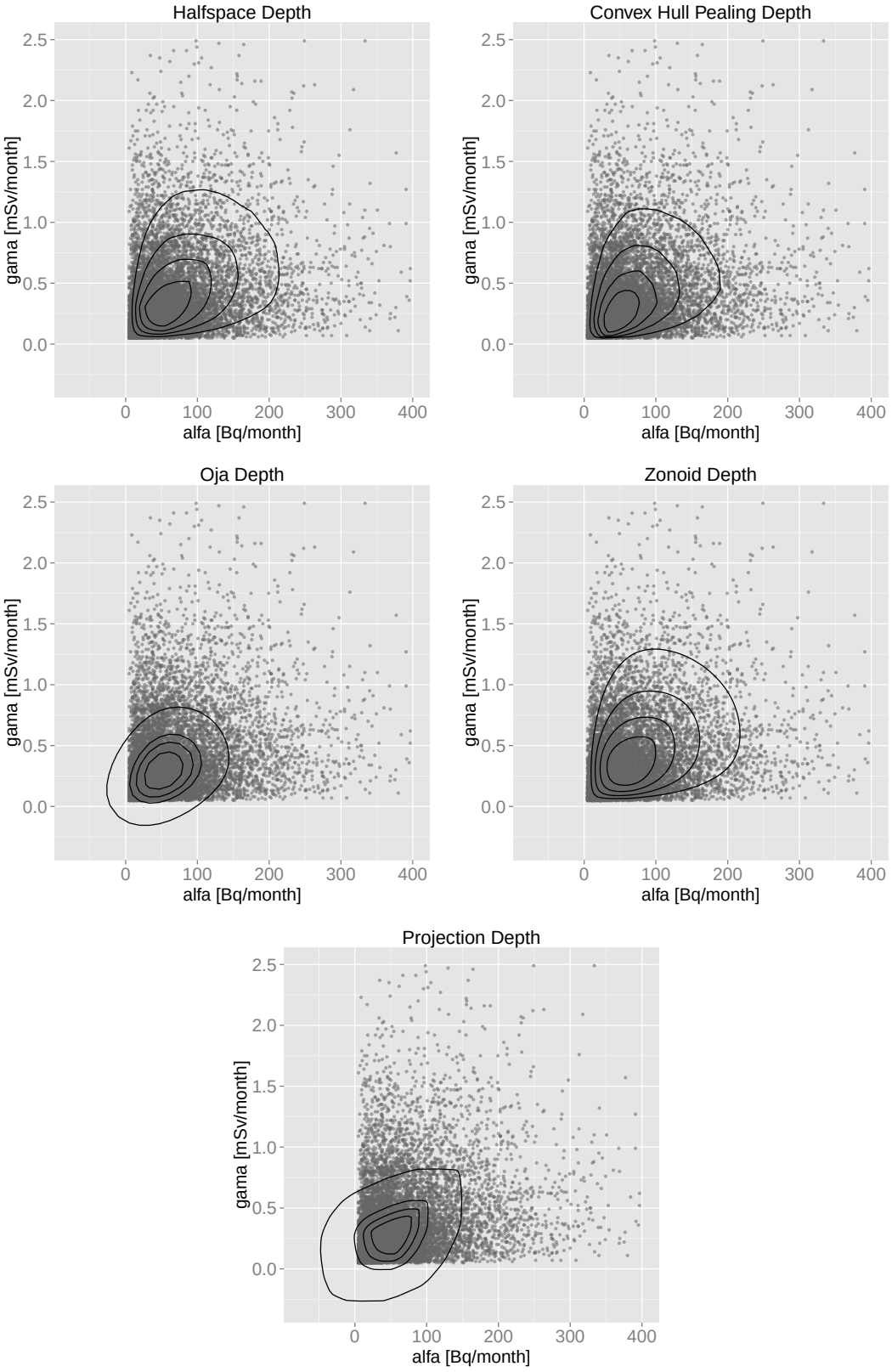



Figure 1.16: Depth functions - contours (20%, 40%, 60% and 80%) of alpha and gamma intakes.

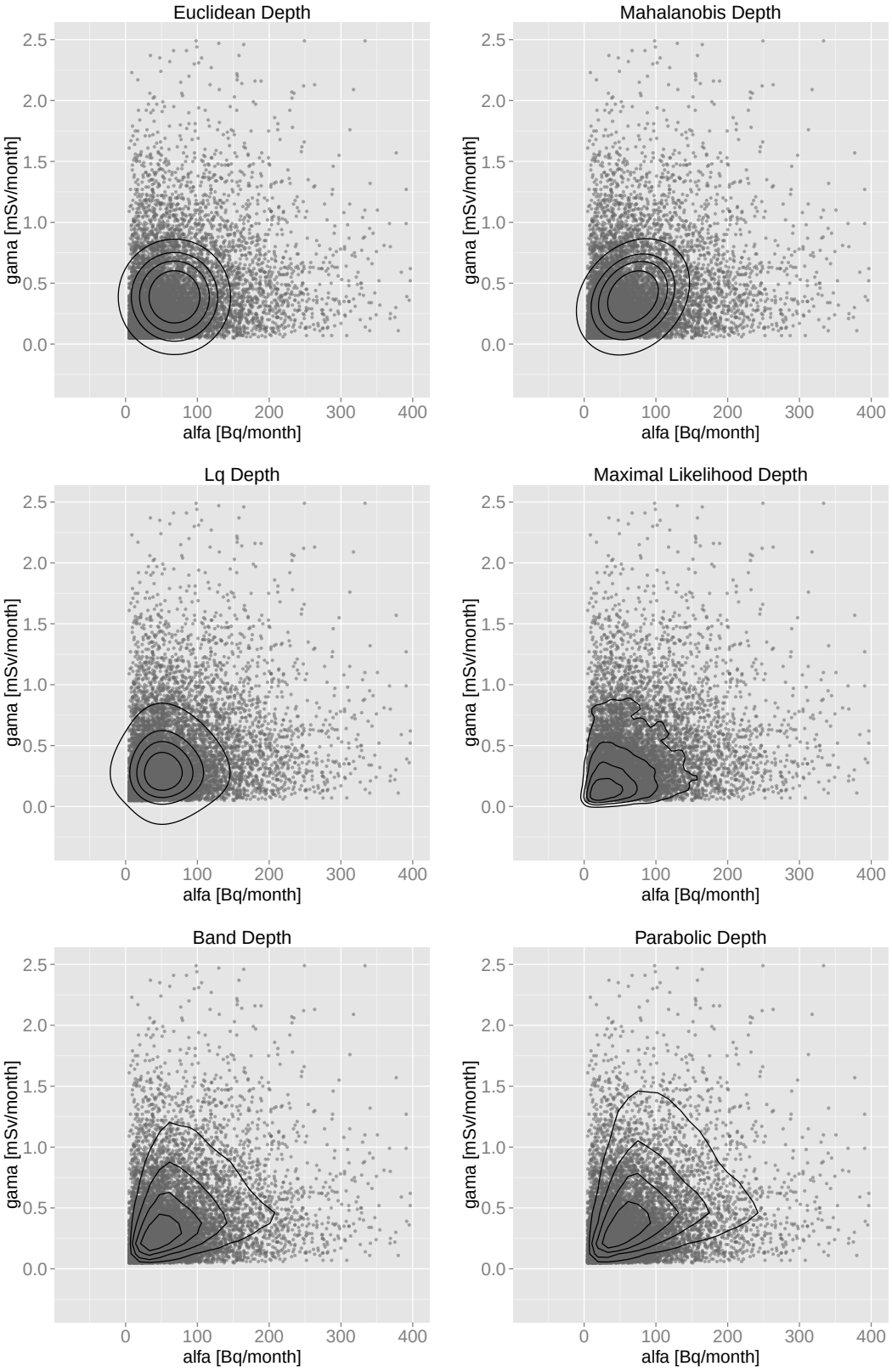



Figure 1.17: Depth functions - contours (20%, 40%, 60% and 80%) of alpha and gamma intakes.

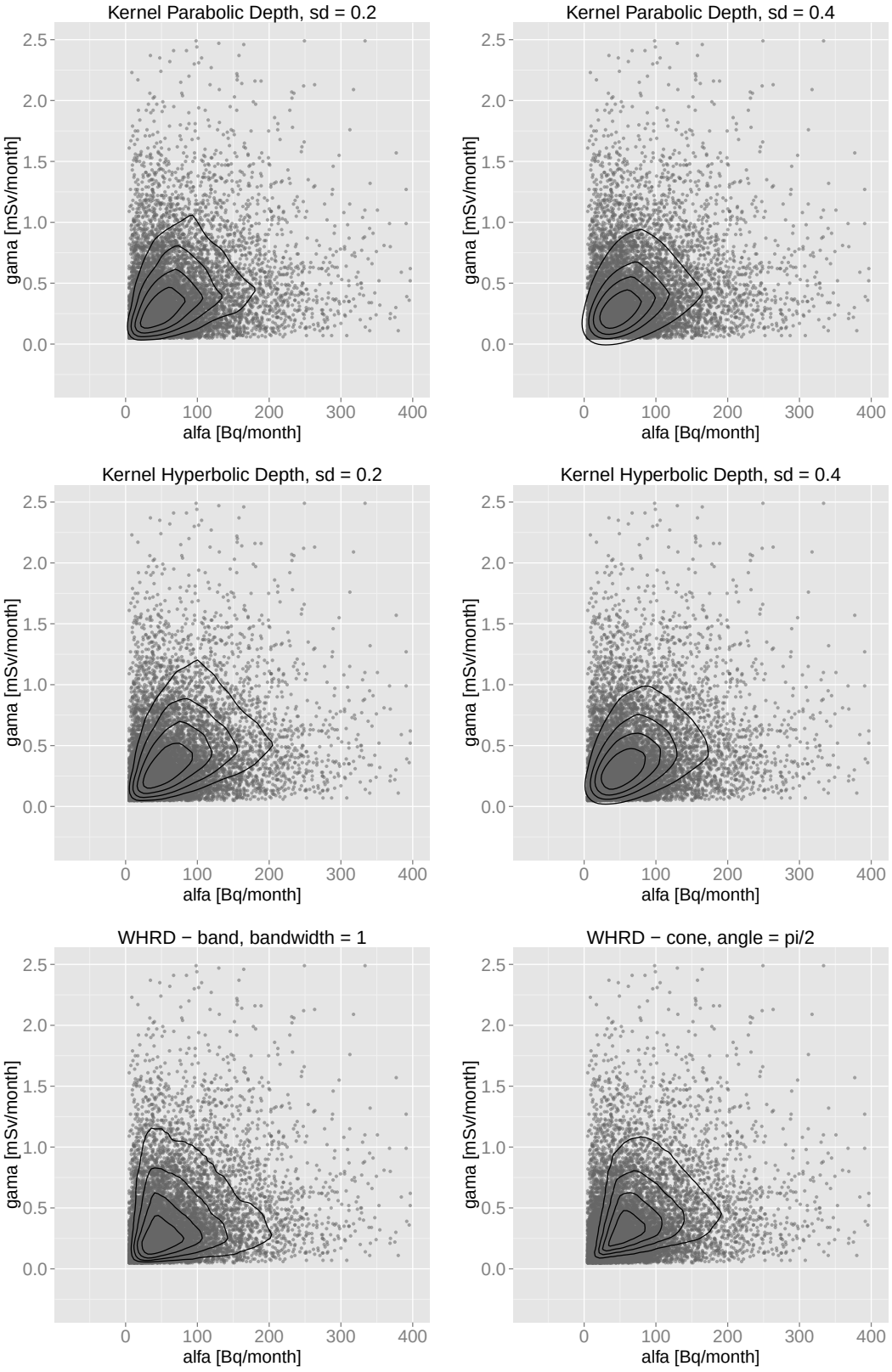



Figure 1.18: Depth functions - contours (20%, 40%, 60% and 80%) of alpha and gamma intakes.

# Chapter 2

# Weighted Depths

This chapter introduces two classes of depth functions that have their origin in the *halfspace depth*. We begin with a straightforward and very natural generalization of the halfspace depth. Some of the results also appeared in [Hlubinka et al., 2010].

## 2.1 Definition of the Generalized Halfspace Depth

Definition 4 may be immediately suggesting to replace the halfspace with *affine boundary* by another "halfspace" with a *more general boundary* (or even without boundary); then it is quite natural to do the next step and to replace the *indicator* in (1.1) by a more general *weight function*. This directly leads us to the definition of the *generalized halfspace depth*.

**Definition 5** (Generalized halfspace depth)**.** Suppose a measurable bounded weight function

$$w : \mathbb{R}^p \times \mathcal{S}^p \longrightarrow \mathbb{R}. \tag{2.1}$$

The *generalized halfspace depth* with respect to the weight function $w$ and probability distribution $P$ of a random vector $\boldsymbol{X} \in \mathbb{R}^p$ of a point $\boldsymbol{x} \in \mathbb{R}^p$ is defined as

$$\mathrm{D}_w(\boldsymbol{x}; P) = \inf_{\|\boldsymbol{u}\|=1} \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) = \inf_{\|\boldsymbol{u}\|=1} \int_{\mathbb{R}^p} w(\boldsymbol{y} - \boldsymbol{x}, \boldsymbol{u}) \ \mathrm{d}P(\boldsymbol{y}). \tag{2.2}$$

The *sample version* is defined as

$$\mathrm{D}_{w,n}(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \frac{1}{n} \sum_{i=1}^{n} w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}).$$

*Note* 1. We remind that if the probability distribution $P$ and the weight function $w$ are clear from the context they will be often omitted from the notation and we will simply write $\mathrm{D}(\boldsymbol{x})$ for a generalized halfspace depth of a point $\boldsymbol{x}$. Further if the weight function depends on some parameter of interest we sometimes use

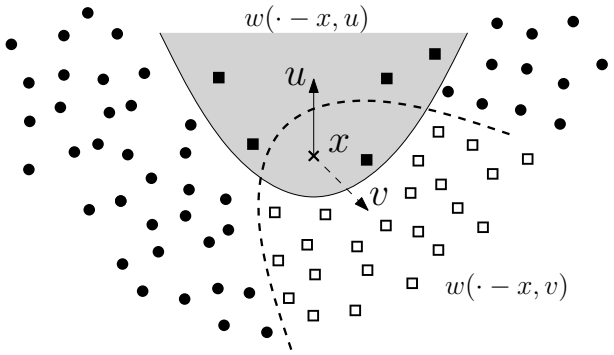



Figure 2.1: The calculation of the generalized halfspace depth at a point $\boldsymbol{x} \in \mathbb{R}^2$ is illustrated. The "banana shaped" distribution is considered, random sample of size $n = 73$ is taken from the distribution. A *parabola* (with focus $\boldsymbol{x}$, $\boldsymbol{u}$ perpendicular to its directrix and a fixed distance between focus and directrix) is used as a boundary of the generalized halfspace and hence an *indicator of closed parabola* is used as the weight function $w$. As $\boldsymbol{u}$ ranges over $\mathcal{S}^2$ the observations lying inside parabola determined by $\boldsymbol{x}$ and $\boldsymbol{u}$ are counted. The parabola covering the minimum number of observations (5 in the example) is shown in gray colour. Hence, the empirical depth of $\boldsymbol{x}$ is 5/73. For an illustration another candidate on minimal parabola (in direction $\boldsymbol{v}$) is also shown in the figure. It contains 24 points hence it is not minimal.

notation with this parameter in subscript instead of subscript denoting the weight function.

*Note* 2. Since we consider only bounded measurable weight functions we may, without loss of generality, consider the weight functions to be *non-negative.* On the other hand, considering nonmeasurable weight function need not to make sense, and unbounded weight functions do not bring any advantage to the generalized halfspace depth.

One can see the illustration of the generalized halfspaces depth idea on Figure 2.1. Similarly as in the definition of the halfspace depth of a point $\boldsymbol{x}$, a vector $\boldsymbol{u} \in \mathcal{S}^2$ represents the direction from $\boldsymbol{x}$ to which we look and count the "number" of observations that "protect" $\boldsymbol{x}$. The depth of the point $\boldsymbol{x}$ is then "number" of observations "protecting" $\boldsymbol{x}$ from the least favorable direction. Of course if, in the previous sentence, our wish is to be more mathematically precise and formal, we have to first specify the weight function $w$ that determines the properties of the generalized depth. Then the vague terms "number" and "protect" can be replaced with something more specific.

The weight function in Figure 2.1 is just an indicator function but it may be quite general function as follows from Definition 5. On the other hand the weight function may be considered to be "regular" for practical applications. The weight function is usually considered to be *symmetric* and *piecewise continuous* (see the discussion below).

**Definition 6.** Multivariate function $w : \mathbb{R}^p \times \mathcal{S}^p \to \mathbb{R}$ is called *piecewise continuous* in what follows if

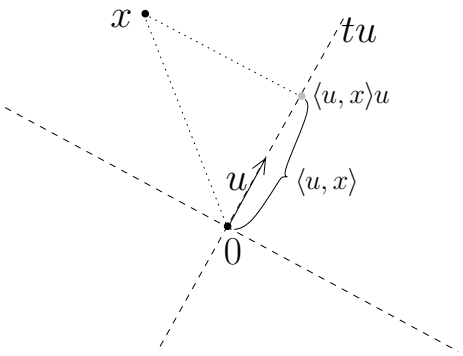



Figure 2.2: Illustration of the definition of the spherical symmetry of the weight function. The weight of $\boldsymbol{x}$ depends only on distances between the orthogonal projection of $\boldsymbol{x}$ to the line $t\boldsymbol{u}$ and points $\boldsymbol{x}$, $\mathbf{0}$ ($\|\boldsymbol{x} - \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{u}\|$, $\langle \boldsymbol{u}, \boldsymbol{x}\rangle$).

1. for any nondegenerate bounded closed interval $I \subset \mathbb{R}^p \times \mathcal{S}^p$ there exist pairwise disjoint open connected sets $G_i, i = 1, \ldots, N$ such that $I = \bigcup_{i=1}^{N} \overline{G}_i$, and $w$ is continuous on $G_i, i = 1, \ldots, N$, and
2. the discontinuity points of $w$ form a subset of a closed set of zero Lebesgue measure.

There are different notions of symmetry for multivariate functions. The most important symmetry for weight functions is the *spherical symmetry*. When we talk about a spherical symmetric weight function we mean the function that satisfies the following definition.

**Definition 7** (Spherical symmetry of the weight functions)**.** We say that a weight function $w : \mathbb{R}^p \times \mathcal{S}^p \to \mathbb{R}$ is *spherically symmetric* if there exists a *piecewise continuous* bounded function $h : [0, +\infty) \times \mathbb{R} \to \mathbb{R}$ such that

$$w(\boldsymbol{x}, \boldsymbol{u}) = h\big(\|\boldsymbol{x} - \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{u}\|, \langle \boldsymbol{u}, \boldsymbol{x}\rangle\big) = h\big(\sqrt{\|\boldsymbol{x}\|^2 - \langle \boldsymbol{u}, \boldsymbol{x}\rangle^2}, \langle \boldsymbol{u}, \boldsymbol{x}\rangle\big). \tag{2.3}$$

The second equality is consequence of the Pythagorean theorem.

*Remark* 3. Fig. 2.2 illustrates the definition of the spherical symmetry. In other words the spherically symmetric weight function $w(\boldsymbol{x}, \boldsymbol{u})$ depends only on the decomposition of the point $\boldsymbol{x}$ to its projection to line $t\boldsymbol{u}, t \in \mathbb{R}$ (point $\langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{u}$, and its norm $\langle \boldsymbol{u}, \boldsymbol{x}\rangle$) and its distance from that line ($\|\boldsymbol{x} - \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{u}\|$). Note that for $\boldsymbol{u} = (0, \ldots, 0, 1)^T$ it holds

$$w\big((x_1, \ldots, x_k, \ldots, x_p)^T, \boldsymbol{u}\big) = w\big((x_1, \ldots, -x_k, \ldots, x_p)^T, \boldsymbol{u}\big), \quad k = 1, \ldots, p-1.$$

One of the main advantages of the generalized halfspace depth is that for a well chosen weight function the depth contours respect the shape of the probability density function contours. In particular, the weighted version of the halfspace depth is more sensitive to nonconvexity or disconnectivity of the underlying distribution (nonconvexity or disconnectivity of its support or of the density function

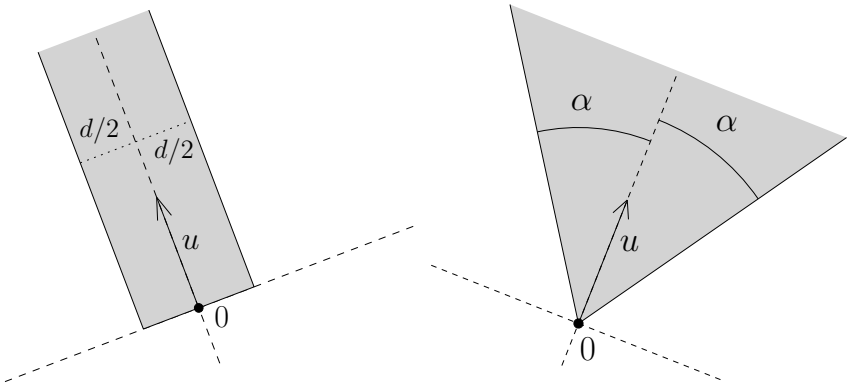



Figure 2.3: The band and the cone weight functions. They are simply the indicators of the band (resp. cone) in given direction $\boldsymbol{u}$.

level-sets). Naturally, the sensitivity depends mostly on the chosen weight function. Hence, one may say that the generalized halfspace depth for a proper weight function is closely related to the structure (geometry) of data.

This property of generalized halfspace depth may be useful if one needs to identify data with non-convex support, mixtures of distributions or for better results of classification or if the problem is to find proper trimming boundary region for a multivariate trimmed mean. An illustration of this fact can be seen on Figure 2.1. The point $\boldsymbol{x}$ is quite deep point of the dataset – from the halfspace depth point of view (but as it was illustrated in Section 1.4 we can expect similar results for most of the depth functions). On the other hand, point $\boldsymbol{x}$ is not so well covered from the direction $\boldsymbol{u}$ and the generalized halfspace depth reveals this fact.

**Example 4** (Weight functions)**.** We introduce a sample of recommended and useful choices of the weight functions. All the weight functions mentioned here are spherically symmetric (see Definition 7) and also piecewise continuous.

- *The halfspace depth:* Let

$$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\left\{\boldsymbol{u}^T\boldsymbol{x} \geq 0\right\} \tag{2.4}$$

  then $\mathrm{D}_w$ is the halfspace depth. See Fig. 1.10, 1.13 and 1.16 for the halfspace depth contours examples.

- *The band (cylinder) depth:* Consider

$$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\left\{\|\boldsymbol{x} - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \boldsymbol{u}\| \leq d/2, \langle \boldsymbol{x}, \boldsymbol{u}\rangle \geq 0\right\}, \tag{2.5}$$

  where $d > 0$ is given bandwidth or diameter of the cylinder base (in higher dimensions), respectively. See the left panel of Fig. 2.3. The depth of $\boldsymbol{x}$ is based on probabilities (empirical probabilities) of (half-)cylinders in the direction of $\boldsymbol{u}$ with radius $d/2$ for which $\boldsymbol{x}$ is in the centre of the cylinder base. Examples of the band depth contours can be seen on Figures 1.11, 1.14 and 1.17.

- *The cone depth:* Let $\alpha \in [0, \pi/2]$ and

$$w(\boldsymbol{x}, \boldsymbol{u}) = \begin{cases} 1, & \text{if } \measuredangle(\boldsymbol{x}, \boldsymbol{u}) \leq \alpha, \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

  It is an indicator of a cone in the direction given by $\boldsymbol{u}$ and with the apex angle $2\alpha$. See also the right panel of Fig. 2.3.

- *The kernel band depth:* A slight modification or generalisation of the band depth is given by the weight function

$$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\{\langle \boldsymbol{x}, \boldsymbol{u} \rangle \geq 0\} \; k(\|\boldsymbol{x} - \langle \boldsymbol{x}, \boldsymbol{u} \rangle \boldsymbol{u}\|), \tag{2.7}$$

  where $k$ is a function $k : \mathbb{R} \to [0, +\infty)$. The usual choice of $k$ is one of the kernel functions used for univariate probability density estimation. See Fig. 2.4 for examples of weight function based on a *triangular kernel* and on a *Gaussian kernel* $k$. The weight does not depend on the distance of $\boldsymbol{x}$ in the direction of $\boldsymbol{u}$ but only on the distance of $\boldsymbol{x}$ from the line $t\boldsymbol{u}, t \in \mathbb{R}$ (see Fig. 2.2). Being derived from the band depth the kernel band depth has similar properties as the band depth. On the other hand, the kernel band depth is more flexible than the band depth, hence it gives usually better results if a suitable kernel function $k$ is chosen. Note that if we choose $k \equiv 1$ we obtain the halfspace depth.

- *The kernel cone depth:* Similarly as the kernel band depth we can define the kernel cone depth weight function as

$$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\{\langle \boldsymbol{x}, \boldsymbol{u} \rangle \geq 0\} \; k(\measuredangle(\boldsymbol{x}, \boldsymbol{u})), \tag{2.8}$$

  where $k : [0, \pi/2] \to [0, +\infty)$ is a kernel function. The weights are assessed according to the angles from directional vector $\boldsymbol{u}$. Kernel functions that are equal to 0 at some finite value should be used. See Fig. 5.1 and also Section 5.2.2 for examples (*triangular* and *tricube kernel*) of such kernels.

- *The conic section depth:* Suppose that $C(\boldsymbol{u}, t)$ is a conic section (*sphere, ellipsoid, paraboloid* or *hyperboloid*) with major axis in the direction given by $\boldsymbol{u} \in \mathcal{S}^p$ and with the focus in the point $t\boldsymbol{u}$. Usually we set parameter $t$ to the point $t\boldsymbol{u}$ be a focus or to apex be in the origin. Then the following weight function may be naturally defined:

$$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\{\boldsymbol{x} \in C(\boldsymbol{u}, t)\}.$$

  For more details about a class of weight functions with focus in $\boldsymbol{x}$ see Chapter 4 Section 4.1 and also see Figure 4.1. Examples of contours obtained by a parabolic weight function with apex in origin can be seen on Figures 1.11, 1.14 and 1.17 - see panels titled "Parabolic Depth".

- *The kernel weighted conic section depth:* The conic section depth may be generalized in a similar way as the band depth is generalized to the kernel band depth. More details are given in Chapter 4. Examples of contours for kernel parabolic and kernel hyperbolic weight functions can be seen on Figures 1.12, 1.15 and 1.18.

- *The local kernel depth:* Consider $w(\boldsymbol{x}, \boldsymbol{u}) = k(\boldsymbol{x})$ for some (non-negative) function $k$ such that
$$\int_{\mathbb{R}^p} k(\boldsymbol{x}) \ \mathrm{d}\boldsymbol{x} = 1,$$
i.e., $k$ is some kernel function. In other words the weight function $w$ does not depend on the directions $\boldsymbol{u} \in \mathcal{S}^p$. The principal difference from the probability density kernel estimation is that the kernel function $k$ is considered fixed with respect to the sample size; the "bandwidth" of $k$ is a fixed value. More details are given in Chapter 4 in the Section 4.2.

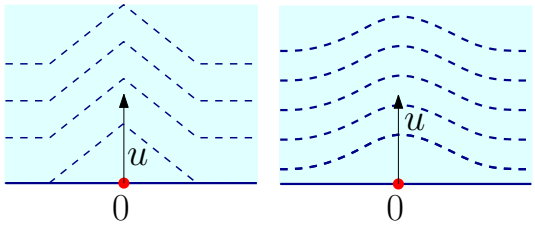



Figure 2.4: The kernel band depth weight functions - triangular and Gaussian kernel.

**Example 5** (Localization of the depth function)**.** The weight function that follows is a slight modification of the weight function in (2.7). If we replace the indicator of a halfspace with a function of the distance from the boundary of the halfspace we can obtain a depth function that is local. Usual choice is the function that decreases with increasing distance from the boundary. More precisely, suppose a function $g: \ \mathbb{R} \to [0, +\infty)$. Further suppose

$$\lim_{|t| \to +\infty} g(t) = 0.$$

Then we can modify the weight function in (2.7) as follows

$$w(\boldsymbol{x}, \boldsymbol{u}) = g(\langle \boldsymbol{x}, \boldsymbol{u} \rangle) \ k(\|\boldsymbol{x} - \langle \boldsymbol{x}, \boldsymbol{u} \rangle \boldsymbol{u}\|).$$

It proposes a class of localized generalized halfspace depth that can only respect a local geometry of the data. For instance if $k \equiv 1$ and

$$g(t) = \mathbb{1}\left\{\langle \boldsymbol{x}, \boldsymbol{u} \rangle \geq 0\right\} e^{-\lambda t^2},$$

where $\lambda > 0$ we obtain the localized halfspace depth with rate of localization $\lambda$. The localized halfspace depth introduced in [Agostinelli & Romanazzi, 2011] and

briefly mentioned in Subsection 1.4.9 is also a member of this class of localized generalized halfspace depth functions.

It is not easy to calculate the depth function for arbitrary point $\boldsymbol{x} \in \mathbb{R}^p$. We show a "classical" example.

**Example 6.** Let $\boldsymbol{X}$ be a two dimensional random vector with Gaussian distribution $\mathcal{N}_2(\mathbf{0}, \mathbf{I}_2)$. Suppose we have a band weight function (2.5) for given $d > 0$.

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

$$\mathrm{D}_w(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) = \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{x}/\|\boldsymbol{x}\|).$$

It means that "minimal" direction is the direction along the ray which connects the origin and $\boldsymbol{x}$. Without loss of generality we can assume that $\boldsymbol{x} = (0, x_2)^T$ (the distribution is symmetric about $\mathbf{0}$ and also about any line containing $\mathbf{0}$). For such point $\boldsymbol{x}$ let $\boldsymbol{u} = (0, 1)^T$. See also Fig. 2.11 for illustration of the following ideas.

One has

$$\begin{aligned}\mathsf{E}\, w\big((x_1, x_2)^T, (0, 1)^T\big) &= \mathsf{P}(X_2 > x_2,\ -d/2 < X_1 < d/2) \\ &= \big(1 - \Phi(x_2)\big)\big(\Phi(d/2) - \Phi(-d/2)\big),\end{aligned}$$

where $\Phi$ is the distribution function of $\mathcal{N}(0, 1)$. For any other direction $\boldsymbol{u} \neq (0, 1)^T$ and $w(\boldsymbol{x}, \boldsymbol{u})$ there exists uniquely determined rotation $\mathbf{A}$ such that $\mathbf{A}\boldsymbol{u} = (0, 1)^T$ and $\mathbf{A}\boldsymbol{X} = \boldsymbol{X}' \sim \mathcal{N}_2(\mathbf{0}, \mathbf{I}_2)$. For $\boldsymbol{x} = (0, x_2)^T$ it holds $\mathbf{A}\boldsymbol{x} = \boldsymbol{x}'$ where $x_2 > x_2'$. It holds that

$$\begin{aligned}\mathsf{E}\, w(\boldsymbol{x}, \boldsymbol{u}) = \mathsf{E}\, w\big(\boldsymbol{x}', (0, 1)^T\big) &= \mathsf{P}(X_2' \geq x_2')\, \mathsf{P}(x_1' - d/2 < X_1' < x_1' + d/2) \\ &= (1 - \Phi(x_2'))(\Phi(x_1' + d/2) - \Phi(x_1' - d/2)).\end{aligned}$$

The expression can be rewritten as follows

$$\begin{aligned}\mathsf{E}\, w(\boldsymbol{x}, \boldsymbol{u}) = \big(1 - \Phi(x_2 \sin\alpha)\big) \\ \cdot \big(\Phi(x_2 \cos\alpha + d/2) - \Phi(x_2 \cos\alpha - d/2)\big),\end{aligned}$$

where $\alpha = \pi/2 - \measuredangle\big(\boldsymbol{u}, (0, 1)^T\big)$. Its derivative with respect to $\alpha$ is equal to 0 for $\alpha = \pi/2$. Hence $\pi/2$ is the point of local extrema. Further the derivative is negative for $\alpha = 0$. It offers that the function decreases on the interval $[0, \pi/2)$. Unfortunately it is hard to show that the derivative is negative for all $(0, \pi/2)$ but numerical computations suggest that the derivative is really negative. Thus the depth of the point $(0, x_2)^T$ most likely is

$$\mathrm{D}_w((0, x_2)^T) = \big(1 - \Phi(x_2)\big)\big(\Phi(d/2) - \Phi(-d/2)\big).$$

Since $\mathbf{A}\boldsymbol{X}$ is $\mathcal{N}_2(\mathbf{0}, \mathbf{I}_2)$ for any orthogonal matrix $\mathbf{A}$ one has that the depth of arbitrary point $\boldsymbol{x} \in \mathbb{R}^p$ is

$$\mathrm{D}_w(\boldsymbol{x}) = \big(1 - \Phi(\|\boldsymbol{x}\|)\big)\big(\Phi(d/2) - \Phi(-d/2)\big).$$

This example is rewritten in Section 2.3 for the *weighted halfspaces ratio depth*. Mentioned depth can be computed exactly without need for numerical computations.

## 2.2 Properties of the Generalized Halfspace Depth

Let us turn our attention to the basic properties of the generalised halfspace depth. All the results that follow hold for an *absolutely continuous* distribution $P$ and also for an *empirical* distribution $P_{\mathrm{n}}$ (and therefore for the empirical depth) as well.

First of all we will show that under a natural condition on the weight function the generalised halfspace depth is vanishing at infinity.

**Theorem 1** (Vanishing at infinity)**.** *Consider a weight function $w$ such that for all $\boldsymbol{u} \in \mathcal{S}^p$ it holds*

$$\lim_{k\to\infty} \sup_{\boldsymbol{x}:\ \langle \boldsymbol{x},\boldsymbol{u}\rangle \le -k} w(\boldsymbol{x},\boldsymbol{u}) = 0. \tag{2.9}$$

*Then for any probability measure $P$ it holds that*

$$\lim_{\|\boldsymbol{x}\|\to\infty} \mathrm{D}_w(\boldsymbol{x}; P) = 0.$$

*Proof.* Suppose a random vector $\boldsymbol{X}$ with distribution $P$. To prove the theorem we use the following two properties.
**1.** Since every probability measure on metric space $\mathbb{R}^p$ is *tight*, i.e., for arbitrary $\varepsilon > 0$ there exists a constant $r_1$ such that $\mathsf{P}(\|\boldsymbol{X}\| > r_1) < \varepsilon$. Hence

$$\mathsf{P}(\langle \boldsymbol{X}, \boldsymbol{x}/\|\boldsymbol{x}\|\rangle > r_1) < \varepsilon, \quad \forall \boldsymbol{x} \in \mathbb{R}^p.$$

**2.** From equation (2.9) it follows (by shifting $w$ to $\boldsymbol{x}$ and by using the definition of limit) that for arbitrary $\varepsilon > 0$ there exist a constant $r_2$ such that $\forall \boldsymbol{x} \in \mathbb{R}^p$

$$\begin{aligned} &\sup_{\boldsymbol{y}:\ \langle \boldsymbol{y}-\boldsymbol{x},\boldsymbol{x}/\|\boldsymbol{x}\|\rangle \le -r_2} w(\boldsymbol{y}-\boldsymbol{x}, \boldsymbol{x}/\|\boldsymbol{x}\|) \\ = &\sup_{\boldsymbol{y}:\ \langle \boldsymbol{y},\boldsymbol{x}/\|\boldsymbol{x}\|\rangle \le \|\boldsymbol{x}\|-r_2} w(\boldsymbol{y}-\boldsymbol{x}, \boldsymbol{x}/\|\boldsymbol{x}\|) < \varepsilon, \quad \forall \boldsymbol{x} \in \mathbb{R}^p. \end{aligned}$$

These facts are sufficient to finish the proof (see Figure 2.5). For arbitrary $\varepsilon > 0$ there exist constants $r_1, r_2$ such that if $\|\boldsymbol{x}\| > r_1 + r_2$ then

$$\mathsf{P}(\langle \boldsymbol{X}, \boldsymbol{x}/\|\boldsymbol{x}\|\rangle > r_1) < \varepsilon,$$

and since $\|\boldsymbol{x}\| - r_2 > r_1$ it holds

$$\sup_{\boldsymbol{y}:\ \langle \boldsymbol{y},\boldsymbol{x}/\|\boldsymbol{x}\|\rangle \le r_1} w(\boldsymbol{y}-\boldsymbol{x}, \boldsymbol{x}/\|\boldsymbol{x}\|) \le \sup_{\boldsymbol{y}:\ \langle \boldsymbol{y},\boldsymbol{x}/\|\boldsymbol{x}\|\rangle \le \|\boldsymbol{x}\|-r_2} w(\boldsymbol{y}-\boldsymbol{x}, \boldsymbol{x}/\|\boldsymbol{x}\|) < \varepsilon.$$

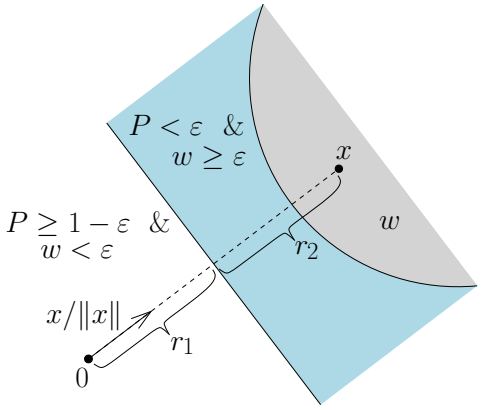



Figure 2.5: Depth is vanishing at infinity. Grey region indicates the weight function for the point $\boldsymbol{x}$ and in the direction $\boldsymbol{x}/\|\boldsymbol{x}\|$. The light blue region has probability smaller than $\varepsilon$ and the weight function there can attain arbitrary values. The white region is the region where the weight function is less than $\varepsilon$.

Indeed: the weight function $w$ is evaluated far from the origin so that $w$ is small enough on a halfspace, $\{\boldsymbol{y} : \langle \boldsymbol{y}, \boldsymbol{x}/\|\boldsymbol{x}\|\rangle \leq r_1\}$, with probability greater than $1-\varepsilon$. Hence, if $\|\boldsymbol{x}\| > r_1 + r_2$ then

$$\begin{aligned} \mathrm{D}_w(\boldsymbol{x}; P) &= \inf_{\|\boldsymbol{u}\|=1} \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u}) \leq \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{x}/\|\boldsymbol{x}\|) \\ &= \int_{\boldsymbol{y}:\ \langle \boldsymbol{y}, \boldsymbol{x}/\|\boldsymbol{x}\|\rangle > r_1} w(\boldsymbol{y}-\boldsymbol{x}, \boldsymbol{x}/\|\boldsymbol{x}\|)\ \mathrm{d}P(\boldsymbol{y}) \\ &\quad + \int_{\boldsymbol{y}:\ \langle \boldsymbol{y}, \boldsymbol{x}/\|\boldsymbol{x}\|\rangle \leq r_1} w(\boldsymbol{y}-\boldsymbol{x}, \boldsymbol{x}/\|\boldsymbol{x}\|)\ \mathrm{d}P(\boldsymbol{y}) \\ &\leq L\, \mathsf{P}(\langle \boldsymbol{X}, \boldsymbol{x}/\|\boldsymbol{x}\|\rangle > r_1) + \varepsilon\, \mathsf{P}(\langle \boldsymbol{X}, \boldsymbol{x}/\|\boldsymbol{x}\|\rangle \leq r_1) \leq L\varepsilon + \varepsilon. \end{aligned} \tag{2.10}$$

where $L$ is constant such that $w \leq L$ ($w$ is bounded). ■

General formulation of the theorem may seem little bit formal. Now we rewrite the theorem for *spherically symmetric* weight functions.

**Corollary 2** (Vanishing at infinity)**.** *Consider a nonnegative spherically symmetric weight function* $w(\boldsymbol{x}, \boldsymbol{u}) = h(\|\boldsymbol{x} - \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{u}\|, \langle \boldsymbol{u}, \boldsymbol{x}\rangle)$ *such that*

$$\lim_{z \to -\infty} \sup_{d \geq 0} h(d, z) = 0 \tag{2.11}$$

*holds. Then*

$$\lim_{\|\boldsymbol{x}\| \to \infty} \mathrm{D}_w(\boldsymbol{x}; P) = 0.$$

*Remark* 4. All the weight functions considered in this thesis satisfy either condition (2.11) or condition (2.9).

*Proof of Corollary 2.* Validity of (2.9) follows from the definition of spherically symmetric weight functions (see Defintion 7) and from (2.11). ■

Vanishing at infinity is one of the so called *key properties* (see Def. 1). The generalized halfspace depth does not in general satisfy the other three key properties. The depth function is not affine invariant but there exists a modification of the depth that satisfies this property. The depth also needs not to attain its maximum in the center of symmetry and needs not to decrease along rays starting in the deepest point. Fig. 1.6 shows an illustration of the case where these properties need not to hold for the weight function whose "width" is not high (e.g. the band weight function (2.5) with $d$ less than distance of the rectangles).

One of the disadvantage of the halfspace depth, simplicial depth and some other known depth functions is that a point outside a support of distribution can attain relatively high values of depth. A properly chosen weight function can overcome this property. For instance consider a weight functions to be as shown on Fig. 2.6 (gray regions). It shows "banana" shaped data. The left panel shows the halfspace depth contours. Even a point $\boldsymbol{x}$ that is quite far from the data cloud has significantly nonzero depth. It leads to convex contours that are going through regions with no data. On the other hand if we choose a weight function shown on the right panel of Fig. 2.6 the depth of a point $\boldsymbol{x}$ that lies on border of the data is equal to zero. It leads to contours that more respect the shape of the data.

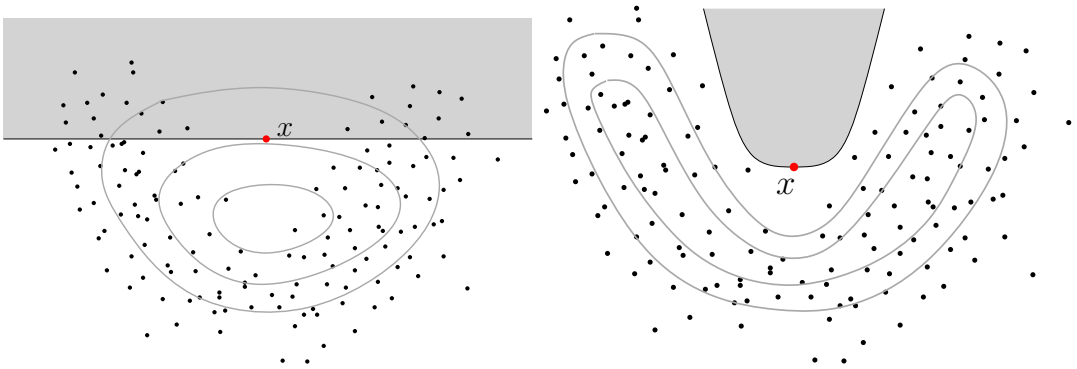



Figure 2.6: In contrast to the halfspace depth there exists a weight function for which a point outside a support attains zero depth value.

The latter we write as a remark without the proof.

*Remark* 5. For a probability distribution satisfying that its zero measure sets' radius is greater than a given value $\delta > 0$ there exists a weight function $w$ such that $\mathrm{D}_w(\boldsymbol{x}) = 0$ for all $\boldsymbol{x}$ outside the support of the distribution and $\mathrm{D}_w(\boldsymbol{x}) > 0$ for all $\boldsymbol{x}$ inside the support. Note that this condition is only sufficient conditions.

The value of depth of any point depends on the weight function. To asses how deep a point is, one may be interested in limits for the depth function.

**Theorem 3.** *Suppose a weight function such that* $0 \leq w \leq L$. *Then for any* $\boldsymbol{x} \in \mathbb{R}^p$ *and any probability measure* $P$ *on* $\mathbb{R}^p$ *it holds that*

$$0 \leq \mathrm{D}_w(\boldsymbol{x}; P) \leq L.$$

*Proof.* It is clear because

$$0 \leq \mathsf{E}_P\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) \leq L \quad \forall \boldsymbol{u} \in \mathcal{S}^p.$$

■

The generalized halfspace depth in general is not *affine invariant.* But it can be shown that it is rotation and translation invariant.

**Theorem 4** (Rotation and translation invariance)**.** *The generalised halfspace depth is translation invariant. If $w$ is moreover a spherically symmetric weight function then the generalised halfspace depth defined on Euclidean space is also rotation invariant, i.e.*

$$\mathrm{D}_w(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}; \mathbf{A}\boldsymbol{X} + \boldsymbol{a}) = \mathrm{D}_w(\boldsymbol{x}; \boldsymbol{X})$$

*for any orthogonal matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ and any vector $\boldsymbol{a} \in \mathbb{R}^p$.*

*Proof.* It is easy to see that the depth is translation invariant for any weight function.

Consider a spherically symmetric weight function and an orthogonal matrix $\mathbf{A}$. It holds (due to the orthogonality of $\mathbf{A}$)

$$\langle \mathbf{A}\boldsymbol{v}, \mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\rangle = \langle \boldsymbol{v}, \boldsymbol{X} - \boldsymbol{x}\rangle, \quad \forall \boldsymbol{v} \in \mathcal{S}^p$$

and

$$\|\mathbf{A}\boldsymbol{y}\| = \|\boldsymbol{y}\|, \quad \forall \boldsymbol{y} \in \mathbb{R}^p.$$

Hence

$$\begin{aligned}
\mathrm{D}_w(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}; \mathbf{A}\boldsymbol{X} + \boldsymbol{a}) &= \inf_{\|\boldsymbol{u}\|=1} \mathsf{E}\, w(\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}), \boldsymbol{u}) \\
&= \inf_{\|\boldsymbol{u}\|=1} \mathsf{E}\, h(\|\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}) - \langle \boldsymbol{u}, \mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\rangle \boldsymbol{u}\|, \langle \boldsymbol{u}, \mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\rangle) \\
&= \inf_{\|\boldsymbol{v}\|=1} \mathsf{E}\, h(\|\mathbf{A}(\boldsymbol{X} - \boldsymbol{x}) - \langle \mathbf{A}\boldsymbol{v}, \mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\rangle \mathbf{A}\boldsymbol{v}\|, \langle \mathbf{A}\boldsymbol{v}, \mathbf{A}(\boldsymbol{X} - \boldsymbol{x})\rangle) \\
&= \inf_{\|\boldsymbol{v}\|=1} \mathsf{E}\, h(\|\mathbf{A}[(\boldsymbol{X} - \boldsymbol{x}) - \langle \boldsymbol{v}, \boldsymbol{X} - \boldsymbol{x}\rangle \boldsymbol{v}]\|, \langle \boldsymbol{v}, \boldsymbol{X} - \boldsymbol{x}\rangle) \\
&= \inf_{\|\boldsymbol{v}\|=1} \mathsf{E}\, h(\|(\boldsymbol{X} - \boldsymbol{x}) - \langle \boldsymbol{v}, \boldsymbol{X} - \boldsymbol{x}\rangle \boldsymbol{v}\|, \langle \boldsymbol{v}, \boldsymbol{X} - \boldsymbol{x}\rangle) = \mathrm{D}_w(\boldsymbol{x}; \boldsymbol{X}).
\end{aligned}$$

■

Note that a spherical symmetry is only a sufficient condition. It is not difficult to find a non-spherically symmetric weight function that gives also rotation invariant depth.

In particular, the generalised halfspace depth in general is not invariant to *scale* transformations (with different scaling of axes). It is well known that also the spatial median is not affine equivariant and therefore there has been developed

so called *transformation/retransformation* technique for constructing affine equivariant spatial median, see e.g. [Ilmonen et al., 2012, Chakraborty et al., 1998]. These methods may be also easily applied to the generalised halfspace depth and we henceforth consider the generalised halfspace depth satisfying conditions of Theorem 4 to be *affine invariant.*

On the other hand the affine invariance may be a source of problems. The depth based classification (in particular with the simple yet natural *maximum depth rule*) usually needs to be modified to give good results. The problems is mainly in the scale invariance of the depth (unlike the probability density function which is not scale invariant at all). More discussion on the necessary modification of the depth-based classifiers may be found in [Li et al., 2012] or in [Hlubinka & Vencalek, 2013].

The affine invariance of the data depth seems to be a natural and desirable property since it is a natural property of an univariate quantile based depth. Recall that:

1. The generalised halfspace depth is translation and rotation invariant.

2. There is a transformation/retransformation technique which makes the proposed depth affine invariant if needed.

But note also that:

3. An univariate depth based on quantiles is even invariant to *any* strictly monotone transformation as the univariate quantiles itself are equivariant with respect to such transformation; for multivariate data it is not true at all – consider , e.g. the log-log transformation and the celebrated halfspace or simplicial depths. We are usually not surprised that the multivariate depth is not invariant to such nonlinear transformations.

4. Recall the "protection of the king" interpretation of the depth. The distance of the "guards" along some axis increases as the coordinates on the axis are multiplied by a factor greater than one. Hence, the king becomes "less protected" from the the direction perpendicular to the axis and his depth decreases.

Hence, we do not consider the full affine invariance (beyond the rotation and translation one) to be the key property of the depth. But as we have mentioned above, it is always possible to have fully affine invariant generalised halfspace depth. In the next definition (one possible) affine *invariant modification of the generalised halfspace depth* is presented.

**Definition 8** (An affine invariant modification of depth)**.** Suppose that the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ of a random vector $\boldsymbol{X}$ exists and that it is nonsingular. Further suppose a spherically symmetric weight function $w$. Then an affine invariant depth function may be defined as

$$\widetilde{\mathrm{D}}_w(\boldsymbol{x};\boldsymbol{X}) = \mathrm{D}_w(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{x};\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{X}) = \inf_{\|\boldsymbol{u}\|=1} \mathsf{E}\, w\big(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}(\boldsymbol{X}-\boldsymbol{x}),\boldsymbol{u}\big), \qquad (2.12)$$

where $\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}$ denotes an inverse square root of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$.

The sample version for known matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ is obtained in the same way as in Definition 5. If we do not know this matrix we use a sample version of the covariance matrix, $\boldsymbol{\Sigma}_n$, instead. In such case, one has

$$\widetilde{\mathrm{D}}_{w,n}(\boldsymbol{x};\boldsymbol{X}) = \inf_{\|\boldsymbol{u}\|=1} \frac{1}{n}\sum_{i=1}^{n} w\left(\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{X}_i - \boldsymbol{x}), \boldsymbol{u}\right).$$

The sample version of matrix has to satisfy affine equivariance property - (2.13). Since the depth is usually robust (depends on the weight function) the sample covariance matrix should also posses this property. The classical sample covariance matrix,

$$\boldsymbol{\Sigma}_n = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)^T,$$

satisfies the equivariance property but it is very sensitive to outliers. See [Stahel, 1981] for overview estimators of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ which satisfy the equivariance property and that are also robust.

Lemma 28 says how we can obtain an inverse square root matrix. First we need to get inverse of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$. Then the lemma is applied for $\mathbf{C} = \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$. Actually, there is even no need to calculate inverse of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ – we only need to calculate $\boldsymbol{\Lambda}^{-1/2} = \mathrm{diag}\left\{\frac{1}{\sqrt{\lambda_1}}, \ldots, \frac{1}{\sqrt{\lambda_p}}\right\}$. It is clear that

$$\mathsf{var}\left(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{X}\right) = \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2} = \mathbf{I}.$$

In R the square root matrix can be computed with function sqrtm() in package expm or by using singular value decomposition - function svd() in R base package.

*Remark* 6. The nonsingularity of matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ is not restrictive assumption. If the covariance matrix is singular, the vector $\boldsymbol{X}$ lies on a lower dimensional hyperplane in $\mathbb{R}^p$ and hence we may transform $\boldsymbol{X}$ to the lower dimensional random vector without loosing any information.

In the previous definition any *affine equivariant dispersion functional* $\mathrm{disp}(\boldsymbol{X})$ may be used instead of the covariance matrix $\mathsf{var}\,\boldsymbol{X}$ in order to obtain affine invariant version of the depth (see proof of the Theorem 5). The functional should satisfy the following properties: for any random vector $\boldsymbol{Z}$ the functional $\mathrm{disp}(\boldsymbol{Z}) \in \mathbb{R}^{p\times p}$ is a positive definite symmetric nonsingular matrix and for any nonsingular matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ it holds that

$$\mathrm{disp}(\mathbf{A}\boldsymbol{Z}) = \mathbf{A}\mathrm{disp}(\boldsymbol{Z})\mathbf{A}^T. \tag{2.13}$$

There is plenty of papers focused on this topic. For details see [Donoho, 1982], [Stahel, 1981], [Rousseeuw, 1985].

**Theorem 5.** *Let the weight function $w$ be spherically symmetric and $\boldsymbol{X}$ be a random vector with positive definite covariance matrix* $\mathsf{var}\,\boldsymbol{X} = \boldsymbol{\Sigma}_{\boldsymbol{X}}^{1/2}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{1/2}$. *Then*

$\widetilde{\mathrm{D}}_w$ *defined as* (2.12) *is affine invariant, i.e. for any nonsingular matrix* $\mathbf{A} \in \mathbb{R}^{p\times p}$, *it holds that*

$$\widetilde{\mathrm{D}}_w(\mathbf{A}\boldsymbol{x}; \mathbf{A}\boldsymbol{X}) = \widetilde{\mathrm{D}}_w(\boldsymbol{x}; \boldsymbol{X}).$$

*Proof.* It is clear that if $w$ is spherically symmetric then there exists a function $\widetilde{w}$ such that

$$w(\boldsymbol{y}, \boldsymbol{u}) = \widetilde{w}(\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{y}),$$

where $\boldsymbol{\Gamma}_{\boldsymbol{u}}$ is an orthogonal matrix such that

$$\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{u} = (0, 0, \ldots, 0, 1)^T. \tag{2.14}$$

Matrix $\boldsymbol{\Gamma}_{\boldsymbol{u}}$ exists but it needs not to be unique. On the other hand $\widetilde{w}(\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{y})$ does not depend on the choice of matrix $\boldsymbol{\Gamma}_{\boldsymbol{u}}$ since

$$\widetilde{w}(\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{y}) = h\left(\sqrt{(\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{y})_1^2 + \ldots + (\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{y})_{p-1}^2},\ (\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{y})_p\right)$$

does not depend on the choice of orthogonal matrix satisfying (2.14). Note that function $h$ is the function from equation (2.3).

Recall that the depth function $\mathrm{D}_w$ is rotation invariant for spherically symmetric weight functions by Theorem 4. Denoting the set of all orthogonal $p \times p$ matrices by $\mathcal{O}_p$, i.e.,

$$\mathcal{O}_p = \{\boldsymbol{\Gamma} \in \mathbb{R}^{p\times p} :\ \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \mathbf{I}\},$$

then

$$\mathrm{D}_w(\boldsymbol{x}; \boldsymbol{X}) = \inf_{\|\boldsymbol{u}\|=1} \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) = \inf_{\boldsymbol{\Gamma}\in\mathcal{O}_p} \mathsf{E}\, \widetilde{w}(\boldsymbol{\Gamma}(\boldsymbol{X} - \boldsymbol{x})).$$

Consider first the case $\mathsf{var}\,\boldsymbol{X} = \boldsymbol{\Sigma}_{\boldsymbol{X}} = \mathbf{I}$ and let $\mathbf{A}$ be a nonsingular $p \times p$ matrix. Then

$$\begin{aligned}\widetilde{\mathrm{D}}_w(\mathbf{A}\boldsymbol{x}; \mathbf{A}\boldsymbol{X}) &= \mathrm{D}_w(\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A}\boldsymbol{x}, \boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A}\boldsymbol{X}) \\ &= \inf_{\boldsymbol{\Gamma}\in\mathcal{O}_p} \mathsf{E}\, \widetilde{w}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A}(\boldsymbol{X} - \boldsymbol{x})).\end{aligned} \tag{2.15}$$

It must be proved that $\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A}$ is an orthogonal matrix. Multiplying the latter matrix by its transpose results in

$$\begin{aligned}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A})^T\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A} &= \mathbf{A}^T(\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2})^T\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A} = \mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1}\mathbf{A} \\ &= \mathbf{A}^T(\mathbf{A}\boldsymbol{\Sigma}_{\boldsymbol{X}}\mathbf{A}^T)^{-1}\mathbf{A} = \mathbf{A}^T(\mathbf{A}^T)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{A}^{-1}\mathbf{A} \\ &= \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1} = \mathbf{I}.\end{aligned} \tag{2.16}$$

Hence, from (2.15), it holds

$$\widetilde{\mathrm{D}}_w(\mathbf{A}\boldsymbol{x}; \mathbf{A}\boldsymbol{X}) = \inf_{\boldsymbol{\Gamma}\in\mathcal{O}_p} \mathsf{E}\, \widetilde{w}(\boldsymbol{\Gamma}(\boldsymbol{X} - \boldsymbol{x})) = \inf_{\boldsymbol{\Gamma}\in\mathcal{O}_p} \mathsf{E}\, \widetilde{w}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}(\boldsymbol{X} - \boldsymbol{x})) = \widetilde{\mathrm{D}}_w(\boldsymbol{x}; \boldsymbol{X}),$$

by the obvious fact that $\{\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A} :\ \mathbf{A} \in \mathcal{O}_p\} = \mathcal{O}_p$.

Consider now a general case with a nonsingular covariance matrix $\mathsf{var}\,\boldsymbol{X} = \boldsymbol{\Sigma}_{\boldsymbol{X}} \neq \mathbf{I}$. There exists a random vector $\boldsymbol{Y}$ and a regular matrix $\mathbf{D}$ which satisfy

$\mathsf{var}\,\boldsymbol{Y} = \mathbf{I}$ and $\boldsymbol{X} = \mathbf{D}\boldsymbol{Y}$. By the first part of the proof one has

$$\widetilde{\mathrm{D}}_w(\boldsymbol{x}; \boldsymbol{X}) = \widetilde{\mathrm{D}}_w(\mathbf{D}\boldsymbol{y}; \mathbf{D}\boldsymbol{Y}) = \widetilde{\mathrm{D}}_w(\boldsymbol{y}; \boldsymbol{Y}),$$

where $\boldsymbol{\xi} = \mathbf{D}^{-1}\boldsymbol{x}$. And again, using the result from the first part of the proof, we finally have

$$\widetilde{\mathrm{D}}_w(\mathbf{A}\boldsymbol{x}; \mathbf{A}\boldsymbol{X}) = \widetilde{\mathrm{D}}_w(\mathbf{A}\mathbf{D}\boldsymbol{y}; \mathbf{A}\mathbf{D}\boldsymbol{Y}) = \widetilde{\mathrm{D}}_w(\boldsymbol{y}; \boldsymbol{Y}).$$

■

We see, that affine invariance is "somehow" connected with elliptical symmetry. Use covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ as a measure of dispersion usually makes sense for elliptically symmetric distributions (in particular normal distribution). For other than elliptically symmetric distributions use of a covariance matrix as a measure of dispersion always need not to make sense. Majority of the depth functions (actually all known depth functions to the author) that are affine invariant does not work well for non-elliptical distributions.

Unfortunately, the empirical depth is not unbiased estimator of the population depth in general. Indeed it is even always non-positively biased estimator of the depth.

**Theorem 6** (Biasness)**.** *For any weight function and for any probability distribution $P$ it holds*

$$\mathsf{E}\,\mathrm{D}_{w,n}(\boldsymbol{x}) \leq \mathrm{D}_w(\boldsymbol{x}; P).$$

*Proof.* Since

$$\mathrm{D}_n(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \frac{1}{n}\sum_{i=1}^{n} w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) \leq \frac{1}{n}\sum_{i=1}^{n} w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}), \quad \forall \boldsymbol{u} : \ \|\boldsymbol{u}\| = 1.$$

then

$$\mathsf{E}\,\mathrm{D}_n(\boldsymbol{x}) \leq \mathsf{E}\,\frac{1}{n}\sum_{i=1}^{n} w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) = \mathsf{E}\,w(\boldsymbol{X}_1 - \boldsymbol{x}, \boldsymbol{u}), \quad \forall \boldsymbol{u} : \ \|\boldsymbol{u}\| = 1.$$

And finally

$$\mathsf{E}\,\mathrm{D}_n(\boldsymbol{x}) \leq \inf_{\|\boldsymbol{u}\|=1} \mathsf{E}\,w(\boldsymbol{X}_1 - \boldsymbol{x}, \boldsymbol{u}) = \mathrm{D}(\boldsymbol{x}).$$

■

The exact calculation of bias of the empirical generalised halfspace depth is typically impossible. Even for the "simplest" halfspace depth it is usually not possible to find explicit form of the expectation of the sample depth.

In some situations not only the expectation but the empirical depth itself cannot be larger than the population depth.

**Example 7.** Consider the *uniform distribution* on the unit circle (with centre in the origin). Let us calculate the *halfspace depth* of the origin. It is clear that $\mathrm{HD}(\mathbf{0}) = 1/2$.

On the other hand let $n = 2k + 1$ be a sample size for $k \in \mathbb{N}$. Then, clearly,

$$\mathrm{HD}_{2k+1}(\mathbf{0}) \leq \frac{k}{2k+1} < \frac{1}{2} = \mathrm{HD}(\mathbf{0}),$$

hence

$$\mathsf{E}\,\mathrm{HD}_{2k+1}(\mathbf{0}) \leq \frac{k}{2k+1} < \frac{1}{2} = \mathrm{HD}(\mathbf{0}).$$

For any even sample size $n = 2k$ it holds

$$\mathrm{HD}_{2k}(\mathbf{0}) \leq \frac{k}{2k} = \frac{1}{2} = \mathrm{HD}(\mathbf{0}).$$

*Uniqueness* of the deepest point cannot be in general ensured for the genralised halfspace depth. While for absolutely continuous probability distributions the halfspace depth gives always the unique deepest point it need not to remain true for the generalised halfpsace depth.

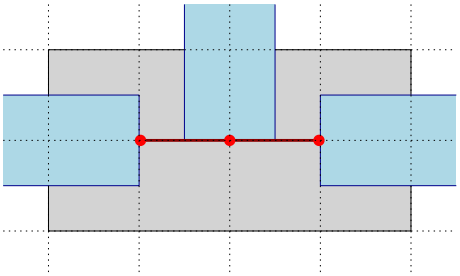


Figure 2.7: The deepest point may not be unique. The dark red line marks the set of all deepest points. Minimal bands for some points are shown in light blue color. Support of the uniform distribution is marked with gray color.

**Example 8** (Non-uniqueness of the deepest point)**.** Suppose we have a band weight function (see (2.5)) of bandwidth $2h$ defined as

$$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\left\{\|\boldsymbol{x} - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \boldsymbol{u}\| \leq h \;\&\; \langle \boldsymbol{x}, \boldsymbol{u}\rangle \geq 0\right\}$$

and set $h = 0.25$.

Consider a *bivariate uniform distribution* on a rectangle $[-1, 1] \times [-0.5, 0.5]$ (see Fig. 2.7). Then it holds that the set of all the deepest points is equal to $\mathcal{A} = \{[-0.5, 0.5] \times \{0\}\}$ (dark red line in the figure). For any $\boldsymbol{x} \in \mathcal{A}$, $\mathrm{D}_w(\boldsymbol{x}) = 0.25$.

**Example 9** (Uniqueness of the deepest point)**.** Consider again the weight function from the previous example. Consider further a *bivariate Gaussian distribution* $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\sigma_{11} > \sigma_{22}$, $\sigma_{12} = \sigma_{21} = 0$. Then $\mathbf{0}$ is the unique deepest point,

$$\mathrm{D}_w(\mathbf{0}) = \frac{2\Phi(0.25/\sigma_{11}) - 1}{2}.$$

Again, it is not hard to show that $\mathrm{D}_w(\boldsymbol{x}) < D_w(\mathbf{0})$ for any $\boldsymbol{x} \neq \mathbf{0}$.

Note that in both examples above we have considered *centrally symmetric* distributions (see Section 1.2). It seems that the main reason why there is the unique deepest point of the Gaussian distribution is its *elliptical symmetry* together with the fact that the center is also the modus. We see that it is not difficult to find *centrally symmetric* distribution such that the point of symmetry is not the deepest point and hence there must exist multiple deepest points. The set of all the deepest points even need not to be connected.

The second reason of possible multiplicity of deepest points is the weight function, in particular the *restricted support* of the band weight function; the support of $w$ is not the whole halfspace $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{u} \rangle \geq 0\}$ and the complement of the support does not have the same geometry as the support. The deepest point of uniform distribution on rectangle is unique if the support of weight function $w$ is the whole halfspace. Even more, the deepest point of uniform distribution on a rectangle is unique for the band weight function as well if the bandwidth $2h$ is wider than the larger side of the rectangle.

Consider a *spherically symmetric weight function* (Definition 7) such that

$$h(t, l) = e^{-t^2/\sigma^2} \mathbb{1}\{l \geq 0\}$$

(the weight function is shown on the right panel of Fig. 2.4) then in Example 8 we obtain the unique deepest point (equal to $\mathbf{0}$) also for the uniform distribution. The reason is that all points in the halfspace are "counted" by the latter weight function in contrast to the band weight function (especially with small bandwidth). The parameter $\sigma$ enables to control a "directional localness." For large values of $\sigma$ the corresponding generalized halfspace depth is similar to the classical halfspace depth while for small values of $\sigma$ the results are more local.

In Section 2.3 a version of the weighted depth with the *unique deepest* in the centre of symmetry is proposed. Another possibility how to obtain a unique "deepest point" is to take the geometric centre (i.e., the mean of deepest points with respect to the Lebesgue measure) of all deepest points. The following theorem shows that such deepest point has symmetric distribution in the case when $P$ is symmetric distribution.

**Theorem 7** (Symmetricity of the deepest point)**.** *If the distribution of $\boldsymbol{X}$ is centrally symmetric about a point $\boldsymbol{\theta}$ (i.e. $\boldsymbol{X} - \boldsymbol{\theta}$ and $-(\boldsymbol{X} - \boldsymbol{\theta})$ are identically distributed), then the distribution of the geometric centre of the set of all the sample deepest points is also symmetric about the population centre $\boldsymbol{\theta}$.*

*Proof.* It follows easily from the symmetry of distribution of $\boldsymbol{X}$, the orthogonal invariance of the generalised halfspace depth (Theorem 4), and the fact that the geometric centre is equivariant to all orthogonal transformations. For more details see also [Liu et al., 1999], Proposition 3.1, where the result is proved for an affine invariant depth function, but it remains the same in the case of invariance to only orthogonal transformations.

Briefly: Without loss of generality suppose $\boldsymbol{\theta} = \mathbf{0}$. Denote by $\widehat{\boldsymbol{\theta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ the geometric center of all sample deepest points based on the random sample

$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. The proof is based on the fact that

$$\widehat{\boldsymbol{\theta}}(-\boldsymbol{X}_1, \ldots, -\boldsymbol{X}_n) = -\widehat{\boldsymbol{\theta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n).$$

This equation is valid because of the symmetry, invariance of the depth against orthogonal transformations and the equivariance of the geometric center against all orthogonal transformations of the set of all sample deepest points.

■

The theorem has one useful corollary.

**Corollary 8.** *Suppose that the conditions of Theorem 7 hold. Then the geometric center of the set of all the sample deepest points is unbiased estimator of a population center* $\boldsymbol{\theta}$.

*Proof.* As a consequence of symmetry (Theorem 7) it holds

$$\mathsf{E}\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) - \boldsymbol{\theta}\right) = -\mathsf{E}\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) - \boldsymbol{\theta}\right).$$

Hence

$$\mathsf{E}\,\widehat{\boldsymbol{\theta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \boldsymbol{\theta}.$$

■

*Remark* 7. In the latter corollary the geometric center functional can be replaced with any functional that assigns an unique point to any set of points and that is affine equivariant against orthogonal transformations of this set.

Theorem 6 and Example 7 show that the depth is biased. Fortunately, biaseness does not affect the "order" of the deepest point and thus the sample deepest point is *unbiased estimator.*

## 2.3 The Weighted Halfspaces Ratio Depth

In this section we propose depth function derived from the generalized halfspace depth which produces the unique deepest point and it still keeps some of the advantages of the generalized halfspace depth. Actually, our first proposal of the depth that allows us to have noncovex contours (and hence eliminate the cases when the deepest point lie outside the support of a distribution) was the depth defined in this section.

Majority of the results derived in this section are consequences of the results for the generalized halfspace depth. The generalized halfspace depth seems to be more intuitive and also posses better properties. This is the reason why we first introduced the generalized halfspace depth.

The main idea of the depth introduced in this section is to use *ratio of weighted probability of opposite halfspaces* instead of only weighted probability of a (half)space. Using same notation as in previous sections we define the *weighted halfsapces ratio depth* (WHRD) in the following manner.

**Definition 9** (WHRD)**.** Suppose a measurable bounded weight function

$$w : \mathbb{R}^p \times \mathcal{S}^p \longrightarrow \mathbb{R}.$$

The *weighted halfsapces ratio depth* of a point $\boldsymbol{x} \in \mathbb{R}^p$ with respect to the weight function $w$ and probability distribution $P$ of a random vector $\boldsymbol{X} \in \mathbb{R}^p$ is defined as

$$\mathrm{RD}_w(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, -\boldsymbol{u})}. \tag{2.17}$$

The *sample version* is defined as

$$\mathrm{RD}_{w,n}(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \frac{\mathsf{E}_n\, w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u})}{\mathsf{E}_n\, w(\boldsymbol{X}-\boldsymbol{x}, -\boldsymbol{u})} = \inf_{\|\boldsymbol{u}\|=1} \frac{\sum_{i=1}^n w(\boldsymbol{X}_i-\boldsymbol{x}, \boldsymbol{u})}{\sum_{i=1}^n w(\boldsymbol{X}_i-\boldsymbol{x}, -\boldsymbol{u})}.$$

The term $0/0$ is in both cases defined to be 1.

The definition, for the band weight function (see (2.5)), is illustrated on Fig. 2.8. Examples of contours for the weighted halfspaces ratio depth for the band and cone weight functions, (2.5) and (2.6), can be seen on Figures 1.12, 1.15 and 1.18.

Similarly, as in the case of generalized halfspace depth, we can obtain the halfspace depth if we choose the weight function to be equal to the indicator of a halfspace. But for this purpose it is better to use an alternative definition of WHRD:

**Definition 10** (WHRD II)**.** Define depth function

$$\widehat{\mathrm{RD}}_w(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u}) + \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, -\boldsymbol{u})}, \tag{2.18}$$

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

The depth functions RD and $\widehat{\mathrm{RD}}$ are equivalent in the sense of the multivariate ordering:

**Theorem 9.** *For any weight function $w$ and for all $\boldsymbol{x}$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \mathbb{R}^p$ the equivalence*

$$\mathrm{RD}_w(\boldsymbol{x}_1) \le \mathrm{RD}_w(\boldsymbol{x}_2) \iff \widehat{\mathrm{RD}}_w(\boldsymbol{x}_1) \le \widehat{\mathrm{RD}}_w(\boldsymbol{x}_2) \tag{2.19}$$

*holds. Moreover,*

$$\widehat{\mathrm{RD}}_w(\boldsymbol{x}) \le \frac{1}{2}, \tag{2.20}$$

*and*

$$\mathrm{RD}_w(\boldsymbol{x}) = \frac{\widehat{\mathrm{RD}}_w(\boldsymbol{x})}{1-\widehat{\mathrm{RD}}_w(\boldsymbol{x})}. \tag{2.21}$$

*Proof.* The proof of (2.20) is very similar to the proof of Theorem 10 that will follow:

$$\widehat{\mathrm{RD}}_w(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u}) + \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x}, -\boldsymbol{u})}$$

$$= \inf_{\|\boldsymbol{u}\|=1} \min \left\{ \frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u}) + \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},-\boldsymbol{u})}, \right.$$
$$\left. \frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},-\boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u}) + \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},-\boldsymbol{u})} \right\} \leq \frac{1}{2}.$$

Denote for fixed $\boldsymbol{u} \in \mathcal{S}^p$

$$v_+ = \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})$$

and

$$v_- = \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},-\boldsymbol{u}).$$

If $v_- > 0$ then

$$\frac{v_+}{v_-} = \frac{v_+}{v_- + v_+} \left( \frac{v_-}{v_- + v_+} \right)^{-1} = \frac{v_+}{v_- + v_+} \left( 1 - \frac{v_+}{v_- + v_+} \right)^{-1}. \tag{2.22}$$

If $v_- = 0$ and $v_+ > 0$ then $v_-$ and $v_+$ in (2.22) may be interchanged (see arguments for (2.20)).

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

$$\frac{v_+}{v_-} = 1,$$
$$\frac{v_+}{v_- + v_+} = \frac{1}{2}.$$

It follows

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

Equation(2.21) now follows from definition of $v_+$ and $v_-$.

Since the function

$$x \mapsto x/(1-x)$$

is increasing in $x$ for $x \in [0, 1/2]$, the equivalence (2.19) follows. ■

*Remark* 8. The previous theorem shows that the depth (2.18) is in some sense a direct generalisation of the halfspace depth. Indeed, the halfspace depth $\mathrm{HD}(\boldsymbol{x})$ is for absolutely continuous distributions equal to $\widehat{\mathrm{RD}}_w(\boldsymbol{x})$ for the weight function defined in equation (2.4) of Example 4.

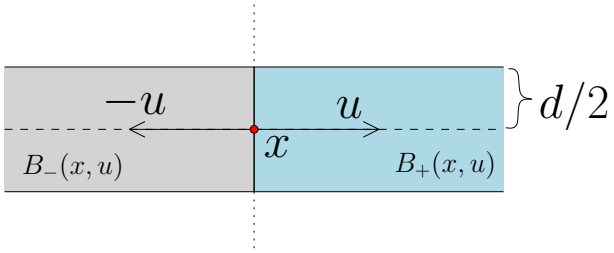



Figure 2.8: The band weight function for the weighted halfspaces ratio depth. Numerator function is in blue colour, denominator function in gray color.

**Example 10.** Consider the *band (cylinder) weight function* (see (2.5) in Example 4). In particular, for $\mathbb{R}^2$ the definition has the following meaning (see also Fig. 2.8). Given a fixed point $\boldsymbol{x} \in \mathbb{R}^2$ and a direction $\boldsymbol{u} \in \mathcal{S}^2$, we consider a line $l = \boldsymbol{x} + t\boldsymbol{u},\ t \in \mathbb{R}$ for which a band with width $d$

$$\mathbb{B}(\boldsymbol{x}, \boldsymbol{u}) = \{\boldsymbol{y} \in \mathbb{R}^2 :\ \mathrm{dist}(\boldsymbol{y}, l) \leq d/2\}$$

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

$$\mathrm{RD}_d(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \frac{p_+(\boldsymbol{x}, \boldsymbol{u})}{p_-(\boldsymbol{x}, \boldsymbol{u})}.$$

The sample version is calculated from the number of observations in $\mathbb{B}_+(\boldsymbol{x}, \boldsymbol{u})$ and $\mathbb{B}_-(\boldsymbol{x}, \boldsymbol{u})$. The depth from the alternative definition (Def. 10) is in the form

$$\widehat{\mathrm{RD}}_d(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \frac{p_+(\boldsymbol{x}, \boldsymbol{u})}{p_-(\boldsymbol{x}, \boldsymbol{u}) + p_+(\boldsymbol{x}, \boldsymbol{u})}.$$

The example is connected with our motivation mentioned in introduction chapter - with directional quantiles, see Section 5.1 for more informations. Suppose that a minimal direction exists and denote it by $\boldsymbol{u}_0$. Consider a representation of a point $(\boldsymbol{y} - \boldsymbol{x}) \in \mathbb{R}^p$ as $(r; \boldsymbol{u})$, in the form

$$\boldsymbol{y} - \boldsymbol{x} = r\boldsymbol{u},$$

where $r \in \mathbb{R}$ and $\boldsymbol{u} \in \mathcal{S}^p \cap \{\boldsymbol{v} :\ v_p \geq 0\}$. Latter representation is similar to a representation in spherical coordinates system. Representation of the random vector $\boldsymbol{X} - \boldsymbol{x}$ (random vector shifted to $\boldsymbol{x}$) we denote by $(R; \boldsymbol{U})$. Then it can be shown that it holds

$$\lim_{d \to 0+} \widehat{\mathrm{RD}}_d(\boldsymbol{x}) = \mathsf{P}\left(R \geq 0 | \boldsymbol{U} = \boldsymbol{u}_0\right).$$

Hence the deepest point is well balanced with respect to the conditional probability mass on lines going through this point. This property was one of our goals

when we searched for suitable centre for directional approach. The same result can be shown using the cone weight function (2.6) when $\alpha \to 0+$.

## 2.4 Properties of the Weighted Halfspaces Ratio Depth

Let us summarise some facts about the depth function RD. There is one advantage in comparison to the *generalized halfspace depth* - WHRD is always between 0 and 1.

**Theorem 10.** *Suppose a weight function $w \geq 0$. Then for any probability measure $P$ on $\mathbb{R}^p$ and any $\boldsymbol{x} \in \mathbb{R}^p$ it holds*

$$0 \leq \mathrm{RD}_w(\boldsymbol{x}; P) \leq 1.$$

*Proof.* The first inequality is clear.

For the second inequality it is not difficult to see that

$$\min\left\{\frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},-\boldsymbol{u})}, \frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},-\boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})}\right\} \leq 1$$

It follows

$$\begin{aligned}\mathrm{RD}_w(\boldsymbol{x}) &= \inf_{\|\boldsymbol{u}\|=1} \frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},-\boldsymbol{u})} \\ &= \inf_{\|\boldsymbol{u}\|=1} \min\left\{\frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},-\boldsymbol{u})}, \frac{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},-\boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})}\right\} \leq 1.\end{aligned}$$

■

Similarly as in the case of the generalized halfspace depth we can prove that WHRD vanishes at infinity. We need to assume additional assumptions.

**Theorem 11** (Vanishing at infinity)**.** *Suppose a continuous distribution and a spherically symmetric weight function. Further suppose that*

$$w(\boldsymbol{x},\boldsymbol{u}) = l\big(\|\boldsymbol{x} - \langle \boldsymbol{u},\boldsymbol{x}\rangle \boldsymbol{u}\|\big) m\big(\langle \boldsymbol{u},\boldsymbol{x}\rangle\big),$$

*where*

1. *$l: [0,+\infty) \longrightarrow [0,+\infty)$ is nonincreasing function, $l$ is bounded and $l(0) > 0$,*
2. *$m: \mathbb{R} \longrightarrow [0,+\infty)$ is nondecreasing function on $[0,+\infty)$, $m$ is bounded and*

$$\lim_{z\to-\infty} m(z) = 0.$$

*Then*

$$\lim_{\|\boldsymbol{x}\|\to+\infty} \mathrm{RD}_w(\boldsymbol{x}) = 0.$$

*Note* 9. Conditions from the theorem are only sufficient conditions.

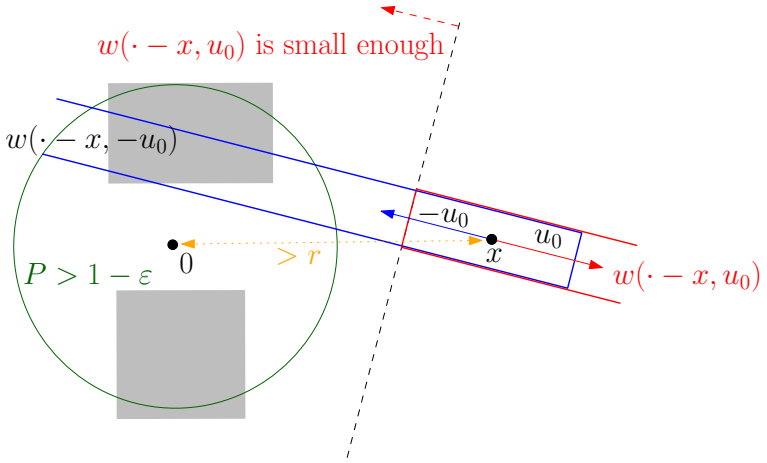



Figure 2.9: WHRD vanishes at infinity. Illustration for the proof of Theorem 11. Support is in gray color. The weight function in a direction $\boldsymbol{u}_0$ (resp. in a direction $-\boldsymbol{u}_0$) is marked with red color (resp. with blue color). The green circle denotes the sphere with probability greater than $1 - \varepsilon$.

*Proof of Theorem 11.* The proof is very similar to the proof of Theorem 1. Thus the proof is not shown here in detail. Figure 2.9 illustrates the ideas of the proof.

Briefly: Under theorem's assumptions the following hold. For any $\varepsilon > 0$ there exists $r = r_1 + r_2 > 0$ satisfying:

(a) $P(\{\boldsymbol{y} : \|\boldsymbol{y}\| > r_1\}) < \varepsilon$,

(b) if $\|\boldsymbol{x}\| > r$ then there exists a direction $\boldsymbol{u}_0$ such that

(c) $\mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, -\boldsymbol{u}_0) \gtrsim O(\varepsilon)$ (Since $m$ is nondecreasing on $[0, +\infty)$ and 1. holds.),

(d) $\mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}_0) \approx o(\varepsilon)$ (Since (a) and 2. hold we can move $\boldsymbol{x}$ until $r_2$ is high enough for validity of this property.).

Fact that upper bound for the depth of $\boldsymbol{x}$ is

$$\frac{\mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}_0)}{\mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, -\boldsymbol{u}_0)}$$

finishes the proof. ■

**Theorem 12.** *The depth function* $\mathrm{RD}_w$ *defined by* (2.17) *is translation invariant.*

*Proof.* It follows directly from the definition that

$$\mathrm{RD}_w(\boldsymbol{x}+\boldsymbol{a};\boldsymbol{X}+\boldsymbol{a}) = \mathrm{RD}_w(\boldsymbol{x};\boldsymbol{X}).$$

■

**Theorem 13.** *Suppose a spherically symmetric weight function. Then the depth function defined by* (2.17) *is rotation invariant.*

*Proof.* The proof follows the lines of the proof of Theorem 4. For the numerator and denominator the same equations as are in the proof hold. ■

We can obtain affine invariance property in the same way as in the case of the generalized halfspace depth (Definition 8 and Theorem 5).

**Theorem 14** (Affine invariant modification of WHRD)**.** *Suppose that the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ of a random vector $\boldsymbol{X}$ exists and that it is positive define. Further suppose a spherically symmetric weight function $w$. Then the depth function defined as*

$$\widetilde{\mathrm{RD}}_w(\boldsymbol{x};\boldsymbol{X}) = \mathrm{RD}_w(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{x};\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{X}) = \inf_{\|\boldsymbol{u}\|=1} \frac{\mathsf{E}\, w(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}(\boldsymbol{X}-\boldsymbol{x}),\boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}(\boldsymbol{X}-\boldsymbol{x}),-\boldsymbol{u})}, \quad (2.23)$$

*where, again, $\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}$ denotes an inverse square root of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ ($\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2} = \boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$), is affine invariant. In other words, for any nonsingular matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ it holds that*

$$\widetilde{\mathrm{RD}}_w(\mathbf{A}\boldsymbol{x};\mathbf{A}\boldsymbol{X}) = \widetilde{\mathrm{RD}}_w(\boldsymbol{x};\boldsymbol{X}).$$

The sample version is defined the same way as the sample version of the affine invariant modification of the generalized halfspace depth - an affine equivariant sample dispersion matrix is substituted for the covariance matrix.

*Proof of Theorem 14.* The proof is almost the same as the proof of Theorem 5. Only one difference is in the denominator. In the first part of the proof we use relation $\boldsymbol{\Gamma}_{-\boldsymbol{u}} = -\boldsymbol{\Gamma}_{\boldsymbol{u}}$. If we follow the lines of the proof, the expression (2.15) will be in the form

$$\widetilde{\mathrm{RD}}_w(\mathbf{A}\boldsymbol{x};\mathbf{A}\boldsymbol{X}) = \inf_{\boldsymbol{\Gamma}\in\mathcal{O}_p} \frac{\mathsf{E}\, \widetilde{w}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A}(\boldsymbol{X}-\boldsymbol{x}))}{\mathsf{E}\, \widetilde{w}(-\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A}(\boldsymbol{X}-\boldsymbol{x}))}.$$

From equation (2.16) it follows that $-\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}}^{-1/2}\mathbf{A}$ is also orthogonal matrix and hence

$$\widetilde{\mathrm{RD}}_w(\mathbf{A}\boldsymbol{x};\mathbf{A}\boldsymbol{X}) = \widetilde{\mathrm{RD}}_w(\boldsymbol{x};\boldsymbol{X}).$$

That finishes the part of the proof when $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \mathbf{I}$.

The part of the proof for a general covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ remains the same. ■

*Remark* 10. Remark 5 and the thoughts in the text preceding the remark also hold here, i.e. for any distribution we can find a weight function such that the depth is zero outside the support of the distribution.

In previous sections we saw that the generalized halfspace depth does not have to attain its maximum value in the center even in the case of symmetric distributions. One of the advantage of WHRD is that it provides us with the deepest point in the center of symmetry (see Theorems 15, 16). There is also usually the unique deepest point, especially for symmetric distributions. On the other hand WHRD in general does not give us a unique deepest point even in a situation of an absolutely continuous distribution with connected support.

**Example 11.** Let us consider a uniform distribution on a set

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

Let us consider the band weight function (see (2.5) and Example 10) with a small $d$, say $d = 1/10$, and the corresponding WHRD 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 $\mathrm{RD}(\boldsymbol{x}) \le 1/9$.

Consider the point $\boldsymbol{z} = (5, 1/2)^T$. After some calculations we get $\mathrm{RD}(\boldsymbol{z}) > 1/9 \ge \mathrm{RD}(\boldsymbol{x})$ for any $\boldsymbol{x} = (x_1, x_1)^T$. Indeed, the lower estimate for $\mathrm{RD}(\boldsymbol{z})$ may be obtained considering a line $l$ connecting $\boldsymbol{z}$ and the point $(0, 10)^T$ together with a band of the width $d$ around $l$ and, on the other hand considering a line $l'$ connecting $\boldsymbol{z}$ and the point $(5, 0)^T$ with the same band around. See Figure 2.10 for a visualisation of this example.

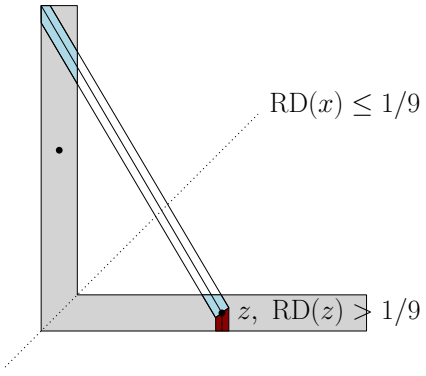



Figure 2.10: The deepest point need not to be unique, see Example 11.

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.

In general the WHRD does not satisfy two of the key properties. The depth is not affine invariant. Although, as shown in the previous theorem, an affine

invariant modification exists. And it needs not to decrease along the rays beginning in the deepest point. In Example 11 there is not any "natural" deepest point. On the other hand, if there exists an intuitive deepest point, like the point of central symmetry, we would like to prove that it is the deepest point for the WHRD. Indeed it is, in contrast to the generalized halfspace depth, the case for a suitable weight function.

**Theorem 15.** *Let the weight function $w$ be spherically symmetric and suppose that the distribution of $\boldsymbol{X}$ is centrally symmetric about a point $\boldsymbol{\theta}$. Then*

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

*Proof.* It can be assumed that $\boldsymbol{\theta} = \mathbf{0}$ without loss of generality (the translation invariance of depth function $\mathrm{RD}_w$). Since $w$ is spherically symmetric it holds

$$w(\boldsymbol{x}, \boldsymbol{u}) = w(-\boldsymbol{x}, -\boldsymbol{u}), \quad \forall \boldsymbol{x} \in \mathbb{R}^p, \ \boldsymbol{u} \in \mathcal{S}^p. \tag{2.24}$$

It follows that

$$\mathsf{E}\, w(\boldsymbol{X}, \boldsymbol{u}) = \mathsf{E}\, w(-\boldsymbol{X}, -\boldsymbol{u}) = \mathsf{E}\, w(\boldsymbol{X}, -\boldsymbol{u}).$$

The first equality is the consequence of (2.24), the second equality is the consequence of the fact that $\boldsymbol{X}$ is centrally symmetric about $\mathbf{0}$. Thus $\mathrm{RD}_w(\mathbf{0}) = 1$. The fact that $\mathrm{RD}_w(\boldsymbol{x}) \leq 1$, $\forall \boldsymbol{x}$, (see Theorem 10) completes the proof. ■

Latter result may be extended to angular symmetric distributions.

**Theorem 16.** *Let $w$ be spherically symmetric and suppose that the distribution of $\boldsymbol{X}$ is angular symmetric about point $\boldsymbol{\theta}$. If $w$ is such that*

$$w(k\boldsymbol{x}, \boldsymbol{u}) = w(\boldsymbol{x}, \boldsymbol{u}), \quad \forall \boldsymbol{x} \in \mathbb{R}^p, \ k \geq 0 \tag{2.25}$$

*then*

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

*Proof.* It is analogous to the proof of Theorem 15. Let $\boldsymbol{\theta} = \mathbf{0}$ without loss of generality. Under the assumption (2.25) it holds

$$\begin{aligned} \mathsf{E}\, w(\boldsymbol{X}, -\boldsymbol{u}) &= \mathsf{E}\, w(-\boldsymbol{X}, \boldsymbol{u}) = \mathsf{E}\, w(-\boldsymbol{X}/\|\boldsymbol{X}\|, \boldsymbol{u}) \\ &= \mathsf{E}\, w(\boldsymbol{X}/\|\boldsymbol{X}\|, \boldsymbol{u}) = \mathsf{E}\, w(\boldsymbol{X}, \boldsymbol{u}) \end{aligned}$$

for all $\boldsymbol{u} \in \mathcal{S}^p$. Hence

$$\mathrm{RD}_w(\mathbf{0}) = \inf_{\|\boldsymbol{u}\|=1} \frac{\mathsf{E}\, w(\boldsymbol{X}, \boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X}, -\boldsymbol{u})} = 1.$$

■

*Remark* 11. Condition (2.25) says that a weight function is constant along the rays beginning in the origin.

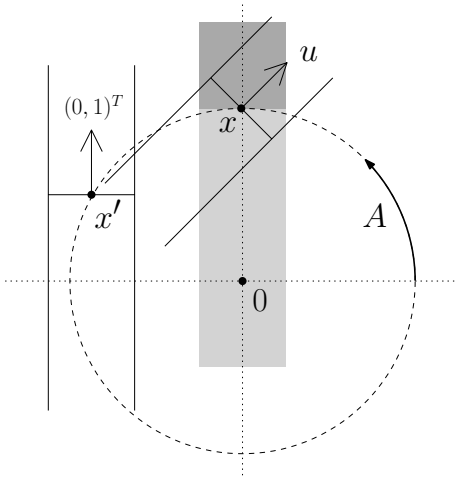



Figure 2.11: The band WHRD for bivariate standard Gaussian distribution. The gray coloured band is the minimal band.

It is not always easy to calculate the depth of arbitrary point $\boldsymbol{x} \in \mathbb{R}^p$. We finish this section with an example similar to Example 6. Now the calculation is easier. Note that it does not hold in general that calculation of WHRD is easier than the generalized halfspace depth calculation.

**Example 12.** Let $\boldsymbol{X}$ be a two dimensional random vector with Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$. Suppose we have the band weight function (2.5) for given $d > 0$. We use the same notation as in Example 10, hence

$$\mathrm{RD}_d(\boldsymbol{x}) = \inf_{\|\boldsymbol{u}\|=1} \frac{p_+(\boldsymbol{x}, \boldsymbol{u})}{p_-(\boldsymbol{x}, \boldsymbol{u})}. \tag{2.26}$$

Remind that $p_+(\boldsymbol{x}, \boldsymbol{u})$ denotes the probability of the band in the direction $\boldsymbol{u}$ from the point $\boldsymbol{x}$ and with width $d$, $p_-(\boldsymbol{x}, \boldsymbol{u})$ denotes the band in opposite direction. First we show that the minimal direction is the direction along the ray connecting the origin with point $\boldsymbol{x}$, i.e. that for an arbitrary point $\boldsymbol{x}$ it holds

$$\mathrm{RD}_d(\boldsymbol{x}) = \min\left\{\frac{p_+(\boldsymbol{x}, \boldsymbol{u}_0)}{p_-(\boldsymbol{x}, \boldsymbol{u}_0)}, \frac{p_-(\boldsymbol{x}, \boldsymbol{u}_0)}{p_+(\boldsymbol{x}, \boldsymbol{u}_0)}\right\}$$

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

$$w((x_1, x_2)^T, (0, 1)^T) = \begin{cases} 1, & \text{if } -d/2 < x_1 < d/2,\ x_2 > 0, \\ 0, & \text{otherwise.} \end{cases}$$

In what follows see Fig. 2.11 as an illustration of ideas used in this example. For any other direction $\boldsymbol{u} \neq (0, 1)^T$ there exists an uniquely determined orthogonal rotation $\mathbf{A} \in \mathbb{R}^{p\times p}$ such that $\mathbf{A}\boldsymbol{u} = (0, 1)^T$ and $\mathbf{A}\boldsymbol{X} = \boldsymbol{X}' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$. For

$\boldsymbol{x} = (0, x_2)^T$ it holds $\mathbf{A}\boldsymbol{x} = \boldsymbol{x}'$ where $x_2 > x_2'$. It holds that

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

where $\Phi$ denotes the distribution function of $\mathcal{N}(0,1)$. 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

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

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

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

The depth does not depend on the value of $d$ and it is equivalent to the halfspace depth (see Theorem 14).

## 2.5 Computational Aspects

We shortly discuss the computational aspects of the sample depth computation. Since the weighted halfspace depths (GHD, WHRD) are defined for a broad class of weight functions, a general fast algorithm for the depth computing does not exist. Also, the theoretical depth of a 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 $w(\boldsymbol{x}, \boldsymbol{u})$ can attain different values for every transformation $\boldsymbol{u} \in \mathcal{S}^p$, which means that possibly uncountable number of values must be considered.

On the other hand, in some special cases the sample depth may be calculated exactly. It is the case when the weight function is piecewise constant. The cone depth (2.6), the band depth (2.5), the halfspace depth (2.4) are, in particular, examples of such depths. The set $\{\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}), \boldsymbol{u}),\ \boldsymbol{u} \in \mathcal{S}^p\}$ is finite for each $\boldsymbol{x}$ in such a case.

Straightforward algorithm is used to calculate the sample depth of a given point $\boldsymbol{x}$. It uses a predefined number of vectors in $\mathcal{S}^p$ which represent directions (usually halfspaces) in which we compute sample weighted probability. For every such vector we calculate the sample weighted probability. The depth is set to the smallest value of weighted probability (GHD) or of portions of sample weighted probabilities in directions $\boldsymbol{u}$ and $-\boldsymbol{u}$ (WHRD). For sample size $n$ the calculation of weighted probability takes $O(n)$ steps. There are $k$ directions (we usually set $k = 1000$, which brings very precise answer), hence computation of the depth of

a given point takes $O(kn)$ steps. If one wants to compute the depth of all points in dataset it takes $O(kn^2)$ steps.

When we use Fortran and C programming languages and when sample size $n = 1000$ then the sample depth calculation of all points in the dataset takes about 20 – 50 seconds on average current (2014) computer. In R, MATLAB and similar programs, the calculation usually takes much longer (in R almost 100 fold increase of computational time). For the sample depth calculation the routines were programmed in Fortran and C and they were compiled as dynamic libraries. These libraries were linked to R via R functions. Hence for work with data the R programming language was used. Source files for Fortran, C and R can be provided by the author if requested. For a large datasets in higher dimension computation speed improvement can be achieved if GPU parallel computing (e.g. NVIDIA CUDA platform) is used.

# Chapter 3

# Consistency and Some Other Asymptotical Properties

The strong consistency of the sample version of the proposed depth functions is discussed in this chapter. First the consistency of the generalized halfspace depth is shown. The consistency of the weighted halfspaces ratio depth is obtained by slight expansion of the results and statements in the proofs of consistency of the generalized halfspace depth. As we will see, compared to former depth, the latter depth need not to be uniformly consistent over $\mathbb{R}^p$.

The proof for both depth functions is based on *Uniform law of large numbers*. For more details about this law see Appendix. Results about strong consistency first appeared in [Kotík, 2009] and in [Hlubinka et al., 2010].

## 3.1 Consistency of the Sample Generalized Halfspace Depth

In this section we prove that $\mathrm{D}_{w,n}(\boldsymbol{x})$ is *uniformly strong consistent*, i.e.,

$$\sup_{\boldsymbol{x}\in\mathbb{R}^p} \left|\mathrm{D}_{w,n}(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})\right| \xrightarrow{n\to\infty} 0 \quad \text{almost sure}$$

for *any* absolutely continuous multivariate distribution, for *regular* weight function (as stated in Theorem 17), and without any assumption on moments.

*Note* 12. The regularity of the weight function needs some continuity assumption. Since the weight function $w$ is defined on $\mathbb{R}^p \times \mathcal{S}^p$ the convergence

$$(\boldsymbol{y}^T, \boldsymbol{u}^T)^T \longrightarrow (\boldsymbol{x}^T, \boldsymbol{v}^T)^T$$

on $\mathbb{R}^p \times \mathcal{S}^p$ is defined in the sense

$$\|\boldsymbol{x} - \boldsymbol{y}\| + \measuredangle(\boldsymbol{u}, \boldsymbol{v}) \longrightarrow 0. \tag{3.1}$$

**Theorem 17** (Uniform strong consistency)**.** *Consider a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from an absolutely continuous distribution $P$ and suppose a weight function*

*$w$ satisfying*

$$\lim_{(\boldsymbol{y}^T,\boldsymbol{u}^T)^T\to(\boldsymbol{x}^T,\boldsymbol{v}^T)^T} w(\boldsymbol{y},\boldsymbol{u}) = w(\boldsymbol{x},\boldsymbol{v}) \quad \textit{for a.a. } \boldsymbol{x}\in\mathbb{R}^p \textit{ and for all } \boldsymbol{v}\in\mathcal{S}^p. \tag{3.2}$$

*Further suppose that*

$$\lim_{k\to\infty} \sup_{\boldsymbol{u}\in\mathcal{S}^p} \sup_{\boldsymbol{x}:\langle\boldsymbol{x},\boldsymbol{u}\rangle\le -k} w(\boldsymbol{x},\boldsymbol{u}) = 0. \tag{3.3}$$

*Then*

$$\sup_{\boldsymbol{x}\in\mathbb{R}^p} \left|\mathrm{D}_{w,n}(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})\right| \xrightarrow{n\to+\infty} 0 \quad \textit{almost surely.} \tag{3.4}$$

*Note* 13. Recall that the weight function is supposed to be measurable and bounded (see Definition 5).

*Note* 14. Equation (3.3) means that for arbitrary $\boldsymbol{u}\in\mathcal{S}^p$ a function $w(\cdot,\boldsymbol{u})$ *uniformly vanishes at infinity* in the halfspace with normal vector given by $-\boldsymbol{u}$. Every weight function should posses this property.

Note also that for the spherically symmetric weight functions the condition (3.3) is equivalent to the condition (2.11), i.e. the convergence is always uniform in $\boldsymbol{u}\in\mathcal{S}^p$.

Finally note that for weight functions like halfspace, band (cylinder), kernel band (see Example 4) it holds $w(x,u)=0$ if $\langle\boldsymbol{x},\boldsymbol{u}\rangle<0$ and (3.3) holds trivially.

*Proof of Theorem 17.* First the uniform strong consistency over a compact set, (3.5), is shown. This result is then extended to whole $\mathbb{R}^p$.

Now we show the uniform strong consistency over any compact set, i.e., for arbitrary compact set $C\subset\mathbb{R}^p$ it holds that

$$\sup_{\boldsymbol{x}\in C} \left|\mathrm{D}_{w,n}(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})\right| \xrightarrow{n\to+\infty} \text{a.s.} \tag{3.5}$$

It is sufficient to show that the class of functions

$$\mathcal{G} = \{\boldsymbol{z}\mapsto w(\boldsymbol{z}-\boldsymbol{x},\boldsymbol{u}):\ \boldsymbol{u}\in\mathcal{S}^p,\ \boldsymbol{x}\in C\} \tag{3.6}$$

satisfies the *Uniform law of large numbers* (ULLN) with respect to the probability measure $P$. Indeed, if the class (3.6) satisfies the ULLN then (see Appendix and Theorem 31 in particular)

$$\sup_{\boldsymbol{u}\in\mathcal{S}^p,\ \boldsymbol{x}\in C} \left|\frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i-\boldsymbol{x},\boldsymbol{u}) - \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})\right| \xrightarrow{n\to+\infty} 0 \quad \text{a.s.} \tag{3.7}$$

and since (see Lemma 29 for the following inequality)

$$\sup_{\boldsymbol{x}\in C} \left|\inf_{\boldsymbol{u}\in\mathcal{S}^p} \frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i-\boldsymbol{x},\boldsymbol{u}) - \inf_{\boldsymbol{u}\in\mathcal{S}^p} \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})\right|$$

$$\leq \sup_{\boldsymbol{x}\in C,\ \boldsymbol{u}\in\mathcal{S}^p} \left| \frac{1}{n}\sum_{i=1}^{n} w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) - \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) \right| \xrightarrow{n\to+\infty} 0 \quad \text{a.s.} \tag{3.8}$$

We will prove that the standard *bracketing* sufficient condition for the ULLN (see Theorem 31) holds for the class (3.6), i.e., that

$$H_{1,B}(\varepsilon, \mathcal{G}, P) < +\infty, \quad \forall \varepsilon > 0,$$

where $H_{1,B}(\varepsilon, \mathcal{G}, P)$ denotes *entropy* with $\varepsilon$*–bracketing* of $\mathcal{G}$ for $\mathcal{L}_1(P)$–metric.

The bracketing functions are constructed as local suprema and infima of (3.6). Define functions $W^U_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{z};\varphi,\xi)$ and $W^L_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{z};\xi,\varphi)$ for all $\boldsymbol{s}\in\mathcal{S}^p$, $\boldsymbol{x}\in C$ and $0 < \varphi < \pi$, $\xi > 0$ as

$$W^U_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{z};\xi,\varphi) = \sup\big\{w(\boldsymbol{z}-\boldsymbol{a},\boldsymbol{u}) :\ \boldsymbol{a}\in\mathbb{R}^p, \|\boldsymbol{a}-\boldsymbol{x}\| \leq \xi;\ \boldsymbol{u}\in\mathcal{S}^p, \measuredangle(\boldsymbol{u},\boldsymbol{s}) \leq \varphi\big\},$$
$$W^L_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{z};\xi,\varphi) = \inf\big\{w(\boldsymbol{z}-\boldsymbol{a},\boldsymbol{u}) :\ \boldsymbol{a}\in\mathbb{R}^p, \|\boldsymbol{a}-\boldsymbol{x}\| \leq \xi;\ \boldsymbol{u}\in\mathcal{S}^p, \measuredangle(\boldsymbol{u},\boldsymbol{s}) \leq \varphi\big\}.$$

It is clear that

$$W^L_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{z};\xi,\varphi) \leq w(\boldsymbol{z}-\boldsymbol{a},\boldsymbol{u}) \leq W^U_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{z};\xi,\varphi) \tag{3.9}$$

holds for all $\boldsymbol{z}\in\mathbb{R}^p$ and for all $\boldsymbol{a},\boldsymbol{u}$ such that $\|\boldsymbol{a}-\boldsymbol{x}\| < \xi$ and $\measuredangle(\boldsymbol{u},\boldsymbol{s}) < \varphi$.

It follows from Lemma 30 that $W^L_{\boldsymbol{x},\boldsymbol{s}}(\cdot;\xi,\varphi)$ and $W^U_{\boldsymbol{x},\boldsymbol{s}}(\cdot;\xi,\varphi)$ are nonnegative *measurable* functions on $\mathbb{R}^p$ for any $\boldsymbol{x}\in\mathbb{R}^p, \boldsymbol{u}\in\mathcal{S}^p, \xi > 0, \varphi > 0$.

The next step is to show that for any $\varepsilon > 0$ and for an arbitrary unit vector $\boldsymbol{s}\in\mathcal{S}^p$ and any $\boldsymbol{x}\in C$ there exist $\xi,\varphi > 0$ such that bracketing functions $W^L_{\boldsymbol{x},\boldsymbol{s}}(\cdot;\xi,\varphi)$ and $W^U_{\boldsymbol{x},\boldsymbol{s}}(\cdot;\xi,\varphi)$ are $\varepsilon$-close to $w(\cdot-\boldsymbol{x},\boldsymbol{s})$ in $\mathcal{L}_1(P)$ metric. Indeed, as $W^U$ and $W^L$ are bounded measurable functions it is possible to write

$$\lim_{\substack{\varphi\to0+\\ \xi\to0+}} \mathsf{E}\, W^U_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{X};\xi,\varphi) = \mathsf{E} \lim_{\substack{\varphi\to0+\\ \xi\to0+}} W^U_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{X};\xi,\varphi) = \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{s}). \tag{3.10}$$

We can change the limit and the integral sign in (3.10) because of boundedness and measurability of $W^U_{\boldsymbol{x},\boldsymbol{s}}(\cdot;\xi,\varphi)$ by use of *Lebesgue's dominated convergence theorem.* The second equation of (3.10) is a consequence of the continuity condition (3.2) (recall that we are considering depth for $P$ being absolutely continuous distribution).

Consider any fixed vectors $\boldsymbol{x},\boldsymbol{s}$. It follows from (3.10) that for all $\varepsilon > 0$ there exists $\varphi_\varepsilon > 0$ and $\xi_\varepsilon > 0$ such that for all $\varphi\in[0,\varphi_\varepsilon)$ and all $\xi\in[0,\xi_\varepsilon)$ it holds

$$\mathsf{E}\left|W^U_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{X};\xi,\varphi) - w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{s})\right| = \left|\mathsf{E}\, W^U_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{X};\xi,\varphi) - \mathsf{E}\, w(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{s})\right| < \varepsilon.$$

The first equality follows from the inequality (3.9) - the difference is greater than 0. An analogous result holds for $W^L_{\boldsymbol{x},\boldsymbol{s}}(\cdot;\xi,\varphi)$ as well.

Hence for arbitrary $\boldsymbol{x} \in C$ and arbitrary $\boldsymbol{s} \in \mathcal{S}^p$ and any $\varepsilon > 0$ there is a pair $\xi_\varepsilon(\boldsymbol{x}, \boldsymbol{s}) > 0$ and $\varphi_\varepsilon(\boldsymbol{x}, \boldsymbol{s}) > 0$ such that

$$\sup_{\xi < \xi_\varepsilon(\boldsymbol{x},\boldsymbol{s}),\ \varphi < \varphi_\varepsilon(\boldsymbol{x},\boldsymbol{s})} \mathsf{E} \left| W^U_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{X}; \xi, \varphi) - W^L_{\boldsymbol{x},\boldsymbol{s}}(\boldsymbol{X}; \xi, \varphi) \right| < \varepsilon. \tag{3.11}$$

Now we show that the entropy with $\varepsilon$-bracketing of $\mathcal{G}$ is finite. Choose any $\varepsilon > 0$, and for all couples $\boldsymbol{x} \in C$, $\boldsymbol{s} \in \mathcal{S}^p$ define sets

$$\mathcal{U}_\varepsilon(\boldsymbol{x}, \boldsymbol{s}) = \{\boldsymbol{a} \in \mathbb{R}^p : \|\boldsymbol{a} - \boldsymbol{x}\| < \xi_\varepsilon(\boldsymbol{x}, \boldsymbol{s})\} \times \{\boldsymbol{u} \in \mathcal{S}^p : \measuredangle(\boldsymbol{u}, \boldsymbol{s}) < \varphi_\varepsilon(\boldsymbol{x}, \boldsymbol{s})\}.$$

The sets $\mathcal{U}_\varepsilon(\boldsymbol{x}, \boldsymbol{s})$, $(\boldsymbol{x}, \boldsymbol{s}) \in C \times \mathcal{S}^p$, form an *open covering* of the set $C \times \mathcal{S}^p$, i.e.,

$$C \times \mathcal{S}^p \subseteq \bigcup_{\boldsymbol{x} \in C,\, \boldsymbol{s} \in \mathcal{S}^p} \mathcal{U}_\varepsilon(\boldsymbol{x}, \boldsymbol{s}),$$

and as $C \times \mathcal{S}^p$ is a *compact* subset of the metric space $(\mathbb{R}^{2p}, d_{2p})$ ($d_{2p}$ denotes the usual *Euclid distance*) it is possible to find a *finite subcovering*. It means that there exists finite $N \in \mathbb{N}$ and points $\boldsymbol{x}_i \in C$, $\boldsymbol{s}_i \in \mathcal{S}^p$, $i = 1, \ldots, N$ such that

$$C \times \mathcal{S}^p \subseteq \bigcup_{i=1,\ldots,N} \mathcal{U}_\varepsilon(\boldsymbol{x}_i, \boldsymbol{s}_i).$$

It follows that for fixed $\varepsilon > 0$ and any function $w(\cdot - \boldsymbol{a}, \boldsymbol{u}) \in \mathcal{G}$ there exists $k \in \{1, \ldots, N\}$ such that $(\boldsymbol{a}, \boldsymbol{u}) \in \mathcal{U}_\varepsilon(\boldsymbol{x}_k, \boldsymbol{s}_k)$. Hence, it holds

$$\begin{aligned} W^L_{\boldsymbol{x}_k,\boldsymbol{s}_k}\big(\boldsymbol{z}; \xi_\varepsilon(\boldsymbol{x}_k, \boldsymbol{s}_k), \varphi_\varepsilon(\boldsymbol{x}_k, \boldsymbol{s}_k)\big) &\le w(\boldsymbol{z} - \boldsymbol{a}, \boldsymbol{u}), \quad \forall \boldsymbol{z} \in \mathbb{R}^p \\ w(\boldsymbol{z} - \boldsymbol{a}, \boldsymbol{u}) &\le W^U_{\boldsymbol{x}_k,\boldsymbol{s}_k}\big(\boldsymbol{z}; \xi_\varepsilon(\boldsymbol{x}_k, \boldsymbol{s}_k), \varphi_\varepsilon(\boldsymbol{x}_k, \boldsymbol{s}_k)\big), \quad \forall \boldsymbol{z} \in \mathbb{R}^p. \end{aligned} \tag{3.12}$$

It follows that the pairs of the functions $(W^L, W^U)$ in (3.12) for $i = 1, \ldots, N$ form an *$\varepsilon$–bracketing* for $\mathcal{G}$, and since

$$H_{1,B}(\varepsilon, \mathcal{G}, P) \le \log N < +\infty$$

it follows that (3.7) (the *Uniform strong consistency* of the sample depth over a compact set $C$) is proved.

Now we prove that (3.7) holds on the whole $\mathbb{R}^p$. *Uniform strong consistency* of the sample depth over the whole space $\mathbb{R}^p$ (3.4) follows from the uniform strong consistency over an arbitrary compact set together with the *tightness* of a probability measure, and condition (3.3).

Choose arbitrary $\varepsilon > 0$. Consider a closed ball $C = \{\boldsymbol{x} : \|\boldsymbol{x}\| \le r\} \subset \mathbb{R}^p$ such that $P(C^c) < \varepsilon$. Take $K > 0$ large enough such that $w(\boldsymbol{x}, \boldsymbol{u}) < \varepsilon$ for any $\boldsymbol{x} \in \mathbb{R}^p$, $\boldsymbol{u} \in \mathcal{S}^p$ such that $\langle \boldsymbol{x}, \boldsymbol{u} \rangle < -K$. Denote $C_K$ the closed $K$ neighborhood of $C$, i.e.

$$C_K = \{\boldsymbol{x} : \|\boldsymbol{x}\| \le r + K\}.$$

Then the following 2 properties hold for all $\omega \in A$, where the set $A$ satisfies $\mathsf{P}(A) = 1$. $\omega$-notation is suppressed in the following lines.

1. There exists $n_1 \in \mathbb{N}$ such that $\sup_{\boldsymbol{x} \in C_K} |\mathrm{D}_{w,n}(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})| < \varepsilon$ for all $n \geq n_1$.
2. There exists $n_2 \in \mathbb{N}$ such that $P_{\mathrm{n}}(C^c) < 2\varepsilon$ for all $n \geq n_2$ since $P(C^c) < \varepsilon$. This follows from the *Glivenko-Cantelli* theorem applied to the random sample $\|\boldsymbol{X}_1\|, \|\boldsymbol{X}_2\|, \ldots, \|\boldsymbol{X}_n\|, \ldots$.

As the weight function is considered bounded, say $w(\boldsymbol{x}, \boldsymbol{u}) \leq b$, it follows that for all $\boldsymbol{x} \in C^c$ it holds that

$$\begin{aligned} \mathrm{D}_{w,n}(\boldsymbol{x}) &\leq b\, P_{\mathrm{n}}(C^c) + \varepsilon, \\ \mathrm{D}_w(\boldsymbol{x}) &\leq b\, P(C^c) + \varepsilon \end{aligned}$$

and therefore there exists some constant $B$ which does not depend on $\varepsilon$ such that it almost sure holds that for all $n \geq n_2$

$$\sup_{x \in C_K^c} \left|\mathrm{D}_{w,n}(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})\right| \leq b\Big(P_{\mathrm{n}}(C^c) + P(C^c)\Big) + 2\varepsilon < B\varepsilon. \tag{3.13}$$

So eventually, for all $\varepsilon > 0$ it holds almost sure: there exist $C_K \subset \mathbb{R}^p$ and $n_0 = \max\{n_1, n_2\}$ such that for all $n \geq n_0$

$$\begin{aligned} &\sup_{\boldsymbol{x} \in \mathbb{R}^p} \left|\mathrm{D}_{w,n}(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})\right| \\ &\qquad \leq \max\left\{ \sup_{\boldsymbol{x} \in C_K} \left|\mathrm{D}_{w,n}(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})\right|, \sup_{\boldsymbol{x} \in C_K^c} \left|\mathrm{D}_{w,n}(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})\right| \right\} \\ &\qquad < B\varepsilon. \end{aligned}$$

These lines finishes the proof of uniform almost sure convergence over $\mathbb{R}^p$, (3.4). ■

The condition (3.2) seems to be quite technical. For usual and reasonable choice of *piecewise continutous* functions the theorem can be rewritten as follows.

**Corollary 18** (Consistency of piecewise continuous weight functions)**.** *Suppose that a weight function $w$ is spherically symmetric (Definition 7) and that a function $h$ from (2.3) is piecewise continuous function (see Definition 6). Further suppose*

$$\lim_{t \to -\infty} \sup_{s \geq 0} h(s,t) = 0. \tag{3.14}$$

*Then conditions (3.2) and (3.3) of Theorem 17 hold and the generalized halfspace depth is uniformly strongly consistent.*

*Proof.* Condition (3.14) immediately implies the validity of condition (3.3). It only remains to show that condition (3.2) holds. Since function

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{pmatrix} \longmapsto \begin{pmatrix} \|\boldsymbol{x} - \langle \boldsymbol{u}, \boldsymbol{x} \rangle \boldsymbol{u}\| \\ \langle \boldsymbol{u}, \boldsymbol{x} \rangle \end{pmatrix}$$

is continuous and also function $h$ is continuous almost everywhere on $[0, +\infty) \times \mathbb{R}$, the set of all discontinuity points of

$$w(\boldsymbol{x}, \boldsymbol{u}) = h(\|\boldsymbol{x} - \langle \boldsymbol{u}, \boldsymbol{x} \rangle \boldsymbol{u}\|, \langle \boldsymbol{u}, \boldsymbol{x} \rangle)$$

has zero Lebesgue measure. ■

*Remark* 15. Uniform strong consistency holds for any weight function mentioned in this thesis, i.e. for the halfspace depth, the kernel density estimates with a fixed bandwidth, the band depth, the (weight) conic section depth and others. All mentioned depths use weight functions that are continuous or piecewise continuous, hence conditions of Corollary 18 hold.

*Remark* 16. If a weight function $w$ satisfies the assumptions from Theorem 17, then for known covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ the affine invariant version of the depth (see Definition 8 and Theorem 5) is also uniformly strongly consistent.

*Proof.* Since

$$\widetilde{\mathrm{D}}_w(\boldsymbol{x}) = \inf_{\boldsymbol{u}:\|\boldsymbol{x}\|=1} \mathsf{E}\, w\left(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}(\boldsymbol{X} - \boldsymbol{x}), \boldsymbol{u}\right),$$

it only remains to realize that the weight function

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{pmatrix} \longmapsto w\left(\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1/2}\boldsymbol{x}, \boldsymbol{u}\right)$$

satisfies the conditions from Theorem 17. ■

## 3.2 Consistency of the Sample Weighted Halfspaces Ratio Depth

We use results from the previous section to prove consistency of the sample version of the weighted halfspaces ratio depth. Let us first denote two important subsets of points. Define

$$\mathcal{H}_1 = \{\boldsymbol{x} : \inf_{\boldsymbol{u}:\ \|\boldsymbol{u}\|=1} \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) > 0\},$$

$$\mathcal{H}_2 = \{\boldsymbol{x} : \exists \delta > 0,\ \exists \boldsymbol{u}_{\boldsymbol{x}} \in \mathcal{S}^p,\ \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}_{\boldsymbol{x}}) = 0 \text{ and } \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, -\boldsymbol{u}_{\boldsymbol{x}}) > \delta\}.$$

A constant $\delta$ does not depend on $\boldsymbol{x}$.

*Remark* 17. The set $\mathcal{H}_1$ contains the interior of support $\mathrm{sp}(P)$, i.e. points whose open neighbourhood is contained in the support of $P$. In the case of absolutely continuous distribution $P(\mathcal{H}_1) = 1$. On the other hand the set $\mathcal{H}_2$ represents points with zero depth and $(\mathrm{csp}(P))^c \subset \mathcal{H}_2$ under very weak conditions on the weight function $w$. Here $\mathrm{csp}(P)$ denotes the *convex closure* of $\mathrm{sp}(P)$. If $\boldsymbol{x} \in \mathcal{H}_1$ then $\mathrm{RD}(\boldsymbol{x}) > 0$ and $\sup_{\boldsymbol{x} \in \mathcal{H}_2} \mathrm{RD}(\boldsymbol{x}) = 0$. Usually, $\mathcal{H}_1 \cup \mathcal{H}_2 = \mathbb{R}^p$.

**Theorem 19** (Consistency of WHRD)**.** *Consider a random sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ from absolutely continuous distribution $P$ and suppose a weight function satisfying the conditions of Theorem 11. Then*

$$\sup_{\boldsymbol{x}\in\mathcal{H}_1\cup\mathcal{H}_2} \left|\mathrm{RD}_{w,n}(\boldsymbol{x}) - \mathrm{RD}_w(\boldsymbol{x})\right| \xrightarrow{n\to+\infty} 0 \quad \textit{almost surely.} \tag{3.15}$$

*Proof.* The proof use results from the proof of the consistency of the generalized halfspace depth which was presented in the previous section. In particular we use ULLN property stated in (3.7) which was proved in the proof of Theorem 17.

For our purposes we will use standard conventions from the 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: $\log 0 := \lim_{x\to 0+} \log x = -\infty$.
Let us, for any $\boldsymbol{x} \in \mathcal{H}_1 \cup \mathcal{H}_2$ and for any $\boldsymbol{u} \in \mathcal{S}^p$, use the notation

$$\widehat{\mathrm{RD}}(\boldsymbol{x}, \boldsymbol{u}) = \frac{\mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u})}{\mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, -\boldsymbol{u})},$$

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

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

$$0 < \mathrm{RD}(\boldsymbol{x}) \le \widehat{\mathrm{RD}}(\boldsymbol{x}, \boldsymbol{u}) \le 1/\mathrm{RD}(\boldsymbol{x}) < +\infty, \quad \forall \boldsymbol{u} \in \mathcal{S}^p.$$

It follows from Lemma 29 that

$$\begin{aligned}
\left|\log \mathrm{RD}_n(\boldsymbol{x}) - \log \mathrm{RD}(\boldsymbol{x})\right| &= \left|\inf_{\boldsymbol{u}\in\mathcal{S}^p} \log \widehat{\mathrm{RD}}_n(\boldsymbol{x}, \boldsymbol{u}) - \inf_{\boldsymbol{u}\in\mathcal{S}^p} \log \widehat{\mathrm{RD}}(\boldsymbol{x}, \boldsymbol{u})\right| \\
&\le \sup_{\boldsymbol{u}\in\mathcal{S}^p} \left|\log \widehat{\mathrm{RD}}_n(\boldsymbol{x}, \boldsymbol{u}) - \log \widehat{\mathrm{RD}}(\boldsymbol{x}, \boldsymbol{u})\right| \\
&\le \sup_{\boldsymbol{u}\in\mathcal{S}^p} \Bigg( \left| \log \frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) - \log \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u})\right| \\
&\qquad + \left|\log \frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, -\boldsymbol{u}) - \log \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, -\boldsymbol{u})\right| \Bigg) \\
&\le 2 \sup_{\boldsymbol{u}\in\mathcal{S}^p} \left|\log \frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) - \log \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u})\right|
\end{aligned}$$

(almost) surely. It follows that

$$\begin{aligned}
&\sup_{\boldsymbol{x}\in\mathcal{H}_1} \left|\log\, \mathrm{RD}_n(\boldsymbol{x}) - \log \mathrm{RD}(\boldsymbol{x})\right| \\
&\le 2 \sup_{\boldsymbol{x}\in\mathcal{H}_1} \sup_{\boldsymbol{u}\in\mathcal{S}^p} \left|\log \frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) - \log \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u})\right|.
\end{aligned}$$

Under theorem assumptions it follows from (3.7) that for arbitrary compact set $C \subset \mathbb{R}^p$,

$$\begin{aligned}
&\sup_{\boldsymbol{x} \in \mathcal{H}_1 \cap C} \left| \log \mathrm{RD}_n(\boldsymbol{x}) - \log \mathrm{RD}(\boldsymbol{x}) \right| \\
&\leq 2 \sup_{\boldsymbol{x} \in \mathcal{H}_1 \cap C,\, \boldsymbol{u} \in \mathcal{S}^p} \left| \log \frac{1}{n} \sum_{i=1}^{n} w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) - \log \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) \right| \\
&\longrightarrow 0 \quad \text{a.s.}
\end{aligned}$$

So eventually

$$\sup_{\boldsymbol{x} \in \mathcal{H}_1 \cap C} \left| \mathrm{RD}_n(\boldsymbol{x}) - \mathrm{RD}(\boldsymbol{x}) \right| \longrightarrow 0 \quad \text{a.s.} \tag{3.16}$$

Now we consider the points from the set $\mathcal{H}_2$. Then

$$\sup_{\boldsymbol{x} \in \mathcal{H}_2} \mathrm{RD}(\boldsymbol{x}) \leq \sup_{\boldsymbol{x} \in \mathcal{H}_2} \widehat{\mathrm{RD}}(\boldsymbol{x}, \boldsymbol{u}_{\boldsymbol{x}}) = 0.$$

It follows

$$\begin{aligned}
\sup_{\boldsymbol{x} \in \mathcal{H}_2} \mathrm{RD}_n(\boldsymbol{x}) &\leq \sup_{\boldsymbol{x} \in \mathcal{H}_2} \widehat{\mathrm{RD}}_n(\boldsymbol{x}, \boldsymbol{u}_{\boldsymbol{x}}) \\
&\leq \frac{1}{\delta} \sup_{\boldsymbol{x} \in \mathcal{H}_2} \frac{1}{n} \sum_{i=1}^{n} w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}_{\boldsymbol{x}}) \\
&\leq \frac{1}{\delta} \sup_{\boldsymbol{x} \in \mathcal{H}_2,\, \boldsymbol{u} \in \mathcal{S}^p} \frac{1}{n} \sum_{i=1}^{n} w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}).
\end{aligned}$$

The latter is consequence of non-negativeness of the weight function together with properties of the set $\mathcal{H}_2$. Again, using (3.7), we obtain that for any compact set $C \subset \mathbb{R}^p$ it holds that

$$\sup_{\boldsymbol{x} \in \mathcal{H}_2 \cap C} \mathrm{RD}_n(\boldsymbol{x}) \longrightarrow 0 \quad \text{a.s.} \tag{3.17}$$

And finally, from (3.16) and (3.17), one has

$$\sup_{\boldsymbol{x} \in (\mathcal{H}_1 \cup \mathcal{H}_2) \cap C} \left| \log \mathrm{RD}_n(\boldsymbol{x}) - \log \mathrm{RD}(\boldsymbol{x}) \right| \longrightarrow 0 \quad \text{a.s.}$$

If we consider the vanishing at infinity property from Theorem 11 then the same arguments as in the last paragraph of the proof of Theorem 17 finishes the proof of the uniform consistency over $\mathcal{H}_1 \cup \mathcal{H}_2$. ■

*Remark* 18. The uniform consistency holds over $\mathcal{H}_1 \cup \mathcal{H}_2$ for any weight function mentioned in this thesis except conic section weight functions if $e < 1$ (see Chapter 4).

We see that in contrary to the generalized halfspace depth the WHRD is not uniformly consistent over whole $\mathbb{R}^p$ in general. There arises a natural question what can be said about the points outside $\mathcal{H}_1 \cup \mathcal{H}_2$ and about the set $\mathcal{H}_1 \cup \mathcal{H}_2$

itself. First of all, let us show two counterexamples to the consistency of the sample depth.

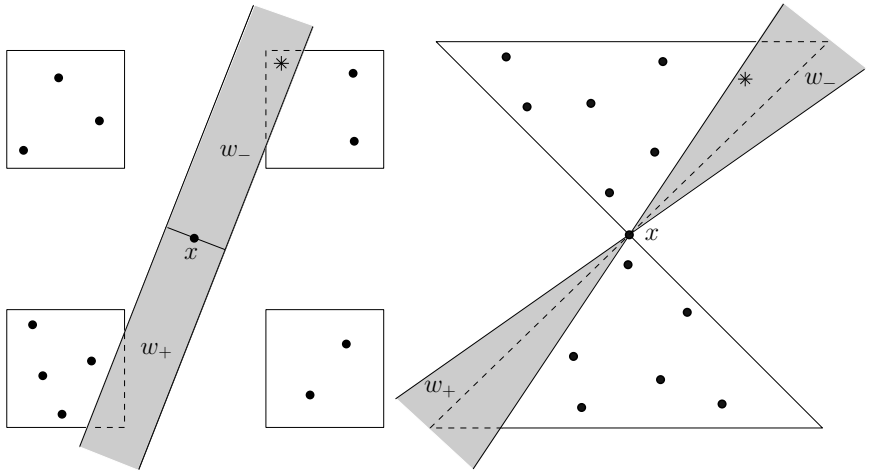



Figure 3.1: The sample depth needs not to be consistent. $w_+$ denotes $w(\cdot - \boldsymbol{x}, \boldsymbol{u})$ and $w_-$ denotes $w(\cdot - \boldsymbol{x}, -\boldsymbol{u})$. For any $n$ there always exists a point, denoted it by "$*$", that causes that the sample depth of the theoretical deepest point $\boldsymbol{x}$ is zero.

**Example 13** (Non-consistency of the sample depth)**.** We consider a uniform distribution on a "hourglass" set, and a uniform distribution on "four tiles". See Figure 3.1 for illustration. The band weight function (2.5) is used for "four tiles" distribution, the cone weight function (2.6) is used for "hourglass" distribution. The width parameter of both functions is chosen such that "0/0" occurs. In both cases the distributions are *symmetric* about a naturally defined central point $\boldsymbol{x}$ and it is exactly the point $\boldsymbol{x}$ where the problem arises. For any sample size $n$ there exists a.s. a direction $\boldsymbol{u}_n \in \mathcal{S}^p$ such that $\mathsf{E}_n\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}_n) = 0$ while $\mathsf{E}_n\, w(\boldsymbol{X} - \boldsymbol{x}, -\boldsymbol{u}_n) > 0$. In both cases the central point $\boldsymbol{x}$ is *the only* point for which the sample depth is not consistent. On the left panel of Fig. 3.2 we see that for the non-center point there exists a direction $\boldsymbol{u}$ such that $\mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) = 0$ and $\mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, -\boldsymbol{u}) > 0$ and hence the sample depth is consistent. It holds $\mathcal{H}_1 \cup \mathcal{H}_2 = \mathbb{R}^p \setminus \{\boldsymbol{x}\}$. Both central points are also points of discontinuity of the depth function. Indeed, the theoretical depth $\mathrm{RD}(\boldsymbol{x}) = 1$ as follows from the symmetry of distribution (see Theorem 15). On the other hand there exists sequence $\boldsymbol{x}_n \to \boldsymbol{x}$ such that $\mathrm{RD}_n(\boldsymbol{x}_n) = 0$ almost sure for all $n$. It's clear that in both examples the sample generalized halfspace depth is consistent even for the point $\boldsymbol{x}$ and $\mathrm{D}(\boldsymbol{x}) = \mathrm{D}_n(\boldsymbol{x}) = 0$. We see that this two depth functions can have completely different behaviour, especially for symmetric distribution when the center does not lie inside the support and simultaneously there is a "free view" outside the support. Note that such distributions are very unusual.

The nature of the problem lies in the limit of "0/0" type. Assume without loss of generality that the central point $\boldsymbol{x}$ is equal to $\mathbf{0}$. In both cases there exists

a direction $\boldsymbol{u}_0 \in \mathcal{S}^p$ and a sequence of directions $\boldsymbol{u}_n \in \mathcal{S}^p$ such that

$$\begin{aligned} &\mathsf{E}\, w(\boldsymbol{X}, \boldsymbol{u}_0) = 0, \quad \mathsf{E}\, w(\boldsymbol{X}, -\boldsymbol{u}_0) = 0, \\ &\mathsf{E}\, w(\boldsymbol{X}, \boldsymbol{u}_n) > 0, \quad \mathsf{E}\, w(\boldsymbol{X}, -\boldsymbol{u}_n) > 0 \quad \forall n, \\ &\mathsf{E}\, w(\boldsymbol{X}, \boldsymbol{u}_n) \to 0, \quad \mathsf{E}\, w(\boldsymbol{X}, -\boldsymbol{u}_n) \to 0 \quad \text{as } n \to +\infty. \end{aligned} \tag{3.18}$$

Note that in this case the generalized halfspace depth of the point $\mathbf{0}$ is equal to 0. This properties hold for the point $\boldsymbol{x}$ on the left panel of Fig. 3.1 but does not hold for the non-central point $\boldsymbol{x}$ shown on the left panel of Fig. 3.2. There exist technical assumptions on the support of probability measure $P$ and on the weight function (beside the regularity condition (3.2)) such that (3.18) 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 $\mathrm{sp}(P)$ coincides with convex closure of $\mathrm{sp}(P)$ then, for an usual choice of the weight functions, $\mathcal{H}_1 \cup \mathcal{H}_2 = \mathbb{R}^p$ and the strong consistency holds for any point. An example may be normal distribution, bivariate exponential distribution, uniform distribution on a convex set and many others.

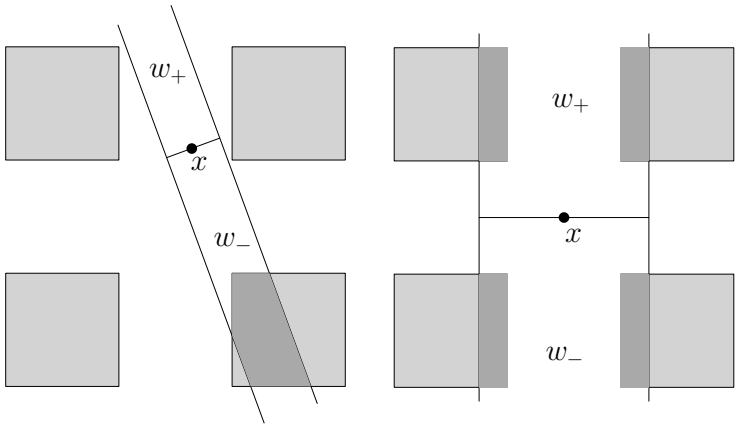



Figure 3.2: In contrary to the situation shown on the left panel of Fig. 3.1 the sample depth of the point $\boldsymbol{x}$ is consistent. The point $\boldsymbol{x}$ shown on the left panel is non-central. The point $\boldsymbol{x}$ shown on the right panel is central and the band width is high enough. In this example these conditions are suficient for consistency of $\mathrm{RD}_n(\boldsymbol{x})$.

As we have mentioned above, there are technical conditions on the support of probability measure $P$ together with the weight function $w$ such that the consistency holds for any $\boldsymbol{y} \in \mathbb{R}^p$. An example (see also Figure 3.2) of such sufficient conditions may be

- There exist $r > 0$ and $\delta_0 > 0$ such that $w(\boldsymbol{y}, \boldsymbol{u}) \geq \delta_0$ if $\|\boldsymbol{x} - \langle \boldsymbol{u}, \boldsymbol{x} \rangle \boldsymbol{u}\| \leq r$.
- The interior of support $\mathrm{sp}(P)$ is a connected set or it is not connected and distance between individual connected sets of the support is strictly less than $r$.

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

$$P\left(\boldsymbol{y} : \mathrm{RD}_n(\boldsymbol{y}) \xrightarrow{n\to\infty} \mathrm{RD}(\boldsymbol{y}),\ \text{a.s.}\right) = 1.$$

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 then the consistency of the sample 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 $\boldsymbol{x}$.

## 3.3 Influence Function and Notes on Limit Distribution

In this section we show some thoughts about the *generalized halfspace depth* that are important for the derivation of the limit distribution. The derivation of limit distribution is not easy for general form of the weight function and arbitrary continuous distribution. This section is rather informal and does not include precise mathematical derivations of the results. Asymptotics for special case of generalized halfspace depth - the halfspace depth - is derived in [Massé, 2004].

We introduce this section with the derivation of the influence function since it, under certain circumstances, can be used while proving asymptotic normality. It is also good measure of robustness. It is very difficult to find the influence function for a general weight function $w$.. Fortunately, when the weight function is an indicator function, the influence function can be derived. Results about the influence function of the halfspace depth can be found in [Romanazzi, 1999, Chen & Tyler, 2002]. The results that are shown here expand the results from these articles.

**Theorem 20** (Influence function for an indicator weight function)**.** *Let us have a point $\boldsymbol{x} \in \mathbb{R}^p$ and an absolutely continuous distribution $P$ and a weight function $w$ such that*

- $$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\left\{\boldsymbol{x} \in A(\boldsymbol{u})\right\},$$
  *where $A(\boldsymbol{u})$ is a non zero Lebesgue measure set for all $\boldsymbol{u} \in \mathcal{S}^p$.*
- *there exists unique minimal direction $\boldsymbol{u}_0 \in \mathcal{S}^p$, i.e.*
  $$\mathrm{D}_w(\boldsymbol{x}; P) = \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}_0) = \mathsf{E}\, \mathbb{1}\left\{\boldsymbol{X} - \boldsymbol{x} \in A(\boldsymbol{u}_0)\right\}.$$

*Then the influence function of generalized halfspace depth of* $\boldsymbol{x}$ *is*

$$\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{z}) = \mathbb{1}\left\{\boldsymbol{z} - \boldsymbol{x} \in A(\boldsymbol{u}_0)\right\} - \mathrm{D}_w(\boldsymbol{x}).$$

*Proof.* Let us denote

$$P_\varepsilon^{\boldsymbol{z}} = (1-\varepsilon)\, P + \varepsilon\delta_{\boldsymbol{z}},$$

where $\delta_{\boldsymbol{z}}(\boldsymbol{z}) = 1$ and $\delta_{\boldsymbol{z}}(B) = 0$ if $\boldsymbol{z} \notin B$. Then

$$\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{z}) = \lim_{\varepsilon \to 0+} \frac{\mathrm{D}_w(\boldsymbol{x}; P_\varepsilon^{\boldsymbol{z}}) - \mathrm{D}_w(\boldsymbol{x}; P)}{\varepsilon}.$$

Denote

$$\begin{aligned}\mathcal{S}_1(\boldsymbol{z}) &= \{\boldsymbol{u} :\ \boldsymbol{z} - \boldsymbol{x} \in A(\boldsymbol{u})\},\\ \mathcal{S}_2(\boldsymbol{z}) &= \{\boldsymbol{u} :\ \boldsymbol{z} - \boldsymbol{x} \notin A(\boldsymbol{u})\}.\end{aligned}$$

The depth with respect to the measure $P_\varepsilon^{\boldsymbol{z}}$ is

$$\mathrm{D}_w(\boldsymbol{x}; P_\varepsilon^{\boldsymbol{z}}) = \min\left\{\inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} \mathsf{E}\,\mathbb{1}\left\{\boldsymbol{X} - \boldsymbol{x} \in A(\boldsymbol{u})\right\},\ \inf_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} \mathsf{E}\,\mathbb{1}\left\{\boldsymbol{X} - \boldsymbol{x} \in A(\boldsymbol{u})\right\}\right\},$$

where

$$\begin{aligned}\inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} \mathsf{E}\,\mathbb{1}\left\{\boldsymbol{X} - \boldsymbol{x} \in A(\boldsymbol{u})\right\} &= (1-\varepsilon)\inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})) + \varepsilon,\\ \inf_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} \mathsf{E}\,\mathbb{1}\left\{\boldsymbol{X} - \boldsymbol{x} \in A(\boldsymbol{u})\right\} &= (1-\varepsilon)\inf_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})).\end{aligned}$$

First consider case

$$\inf_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})) \le \inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})).$$

Then

$$\mathrm{D}_w(\boldsymbol{x}; P_\varepsilon^{\boldsymbol{z}}) = (1-\varepsilon)\inf_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})) = (1-\varepsilon)\mathrm{D}_w(\boldsymbol{x}; P), \quad \forall \varepsilon \ge 0.$$

Now consider

$$\inf_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})) > \inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})). \tag{3.19}$$

If there exists $\varepsilon_0 > 0$ such that

$$(1-\varepsilon_0)\inf_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})) \ge (1-\varepsilon_0)\inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})) + \varepsilon_0$$

then also latter inequality holds for any $\varepsilon,\ 0 \le \varepsilon \le \varepsilon_0$. Hence

$$\mathrm{D}_w(\boldsymbol{x}; P_\varepsilon^{\boldsymbol{z}}) = (1-\varepsilon)\inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} P(\boldsymbol{x} + A(\boldsymbol{u})) + \varepsilon = (1-\varepsilon)\mathrm{D}_w(\boldsymbol{x}; P) + \varepsilon, \quad 0 \le \varepsilon \le \varepsilon_0.$$

Finally if

$$(1-\varepsilon)\inf_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} P(\boldsymbol{x}+A(\boldsymbol{u})) < (1-\varepsilon)\inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} P(\boldsymbol{x}+A(\boldsymbol{u})) + \varepsilon$$

then there exists $\varepsilon_0 > 0$ (since (3.19) holds) such that

$$\mathrm{D}_w(\boldsymbol{x}; P_\varepsilon^{\boldsymbol{z}}) = (1-\varepsilon)\inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} P(\boldsymbol{x}+A(\boldsymbol{u})) + \varepsilon = (1-\varepsilon)\mathrm{D}_w(\boldsymbol{x}; P) + \varepsilon, \quad 0 \le \varepsilon \le \varepsilon_0.$$

So eventually, for small enough $\varepsilon$ one has

$$\mathrm{D}_w(\boldsymbol{x}; P_\varepsilon^{\boldsymbol{z}}) = \begin{cases} (1-\varepsilon)\mathrm{D}_w(\boldsymbol{x}; P), & \text{if } \inf\limits_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} P(\boldsymbol{x}+A(\boldsymbol{u})) \le \inf\limits_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} P(\boldsymbol{x}+A(\boldsymbol{u})), \\ (1-\varepsilon)\mathrm{D}_w(\boldsymbol{x}; P) + \varepsilon, & \text{otherwise.} \end{cases}$$

The first condition in latter function holds if $\boldsymbol{z} \notin \boldsymbol{x} + A(\boldsymbol{u}_0)$. The derivative of $\mathrm{D}_w(\boldsymbol{x}; P_\varepsilon^{\boldsymbol{z}})$ with respect to $\varepsilon$ completes the proof. ■

*Remark* 19. In the last lines of the proof we can see that the assumption that the unique minimal direction $\boldsymbol{u}_0$ exists is not necessary. It was used only for simplicity in the theorem. Without these assumption it holds that

$$\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{z}) = \mathbb{1}\left\{\inf_{\boldsymbol{u}\in\mathcal{S}_2(\boldsymbol{z})} P(\boldsymbol{x}+A(\boldsymbol{u})) \le \inf_{\boldsymbol{u}\in\mathcal{S}_1(\boldsymbol{z})} P(\boldsymbol{x}+A(\boldsymbol{u}))\right\} - \mathrm{D}_w(\boldsymbol{x}). \tag{3.20}$$

*Note* 20. Influence function for a general weight function $w$ is usually in form

$$\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{z}) = w(\boldsymbol{z}-\boldsymbol{x}, \boldsymbol{u}_0) - \mathrm{D}_w(\boldsymbol{x}),$$

where $\boldsymbol{u}_0$ is a minimal direction.

Under some circumstances the influence function can be used for derivation of asymptotic normal distribution. For this purpose the expectation and the variance is needed.

*Remark* 21. Let the assumptions of Theorem 20 hold. Then

$$\mathsf{E}\,\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{X}) = 0$$

and

$$\mathsf{var}\,\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{X}) = \mathrm{D}_w(\boldsymbol{x})(1-\mathrm{D}_w(\boldsymbol{x})).$$

*Proof.*

$$\mathsf{E}\,\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{X}) = \mathsf{E}\big(w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u}_0) - \mathrm{D}_w(\boldsymbol{x})\big) = \mathrm{D}_w(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x}) = 0.$$

$$\begin{aligned}\mathsf{var}\,\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{X}) &= \mathsf{E}\,\mathrm{IF}_{\boldsymbol{x}}^2(\boldsymbol{X}) = \mathsf{E}\big(w^2(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u}_0) - 2\mathrm{D}_w(\boldsymbol{x})w(\boldsymbol{X}-\boldsymbol{x}, \boldsymbol{u}_0) + \mathrm{D}_w^2(\boldsymbol{x})\big) \\ &= \mathsf{E}\,w^2(\boldsymbol{X}-\boldsymbol{x}) - \mathrm{D}_w^2(\boldsymbol{x}) = \mathrm{D}_w(\boldsymbol{x})(1-\mathrm{D}_w(\boldsymbol{x})).\end{aligned}$$

■

*Note* 22. In the case of general form of $w$ the variance usually is

$$\mathsf{var}\,\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{X}) = \mathsf{E}\,w^2(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u}_0) - \mathrm{D}_w^2(\boldsymbol{x}).$$

The mean value remains the same.

Now we list some basic properties of the influence function (3.20).

1. The influence function is bounded, i.e.

$$-1 \le \mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{z}) \le 1, \quad \forall \boldsymbol{z} \in \mathbb{R}^p.$$

   It follows directly from (3.20).

2. The local shift sensitivity of generalized halfspace depth for an indicator weight function is infinite. To see this, let us realize that $\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{z})$ is equal to $1-\mathrm{D}_w(\boldsymbol{x})$ if $\boldsymbol{z}$ is in minimal direction from $\boldsymbol{x}$ and equal to $-\mathrm{D}_w(\boldsymbol{x})$ otherwise. Hence the influence function is discontinuous on the boundary between minimal indicator set $\boldsymbol{x}+A(\boldsymbol{u}_0)$ and its complement. The jump is equal to 1 irrespective of the value of $\mathrm{D}_w(\boldsymbol{x})$.

3. The depth of peripheral points is more exposed to contamination of the distribution than the depth of central points. To see this consider a distribution with a bounded connected support. Often $\sup_{\boldsymbol{z}\in\mathbb{R}^p}|\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{z})|$ is equal to 1 for $\boldsymbol{x}$ lying on the border and equal to $\max\{1-\mathrm{D}_w(\boldsymbol{x}),\mathrm{D}_w(\boldsymbol{x})\}$ for $\boldsymbol{x}$ near the center.

To derive limit distribution suppose assumption of Theorem 20. Under suitable regularity condition and if the minimal direction $\boldsymbol{u}_0$ is unique, one has (*von Mises expansion*)

$$\mathrm{D}_w(\boldsymbol{x};P_{\mathrm{n}}) - \mathrm{D}_w(\boldsymbol{x};P) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{IF}_{\boldsymbol{x}}(\boldsymbol{X}_i) + R_n,$$

where the remainder term is

$$R_n = \min_{\boldsymbol{u}\in\mathcal{S}^p}\frac{1}{n}\sum_{i=1}^{n} w(\boldsymbol{X}_i-\boldsymbol{x},\boldsymbol{u}) - \frac{1}{n}\sum_{i=1}^{n} w(\boldsymbol{X}_i-\boldsymbol{x},\boldsymbol{u}_0).$$

Note that this expansion and the remainder term remain the same if we use general weight function instead. The distribution of $\sqrt{n}\big(\mathrm{D}_w(\boldsymbol{x};P_{\mathrm{n}}) - \mathrm{D}_w(\boldsymbol{x};P)\big)$ is asymptotically normally distributed if

$$\sqrt{n}R_n \xrightarrow{P} 0 \quad \text{if } n\to+\infty. \tag{3.21}$$

Since

$$\mathsf{E}\,\sqrt{n}R_n = \sqrt{n}\big(\mathrm{D}_n(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})\big)$$

one has (by using Markov inequality) that (3.21) holds if

$$\sqrt{n}\big(\mathrm{D}_n(\boldsymbol{x}) - \mathrm{D}_w(\boldsymbol{x})\big) \xrightarrow{n\to\infty} 0 \tag{3.22}$$

in probability. It is not easy to show that neither (3.21) or (3.22) hold. [Romanazzi, 1999] claims that (3.22) holds and therefore if unique minimal direction (halfspace) exists the sample halfspace depth is asymptotically normal. Under certain regularity conditions on $w$ and on $P$ the same result should also holds here, i.e.

$$\sqrt{n}\big(\mathrm{D}_w(\boldsymbol{x}; P_\mathrm{n}) - \mathrm{D}_w(\boldsymbol{x}; P)\big) \xrightarrow[\text{in Law}]{n\to\infty} \mathcal{N}\big(0, \mathsf{E}\, w^2(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}_0) - \mathrm{D}_w^2(\boldsymbol{x})\big). \tag{3.23}$$

Anyway latter consideration are only valid if the unique minimal direction $\boldsymbol{u}_0$ exists. This is quite usual assumption for non-central points. If $\boldsymbol{x}$ is the center of *central symmetry*, the limit distribution is not Gaussian. Surprisingly, in such a case, the limit distribution can be found relatively easily.

*Note* 23. Suppose a centrally symmetric distribution $P$ (see Section 1.2 for the definition of central symmetry) and a spherically symmetric weight function $w$. If $\boldsymbol{x} \in \mathbb{R}^p$ is the center of the symmetry then

$$\sqrt{n}\big(\mathrm{D}_w(\boldsymbol{x}; P_\mathrm{n}) - \mathrm{D}_w(\boldsymbol{x}; P)\big) \xrightarrow[\text{in Law}]{n\to\infty} \inf_{\boldsymbol{u}\in\mathcal{S}^p} Y_{\boldsymbol{u}}, \tag{3.24}$$

where $Y_{\boldsymbol{u}}$ is a Gaussian process such that

$$\mathsf{E}\, Y_{\boldsymbol{u}} = 0 \text{ for any } \boldsymbol{u} \in \mathcal{S}^p$$

and

$$\mathsf{cov}(Y_{\boldsymbol{u}}, Y_{\boldsymbol{v}}) = \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{v}) - \mathrm{D}_w^2(\boldsymbol{x}) \text{ for any } \boldsymbol{u}, \boldsymbol{v} \in \mathcal{S}^p.$$

*Proof.* It holds that

$$\begin{aligned}
&\sqrt{n}\big(\mathrm{D}_w(\boldsymbol{x}; P_\mathrm{n}) - \mathrm{D}_w(\boldsymbol{x}; P)\big) \\
&= \sqrt{n}\left(\inf_{\boldsymbol{u}\in\mathcal{S}^p} \frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) - \mathrm{D}_w(\boldsymbol{x}; P)\right) \\
&= \inf_{\boldsymbol{u}\in\mathcal{S}^p} \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) - \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u})\right).
\end{aligned}$$

In last equation we used the fact that

$$\mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) = \mathrm{D}_w(\boldsymbol{x}; P) \text{ for any } \boldsymbol{u} \in \mathcal{S}^p.$$

Let us denote

$$Y_{\boldsymbol{u}}^n = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) - \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u})\right).$$

Then

1. $\mathsf{E}\, Y_{\boldsymbol{u}}^n = 0$ for any $\boldsymbol{u} \in \mathcal{S}^p$.

2. If we denote
$$\mu(\boldsymbol{u}) = \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}),$$
then for any $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{S}^p$
$$\begin{aligned} R(\boldsymbol{u}, \boldsymbol{v}) &:= \mathsf{cov}(Y_{\boldsymbol{u}}^n, Y_{\boldsymbol{v}}^n) = n\,\mathsf{E}\left(\frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u})\frac{1}{n}\sum_{j=1}^n w(\boldsymbol{X}_j - \boldsymbol{x}, \boldsymbol{v})\right. \\ &\left. -\frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u})\mu(\boldsymbol{v}) - \frac{1}{n}\sum_{i=1}^n w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{v})\mu(\boldsymbol{u}) + \mu(\boldsymbol{u})\mu(\boldsymbol{v})\right) \\ &= n\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathsf{E}\, w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) w(\boldsymbol{X}_j - \boldsymbol{x}, \boldsymbol{v}) - \mu(\boldsymbol{u})\mu(\boldsymbol{v})\right) \\ &= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \Big(\mathsf{E}\, w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}) w(\boldsymbol{X}_j - \boldsymbol{x}, \boldsymbol{v}) - \mu(\boldsymbol{u})\mu(\boldsymbol{v})\Big) \\ &= \frac{1}{n}\sum_{i=1}^n \mathsf{cov}\big(w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}), w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{v})\big) \\ &+ \frac{1}{n}\sum_{i\neq j} \mathsf{cov}\big(w(\boldsymbol{X}_i - \boldsymbol{x}, \boldsymbol{u}), w(\boldsymbol{X}_j - \boldsymbol{x}, \boldsymbol{v})\big) \qquad (3.25) \\ &= \mathsf{cov}\big(w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}), w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{v})\big) + 0 \\ &= \mathsf{E}\, w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{u}) w(\boldsymbol{X} - \boldsymbol{x}, \boldsymbol{v}) - \mathrm{D}_w^2(\boldsymbol{x}). \end{aligned}$$
Because of independence of $\boldsymbol{X}_i, \boldsymbol{X}_j$, $i \neq j$, the term in (3.25) is equal to 0.

3. For any $\boldsymbol{u} \in \mathcal{S}^p$
$$Y_{\boldsymbol{u}}^n \xrightarrow[\text{in Law}]{n\to\infty} Y_{\boldsymbol{u}} \sim \mathcal{N}(0, R(\boldsymbol{u}, \boldsymbol{u})).$$
It follows directly from the *Central Limit Theorem.*

Using *Donsker Theorem* it follows that
$$Y^n \xrightarrow[\text{in Law}]{n\to\infty} Y.$$
Finally, using *Continuous Mapping Theorem*, one has
$$\inf_{\boldsymbol{u}\in\mathcal{S}^p} Y_{\boldsymbol{u}}^n \xrightarrow[\text{in Law}]{n\to\infty} \inf_{\boldsymbol{u}\in\mathcal{S}^p} Y_{\boldsymbol{u}}.$$
■

The covariance structure of the limit Gaussian process depends on the choice of the weight function and also on the distribution $P$.

**Example 14** (Limit distribution of the halfspace depth)**.** Consider bivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ and the halfspace depth, i.e. the generalized halfspace depth with the weight function (2.4). The bottom panel of Fig. 3.3 shows the histogram and kernel density estimate of the sample depth distribution (left

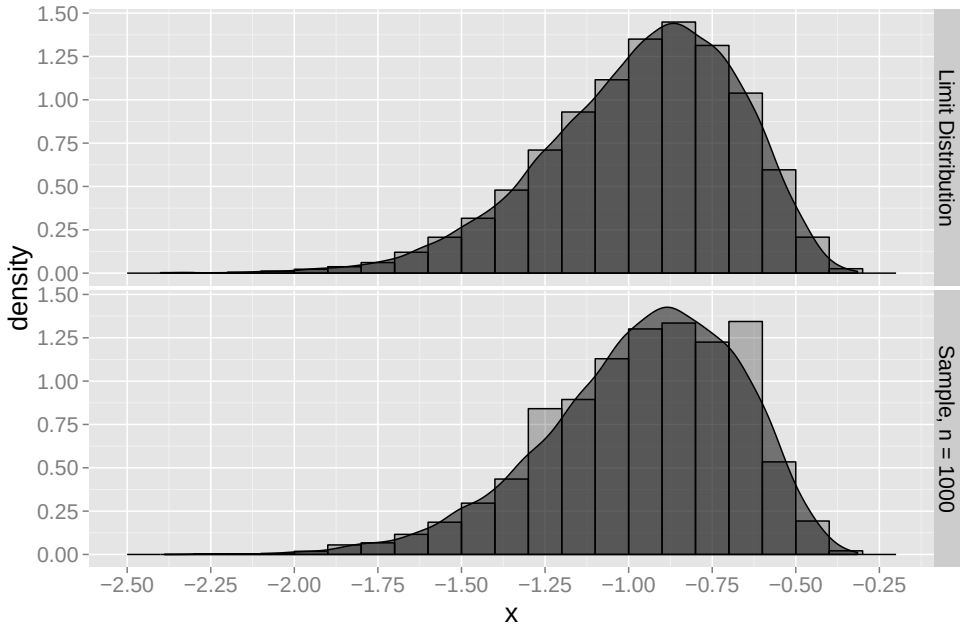



Figure 3.3: The limit distribution of the center of symmetry. The top panel shows the simulated limit distribution. Bottom panel shows distribution of $\sqrt{n}\big(\mathrm{HD}(\mathbf{0}; P_{\mathrm{n}}) - 1/2\big)$ for $n = 1000$.

hand side of (3.24)) of the center of symmetry - $\mathbf{0}$ - and the sample size $n = 1000$. The sample distribution is obtained for 10 000 simulations. Histogram and kernel density estimate of simulated limit distribution (right hand side of (3.24)) is shown on the top panel of Fig. 3.3.

Note that $\mathrm{HD}(\mathbf{0}) = 1/2$ and the covariance function of $Y_{\boldsymbol{u}}$ is

$$\mathsf{cov}(Y_{\boldsymbol{u}}, Y_{\boldsymbol{v}}) = \frac{\pi - \measuredangle(\boldsymbol{u}, \boldsymbol{v})}{2\pi} - \frac{1}{4}.$$

Now consider the point $\boldsymbol{x} = (0,1)^T$. The Halfspace Depth of this point is $1 - \Phi(1) \doteq 0.1587$, where $\Phi$ is the distribution function of $\mathcal{N}(0,1)$. Asymptotic variance is $\mathrm{HD}((0,1)^T)(1 - \mathrm{HD}((0,1)^T)) \doteq 0.1335$. The limit distribution, (3.23), is the Gaussian distribution. It is plotted as the black curve in Figure 3.4. Histograms show the distributions of $\sqrt{n}\Big(\mathrm{HD}\big((0,1)^T; P_{\mathrm{n}}\big) - \mathrm{HD}\big((0,1)^T\big)\Big)$ for $n = 500, 1000, 10000$. Again, 10 000 simulations were made to obtain this data.

The rate of convergence depends on the choice of the weight function. Usually, the limit approximation can be used for smaller $n$ if the weight function is not "focused" on small area (e.g. it is the case of the halfspace depth). If the point is not the center then the sample depth is biased and the distribution of $\mathrm{D}_{w,n}$ is shifted to the left. The shift is even visible for considerable sample size - see the bottom panel of Figure 3.4. Also the distribution is lightly skewed if $n < 1000$.

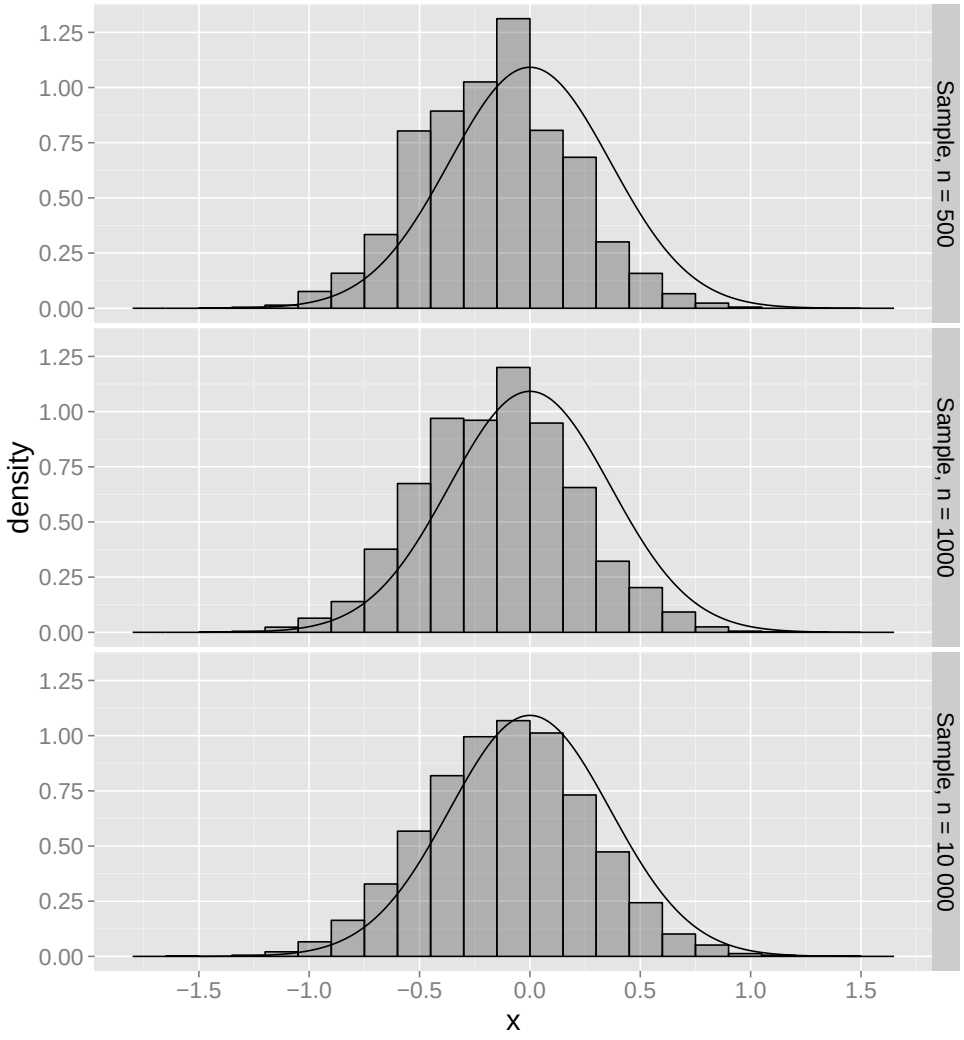



Figure 3.4: Limit distribution of non-central point. The distribution of $\sqrt{n}\Big(\mathrm{HD}\big((0,1)^T; P_{\mathrm{n}}\big) - \mathrm{HD}\big((0,1)^T\big)\Big)$ for $n = 500, 1000, 10000$. Curves show the density of limit normal distribution.

# Chapter 4

# From the Halfspace Depth to the Kernel Density Estimate

The idea of kernel (band) weight function for the generalized halfspace depth may suggest to study weight functions connecting the halfspace depth on one hand and local probability density function estimates on the other hand. Note that the weight function $w$ with

- bounded support, i.e. $\{\boldsymbol{x} : \ w(\boldsymbol{x}, \boldsymbol{u}) > 0\}$ is bounded subset of $\mathbb{R}^p$, or with
- bounded integral, i.e. $\int_{\mathbb{R}^p} w(\boldsymbol{x}, \boldsymbol{u}) \ \mathrm{d}\boldsymbol{x} = c$ for some constant $c$ (we choose usually $c = 1$),

is getting quite close to the idea of *local smoothing techniques.* Let us introduce a special class of weight functions to illustrate this idea. Note that all the theorems from Section 2.2, including the uniform strong consistency, hold for weight functions from this class.

## 4.1 Indicator of a Conic Section as a Weight Function

Recall that the choice of a weight function $w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\{\langle \boldsymbol{x}, \boldsymbol{u} \rangle \geq 0\}$ leads to the halfspace depth. We will use the class of conic section functions as a generalisation of the boundary line to obtain a family of weight functions connecting the halfspace depth and a certain kernel density estimate. The interior of conic section functions provides a naturally and simply defined class of the generalized halfspaces.

**Definition 11.** For the given eccentricity parameter $e \geq 0$ the function $r_e : \mathbb{R}^p \times \mathcal{S}^p \ \to \ \mathbb{R}$,

$$r_e(\boldsymbol{x}, \boldsymbol{u}) = \|\boldsymbol{x}\| - e\langle \boldsymbol{u}, \boldsymbol{x} \rangle, \tag{4.1}$$

is called the *conic section radius function* of the eccentricity $e$.

The set

$$\mathcal{R} = \mathcal{R}_{e,\boldsymbol{u},l} = \{\boldsymbol{x} : \ r_e(\boldsymbol{x}, \boldsymbol{u}) \leq l\} \tag{4.2}$$

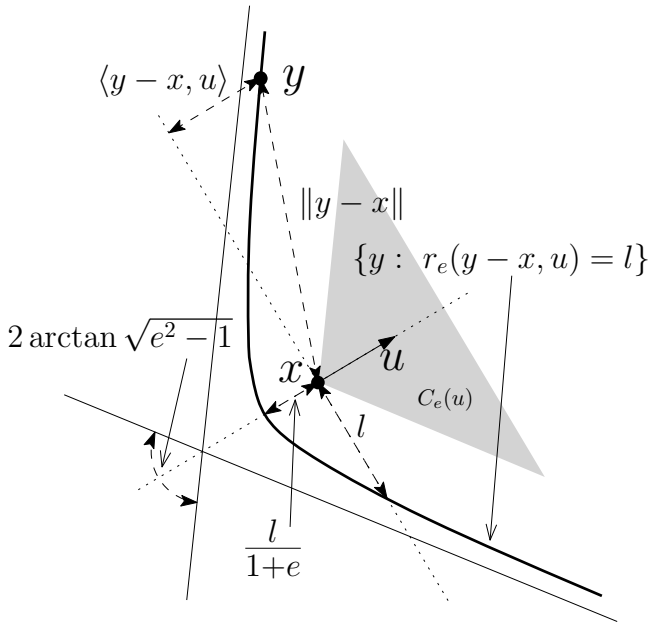



Figure 4.1: A contour (bold curve) of a *conic section* $r_e(\boldsymbol{y} - \boldsymbol{x}, \boldsymbol{u}) = l$ is shown. One branch of a hyperbola (with focus $\boldsymbol{x}$, directional vector $\boldsymbol{u}$ perpendicular to the conjugate axis, fixed eccentricity $e > 1$ and a fixed radius $l > 0$) is used as the boundary of the generalized halfspace. The gray cone indicates the area where $r_e(\boldsymbol{x}, \boldsymbol{u})$ is negative.

is the *conic section* with *eccentricity* $e$, $\boldsymbol{u}$ being normal vector to the *directrix*, and $l$ is the *parameter* ("radius") of the conic section.

We have two parameters, one of them - *eccentricity* $e$ - controls the shape of $\mathcal{R}_{e,\boldsymbol{u},l}$. Second parameter $l$ controls the "width" of $\mathcal{R}_{e,\boldsymbol{u},l}$. See Figure 4.1. The distance from origin $\mathbf{0}$ (resp. the point whose depth we calculate) to the vertex of the conic section is equal to $l/(1+e)$. The distance between the origin and the intersection of the conic section with the hyperplane $\{\boldsymbol{y} : \ \langle \boldsymbol{u}, \boldsymbol{y} \rangle = 0\}$ is equal to $l$. In the case of *ellipse* ($0 \leq e < 1$) the distance from origin to the other vertex is equal to $l/(1-e)$. Boundaries of $\mathcal{R}_{e,\boldsymbol{u},l}$ for various values of *eccentricity* $e$ is shown on Fig. 4.2. The *conic section radius* $l$ is set to 1. Hence the graphs of implicit function $r_e(\boldsymbol{x}, \boldsymbol{u}) = 1$ are shown. The role of the eccentricity $e$ is also explained in Table 4.1.

*Remark* 24. For arbitrary $\boldsymbol{u} \in \mathcal{S}^p$ and for $0 \leq e \leq 1$ (sphere, ellipsoid and paraboloid) it holds that

$$r_e(\boldsymbol{x}, \boldsymbol{u}) \geq 0 \quad \text{for all} \quad \boldsymbol{x} \in \mathbb{R}^p.$$

If $e > 1$ (hyperboloid) for any $\boldsymbol{u} \in \mathcal{S}^p$ it holds that

$$r_e(\boldsymbol{x}, \boldsymbol{u}) \leq 0 \quad \text{iff} \quad \boldsymbol{x} \in \mathcal{C}_e(\boldsymbol{u}),$$

where $\mathcal{C}_e(\boldsymbol{u})$ is the closed cone with the vertex in $\mathbf{0}$, axis in the direction of $\boldsymbol{u}$ and aperture (maximum angle between lines going through vertex) equal to

| $e = 0$ | $0 < e < 1$ | $e = 1$ | $e > 1$ | $e = +\infty$ |
|---|---|---|---|---|
| **Sphere:** radius $l$, centre $\mathbf{0}$) | **Ellipsoid:** one focus $\mathbf{0}$, length of major axis is $2l/(1-e^2)$, length of minor axes is $2l/\sqrt{1-e^2}$ | **Paraboloid:** focus $\mathbf{0}$, focus–vertex distance is $l/(1+e)$ | **Hyperboloid:** focus $\mathbf{0}$, focus–vertex distance is $l/(1+e)$, aperture of the asymptotic cone is $2\arctan\sqrt{e^2-1}$ | **Halfspace:** $\mathbf{0}$ on boundary, $\boldsymbol{u}$ is the normal vector to the boundary hyperplane |
| *"Kernel density estimate"* | *Localised depth* | *Depth with possibly non–convex contours* | | *The halfspace depth* |
| *local characteristic* | | *global characteristic* | | |

Table 4.1: Conic sections with respect to the eccentricity $e$, i.e. a characteristic of implicit functions $r_e(\boldsymbol{x}, \boldsymbol{u}) = l$ for a fixed value $l$, $l > 0$. A graphic illustration of the role of the eccentricity is shown on Fig. 4.2.

$2\arctan\sqrt{e^2-1}$. Greater eccentricity $e$ implies wider cone (i.e. greater aperture). On Fig. 4.1 the cone $\mathcal{C}_e(\boldsymbol{u})$ is marked as the gray coloured area.

**Definition 12** (Conic section weight function I)**.** For given radius $l > 0$ and eccentricity $e \geq 0$ we define the *indicator conic section weight function* as

$$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\left\{r_e(\boldsymbol{x}, \boldsymbol{u}) \leq l\right\} = \mathbb{1}\left\{\boldsymbol{x} \in \mathcal{R}_{e,\boldsymbol{u},l}\right\}. \tag{4.3}$$

In other words this weight function is an *indicator of a conic section* of eccentricity $e$, radius $l$, with the major axis in the direction given by $\boldsymbol{u}$, and with the focus in the origin.

*Note* 25. The sphere $\mathcal{R}_{0,\boldsymbol{u},l}$ clearly does not depend on $\boldsymbol{u}$. Hence the corresponding depth is simply

$$\mathrm{D}_w(\boldsymbol{x}) = \mathsf{P}(\|\boldsymbol{X} - \boldsymbol{x}\| \leq l).$$

If we calculate the (conic section) depth of a point $\boldsymbol{x} \in \mathbb{R}^p$, the weight function $w$ of (4.3) is in the form

$$w(\boldsymbol{y} - \boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\left\{r_e(\boldsymbol{y} - \boldsymbol{x}, \boldsymbol{u}) \leq l\right\}. \tag{4.4}$$

Eccentricity $0 \leq e < 1$ provides us with an ellipsoid with one focus in $\boldsymbol{x}$. If $e = 1$ we obtain a *parabola* with focus in $\boldsymbol{x}$ and finally, $e > 1$ gives us a *hyperbola* again with one focus in the point $\boldsymbol{x}$. Choosing $e = 0$ the weight function $w$ becomes *indicator of a sphere*, hence it does not depend on the direction $\boldsymbol{u} \in \mathcal{S}^p$. The weight function may be rescaled to have *unit integral* with respect to the Lebesgue measure and hence a kernel density estimate with kernel

$$k(\boldsymbol{x}) = \frac{\Gamma\left(p/2+1\right)}{\pi^{p/2} l^p} \mathbb{1}\left\{\|\boldsymbol{x}\| \leq l\right\} \tag{4.5}$$

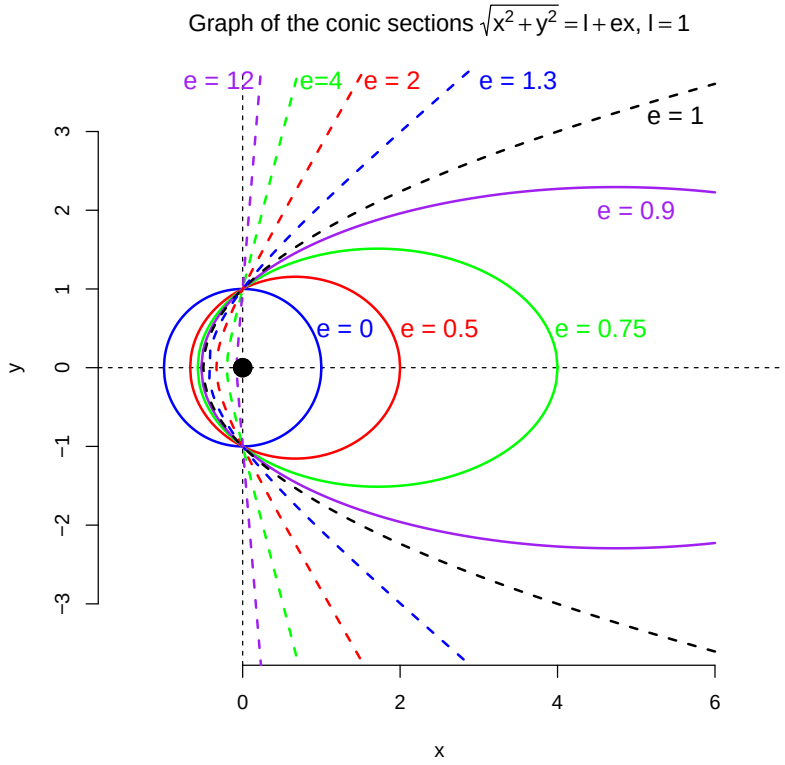



Figure 4.2: Graphs of conic sections (graphs of implicit functions $r_e(\boldsymbol{x}, \boldsymbol{u}) = l$) in $\mathbb{R}^2$ for different values of the eccentricity $e$. The radius is set to $l = 1$ and $\boldsymbol{u} = (1, 0)^T$, hence all the contours go through the points $(0, 1)^T$ and $(0, -1)^T$.

is obtained. Note that it holds

$$\mathrm{Vol}(\{\boldsymbol{y} : \ \|\boldsymbol{y}\| \leq l\}) = \frac{\pi^{p/2} l^p}{\Gamma\left(p/2 + 1\right)}.$$

Hence the first term from the right hand side of (4.5) is equal to the inverse of volume of a (hyper)sphere in $\mathbb{R}^p$ with radius $l$.

To conclude this section it remains to show that the weight function $w$ defined in (4.4) converges to the indicator of a halfspace if $e \to +\infty$.

**Theorem 21.** *For arbitrary $\boldsymbol{u} \in \mathcal{S}^p$ and arbitrary $l > 0$ it holds that*

$$\lim_{e \to +\infty} \mathbb{1}\left\{r_e(\boldsymbol{x}, \boldsymbol{u}) \leq l\right\} = \mathbb{1}\left\{\boldsymbol{x} \in \mathcal{H}_{\boldsymbol{u},l}\right\},$$

*where*

$$\mathcal{H}_{\boldsymbol{u},l} = \{\boldsymbol{x} : \ \langle \boldsymbol{x}, \boldsymbol{u} \rangle > 0\} \cup \{\boldsymbol{x} : \ \langle \boldsymbol{x}, \boldsymbol{u} \rangle = 0, \ \|\boldsymbol{x}\| \leq l\}. \tag{4.6}$$

The set $\mathcal{H}_{\boldsymbol{u},l}$ is "almost" halfspace with normal vector $\boldsymbol{u}$; the only difference is in a part of the boundary line $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{u} \rangle = 0\}$. Since we assume absolutely continuous distributions only we get the halfspace depth if $e \to +\infty$.

*Proof of Theorem 21.* Without loss of generality consider $\boldsymbol{u} = (0, \dots, 0, 1)^T$. The proof is divided into three parts according to the values of $\langle \boldsymbol{u}, \boldsymbol{y} \rangle = y_p$.
**Case 1:** Consider $y_p > 0$. It holds

$$\langle \boldsymbol{y}, \boldsymbol{u} \rangle > 0 \Rightarrow r_e(\boldsymbol{y}, \boldsymbol{u}) \xrightarrow{e \to +\infty} -\infty < l,$$

and therefore $\boldsymbol{y} \in H_{\boldsymbol{u},l}$.
**Case 2:** Consider $y_p = 0$. It holds

$$\langle \boldsymbol{y}, \boldsymbol{u} \rangle = 0 \Rightarrow r_e(\boldsymbol{y}, \boldsymbol{u}) = \|\boldsymbol{y}\|,$$

and therefore $\boldsymbol{y} \in H_{\boldsymbol{u},l}$.
**Case 3:** Consider $y_p < 0$. It holds

$$\langle \boldsymbol{y}, \boldsymbol{u} \rangle < 0 \Rightarrow r_e(\boldsymbol{y}, \boldsymbol{u}) = \xrightarrow{e \to +\infty} +\infty > l,$$

and therefore $\boldsymbol{y} \notin H_{\boldsymbol{u},l}$. ■

## 4.2 Kernel Weighted Conic Functions

The conic section indicator weight function from the previous section may be combined with a (univariate) weight function. A natural selection of such a weight function is any kernel function well known from *kernel smoothing* techniques.

One possible class of the weight functions that have their origin in the *kernel density estimate* is presented below. This class covers broad and relatively natural choices of the weight functions.

**Definition 13.** Suppose a non-increasing bounded function

$$k : \mathbb{R} \longrightarrow [0, +\infty), \quad \text{such that} \quad k(l) = k(0) \text{ if } l \leq 0,$$

and

$$\int_{\mathbb{R}^p} k(\|\boldsymbol{x}\|) \, \mathrm{d}\boldsymbol{x} = 1.$$

Then the *kernel weighted conic function* is defined as

$$w(\boldsymbol{x}, \boldsymbol{u}) = k(r_e(\boldsymbol{x}, \boldsymbol{u})).$$

*Note* 26. In contrast to the classical kernels the function $k$ from Definition 13 satisfies $k(l) = k(0)$ for $l < 0$. It is mentioned above that $r_e(\boldsymbol{x}, \boldsymbol{u})$ may be negative only for hyperboloids ($e > 1$) if $\boldsymbol{x} \in \mathcal{C}_e$. Hence, the weight function $w$ is constant on $\mathcal{C}_e$.

**Example 15** (Choices of kernels)**.** Let us show two possible kernels:

**1.** *Gaussian kernel:* Suppose given parameter $\sigma > 0$ then

$$k(l) = \begin{cases} (2\pi)^{-p/2}\sigma^{-p} \exp\left(-\frac{l^2}{2\sigma^2}\right), & \text{if } l \geq 0, \\ (2\pi)^{-p/2}\sigma^{-p}, & \text{if } l < 0, \end{cases}$$

satisfies conditions from Definition 13. Hence the kernel conic weight function is in the form

$$w(\boldsymbol{x}, \boldsymbol{u}) = \begin{cases} (2\pi)^{-p/2}\sigma^{p} \exp\left(-\frac{(\|\boldsymbol{x}\| - e\langle \boldsymbol{x}, \boldsymbol{u}\rangle)^2}{2\sigma^2}\right), & \text{if } \boldsymbol{x} \in \mathbb{R}^p \setminus \mathcal{C}_e(\boldsymbol{u}), \\ (2\pi)^{-p/2}\sigma^{p}, & \text{if } \boldsymbol{x} \in \mathcal{C}_e(\boldsymbol{u}). \end{cases}$$

Graphs of this function are on Figure 4.3. Examples of contours for kernel weighted parabolic and hyperbolic functions with the Gaussian kernel can be seen in Figures 1.12, 1.15 and 1.18.

**2.** *Triangular kernel:* For a given parameter $\sigma > 0$ we can define the triangular kernel function as

$$k(l) = \begin{cases} \frac{p\Gamma(p/2+1)}{\pi^{p/2}\sigma^p}, & \text{if } l < 0, \\ \frac{p\Gamma(p/2+1)}{\pi^{p/2}\sigma^p}\left(1 - \frac{l}{\sigma}\right), & \text{if } 0 \leq l \leq \sigma, \\ 0, & \text{if } l > \sigma. \end{cases}$$

Again, it can be shown that $k$ satisfies condition from Definition 13.

*Remark* 27. Other kernels well known from the literature (e.g. [Silverman, 2000, Schimek, 2000]) may be easily applied to the kernel weighted halfspace depth as well.

The halfspace depth is again the limiting case of the kernel weighted depth if $e \to +\infty$.

**Theorem 22.** *Let $w$ be a kernel weighted conic function from Definition 13. Denote by $\mathrm{D}_e$ the depth function using this weight function for a given eccentricity $e$. Then, for arbitrary $\boldsymbol{x} \in \mathbb{R}^p$ it holds that*

$$\lim_{e \to +\infty} \mathrm{D}_e(\boldsymbol{x}) = k(0)\mathrm{HD}(\boldsymbol{x}).$$

*Proof.* Since $k$ is bounded, we suppose an absolutely continuous distribution and (see the proof of Theorem 21)

$$\lim_{e \to +\infty} k(r_e(\boldsymbol{x}, \boldsymbol{u})) = \begin{cases} k(0), & \text{if } \langle \boldsymbol{x}, \boldsymbol{u}\rangle > 0, \\ k(\|\boldsymbol{x}\|), & \text{if } \langle \boldsymbol{x}, \boldsymbol{u}\rangle = 0, \\ 0, & \text{if } \langle \boldsymbol{x}, \boldsymbol{u}\rangle < 0. \end{cases}$$

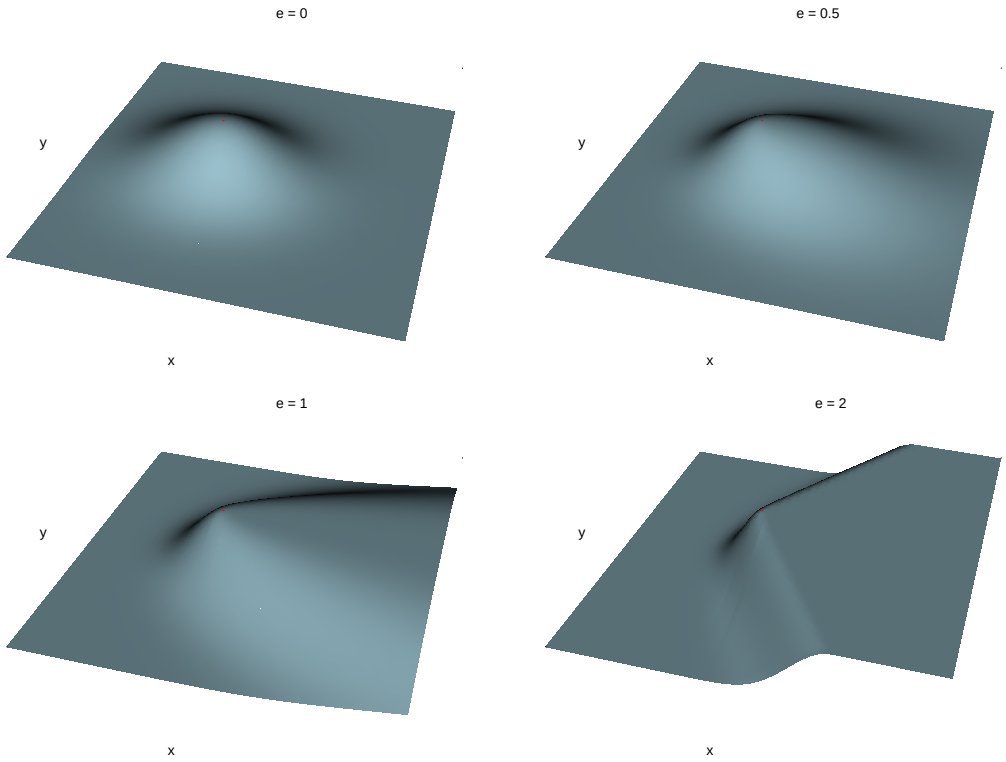



Figure 4.3: Kernel weighted conic functions for values of eccentricity $e = 0, 0.5, 1, 2$ and for the *Gaussian kernel* with $\sigma = 1$.

Then using Lebesgue theorem one has

$$\begin{aligned}\lim_{e\to+\infty} \mathrm{D}_e(\boldsymbol{x}) &= \inf_{\boldsymbol{u}\in\mathcal{S}^p} \mathsf{E} \lim_{e\to+\infty} k(r_e(\boldsymbol{X}-\boldsymbol{x},\boldsymbol{u})) \\ &= k(0) \inf_{\boldsymbol{u}\in\mathcal{S}^p} \mathsf{E}\, \mathbb{1}\{\langle \boldsymbol{u}, \boldsymbol{X}-\boldsymbol{x}\rangle > 0\} = k(0)\mathrm{HD}(\boldsymbol{x}).\end{aligned}$$

■

*Remark* 28 (Bandwidth selection). There is a question how to choose the *bandwidth* parameter. For instance, consider the Gaussian kernel from Example 15. The bandwidth parameter is denoted by $\sigma$ in the example. A recommended strategy to choose $\sigma$ is to start with $e = 0$, i.e., with kernel density estimation. Hence, the bandwidth parameter $\sigma$ can be chosen with help of the well known *plug–in* or *cross validation* estimate of *mean integrated square error*. Such estimates can be obtained using R library called ks. Setting diagonal bandwidth matrix and using the average of its diagonal components as our bandwidth parameter $\sigma$ works fine here. For $e > 0$ our recommendation is to use the same bandwidth parameter as for $e = 0$.

**Example 16.** A random sample was simulated from a mixture of distributions $P = 0.5\,P_1 + 0.5\,P_2$, where $P_1 \sim \mathcal{N}\left((0,0)^T, \mathbf{I}\right)$, and $P_2 \sim \mathcal{N}\left((4,0)^T, \mathbf{I}\right)$. The

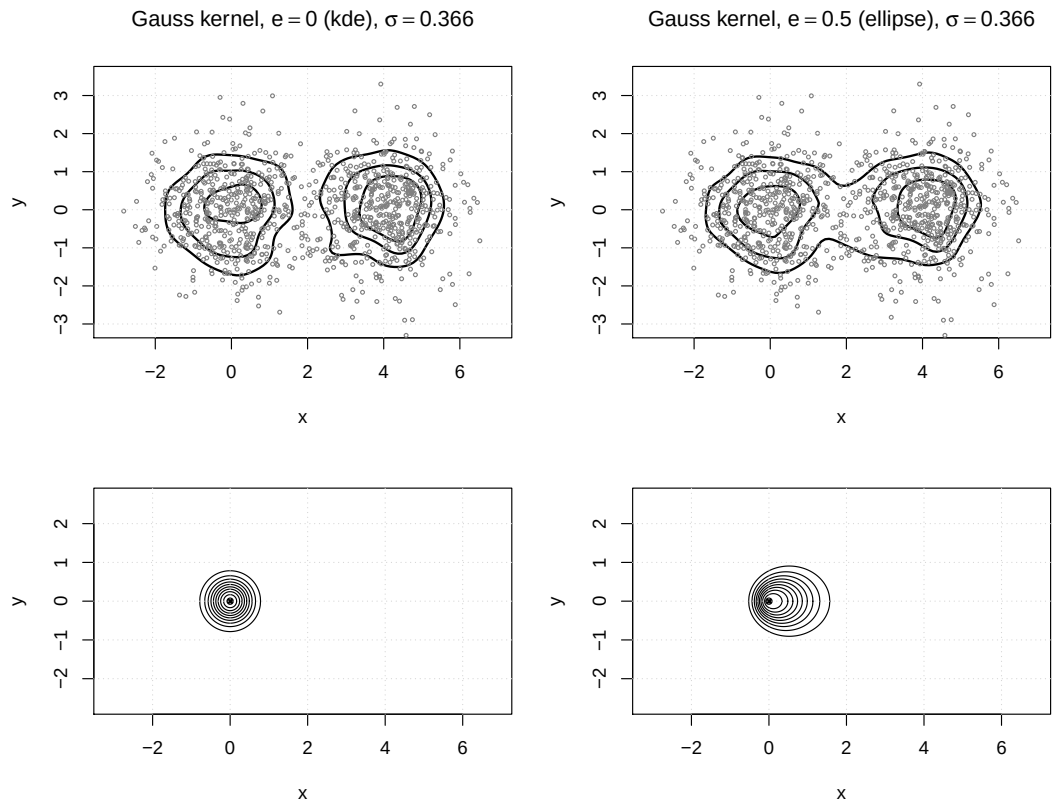



Figure 4.4: 25%, 50%, and 75% depth contours of the kernel weighted conic function for values of eccentricity $e = 0$ (upper left panel), and $e = 0.5$ (upper right panel). In the bottom panels contours of the corresponding weight functions for the direction $\boldsymbol{u} = (1, 0)^T$ are plotted (for values $\frac{i}{10} w(0, \boldsymbol{u})$, $i = 1, \ldots, 10$).

Gaussian kernel (Example 15) was used. The eccentricity parameters were chosen to be:

- $e = 0$ (kernel density estimate),
- $e = 0.5$ (elliptic weight function),
- $e = 1$ (parabolic weight function),
- $e = 1.5$ (hyperbolic weight function), and
- $e = +\infty$ (the halfspace depth).

The *bandwidth* parameter was chosen for $e = 0$ using plug–in estimate as $\sigma = 0.366$. The same bandwidth is used for all values of the eccentricity.

Depth contours for 25%, 50%, and 75% probability contents are shown in Figures 4.4, 4.5, and 4.6. Clearly the smaller the value of the eccentricity $e$ is the more local characterisation of the underlying distribution we get and vice-versa.

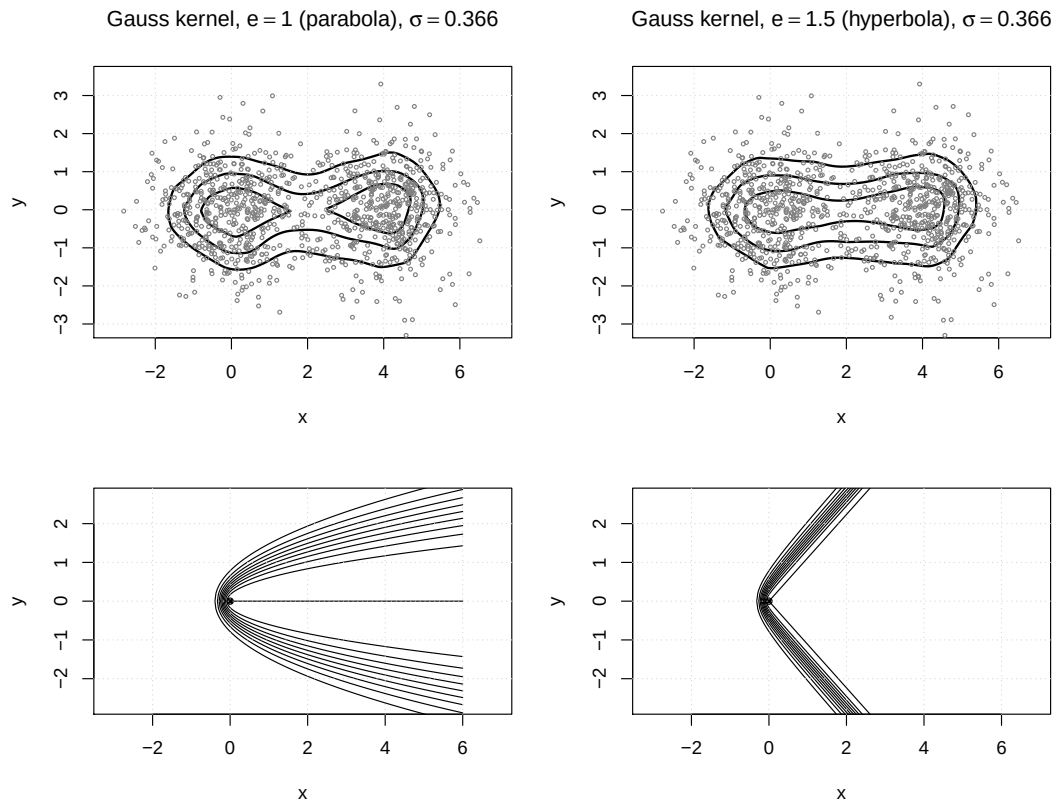



Figure 4.5: 25%, 50%, and 75% depth contours of the kernel weighted conic function for values of eccentricity $e = 1$ (upper left panel), and $e = 1.5$ (upper right panel). In the bottom panels contours of the corresponding weight functions for the direction $\boldsymbol{u} = (1, 0)^T$ are plotted (for values $\frac{i}{10} w(0, \boldsymbol{u})$, $i = 1, \ldots, 10$).

The weight function contours are also shown in Figures 4.4, and 4.5. The maximum of the kernel is attained at 0, and it holds that $w(0, \boldsymbol{u}) = k(0) \doteq 1.19$. The weight function contours $i \times 0.119$, $i = 1, \ldots, 10$, are plotted.

*Remark* 29 (Maximum of the kernel weighted conic function). We will illustrate the general fact about the kernel weighted conic functions just for the bivariate setting ($p = 2$). Consider a direction $\boldsymbol{u} = (1, 0)^T$. Then

(i) If $e < 1$ the "inner" contour $\mathcal{I}$ (the set of points where the weight function attains its maximum) consists of exactly one point, in particular $\mathcal{I} = \{(0, 0)\}$.

(ii) If $e = 1$ then $\mathcal{I} = (t, 0)^T$, $t \geq 0$.

(ii) If $e > 1$ then $\mathcal{I} = \mathcal{C}_e(\boldsymbol{u})$ (see Remark 24 for definition of $\mathcal{C}_e$).

Similar results hold for $p > 2$ as well.

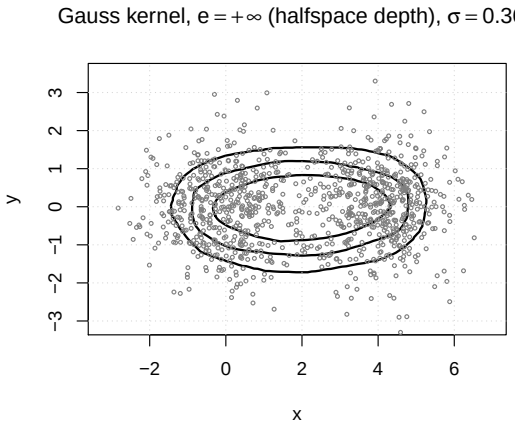



Figure 4.6: 25%, 50%, and 75% depth contours of the *halfspace depth*, i.e., the kernel weighted conic function for the limiting eccentricity $e = +\infty$.

## 4.3 Characterization of Some Conic Section Depth Functions

Paper of [Carrizosa, 1996] shows a characterization of the halfspace depth. The deepest point is so called *Simpson point*. It is known in econometrics and operational research. In the paper it is shown that

$$\mathrm{HD}(\boldsymbol{x}) = \inf_{\boldsymbol{y}\in\mathbb{R}^p} P(\{\boldsymbol{a} : \ \|\boldsymbol{a}-\boldsymbol{y}\| \geq \|\boldsymbol{a}-\boldsymbol{x}\|\}).$$

Suppose two firms want enter the market. Consumers are assumed to be distributed in $\mathbb{R}^2$ according to probability measure $P$. Location of the facilities of both firm we denote by $\boldsymbol{F}_1$, $\boldsymbol{F}_2$. Further assume that consumers will use their closest facility, therefore the first firm captures consumers of those $\boldsymbol{a}$ which

$$\|\boldsymbol{a}-\boldsymbol{F}_1\| \leq \|\boldsymbol{a}-\boldsymbol{F}_2\|.$$

If the first firm locates its facility in $\boldsymbol{F}_1$ then the second firm will locate its facility at

$$\boldsymbol{F}_2 = \arg\max_{\boldsymbol{y}\in\mathbb{R}^2} P(\{\boldsymbol{a} : \ \|\boldsymbol{a}-\boldsymbol{y}\| \leq \|\boldsymbol{a}-\boldsymbol{F}_1\|\})$$

to obtain the best possible market share. Hence the first firm should locate its facility at the position which minimizes the market share of the other firm entering the market, i.e.

$$\boldsymbol{F}_1 = \arg\max_{\boldsymbol{F}\in\mathbb{R}^2} \inf_{\boldsymbol{y}\in\mathbb{R}^2} P(\{\boldsymbol{a} : \ \|\boldsymbol{a}-\boldsymbol{F}\| \leq \|\boldsymbol{a}-\boldsymbol{y}\|\}).$$

This is actually the *deepest point of the halfspace depth.* These considerations lead to the situation where the best location to place a new facility is exactly the

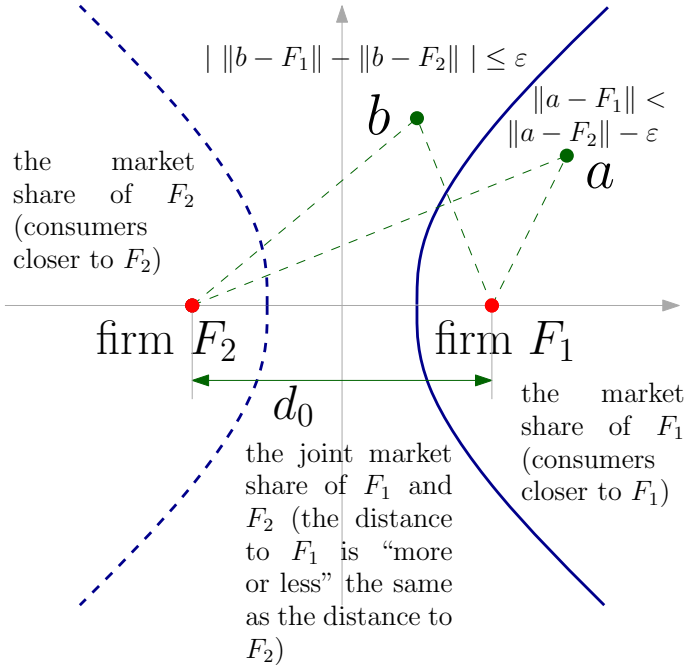



Figure 4.7: The market share. Facilities cannot be placed closer than $d_0$ to each other. Consumers with $\pm\varepsilon$ same distance to the both facilities choose the facility randomly.

location where other existing facilities are located.

Now we consider situation where a new facility can be located in the *distance not less than a given value* and where both firms have a *joint market share*, i.e. there exists an area such that consumers have more or less the same distance to the both firms. Formally - suppose that the second firm (on location $\boldsymbol{F}_2$) will not place its facility closer to the first firm (location $\boldsymbol{F}_1$) than $d_0$, further suppose that for a "shared distance" $\varepsilon > 0$ the first firm captures a consumer on a location $\boldsymbol{a}$ if

$$\|\boldsymbol{a} - \boldsymbol{F}_1\| < \|\boldsymbol{a} - \boldsymbol{F}_2\| - \varepsilon,$$

similarly the second firm. If

$$\big|\|\boldsymbol{b} - \boldsymbol{F}_1\| - \|\boldsymbol{b} - \boldsymbol{F}_2\|\big| \leq \varepsilon$$

a consumer on a position $\boldsymbol{b}$ chooses between $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ randomly. This can be the case of two mobile network operators and their transmitter facilities. See Figure 4.7 for an illustration. The *market share* belongs only to the firm in a position $\boldsymbol{F}_1$ - in the worst case scenario (if $\boldsymbol{F}_2$ is in optimal position) - is

$$\mathrm{MS}_{d_0,\varepsilon}(\boldsymbol{F}_1) = \inf_{\boldsymbol{F}_2:\, \|\boldsymbol{F}_2 - \boldsymbol{F}_1\| \geq d_0} \mathsf{P}\big(\{\boldsymbol{a} :\ \|\boldsymbol{a} - \boldsymbol{F}_1\| < \|\boldsymbol{a} - \boldsymbol{F}_2\| - \varepsilon\}\big).$$

In the following theorem it is shown how is this market share measure connected with the generalized halfspace depth.

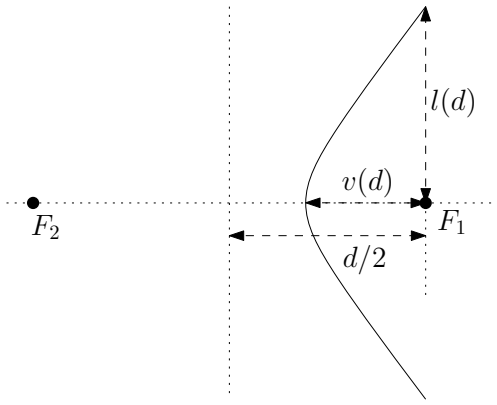



Figure 4.8: Dependence of hyperboloid parameters on the distance between focuses.

**Theorem 23.** *Let us have $d_0 > 0$ and $\varepsilon > 0$. Then for any $\boldsymbol{F}_1 \in \mathbb{R}^p$*

$$\mathrm{MS}_{d_0,\varepsilon}(\boldsymbol{F}_1) = \mathrm{D}_w(\boldsymbol{F}_1),$$

*where*

$$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\left\{ r_{d_0/\varepsilon}(\boldsymbol{x}, \boldsymbol{u}) < \frac{d_0^2 - \varepsilon^2}{2\varepsilon} \right\}.$$

*Hence*

$$\mathrm{MS}_{d_0,\varepsilon}(\boldsymbol{F}_1) = \inf_{\boldsymbol{u} \in \mathcal{S}^p} \mathsf{E}\, \mathbb{1}\left\{ r_{d_0/\varepsilon}(\boldsymbol{X} - \boldsymbol{F}_1, \boldsymbol{u}) < \frac{d_0^2 - \varepsilon^2}{2\varepsilon} \right\},$$

*where the random vector $\boldsymbol{X}$ describes distribution of the consumers.*

*Proof.* See Figure 4.8. Figures 4.1 and 4.7 may be also handy. Suppose fixed $\varepsilon > 0$ and consider $\boldsymbol{F}_1, \boldsymbol{F}_2$, $\|\boldsymbol{F}_1 - \boldsymbol{F}_2\| = d$. Then the set of points $\boldsymbol{a}$ satisfying

$$\big| \|\boldsymbol{a} - \boldsymbol{F}_1\| - \|\boldsymbol{a} - \boldsymbol{F}_2\| \big| = \varepsilon$$

is a hyperboloid with focuses in $\boldsymbol{F}_1, \boldsymbol{F}_2$. It is well known that for a hyperboloid holds that

$$\begin{aligned} e(d) &= \frac{d/2}{d/2 - v(d)}, \\ \varepsilon &= \big(d - v(d)\big) - v(d), \\ v(d) &= \frac{l(d)}{1 + e}. \end{aligned}$$

Where in parenthesis we emphasize the fact that values of these parameters depend on value of parameter $d$. It follows

$$\begin{aligned} e(d) &= \frac{d}{\varepsilon}, \\ l(d) &= \frac{d^2 - \varepsilon^2}{2\varepsilon}, \end{aligned}$$

$$v(d) = \frac{d - \varepsilon}{2}.$$

Then if $d > d_0$ it follows

$$v(d) > v(d_0),$$
$$l(d) > l(d_0).$$

Hence the part of the hyperboloid closer to $\boldsymbol{F}_1$ for $d_0$ is nested in the part of hyperboloid for $d > d_0$ (if the direction of axis remains the same). Therefore the infimum occurs for $\|\boldsymbol{F}_1 - \boldsymbol{F}_2\| = d_0$, i.e. for "fixed" shape of hyperboloid with the focus in $\boldsymbol{F}_1$, conic section radius $l = l(d_0)$ and eccentricity $e = e(d_0)$. ■

*Note* 30. Without the joint market share (the case $\varepsilon = 0$) similar thoughts lead to a market share

$$\mathrm{MS}_{d_0}(\boldsymbol{F}_1) = \mathrm{D}_w(\boldsymbol{F}_1),$$

where

$$w(\boldsymbol{x}, \boldsymbol{u}) = \mathbb{1}\left\{\langle \boldsymbol{x}, \boldsymbol{u}\rangle \geq -d_0/2\right\}.$$

Hence the consumers are divided into 2 hyperplanes where the boundary lies at the half distance between facilities' locations $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$.

# Chapter 5

# Other Depth Approaches

This chapter describes some ideas that are closely related to the data depth and where the proposed depth functions can be used and may bring some improvements. First a *directional approach* is introduced. There exist more possibilities how to define directional quantiles. One of the most promising approach is to look for univariate quantiles of the distance from some point given direction. The depth functions proposed in the thesis provide us with very good and natural candidate for such point.

The second section deals with *regression depth* that is very promising method when dealing with noisy regression data. Unfortunately not a big development of this methodology have been done so far. The most known definition is definition according to [Rousseeuw & Hubert, 1999]. Other definition of the regression depth is connected with the classical data depth. We can use a classical data depth function to determine the depth of a regression plane. Our proposed depth can also be used for this purpose. We also show a new result about a non-fit position of quadratic regression fit.

Finally we mention *functional data depth*. Now it is a very popular topic and it is still under massive development since it provides us with very good possibility of functional data interference. Popular integral depth can be extended if we consider also derivatives of functions. Then instead of univariate depth integrand we use the classical data depth function. Again, use of the depth function proposed in this thesis can bring improvements since it can "more respect" the shape of data.

## 5.1 Directional Quantiles - Directional Depth - Introduction

This section is about the directional quantiles approach that is in some sense well connected with the depth and that brings generalization of quantiles to the multivariate case. *Univariate quantiles* are one of the main concepts in the statistics. Unfortunately no direct generalization to multivariate case exists. It is due to the lack of natural ordering in multidimensional space. Although a lot of

approaches have appeared in recent 50 years, none of them has become broadly preferred or used.

There exist many definitions of the directional quantiles in multivariate case. Master thesis [Hasil, 2004] introduces idea of directional quantile in the multivariate settings. Master thesis [Kotík, 2007] shows properties and possibilities of estimation of the directional quantiles. The idea is based on finding univariate quantiles on rays beginning in the "center". This section is dedicated primarily on this approach since the results has not been published yet. First we briefly recapitulate the other directional approaches.

The easiest way to define *multivariate quantiles* arises from the "classical" assumption of *normality.* Multivariate $\tau$-quantiles can be defined as points laying on the border of an ellipsoid with the probability $\tau$. Parameters of this ellipsoid depend on the covariance matrix of the underlying distribution. The assumption of normality (or of the *elliptical symmetry*) is often too strong and it need not to be valid in many situations. In this case one would like to use nonparametric approach. Various ideas have been shown in the literature. An approach based on the *depth functions* can be seen e.g. in [Liu et al., 1999]. border of a central region with a probability $\tau$ can be considered as a set of all multivariate $\tau$ quantiles. Other approach was proposed in [DasGupta et al., 1995]. Their construction of central sets is based on *scaling* of a suitable a priori given set.

Paper [Wei, 2008] uses similar approach to the approach proposed in this section. Having $\boldsymbol{\theta} \in \mathbb{R}^p$ as a center of the distribution the $\tau 100\%$ directional interval of $\boldsymbol{X}$ along the line $a(\boldsymbol{u}) = \{\boldsymbol{\theta} + t\boldsymbol{u},\ t \in \mathbb{R}\}$ is defined as the closed interval $[l_\tau(\boldsymbol{u}), u_\tau(\boldsymbol{u})]$ which contains all of the points $\boldsymbol{x} \in a(\boldsymbol{u})$ satisfying

$$\mathsf{P}\big(\langle \boldsymbol{X}, \boldsymbol{u}\rangle \leq \boldsymbol{x} | \boldsymbol{X} \in a(\boldsymbol{u})\big) \leq \frac{1+\tau}{2}$$

and

$$\mathsf{P}\big(\langle \boldsymbol{X}, \boldsymbol{u}\rangle \leq \boldsymbol{x} | \boldsymbol{X} \in a(\boldsymbol{u})\big) \geq \frac{1-\tau}{2}.$$

The $\tau$-central region is defined as

$$\mathcal{C}(\tau) = \bigcup_{\boldsymbol{u} \in \mathcal{S}^p} [l_\tau(\boldsymbol{u}), u_\tau(\boldsymbol{u})].$$

It holds that for continuous distributions the $\tau$-central region has probability $\tau$. In other (Wei) words the $\tau$-central region consists of a family of directional reference intervals, each of which reflects the outlyingness of measurements from $\boldsymbol{\theta}$ along one spatial direction. Clearly, if $p = 1$ then $\mathcal{C}(\tau) = \big[F^{-1}(\frac{1-\tau}{2}), F^{-1}(\frac{1+\tau}{2})\big]$, where $F^{-1}$ is the quantile function. Wei uses this approach to construction of covariate-dependent quantile contours of conditional growth charts.

Paper [Kong & Mizera, 2012] shows another definition of the directional quantiles. Authors do not suppose a known central point $\boldsymbol{\theta}$. They also do not work with conditional distribution. They define $\tau$-directional quantile along direction

$\boldsymbol{u} \in \mathcal{S}^p$ as

$$Q(\tau, \boldsymbol{u}) = \inf\{t : \mathsf{P}(\langle \boldsymbol{u}, \boldsymbol{X} \rangle \leq t) \geq \tau\}.$$

Obviously, $\langle \boldsymbol{u}, \boldsymbol{X} \rangle = t$ is the hyperplane with normal vector $\boldsymbol{u}$ and in the distance $t$ from the origin. Thus $Q(\tau, \boldsymbol{u})$ is the quantile of the projection of $\boldsymbol{X}$ into $t\boldsymbol{u}, t \in \mathbb{R}$. Main interest of the authors is not the directional quantiles themselves but the *directional quantiles envelopes*. The definition is

$$\mathcal{C}(\tau) = \bigcap_{\boldsymbol{u} \in \mathcal{S}^p} H(\boldsymbol{u}, Q(\tau, \boldsymbol{u})),$$

where $H(\boldsymbol{u}, Q(\tau, \boldsymbol{u})) = \{\boldsymbol{x} : \langle \boldsymbol{u}, \boldsymbol{x} \rangle \geq \tau\}$. Theorem 2 from the paper shows that $\mathcal{C}(\tau)$ corresponds to the $\tau$-central halfspace depth region from Definition 3, i.e. it holds that

$$\mathcal{C}(\tau) = \{\boldsymbol{x} : \mathrm{HD}(\boldsymbol{x}) \geq \tau\}.$$

It is a completely different approach from Wei's definition of the directional quantiles. It inherits the properties of the halfspace depth, hence the directional quantiles envelopes are always convex sets and also border of an envelope can lie outside the support for some directions. The Wei's approach considers only the conditional probability on the lines going through the central point hence the border of any $\tau$-central set lies inside the support for any direction $\boldsymbol{u} \in \mathcal{S}^p$ if the central point lies inside the support.

Our definition ([Hasil, 2004, Kotík, 2007]) is similar to [Wei, 2008]. Instead of lines going through a center we consider rays beginning in the center. This seems to be more appropriate in the case of *star-shaped* (see Definition 16) central regions than the Wei's approach. On the other hand it is more sensitive to the proper choice of the central point. If a naturally given central point exists (such as center of a symmetry) the proposed directional approach is quite natural.

## 5.2 Directional Quantiles along Rays

Suppose a univariate random variable $X$ with a distribution function $F$. Then the quantile function $F^{-1}$ is defined in the following way

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad 0 < u < 1.$$

The median $\hat{\mu}$ is any point that satisfies $\hat{\mu} = F^{-1}(1/2)$. So it can be though as a central point of the underlying distribution. From this center we can look to the direction left and right and calculate there conditional quantiles. More formally, we consider conditional distribution functions

$$\begin{aligned} F_r(x) &= \mathsf{P}(X \leq x | X > \hat{\mu}) = 2F(x) - 1, \quad \text{for } x > \hat{\mu}, \\ F_l(x) &= \mathsf{P}(X > x | X \leq \hat{\mu}) = 1 - 2F(x), \quad \text{for } x \leq \hat{\mu}. \end{aligned}$$

Conditional quantiles "to the direction left" and "to the right" are $F_r^{-1}(\tau) = F^{-1}(\frac{1+\tau}{2})$ and $F_l^{-1}(\tau) = F^{-1}(\frac{1-\tau}{2})$. It holds that

$$\mathsf{P}\big(F_l^{-1}(\tau) \le X \le F_r^{-1}(\tau)\big) = \tau,$$

if $X$ has a continuous distribution. Thus we get the $\tau$-confidence interval that is the same as the $\tau$-confidence interval obtained directly by $F^{-1}$, because both direction from $\hat{\mu}$ have the same probability. We use similar approach for random vectors.

In what follows we suppose a distribution with a density $f$. Our aim is to find the conditional quantiles in all directions (rays) from a central point. So there is a need of existence of a naturally given point that represents the *center*. Note that the choice of the center may have an influence on the properties of the directional quantiles so if there does not exist a naturally preferred central point we suggest to use some generalization of the *univariate median* based on the *data depth* - i.e. the *deepest point*. The central point needs to be well balanced in the sense that it is well surrounded by data. It is desired that in any direction from the central point there lies as much probability mass as possible.

Now consider multivariate case. Let $\boldsymbol{\mu}$ be a center of the underlying distribution. For each ray starting in $\boldsymbol{\mu}$ we find the univariate quantiles of the variable measuring the distance from this point. Since every such ray has zero probability in $\mathbb{R}^p$, we need first consider how to settle with conditional probability of the distance from the central point given the direction. A nice representation of the distance from $\boldsymbol{\mu}$ and directions in $\mathbb{R}^p$ is provided by using *hyperspherical coordinates* $(r, \boldsymbol{\varphi}) = (r, (\varphi_1, \ldots, \varphi_{p-1})^T)$. In what follows we denote the transformation from hyperspherical coordinates back to *Cartesian coordinates* by $\psi$.

$$\begin{aligned} \psi:\ (r,\boldsymbol{\varphi}) \ \mapsto \quad & \begin{aligned} x_1 &= \mu_1 + r\sin\varphi_1\sin\varphi_2\ldots\sin\varphi_{p-2}\sin\varphi_{p-1},\\ x_2 &= \mu_2 + r\sin\varphi_1\sin\varphi_2\ldots\sin\varphi_{p-2}\cos\varphi_{p-1},\\ &\vdots\\ x_{p-1} &= \mu_{p-1} + r\sin\varphi_1\cos\varphi_2,\\ x_p &= \mu_p + r\cos\varphi_1. \end{aligned} \end{aligned} \tag{5.1}$$

Jacobian of the latter transformation is

$$J(r,\boldsymbol{\varphi}) = r^{p-1}\ \sin^{p-2}\varphi_1\ldots\sin\varphi_{p-2}.$$

The function $\psi$ is one to one mapping for $\varphi_i \in (0,\pi),\ i = 1,\ldots,p-2,\ \varphi_{p-1} \in (0,2\pi),\ r > 0$. Each direction is determined by a vector of angles.

Our aim is to find quantiles of the *radius* (random variable $\rho$) for a given vector of *angles* (random vector $\boldsymbol{\phi}$). For this, we use the conditional distribution that we obtain by using the *transformation theorem*. The joint density of radius and angles $(\rho, \boldsymbol{\phi})$ (function $p$) the marginal density of $\boldsymbol{\phi}$ (function $s$) and the

*conditional density* for given $\boldsymbol{\phi} = \boldsymbol{\varphi}$ (function $q$) are

$$\begin{aligned} p(r, \boldsymbol{\varphi}) &= \left|J(r, \boldsymbol{\varphi})\right| \; f\big(\psi(r, \boldsymbol{\varphi})\big) = r^{p-1} \; \sin^{p-2} \varphi_1 \dots \sin \varphi_{p-2} \; f\big(\psi(r, \boldsymbol{\varphi})\big), \\ s(\boldsymbol{\varphi}) &= \int_0^{+\infty} p(u, \boldsymbol{\varphi}) \; \mathrm{d}u, \\ q(r|\boldsymbol{\varphi}) &= \begin{cases} \dfrac{p(r, \boldsymbol{\varphi})}{s(\boldsymbol{\varphi})} & \text{if } s(\boldsymbol{\varphi}) \neq 0, \\ 0 & \text{if } s(\boldsymbol{\varphi}) = 0. \end{cases} \end{aligned} \tag{5.2}$$

In what follows we often work with the conditional distribution function of the radius variable given angle:

$$\mathsf{P}(\rho \leq r|\boldsymbol{\phi} = \boldsymbol{\varphi}) = \int_0^r q(u|\boldsymbol{\varphi}) \; \mathrm{d}u.$$

### 5.2.1 Directional Quantiles in $\mathbb{R}^p$

In this section we define a population version of the directional quantiles and we show some of its basic properties. We propose two different definitions. First definition is based on the transformation to hyperspherical coordinates. The second, that can help to better understand our approach and that may seem to be more straightforward, is based on evaluation of limit of the conditional probability of a *cone* in given direction.

**Definition 14.** Suppose a random vector $\boldsymbol{X} \in \mathbb{R}^p$ and its representation $(\rho, \boldsymbol{\phi})$ in hyperspherical coordinates shown in (5.1). Then, for any $\boldsymbol{\varphi}$, $\varphi_i \in (0, \pi)$, $i = 1, \dots, p-2$, $\varphi_{p-1} \in (0, 2\pi)$, we define the conditional distribution and the quantile function of the radius for a given angle (direction) as

$$Q(r|\boldsymbol{\varphi}) = \mathsf{P}(\rho \leq r|\boldsymbol{\phi} = \boldsymbol{\varphi})$$

and

$$Q_\tau^{-1}(\boldsymbol{\varphi}) = \inf\{r \geq 0 : \; Q(r|\boldsymbol{\varphi}) \geq \tau\}.$$

Then the *$\tau$-directional quantile* in the direction represented by the angle $\boldsymbol{\varphi}$ is defined as

$$\boldsymbol{\theta}_\tau(\boldsymbol{\varphi}) = \psi\big(Q_\tau^{-1}(\boldsymbol{\varphi}), \boldsymbol{\varphi}\big).$$

In other words $\boldsymbol{\theta}_\tau(\boldsymbol{\varphi})$ is a point that lies in the distance of $Q_\tau^{-1}(\boldsymbol{\varphi})$ from the center in the direction represented by the angle $\boldsymbol{\varphi}$.

*Note* 31. At the endpoints of the interval for angles (values 0, $\pi$, $2\pi$) we can define the $\tau$-directional quantile as a limit of $\boldsymbol{\theta}_\tau(\cdot)$ at these values.

We show another way how to define the directional quantiles. For any direction $\boldsymbol{s} \in \mathcal{S}^p$, we denote by

$$\mathcal{A}_{\boldsymbol{\mu}}^{\delta}(\boldsymbol{s}) = \{\boldsymbol{x} \in \mathbb{R}^p : \; \measuredangle(\boldsymbol{x} - \boldsymbol{\mu}, \boldsymbol{s}) \leq \delta\}$$

an *infinite cone* with the apex $\boldsymbol{\mu}$, with the axis in the direction given by $\boldsymbol{s}$ and with an aperture $2\delta$. Finally, we denote the *cone* of height $r$ by

$$\mathcal{A}_{\boldsymbol{\mu}}^{\delta}(\boldsymbol{s}, r) = \{\boldsymbol{x} \in \mathbb{R}^p : \ \measuredangle(\boldsymbol{x} - \boldsymbol{\mu}, \boldsymbol{s}) \leq \delta, \ \|\boldsymbol{x} - \boldsymbol{\mu}\| \leq r\}.$$

Using this notation we can define the conditional distribution function in a different manner.

**Definition 15** (Alternative definition)**.** The conditional distribution function of the radius (distance from center) for a given direction $\boldsymbol{s} \in \mathcal{S}^p$ is defined as

$$F(r|\boldsymbol{s}) = \lim_{\delta \to 0+} \mathsf{P}\big(\boldsymbol{X} \in \mathcal{A}_{\boldsymbol{\mu}}^{\delta}(\boldsymbol{s}, r) \ \mid \ \boldsymbol{X} \in \mathcal{A}_{\boldsymbol{\mu}}^{\delta}(\boldsymbol{s})\big).$$

We define the *$\tau$-directional quantile* in direction $\boldsymbol{s}$ in the same way as in Def. 14, i.e. as the point

$$\boldsymbol{\theta}_{\tau}(\boldsymbol{s}) = \boldsymbol{\mu} + F^{-1}(\tau|\boldsymbol{s}) \cdot \boldsymbol{s},$$

where again $F^{-1}(\cdot|\boldsymbol{s})$ is the conditional quantile function.

*Note* 32. There is one to one correspondence between vectors of angles $\boldsymbol{\varphi}$ and vectors of directions $\boldsymbol{s} \in \mathcal{S}^p$, (since $\psi(1, \boldsymbol{\varphi}) - \boldsymbol{\mu} \in \mathcal{S}^p$). Hence according to circumstances, we can use the notation $\boldsymbol{\theta}_{\tau}(\boldsymbol{\varphi})$ or $\boldsymbol{\theta}_{\tau}(\boldsymbol{s})$.

The next theorem claims when the latter and the former definitions are equivalent. For this purpose we need one additional assumption.

**Theorem 24.** *Suppose that the density $f$ of a random vector $\boldsymbol{X}$ is bounded and continuous and that condition*

$$\lim_{r \to +\infty} \sup_{\boldsymbol{s} \in \mathcal{S}^p} \int_r^{+\infty} t^{p-1} f(\boldsymbol{\mu} + t\boldsymbol{s}) \ \mathrm{d}t = 0 \tag{5.3}$$

*holds. Then Def. 14 and Def. 15 are equivalent.*

The proof is very similar to the proof of Theorem 26 and it uses results from the proof of Theorem 25. Therefore the proof of the theorem will follow later.

Condition (5.3) states the *uniform integrability* and it can be replaced with condition (5.4) that is easier to check:

$$\exists \gamma > 0 : \quad \sup_{\boldsymbol{s} \in \mathcal{S}^p} \int_0^{+\infty} t^{p-1+\gamma} f(\boldsymbol{\mu} + t\boldsymbol{s}) \ \mathrm{d}t < +\infty. \tag{5.4}$$

Note that it is not sufficient to assume that the components of $\boldsymbol{X}$ posses finite moments.

**Theorem 25.** *Suppose that (5.4) holds then (5.3) also holds.*

*Proof.* Suppose, without lost of generality, that $\boldsymbol{\mu} = \boldsymbol{0}$. It holds that

$$\int_r^{+\infty} t^{p-1} f(t\boldsymbol{s}) \ \mathrm{d}t \leq \int_r^{+\infty} t^{p-1} \frac{t^{\gamma}}{r^{\gamma}} f(t\boldsymbol{s}) \ \mathrm{d}t \leq \frac{1}{r^{\gamma}} \int_0^{+\infty} t^{p-1+\gamma} f(t\boldsymbol{s}) \ \mathrm{d}t$$

$$\leq \frac{1}{r^\gamma} \sup_{\boldsymbol{s}\in\mathcal{S}^p} \int_0^{+\infty} t^{p-1+\gamma} f(t\boldsymbol{s}) \ \mathrm{d}t \xrightarrow{r\to+\infty} 0,$$

where last term goes to 0 because we have uniform finiteness of integrals (5.4). Hence (5.3) holds. ■

Assumption (5.4) holds for majority of multivariate distributions. Let us present a simple and classical example.

**Example 17.** Consider a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with density $f$, where $\boldsymbol{\mu}$ will also represents the center. Further without loss of generality we suppose $\boldsymbol{\mu} = \mathbf{0}$. Set $\gamma = 1$. Then

$$\sup_{\boldsymbol{s}\in\mathcal{S}^p} \int_0^{+\infty} t^{p-1+\gamma} f(t\boldsymbol{s}) \ \mathrm{d}t < +\infty,$$

since $\langle \boldsymbol{s}, \boldsymbol{X} \rangle$ is also Gaussian and hence it possesses finite moments. It follows that (5.4) holds and so does Theorem 24.

The set of all $\tau$-directional quantiles, $\{\boldsymbol{\theta}_\tau(\boldsymbol{s}) : \ \boldsymbol{s} \in \mathcal{S}^p\}$, give us a surface in $\mathbb{R}^p$. The next theorem shows when this surface is continuous. Let us denote by $\lambda_1$ *Lebesgue measure* in $\mathbb{R}$. Suppose that the *support* $\mathcal{M} = \mathrm{sp}(P)$ is *star–shaped* about the central point $\boldsymbol{\mu}$ in the sense of the following definition.

**Definition 16.** We say that a set $\mathcal{M}$ is *star–shaped* about a point $\boldsymbol{\mu} \in \mathcal{M}$ if, for every $\boldsymbol{x} \in \mathcal{M}$, the line segment joining $\boldsymbol{\mu}$ and $\boldsymbol{x}$ is completely contained in $\mathcal{M}$.

**Theorem 26.** *Let $\boldsymbol{\mu} \in \mathsf{int}(\mathcal{M})$ and let $\mathcal{M}$ be star shaped about $\boldsymbol{\mu}$. Further suppose that condition (5.3) or (5.4) holds, the density $f$ is bounded and continuous on $\mathcal{M}$ and*

$$\lambda_1(\partial\mathcal{M} \cap l) = 0, \quad \textit{for arbitrary line } l \textit{ such that } \boldsymbol{\mu} \in l. \tag{5.5}$$

*Then $\boldsymbol{s} \mapsto \boldsymbol{\theta}_\tau(\boldsymbol{s})$ is continuous function for any $\tau \in (0, 1)$.*

*Proof.* First we show that for each vector of angles $\boldsymbol{\varphi}_0$ and any sequence of vectors of angles $\{\boldsymbol{\varphi}_n\}$, such that $\boldsymbol{\varphi}_n \xrightarrow{n\to+\infty} \boldsymbol{\varphi}_0$, the conditional distribution function

$$Q_n(x) = Q(x|\boldsymbol{\varphi}_n) = \int_0^x q(t|\boldsymbol{\varphi}_n) \ \mathrm{d}t = \frac{1}{s(\boldsymbol{\varphi}_n)} \int_0^x p(t, \boldsymbol{\varphi}_n) \ \mathrm{d}t$$

converges weakly to $Q_0(x) = Q(x|\boldsymbol{\varphi}_0)$. This means to show that for almost all $r > 0$ we can change the limit and the integral sign.

Assumption $\boldsymbol{\mu} \in \mathsf{int}(\mathcal{M})$ together with the continuity of $f$ imply the continuity of the joint density $p$ (see (5.2)) on $\{(r, \boldsymbol{\varphi}) : \ r \geq 0, \ \psi(r, \boldsymbol{\varphi}) \in \mathsf{int}(\mathcal{M})\}$. Further, condition (5.5) guarantees that all the limits of $\boldsymbol{\varphi} \mapsto p(r, \boldsymbol{\varphi})$ exist (in other words we can informally say that no "jumps" of the quantile $\boldsymbol{\theta}_\tau(\cdot)$ will occur). Since $f$ is bounded we can use *Lebesgue convergence dominating theorem* in the following equation. For any $r > 0$ it holds that

$$\lim_{n\to+\infty} \int_0^r p(t, \boldsymbol{\varphi}_n) \ \mathrm{d}t$$

$$= \lim_{n\to+\infty} \int_0^r t^{p-1} \sin^{p-2}\varphi_{n,1} \dots \sin\varphi_{n,p-2}\ f(\psi(t,\boldsymbol{\varphi}_n))\ \mathrm{d}t \qquad (5.6)$$
$$= \int_0^r p(t,\boldsymbol{\varphi}_0)\ \mathrm{d}t.$$

To finish the proof of the weak convergence of $Q_n$ we need to show similar result to (5.6) for the marginal density

$$s(\boldsymbol{\varphi}) = \int_0^{+\infty} p(t,\boldsymbol{\varphi})\ \mathrm{d}t = \int_0^{+\infty} t^{p-1} \sin^{p-2}\varphi_1 \dots \sin\varphi_{p-2}\ f(\psi(t,\boldsymbol{\varphi}))\ \mathrm{d}t.$$

Suppose $\varepsilon > 0$. Then, since (5.3) (or (5.4) which is a sufficient condition for (5.3)) holds, there exists $r_0 > 0$ such that

$$\left|\sup_{\boldsymbol{\varphi}} \int_{r_0}^{+\infty} p(t,\boldsymbol{\varphi})\ \mathrm{d}t\right| \le \sup_{\boldsymbol{\varphi}} \int_{r_0}^{+\infty} t^{p-1} f(\psi(t,\boldsymbol{\varphi}))\ \mathrm{d}t < \varepsilon. \qquad (5.7)$$

And so eventually, since (5.6) holds, there exists $n_0$ such that if $n > n_0$ then

$$|s(\boldsymbol{\varphi}_n) - s(\boldsymbol{\varphi}_0)| \le \left|\int_0^{r_0} p(t,\boldsymbol{\varphi}_n)\ \mathrm{d}t - \int_0^{r_0} p(t,\boldsymbol{\varphi}_0)\ \mathrm{d}t\right|$$
$$+ \left|\int_{r_0}^{+\infty} p(t,\boldsymbol{\varphi}_n))\ \mathrm{d}t\right| + \left|\int_{r_0}^{+\infty} p(t,\boldsymbol{\varphi}_0))\ \mathrm{d}t\right| < 3\varepsilon.$$

The inequality for last two terms is obtained by using (5.7).

Hence $Q_n$ converges weakly to $Q_0$. According to Lemma 8.3.1 in [Resnick, 1999], $Q_\tau^{-1}(\boldsymbol{\varphi}_n) \xrightarrow{n\to+\infty} Q_\tau^{-1}(\boldsymbol{\varphi}_0)$ for arbitrary $\tau \in (0,1)$. *Heine definition of continuity* finishes the proof. ■

*Proof of Theorem 24.* The proof is very similar to the proof of Theorem 26. Without loss of generality we assume $\boldsymbol{s} = (0,\dots,0,1)^T$. Such direction has hyperspherical coordinates representation as the vector of angles $\widetilde{\boldsymbol{\varphi}} = (0,\dots,0)^T$. According to Note 31 and according to the proof of the weak convergence of $Q_n$ from the proof of Theorem 25, one has

$$\lim_{\boldsymbol{\varphi}\to\widetilde{\boldsymbol{\varphi}}} Q(r|\boldsymbol{\varphi}) = Q(r|\widetilde{\boldsymbol{\varphi}}), \quad \forall r > 0.$$

We now show that this limit is equal to $F(r|\boldsymbol{s})$.

$$F(r|\boldsymbol{s}) = \lim_{\delta\to0+} \mathsf{P}(\boldsymbol{X} \in \mathcal{A}_{\boldsymbol{\mu}}^{\delta}(\boldsymbol{s},r)|\ \boldsymbol{X} \in \mathcal{A}_{\boldsymbol{\mu}}^{\delta}(\boldsymbol{s}))$$
$$= \lim_{\delta\to0+} \frac{\int_0^r \int_0^\delta \cdots \int_0^\delta \int_0^{2\pi} p(t,\boldsymbol{\varphi})\ \mathrm{d}\varphi_{p-1} \dots\ \mathrm{d}\varphi_1\ \mathrm{d}t}{\int_0^{+\infty} \int_0^\delta \cdots \int_0^\delta \int_0^{2\pi} p(t,\boldsymbol{\varphi})\ \mathrm{d}\varphi_{p-1} \dots\ \mathrm{d}\varphi_1\ \mathrm{d}t}$$
$$= \lim_{\delta\to0+} \frac{\int_0^r 2\pi\delta^{p-2} p(t,\boldsymbol{\varphi}_{t,\delta}^*)\ \mathrm{d}t}{\int_0^{+\infty} 2\pi\delta^{p-2} p(t,\boldsymbol{\varphi}_{t,\delta}^*)\ \mathrm{d}t} = \lim_{\delta\to0+} \frac{\int_0^r p(t,\boldsymbol{\varphi}_{t,\delta}^*)\ \mathrm{d}t}{\int_0^{+\infty} p(t,\boldsymbol{\varphi}_{t,\delta}^*)\ \mathrm{d}t} \qquad (5.8)$$

$$= \frac{\lim_{\delta\to 0+} \int_0^r p(t, \boldsymbol{\varphi}^*_{t,\delta})\ \mathrm{d}t}{\lim_{\delta\to 0+} \int_0^{+\infty} p(t, \boldsymbol{\varphi}^*_{t,\delta})\ \mathrm{d}t} = \frac{\int_0^r p(t, \widetilde{\boldsymbol{\varphi}})\ \mathrm{d}t}{\int_0^{+\infty} p(t, \widetilde{\boldsymbol{\varphi}})\ \mathrm{d}t} = Q(r|\widetilde{\boldsymbol{\varphi}}). \tag{5.9}$$

In equation (5.8) we used the *mean value theorem for integration*, where for any $t > 0$ elements of $\boldsymbol{\varphi}^*_{t,\delta}$ satisfy $(\varphi^*_{t,\delta})_i \in [0, \delta),\ i = 1, \ldots, p-2,\ (\varphi^*_{t,\delta})_{p-1} = 0$. The change of limit and integration sign in (5.9) is possible using the same arguments as in the proof of Theorem 26. And finally, for arbitrary $u > 0$, $\lim_{\delta\to 0+} p(t, \boldsymbol{\varphi}^*_{t,\delta}) = p(t, \widetilde{\boldsymbol{\varphi}})$, since $p(t, \cdot)$ is bounded and continuous function. ■

**Example 18.** Do we really need assumption (5.3) to hold? Consider function

$$h(t) = \begin{cases} \dfrac{1}{4t^2} & \text{if } |t| > 1, \\ \dfrac{1}{4} & \text{if } |t| \le 1 \end{cases}$$

and define $f(x_1, x_2) = h(x_1)h(x_2)$. Then $f$ is a density of a random vector with spherically symmetric distribution. We choose the center of symmetry $(0, 0)^T$ as the center for the directional quantiles. Since

$$\lim_{r\to+\infty} \sup_{\boldsymbol{s}\in\mathcal{S}^p} \int_r^{+\infty} t f(t\boldsymbol{s})\ \mathrm{d}t = \lim_{r\to+\infty} \int_r^{+\infty} t f(t, 0)\ \mathrm{d}t$$
$$= \lim_{r\to+\infty} \int_r^{+\infty} \frac{1}{16t}\ \mathrm{d}t = +\infty,$$

assumption (5.3) does not hold. And, indeed, it can be shown that $\tau$-directional quantiles $\boldsymbol{\theta}_\tau(\cdot)$ at directions that correspond to the angles $0,\ \pi/2,\ \pi,\ 3\pi/2$ tends to infinity for any $\tau \in (0, 1)$.

Univariate quantiles are often used for construction of *confidence intervals*. One would like to know if it makes sense to use directional quantiles for such a purpose in the multivariate case. In the following, for any random vector $\boldsymbol{Z}$ and its $\tau$-directional quantiles $\boldsymbol{\theta}_\tau(\boldsymbol{s})$, we denote $\tau$-central region as

$$\mathcal{C}_{\boldsymbol{Z}}(\tau) = \{\boldsymbol{\mu} + t\boldsymbol{s} :\ 0 \le t \le F^{-1}(\tau|\boldsymbol{s}),\ \boldsymbol{s} \in \mathcal{S}^p\}.$$

**Theorem 27.** *Let $\boldsymbol{X}$ be a random vector with a continuous distribution. Then*

*(i)* $\mathsf{P}\big(\boldsymbol{X} \in \mathcal{C}_{\boldsymbol{X}}(\tau)\big) = \tau, \quad \forall \tau \in (0, 1)$,

*(ii)* $\mathcal{C}_{\boldsymbol{X}}(\alpha) \subseteq \mathcal{C}_{\boldsymbol{X}}(\beta)$ *if* $0 < \alpha \le \beta < 1$,

*(iii)* $\mathcal{C}_{\mathbf{A}\boldsymbol{X}}(\tau) = \mathbf{A}\mathcal{C}_{\boldsymbol{X}}(\tau)$ *for an arbitrary nonsingular matrix* $\mathbf{A}$*, if the center of the random vector* $\mathbf{A}\boldsymbol{X}$ *is equal to* $\mathbf{A}\boldsymbol{\mu}$.

*Proof.* (i): Using hyperspherical representation of the set $\mathcal{C}_{\boldsymbol{X}}(\tau)$ one has

$$\begin{aligned}\mathsf{P}\big(\boldsymbol{X} \in \mathcal{C}_{\boldsymbol{X}}(\tau)\big) &= \int_0^{2\pi}\int_0^{\pi}\dots\int_0^{\pi}\int_0^{Q_\tau^{-1}(\boldsymbol{\varphi})} q(r|\boldsymbol{\varphi})s(\boldsymbol{\varphi})\ \mathrm{d}r\ \mathrm{d}\varphi_1\dots\ \mathrm{d}\varphi_{p-1}\\ &= \int_0^{2\pi}\int_0^{\pi}\dots\int_0^{\pi}\tau\ s(\boldsymbol{\varphi})\ \ \mathrm{d}\varphi_1\dots\ \mathrm{d}\varphi_{p-1}\\ &= \tau\int_0^{2\pi}\int_0^{\pi}\dots\int_0^{\pi} s(\boldsymbol{\varphi})\ \ \mathrm{d}\varphi_1\dots\ \mathrm{d}\varphi_{p-1} = \tau.\end{aligned}$$

The proof of property (ii) is clear. Property (iii) is an obvious consequence of the fact that an affine transformation maps an arbitrary line to a line. ■

### 5.2.2 Estimation

Now we show how we can estimate the directional quantiles. The estimate is based on Def. 15.

For now we suppose that we already know the central point of our data. It may mean that a position of the center arises naturally from a character of data. If we do not know it (actually more realistic situation) we assume that we have already estimated such point. In both cases we denote the center used for estimation of the directional quantiles by $\hat{\boldsymbol{\mu}}$.

We estimate the $\tau$-directional quantile in direction given by $\boldsymbol{s} \in \mathcal{S}^p$ in straightforward way according to Def. 15. We simply estimate the conditional distribution only from observations that lie in a cone of some predefined aperture $2\alpha$. The parameter (angle) $\alpha$ here represents the *"kernel width"* and we first have to choose it. Choice of the kernel width can be done in the same way as a choice of a bandwidth in kernel density estimates or smoothing regression methods. After choosing the width of the kernel, the estimate for the conditional distribution function $F(r|\boldsymbol{s})$ can be obtained by employing, in Def. 15, the *empirical probability measure* $P_{\mathrm{n}}(A) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{\boldsymbol{X}_i \in A\}$. Now, the estimate is defined as

$$\begin{aligned}F_{\alpha,n}(r|\boldsymbol{s}) = P_{\mathrm{n}}(\mathcal{A}_{\boldsymbol{\mu}}^{\alpha}(\boldsymbol{s},r)\,|\,\mathcal{A}_{\boldsymbol{\mu}}^{\alpha}(\boldsymbol{s})) &= \frac{P_n(\mathcal{A}_{\boldsymbol{\mu}}^{\alpha}(\boldsymbol{s},r))}{P_n(\mathcal{A}_{\boldsymbol{\mu}}^{\alpha}(\boldsymbol{s}))}\\ &= \frac{\sum_{i=1}^{n}\mathbb{1}\left\{\boldsymbol{X}_i \in \mathcal{A}_{\boldsymbol{\mu}}^{\alpha}(\boldsymbol{s},r)\right\}}{\sum_{i=1}^{n}\mathbb{1}\left\{\boldsymbol{X}_i \in \mathcal{A}_{\boldsymbol{\mu}}^{\alpha}(\boldsymbol{s})\right\}}.\end{aligned} \tag{5.10}$$

The $\tau$-directional quantile in a direction given by $\boldsymbol{s}$ is then estimated in a standard way by

$$F_{\alpha,n}^{-1}(\tau|\boldsymbol{s}) = \inf\{r \geq 0 :\ F_{\alpha,n}(r|\boldsymbol{s}) \geq \tau\}.$$

This is an easy and computationally non–demanding way how to obtain an estimate. If we choose $\alpha \to 0$ (sufficiently slowly) with increasing sample size $n$ then it can be shown that under certain conditions $F_{\alpha,n}$ is an *consistent estimate.*

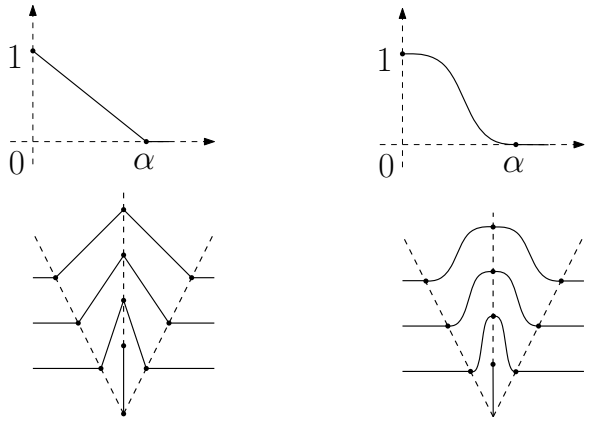



Figure 5.1: Left panel illustrates a *triangular* kernel function. Right panel illustrates a *tricube* kernel (5.12).

As well as in the kernel density estimates we can in (5.10) use a *kernel "cone" function* instead of indicator of a cone. Suppose a "cone" kernel function

$$w: \ [0,\pi] \longrightarrow [0,+\infty), \ \text{where } w(\beta) = 0 \text{ if } \beta \in (\pi/2, \pi].$$

and define a kernel estimate of the conditional distribution function as

$$F_{w,n}(r|\boldsymbol{s}) = \frac{\sum_{i=1}^{n} w(\measuredangle(\boldsymbol{X}_i - \boldsymbol{\mu}, \boldsymbol{s}))\ \mathbb{1}\{\|\boldsymbol{X}_i - \boldsymbol{\mu}\| \le r\}}{\sum_{i=1}^{n} w(\measuredangle(\boldsymbol{X}_i - \boldsymbol{\mu}, \boldsymbol{s}))}. \tag{5.11}$$

Again, the conditional quantiles $F_{w,n}^{-1}$ are defined in the same way as before. Equality (5.11) states more general form of (5.10). For instance if we choose $w(\beta) = \mathbb{1}\{\beta \le \alpha\}$ then we get the same result as in (5.10). $\alpha$ is the bandwidth parameter. Another example is the *triangular kernel*. See the left panel of Fig. 5.1 for illustration of this kernel function and the left panel of Fig. 5.2 for application of this kernel. This kernel linearly decreases from 1 to 0 on the interval $[0, \alpha]$, where $\alpha \in (0, \pi/2]$. Next useful kernel is the *tricube kernel* (see the right panel of Fig. 5.1) $w(\beta) = k(\beta/\alpha)$, where

$$k(u) = \frac{70}{80}\left(1 - |u|^3\right)^3 \mathbb{1}\{|u| \le 1\} \tag{5.12}$$

for bandwidth parameter $\alpha$. The tricube kernel usually leads to good results since it smoothly decreases to zero value in the predefined point.

In the following remark we list the properties that are known from kernel estimates theory or that are clear. So they are stated without the proof in the following remark. We denote the sample version of the set $\mathcal{C}_{\boldsymbol{X}}(\tau)$ for a kernel function $w$ by $\mathcal{C}_{w,n}(\tau)$, i.e.

$$\mathcal{C}_{w,n}(\tau) = \{\boldsymbol{\mu} + t\boldsymbol{s}: \ 0 \le t \le F_{w,n}^{-1}(\tau|\boldsymbol{s}), \ \boldsymbol{s} \in \mathcal{S}^p\}.$$

*Remark* 33. The following properties hold.

(i) According to the sample size we can choose a kernel $w_n$ such that $F^{-1}_{w_n,n}(\tau|\boldsymbol{s})$ is a consistent estimate of $F^{-1}(\tau|\boldsymbol{s})$, thus for sufficiently large $n$ approximately $\lfloor n\tau \rfloor$ points lie in the confidence set $\mathcal{C}_{w_n,n}(\tau)$.

(ii) $\mathcal{C}_{w,n}(\tau)$ is rotational invariant for any $w$.

(iii) $\mathcal{C}_{w,n}(\tau_1) \subseteq \mathcal{C}_{w,n}(\tau_2)$ for any $w$ if $\tau_1 \leq \tau_2$.

When $p = 2$, i.e. in two dimensional space it is also possible to estimate directional quantiles using *$\tau$-quantile regression* (see [Koenker, 2005]) for doublets $(R_i, F_i),\ i = 1, \ldots, n$. Here $(R_i, F_i),\ i = 1, \ldots, n$, is a random sample obtained by inverse polar transformation $\psi^{-1}$ of random sample $\boldsymbol{X}_i,\ i = 1, \ldots, n$ (see (5.1)). We can apply $\tau$-quantile regression estimate of $R_i$ on

$$1, \sin F_i, \cos F_i, \sin 2F_i, \cos 2F_i, \ldots, \sin LF_i, \cos LF_i,$$

where $L$ is the order of *Fourier series expansion.* For more information about Fourier series see e.g. [Tolstov & Silverman, 1976]. Parameter $L$ has to be chosen in advance. Any usual technique such as *cross-validation* can be used. Higher values (relative to sample size $n$) of $L$ lead to "rugged" estimate, lower values lead to "oversmooth" estimate. If we denote estimated coefficients of finite Fourier series expansion by $\widehat{\alpha}_0, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_L, \widehat{\beta}_1, \ldots, \widehat{\beta}_L$ we obtain continuous $2\pi$-periodical function

$$r_\tau(\varphi) = \widehat{\alpha}_0 + \sum_{i=1}^{L} \widehat{\alpha}_i \sin i\varphi + \widehat{\beta}_i \cos i\varphi. \tag{5.13}$$

After transformation back to Cartesian coordinates we obtain an estimate of quantile contour. We can proceed in similar way if $p > 2$, but a non-smoothness issues for certain values of angles has to be solved and hence other expansion such as *spherical harmonics* expansions are recommended to use.

### 5.2.3 Selection of the Smoothing Parameter

The choice of smoothing parameter in (5.10) and (5.11) or the choice of order of expansion in (5.13) plays an important role. We suggest to use the *$K$-fold cross-validation,* where $K = 10$, to deal with bias–variance trade-off. Let $\kappa : \{1, \ldots, n\} \to \{1, \ldots, K\}$ be an indexing function that indicates the partition to which observation $i$ is allocated by the randomization. For kernel estimate we choose smoothing parameter $\alpha$ (or even a kernel $w$) according to criterion

$$CV(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \rho_\tau\Big( \|\boldsymbol{X}_i - \boldsymbol{\mu}\| - F^{-1}_{w,n,\kappa(i)}\big(\tau \mid \boldsymbol{X}_i/\|\boldsymbol{X}_i\|\big) \Big),$$

where $F^{-1}_{w,n,k}$ is an estimate based on (5.11) or (5.10) computed with the $k$–th part of the data removed. The loss function $\rho_\tau$ is the quantile regression loss function, i.e.

$$\rho_\tau(u) = u(\tau - \mathbb{1}\{u < 0\}).$$

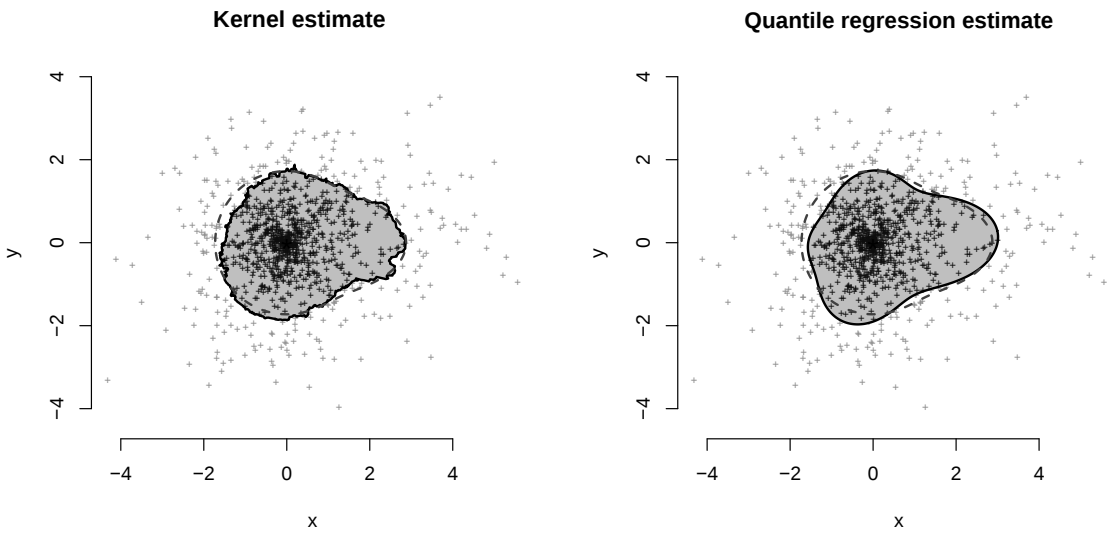



Figure 5.2: Estimated 75% reference quantile contours for artificial data of the sample size $n = 1000$. On the left picture the estimate obtained by (5.11) for the *triangular kernel* with $\alpha_0 = \pi/7 \doteq 0.449$ is plotted (chosen by cross validation - see Fig. 5.3). 749 points lie inside the estimated central region. The *Fourier series estimate* (5.13) for $L = 4$ is on the right picture. 751 measurements lie inside. The dashed contour denotes the theoretical contour with 760 points lying inside.

Tenfold cross-validation was used in Example 19, see also Fig. 5.3. It seems that for the triangular kernel we should choose value of bandwidth $\alpha$ between 0.35 and 0.5.

### 5.2.4 Choice of the Center

To know a position of a center is more a wish than a reality. Of course, sometimes there exists a naturally given center (e.g. when we measure deviations, it may sound reasonable to choose the origin $\mathbf{0}$ as a center). In other cases we need to choose somehow such point. We suggest to use some multivariate generalization of the *univariate median.* Most of these generalizations satisfy the desirable property that the deepest point is in some sense the most "surrounded" by data.

We recommend to use some of the data depth function showed in the previous chapters (actually to find appropriate center point for directional quantiles was one of motivation to work on data depth). In particular use of the *generalized halfspace depth* leads to good central points. The most appropriate candidate points for a center we obtain by using the *kernel cone depth* for the weight function defined in (2.8). The best results we get if the kernel cone weight function (and bandwidth parameter) is the same as in the estimation of directional quantiles in (5.11). But it is tricky to properly choose the bandwidth parameter if we do not know the position of center.

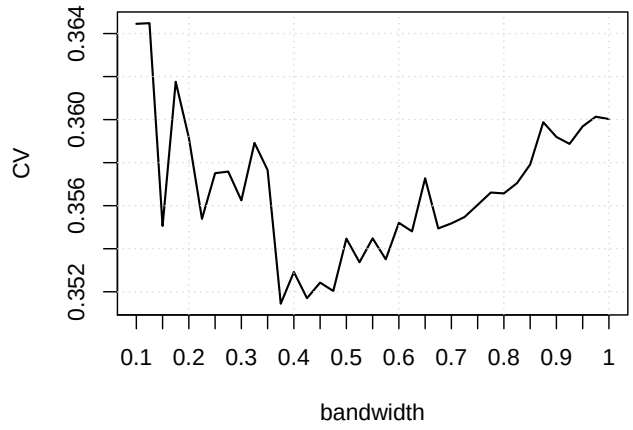



Figure 5.3: Tenfold cross-validation curve for data from Example 19 as a function of the bandwidth parameter $\alpha$. Bumpiness occurs due to the fact that for each value of the bandwidth $\alpha$ we use different (randomly chosen) division into the training sets and the validation set. It also indicates variance of the cross-validation procedure. Due to the variance there does not exist unique "best" choice of the bandwidth but there are many values of bandwidth parameters that are not statistically different.

The deepest point obtained by kernel cone depth is the most appropriate choice of center for estimating the directional quantiles because the radius is the most balanced in the sense that probabilities along the least favourable direction is the highest possible.

### 5.2.5 Examples

**Example 19** (The kernel estimate and the Fourier series estimate applied on artificial data)**.** See Fig. 5.2. Dataset was simulated from the model defined as

$$X = R\cos F,\ Y = R\sin F,$$
$$\text{where } F \sim \mathcal{U}(0, 2\pi) \text{ and } R|F \sim \left|\mathcal{N}\left(0, \left(3/2 + \cos^{30}(F/2)\right)^2\right)\right|.$$

Simulated data $(X_i, Y_i),\ i = 1, \ldots, 1000$ are shown on the figure along with the estimated 75% reference quantile contours. The theoretical center is the origin $\mathbf{0}$. The estimate for the center $(0.005, 0.004)^T$ was obtained by using the kernel cone depth weight function (2.8). The *triangular cone weight function* (see the left panel of Fig. 5.1) with $\alpha_0 = \pi/7 \doteq 0.45$ was used. The same weight function was used as a kernel while estimating the contour shown on the left panel of Fig. 5.2. On Fig. 5.3 we can see the tenfold *cross-validation* curve. It seems that we should choose value of bandwidth between 0.35 and 0.5. 749 points lie inside the estimated contour.

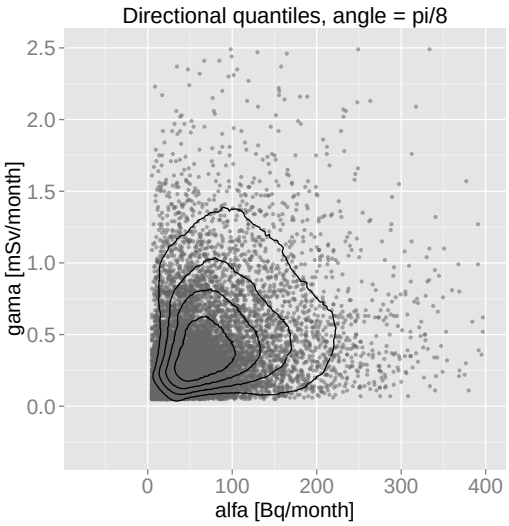



Figure 5.4: Kernel directional quantiles estimate of 20%, 40%, 60%, 80% contours for alpha and gamma intakes.

The Fourier series estimate for $p = 4$ is on the right panel of Fig. 5.2. 751 measurements lie inside. The dotted contour denotes the theoretical contour with 760 points lying inside.

**Example 20** (The alpha and gamma intakes in the Rožná mine)**.** Here we show contours for data from Example 3. 20%, 40%, 60%, 80% contours are shown on the Figure 5.4. The *triangular cone weight function* was used for the deepest point (center) calculation and for kernel estimates of directional quantiles. The bandwidth parameter $\alpha = \pi/16$ was chosen. The central point is $(62.5 \text{ Bq/month}, 0.35 \text{ mSv/month})^T$.

### 5.2.6 Relation to the Data Depth

Besides the fact that the deepest point leads to choice of a good central point the directional quantiles can be considered as a data depth function. Given a central point $\boldsymbol{\theta}$ any point $\boldsymbol{x} \in \mathbb{R}^p$ can be represented as

$$\boldsymbol{x} = \boldsymbol{\theta} + r(\boldsymbol{x})\boldsymbol{s}(\boldsymbol{x}),$$

where $r(\boldsymbol{x}) \geq 0$ and $\boldsymbol{s}(\boldsymbol{x}) \in \mathcal{S}^p$. Then $F\big(r(\boldsymbol{x})|\boldsymbol{s}(\boldsymbol{x})\big)$ can be thought as a measure of non-centrality (outlyingness) of $\boldsymbol{x}$ and hence the function

$$\mathrm{DQ}(\boldsymbol{x}) = \frac{1}{1 + F\big(r(\boldsymbol{x})|\boldsymbol{s}(\boldsymbol{x})\big)}$$

can be though as a data depth function. It is clear that DQ satisfies all the properties of Definition 1 (key properties). Note that the central sets need not

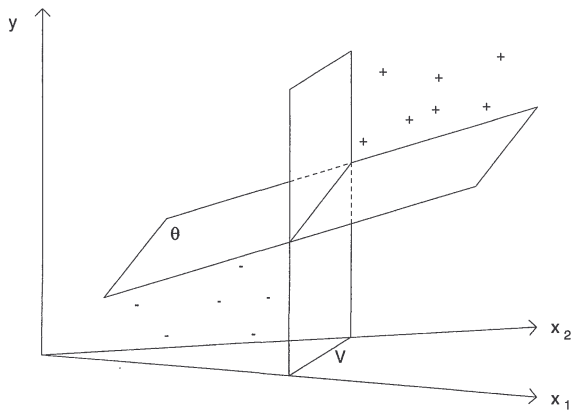



Figure 5.5: Regression plane in nonfit position. The figure was reproduced from [Rousseeuw & Hubert, 1999].

to be convex and that the sample version does not satisfy key property 1 - affine invariance.

## 5.3 Regression Depth

There were attempts to apply concept of the data depth to other than Euclidean spaces. Papers [Rousseeuw & Hubert, 1999, Rousseeuw et al., 1999] introduce a notion of depth in the regression setting. This section shows brief overview of this concept and shows new finding about a regression depth (see Def. 17) in regression setting with quadratic (polynomial) transformation of the independent variable.

Consider a linear regression

$$\mathsf{E}\,\boldsymbol{Y} = \beta_0 + \mathbf{X}\boldsymbol{\beta},$$

where $\boldsymbol{Y} \in \mathbb{R}^n$ is a random vector, $\mathbf{X} \in \mathbb{R}^{n\times p}$ is a dataset matrix and $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of regression coefficients. Regression depth assigns values to regression functions $\theta_0 + \boldsymbol{x}^T\boldsymbol{\theta}$, $\theta_0 \in \mathbb{R}, \boldsymbol{\theta} \in \mathbb{R}^p$, according to their *centrality* with respect to dataset $\mathbf{X}$.

Univariate median is a point with property that the number of removed observation (resp. probability mass of removed halfline in population case) to make this point lying outside the dataset is the highest possible. Similar though is used for definition of regression depth. Informally the regression depth of a regression plane is the smallest number of observations $(Y_i, \boldsymbol{X}_i)$ that have to be removed to guarantee that no points lie between this plane and a vertical plane. Fig. 5.5 shows plane, represented by vector of regression coefficient $\boldsymbol{\theta}$, where no points have to removed.

For formal definition we first need to define *nonfit* position of a plane. Denote by $y_i \in \mathbb{R}$ and $\boldsymbol{x}_i \in \mathbb{R}^p$, $i = 1, \ldots, n$, observed dependent variable and independent variables. Further for any vector of regression coefficients $\theta_0$, $\boldsymbol{\theta} =$

$(\theta_1, \ldots, \theta_p)^T$ denote by

$$r_i(\theta_0, \boldsymbol{\theta}) = y_i - \theta_0 - \boldsymbol{x}_i^T \boldsymbol{\theta}, \quad i = 1, \ldots, n,$$

the residuals of the fit.

**Definition 17.** Regression coefficients $\theta_0$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ are called a *nonfit* to $\mathbf{X}$ (matrix with $\boldsymbol{x}_i$ as rows) iff there exists an affine hyperplane $V$ in independent variables space $\mathbb{R}^p$ such that no $\boldsymbol{x}_i$ belongs to $V$ and such that $r_i(\theta_0, \boldsymbol{\theta}) > 0$ for all $\boldsymbol{x}_i$ in one of its open halfspaces and $r_i(\theta_0, \boldsymbol{\theta}) < 0$ for all $\boldsymbol{x}_i$ in the other open halfspace.

See Fig. 5.5 for illustration of a nonfit. Now we can proceed to the definition of regression depth.

**Definition 18.** The depth $\mathrm{RegD}(\theta_0, \boldsymbol{\theta})$ is the smallest number of observations $(y_i, \boldsymbol{x}_i)$ that would be removed to make a $\theta_0, \boldsymbol{\theta}$ nonfit.

The depth function RegD provides good center–outward ordering of regression planes. The deepest point is an estimate of $\beta_0, \boldsymbol{\beta}$. It posses good properties - almost identical properties as the median regression estimates plus it is very robust to leverages. Unfortunately it is computationally demanding. For computational aspects see [Rousseeuw & Struyf, 1998]. Only "naive" $O(n^{p+2})$ algorithm exists, small sample distribution is hard to determine and the asymptotic distribution is not in explicit form and it need to obtained via simulations (see [Bai Zhi-Dong, 1999]). Also no implementations in statistical software exist. There is a room for further investigations on this very promising statistical concept.

Often we work with quadratic transpose of an univariate dependent variable $x$. Suppose model

$$\mathsf{E}[Y|x] = \beta_0 + \beta_1 x + \beta_2 x^2,$$

where $\beta_0, \beta_1, \beta_2, x \in \mathbb{R}$. The nonfit and the regression depth is defined the same way as in Def. 17 and 18 with $\boldsymbol{x}_i = (x_i, x_i^2)^T$. There arises a question how the depth can be interpreted in original univariate variable $x$ space.

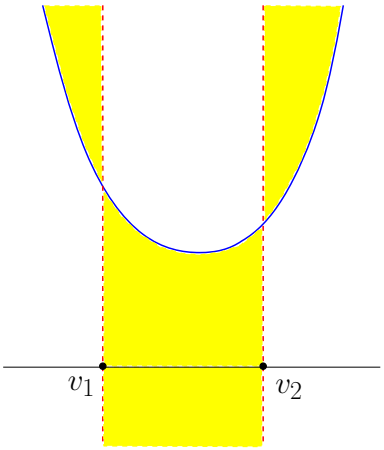



Figure 5.6: Regression depth in quadratic regression setting is equal to the minimum (according to all positions of $v_1 \leq v_2$) number of observations lying only in yellow or only in white areas.

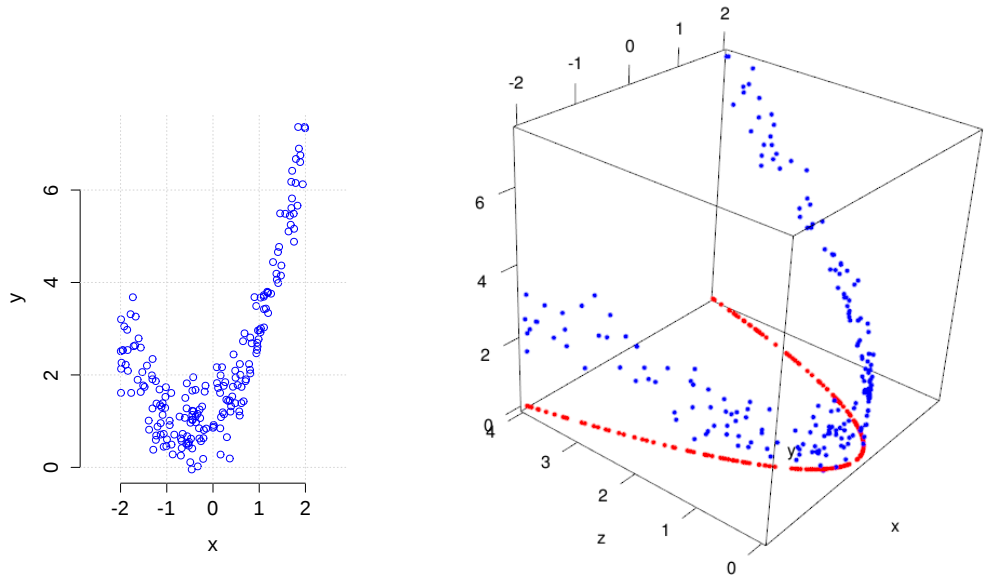



Figure 5.7: Illustration of quadratic regression model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$. *Left panel* shows dependent variable plotted against independent variable $x$. We see quadratic dependence. *Right panel* shows the same dataset where the dependent variable is plotted against $x$ and $x^2$ (axis denoted by $z$). Observed points are plotted in blue color, their projection to the independent $x, x^2$-variables space are plotted in red color. *Left panel* also shows the orthogonal projection to the $x, y$-space.

*Remark* 34. Regression coefficients $\theta_0$ (intercept), $\theta_1$ (linear term), $\theta_2$ (quadratic term) in quadratic regression setting are nonfit if there exist points $v_1 \leq v_2$ such that at least one of the following properties holds

1. $r_i(\theta_0, \theta_1, \theta_2) > 0$ for $x_i \in (v_1, v_2)$ and $r_i(\theta_0, \theta_1, \theta_2) < 0$ for $x_i \in (-\infty, v_1) \cup (v_2, +\infty)$.
2. $r_i(\theta_0, \theta_1, \theta_2) < 0$ for $x_i \in (v_1, v_2)$ and $r_i(\theta_0, \theta_1, \theta_2) > 0$ for $x_i \in (-\infty, v_1) \cup (v_2, +\infty)$.

Regression depth is the smallest number of observations $(y_i, x_i, x_i^2)$ that would be removed to make $\theta_0, \theta_1, \theta_2$ nonfit.

Fig. 5.6 illustrates the nonfit for quadratic regression. To make a quadratic curve nonfit observations only in the yellow areas or only in the white areas need to be removed.

*Proof of Remark 34.* We only need to realize the geometry of $y, x, x^2$ space - see Fig. 5.7, where the same data are plotted in two ways. Consider a plane in $\mathbb{R}^3$. To be consistent with the right panel of the figure we denote axis by $x, z, y$. The plane can be represented as a set

$$\{(x, z, y) : \ y - \theta_0 - \theta_1 x - \theta_2 z = 0\} \tag{5.14}$$

where $\theta_0, \theta_1, \theta_2$ are "regression" plane coefficients. Observations lie in the set $\{(x, z, y) : \ z = x^2\}$ (see red points in the figure). The projection of observation into the $x, y$ plane ($\{(x, z, y) : \ z = 0\}$) gives us the original space that consists of only $x, y$ coordinates - see the left panel of the figure. The same way we obtain view of a quadratic fit (5.14) - the fitted quadratic function $f(x) = \theta_0 + \theta_1 x + \theta_2 x^2$ is projection of the set

$$\{(x, z, y) : \ z = x^2\} \cap \{(x, z, y) : \ y - \theta_0 - \theta_1 x - \theta_2 z = 0\}$$

into $x, y$ plane. There are 2 possibilities. The first possibility is case when a line in $x, z$ space crosses the parabolic function $z = x^2$ in two points (on Fig. 5.6 these points have $x$ coordinate equal to $v_1$ and $v_2$) and divides the parabola into area above the line and the area under the line. Nonfit means that residuals of all observations belonging to one area of the parabola are positive and residuals of observations that belong to the other area are negative. After projection to $x, y$ plane it means that for the projected parabolic function the property 1. or 2. must hold. The second possibility is the case when a line is parallel with $z$ axis - the line crosses the parabola in only one point and thus after projection to $x, y$ plane there is only one point $v_1$ which divides the parabola into two parts. It is the case why we allow that $v_1 = v_2$ to be hold. ■

The same way we can proceed when using a polynomials of higher orders. But we obtain much more points that divided the line to areas with only negative or only positive residuals. One rule holds - an area with only negative residuals is always followed with an area with only positive residuals or vice versa.

We can also define the regression depth using a "classical" data depth function. Any depth function can be extended to the regression framework if we consider the depth of all regression coefficients of all the planes going through

$p+1$ points from the dataset. The deepest point is an estimate of regression coefficients. This is a multiple regression modification of *Passing-Bablok* regression (see [Passing & Bablok, 1983] or [Leonhardt & Zawta, year unknown]) that is often used in *error in variable* regression problems and in measurements method comparisons because it is pretty robust with respect to outlying observations both in dependent and independent variables.

## 5.4 Functional Data Depth

Functional data has become very popular in recent years. In conventional statistical analyses data are considered as scalars or vectors. In functional data analysis, each observations is considered as a real function, $X_i(t),\ i = 1, \ldots, n,\ t \in I$, where $I \subset \mathbb{R}$ is an interval. Usually, $X_i$ are supposed to be continuous and smooth. This statistical concept arises quite naturally in many areas of research. In biology, economics, medicine, the observed data are often stochastic functions or curves. $t$ is usually time variable. Even some problems that can be solved using techniques of multivariate data analysis is better to approach via functional data, e.g. if a data generating process depends on time and multivariate observations are observed in different times for different subjects. Very comprehensive overview of methods and its applications can be seen in [Ramsay & Silverman, 2005, Ramsay & Silverman, 2002].

Very important task in the functional data analysis is to define a center-outward ordering (and thus rank statistics) and robust estimates such as the median curve or the trimmed mean curve. This is why the functional data depth is nowadays under intensive investigation.

The straightforward way how to define the functional data depth is to express (approximate) functional observations as a finite linear combination of some basis functions and then apply an "ordinary" depth function (see Chapters 1 and 2) on obtained coefficients. This approach seems to be not appropriate because dimension (number of coefficients) is relatively high compared to sample size and *curse of dimensionality* occurs (see e.g. [Hastie et al., 2009], Chapter 1, Section 2.5) - majority of points are lying on the "border" of data cloud and hence they have almost zero depth. In some cases an improvement can be achieved if we reduce dimension with principal components.

[Fraiman & Muniz, 2001] defined an integral depth. It's now called the *Fraiman-Muniz depth* (FD). Suppose an one dimensional depth function D, then the depth of functional observation $x$ is

$$\mathrm{FD}(x) = \int_I \mathrm{D}(x(t))\ \mathrm{d}t. \tag{5.15}$$

Usual choice of D is the halfspace depth, i.e. the minimum of probability mass of the halfline beginning at $x(t)$. This depth provides us with sensible center-outward ordering.

[Hlubinka & Nagy, 2012] proposes to also consider derivatives while making the depth calculations because two very different (e.g.. increasing and decreasing)

functions can have the same depth in some cases. Instead of only $x(t)$ in (5.15) the paper proposes to use, for a given order of derivation $K \geq 1$, a vector of

$$\left(x(t), \frac{\partial}{\partial t}x(t), \ldots, \frac{\partial^K}{\partial t^K}x(t)\right)^T.$$

In this case, D is a *multivariate depth function* - e.g. any depth function mentioned in this thesis. The same approach can be used when dealing with multivariate functional data, i.e. when $\boldsymbol{X}_i(t) = (X_{i1}(t), \ldots, X_{ip}(t))^T,\ t \in I,\ i = 1, \ldots, n$.

Multivariate functional data represent often trajectories. Application of functional data depth in this case can be used in computer / mobile game industry if we want to compare gaming style of one player with gaming styles of other players - we are looking for outlyingness of trajectory of the player. Since integral depth usually leads to good results and there does not exist many methods for analyzing functional data, we can expect that "ordinary" depth functions will found applications in this field in the future.

Another definition was shown in [López-Pintado & Romo, 2007] and [López-Pintado & Romo, 2009]. The depth proposed there is called the *band depth*. Suppose that $P$ is a distribution of functional data of interest, i.e. $P$ is a probability measure on a space of continuous (and usually also smooth) functions. The *functional band depth of* $J \geq 2$ *order* is defined as

$$\mathrm{BD}(x) = \frac{1}{J-1}\sum_{j=2}^{J} \mathsf{P}\big(\inf\{X_1, \ldots, X_j\} \leq x \leq \sup\{X_1, \ldots, X_j\}\big),$$

where $X_1, \ldots, X_j$ is a random sample from $P$. In proposed papers authors show that this depth satisfies the key properties of Def. 1. But it is very weak if the observations are noisy and in comparison to Fraiman-Muniz depth it is very computation demanding ($\mathcal{O}(n^J)$).

# Chapter 6

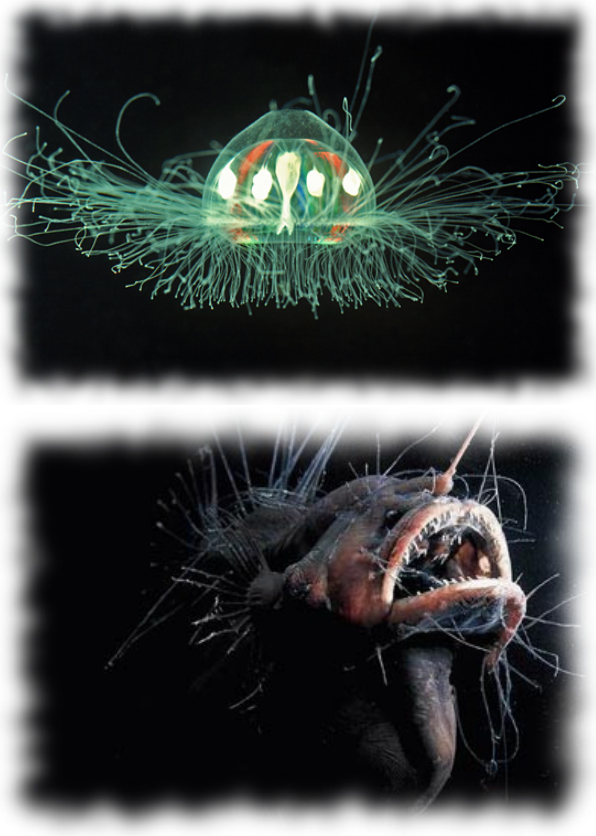
# Conclusion

This thesis brings two new classes of statistical depth functions based on weighted probabilities. Both classes generalize the well known *halfspace depth.* In both cases instead of considering only indicators of halfspaces we consider a more general class of *weighted (half)spaces.* This allows us to more respect the geometric structure of data especially when non-convexly supported distributions are considered. When we deal with non-convexly supported, bimodal or mixed distributions the majority of statistical depth functions often lead to results when big attention is given to observations lying far away from the rest of the data. It is undesirable property, even for the cases where one to one correspondence between depth of the points and the probability distribution exists - consider e.g. trimmed mean, outlier detection or rank test. In these cases it is very important to assign weights (ranks or values of outlyingness) according to the geometric structure of the data. It means that points outside the support or outlying points in the areas with low density should not attain high depth values (and thus high rank or low values of outlyingness). If the latter does not hold the depth based procedures such as rank tests or trimmed mean estimates lost their robustness even in the case when the used depth function is very robust. Proposed depth functions does not suffer from this when suitable weight function is used.

Properties of both proposed classes of depth functions are proved and discussed in detail. We also defined an affine invariant version of proposed depth functions. The uniform strong consistency over $\mathbb{R}^p$ is also proved and some results about the limit distribution are established and discussed. It seems that convergence is very slow, mainly due to the fact that the sample version of the depth is biased. It limits use of the depth in cases when the limit distribution is needed. The generalized halfspace depth posses many good properties and brings a natural link between very local view of data (kernel density estimate) and global view of the data (the halfspace depth). In general it does not provide us with the unique deepest point and the deepest point needs not to lie in the center of symmetry, especially in cases when the center does not lie inside the support. If there exists a need to avoid this the weighted halfspaces ratio depth should be used.

The use of data depth is very sensitive to the what *to be deep* means in specific situations. Badly chosen depth function (wrong definition of the meaning to be

deep) can lead to wrong results. So in advance to depth application we have to properly choose the depth functions according to the character of our problem. For instance the halfspace depth, the Liu depth, the projection depth and many other depth functions are good tools when we work with elliptically symmetric data (especially with multivariate normal distribution) with presence of outlying observations. In this case application of these depths is very effective since it is robust and not a big amount of observations is needed. On other hand when we work with non-elliptically symmetric distribution these depths can lead to biased and misleading results. The depth functions proposed in this thesis can avoid this but in the additional cost of need for a bigger dataset. If it is used for smaller datasets it can happen that the majority of points is marked as points on the border of dataset (i.e. the points with low depth values), especially when a local weight function is used.



*To be deep or not to be deep?*

The thesis also discusses some "classical" data depth related concepts such as regression depth, directional quantiles and functional data depth. In all these fields the proposed depth can bring some improvements.

Nowadays a lot of multivariate data is collected. Hence the use of data depth functions can play important role in the future. Unfortunately the data depth methodology is still not fully developed. Also fast algorithms for higher dimensional data does not exist and contemporary computation power needs not to be sufficient for straightforward computation. The proposed depth functions are

quite computational demanding. So it can limits its use, especially in data with dimension higher than 3, when the computation is very slow. This fact together with need of bigger sample size are the biggest disadvantages of the proposed depth functions.

# Appendix

## Additional Lemmas

**Lemma 28.** *Let* $\mathbf{C}$ *be* $p \times p$ *positive definite matrix. Denote its eigenvalues by* $\lambda_1 \geq \cdots \geq \lambda_p$. *Then* $\lambda_i > 0,\ i = 1, \ldots, p$. *Further denote by* $\mathbf{U}$ *matrix, whose* $k$*-th column is equal to eigenvector* $\boldsymbol{x}_k$, *which corresponds to eigenvalue* $\lambda_k$ *and satisfies* $\boldsymbol{x}_k^T \boldsymbol{x} = 1$. *Let* $\boldsymbol{\Lambda}^{\frac{1}{2}} = diag\{\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_p}\}$. *Then*

$$\mathbf{C} = \mathbf{L}\mathbf{L}, \quad \text{where} \quad \mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^T.$$

The proof is in [Rao, 1973].

**Lemma 29.** *Consider two bounded functions* $f,\ g:\ M \to \mathbb{R}$ *where* $M$ *is subset of a space equipped with a norm or with a distance. Then*

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

*Proof.* 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 $\varepsilon,\ 0 < \varepsilon < \varepsilon_0$, 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 $\varepsilon,\ 0 < \varepsilon < \varepsilon_0$. This completes the proof of the lemma. ■

**Lemma 30.** *Let* $f$ *be a measurable function,* $f:\ \mathcal{M}_1 \times \mathcal{M}_2 \to [0, +\infty)$. *Further suppose that*

$$\lim_{(x_1, x_2) \to (\xi, \nu)} f(x_1, x_2) = f(\xi, \nu)$$

*for almost all* $\xi \in \mathcal{M}_1$ *and all* $\nu \in \mathcal{M}_2$. *Consider a fixed* $\xi_0 \in \mathcal{M}_1$ *and a fixed* $\nu_0 \in \mathcal{M}_2$. *Then for any* $\varepsilon_1 > 0$ *and* $\varepsilon_2 > 0$ *the function*

$$z \mapsto \sup\{f(z - x_1, x_2) :\ x_1 \in \mathcal{M}_1, d_1(x_1, \xi_0) \leq \varepsilon_1;\ x_2 \in \mathcal{M}_2, d_2(x_2, \nu_0) \leq \varepsilon_2\}$$

*is universally measurable.*

*Proof.* The problem is equivalent to a problem of measurability of a function

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

if $h : \mathcal{M}_1 \times \mathcal{M} \to [0, +\infty)$ is a measurable function, where $(\mathcal{M}, \tau)$ is a separable metric space. To see the equivalence let $(\mathcal{M}, \tau) = (\mathcal{M}_1, d_1) \times (\mathcal{M}_2, d_2)$ and $s = (\xi_0, \nu_0)$. Denote $B^a = \{(z,t) : \ h(z,t) > a\}$ and note that $B^a$ is a Borel set for any $a$ due to the measurability of $h$. Denote $C^a := \{(z,s) : \ g(z,s) > a\}$. It is clear that for any $z$

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

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

$$D^{a,e} = \{(z,s,t) \in \mathcal{M}_1 \times \mathcal{M} \times \mathcal{M} : \ (z,t) \in B^a, \ \tau(s,t) \leq e\}$$

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. ■

*Remark* 35. If a function $g$ is universally measurable then for any finite Borel measure $\mu$ on $\mathbb{R} \times \mathbb{R}$ (in particular for any probability measure) there exists 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 $g_2 = g_1$ $\mu$-almost surely. Hence the Lebesgue integral of universally measurable function is well defined.

# Uniform Law of Large Numbers and Entropy

In this section we briefly introduce some basic notions from the empirical process theory. Suppose a measure $Q$ and a class $\mathcal{G}$ of real valued functions on $\mathbb{R}^p$ such that

$$\mathcal{G} \subset \mathcal{L}_q(Q) = \left\{ f : \ \int_{\mathbb{R}^p} |f(\boldsymbol{x})|^q \ \mathrm{d}Q(\boldsymbol{x}) < +\infty \right\}$$

for $q \geq 1$. The $\mathcal{L}_q(Q)$ norm of a function $f$ is defined as

$$\|f\|_{q,Q} = \left( \int_{R^p} |f(\boldsymbol{x})|^q \ \mathrm{d}Q(\boldsymbol{x}) \right)^{1/q}.$$

The *$\delta$–entropy with bracketing* of a class $\mathcal{G}$ is the logarithm of the minimum number $N$ of pairs of functions $\{g_j^L, \ g_j^U\}_{j=1,\ldots,N}$ such that for all $j = 1, \ldots, N$

$$\|g_j^L - g_j^U\|_{q,Q} \leq \delta$$

and such that for each $g \in \mathcal{G}$ there exists a $k \in \{1, \ldots, N\}$ such that

$$g_k^L(\boldsymbol{x}) \leq g(\boldsymbol{x}) \leq g_k^U(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{R}^p.$$

Latter number of pairs we denote by $H_{q,B}(\delta, \mathcal{G}, Q)$. We set $H_{q,B}(\delta, \mathcal{G}, Q) = +\infty$ if no finite set of such pairs, that covers $\mathcal{G}$, exists. Here $\mathcal{G}$ represents a set of transformations of a random vector $\boldsymbol{X} \in \mathbb{R}^p$ with distribution given by $Q$. The following theorem says when the empirical mean converges uniformly over the class of functions $\mathcal{G}$.

**Theorem 31.** *Let $\boldsymbol{X} \in \mathbb{R}^p$ be a random vector with distribution given by $Q$ and let $\mathcal{G}$ be class of real valued functions on $\mathbb{R}^p$ such that $\mathcal{G} \subset \mathcal{L}_1(Q)$. If $H_{1,B}(\delta, \mathcal{G}, Q) < +\infty, \ \forall \delta > 0$, then*

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{X}_i) - \mathsf{E}\, g(\boldsymbol{X}) \right| \overset{n \to +\infty}{\longrightarrow} 0 \quad \textit{almost surely.} \tag{6.1}$$

*We say that $\mathcal{G}$ satisfies the uniform law of large numbers if (6.1) holds.*

The theorem is a corollary of the ordinary law of large numbers and its proof can be found in [Geer, 2000] or in [Pollard, 1984].

# Bibliography


[Agostinelli & Romanazzi, 2011] Agostinelli, C. & Romanazzi, M. (2011). Local depth. *Journal of Statistical Planning and Inference*, 141(2), 817 – 830.

[Bai Zhi-Dong, 1999] Bai Zhi-Dong, H. X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *The Annals of Statistics*, 27(5).

[Barnett, 1976] Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*, (pp. 318–355).

[Burr et al., 2003] Burr, M., Rafalin, E., & Souvaine, D. (2003). Simplicial depth: An improved definition, analysis, and efficiency for the finite sample case. In R. e. a. Liu (Ed.), *DIMACS - Data Depth: Robust Multivariate Analyses, Computational Geometry and Applications*: American Mathematical Society.

[Carrizosa, 1996] Carrizosa, E. (1996). A characterization of halfspace depth. *Journal of Multivariate Analysis*, 58, 21–26.

[Chakraborty et al., 1998] Chakraborty, B., Chaudhuri, P., & Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statistica Sinica.*, 8, 767–784.

[Chen & Morin, 2013] Chen, D. & Morin, P. (2013). Approximating majority depth. *Comput. Geom. Theory Appl.*, 46(9), 1059–1064.

[Chen & Tyler, 2002] Chen, Z. & Tyler, D. E. (2002). The influence function and maximum bias of Tukey's median. *The Annals of Statistics*, 30(6), 1737–1759.

[DasGupta et al., 1995] DasGupta, A., Ghosh, J., & Zen, M. (1995). A new general method for constructing confidence sets in arbitrary dimensions with applications. *Annals of Statistics*, 23(4), 1408–1432.

[Donoho & Gasko, 1992] Donoho, D. & Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Annals of Statistics.*, 20(4), 1803–1827.

[Donoho, 1982] Donoho, D. L. (1982). *Breakdown properties of multivariate location estimators*. PhD thesis, Harvard University.

[Dutta et al., 2011] Dutta, S., Ghosh, A. K., Chaudhuri, P., et al. (2011). Some intriguing properties of Tukey's half-space depth. *Bernoulli*, 17(4), 1420–1434.

[Fraiman et al., 1999] Fraiman, R., Meloche, J., García-Escudero, L. A., Gordaliza, A., He, X., Maronna, R., Yohai, V. J., Sheather, S. J., McKean, J. W., Small, C. G., et al. (1999). Multivariate L-estimation. *Test*, 8(2), 255–317.

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

[Geer, 2000] Geer, S. (2000). *Empirical Processes in M-Estimation*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[Hasil, 2004] Hasil, J. (2004). Problém kvantilu ve více rozměrech (quantiles in multivariate setting). Master's thesis (in Czech), Charles University.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer.

[Hlubinka et al., 2010] Hlubinka, D., Kotík, L., & Vencálek, O. (2010). Weighted halfspace depth. *Kybernetika*, 46(1), 125–148.

[Hlubinka & Nagy, 2012] Hlubinka, D. & Nagy, S. (2012). Functional data depth and classification. *Comput. Stat.* In submission.

[Hlubinka & Vencalek, 2013] Hlubinka, D. & Vencalek, O. (2013). Depth-based classification for distributions with nonconvex support. *Journal of Probability and Statistics*, 2013.

[Ilmonen et al., 2012] Ilmonen, P., Oja, H., & Serfling, R. (2012). On invariant coordinate system (ics) functionals. *International Statistical Review*, 80(1), 93–110.

[Koenker, 2005] Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

[Kong & Mizera, 2012] Kong, L. & Mizera, I. (2012). Quantile tomography: Using quantiles with multivariate data. *Statistica Sinica*, 22, 1589–1610.

[Kong & Zuo, 2010] Kong, L. & Zuo, Y. (2010). Smooth depth contours characterize the underlying distribution. *Journal of Multivariate Analysis*, 101(9), 2222–2226.

[Koshevoy & Mosler, 1997] Koshevoy, G. & Mosler, K. (1997). Zonoid trimming for multivariate distributions. *The Annals of Statistics*, (pp. 1998–2017).

[Kotík, 2007] Kotík, L. (2007). Periodické regresní kvantily (periodic regression quantiles). Master's thesis (in Czech), Charles University.

[Kotík, 2009] Kotík, L. (2009). Uniform strong consistency of halfspace-like depths. In *WDS'09 Proceedings of contributed papers*, part 1, 89–94.

[Kotík, 2014] Kotík, L. (2014). Directional quantiles. *In submission*.

[Kotík & Hlubinka, to appear soon] Kotík, L. & Hlubinka, D. (to appear soon). Generalized halfspace depth.

[Lange et al., 2014] Lange, T., Mosler, K., & Mozharovskyi, P. (2014). Fast nonparametric classification based on data depth. *Statistical Papers*, 55(1), 49–69.

[Leonhardt & Zawta, year unknown] Leonhardt, W. & Zawta, B. (year unknown). Fundamentals of laboratory diagnostics, quality assurance: Introduction to statistics and important definitions.

[Li et al., 2012] Li, J., Cuesta-Albertos, J. A., & Liu, R. Y. (2012). DD-classifier: Nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association*, 107(498), 737–753.

[Liu et al., 2003] Liu, R., Serfling, R., & Souvaine, D. (2003). *Data Depth: Robust Multivariate Analysis, Computational Geometry, and Applications*. DIMACS series in discrete mathematics and theoretical computer science. American Mathematical Soc.

[Liu, 1990] Liu, R. Y. (1990). On a notion of simplicial depth. *The Annals of Statistics*, 18(1), 405–414.

[Liu et al., 1999] Liu, R. Y., Parelius, J. M., & Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussions and rejoinder). *The Annals of Statistics*, 27(3), 783–858.

[Liu & Singh, 2006] Liu, R. Y. & Singh, K. (2006). Rank tests for multivariate scale difference based on data depth. *DIMACS series in discrete mathematics and theoretical computer science*, 72, 17.

[López-Pintado & Romo, 2007] López-Pintado, S. & Romo, J. (2007). Depth-based inference for functional data. *Comput. Stat. Data Anal.*, 51, 4957–4968.

[López-Pintado & Romo, 2009] López-Pintado, S. & Romo, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.*, 486(104), 718–734.

[Mahalanobis, 1936] Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 49–55.

[Massé, 2004] Massé, J.-C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli*, 20(3), 397–419.

[Matoušek, 1992] Matoušek, J. (1992). Computing the center of planar point sets. In J. E. e. a. Goodman (Ed.), *DIMACS: Discrete and Computational Geometry.*: American Mathematical Society.

[Oja, 1983] Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1, 327–332.

[Paindaveine & Van bever, 2013] Paindaveine, D. & Van bever, G. (2013). From depth to local depth: A focus on centrality. *Journal of the American Statistical Association*, 108(503), 1105–1119.

[Passing & Bablok, 1983] Passing, H. & Bablok, W. (1983). A new biometrical procedure for testing the equality of measurements from two different analytical methods. In *Application of linear regression procedures for methods comparison studies in clinical chemistry. Part I.*, volume 21 of *J. Clin. Chem. Clin. Biochem.* (pp. 709–720).

[Pollard, 1984] Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer–Verlag.

[Porzio & Ragozini, 2007] Porzio, G. C. & Ragozini, G. (2007). Convex hull probability depth. *International Workshop on Robust and Nonparametric Statistical Inference. Hejnice, Czech Republic*.

[R Core Team, 2014] R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[Ramsay & Silverman, 2002] Ramsay, J. & Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer.

[Ramsay & Silverman, 2005] Ramsay, J. & Silverman, B. (2005). *Functional Data Analysis*. Springer.

[Rao, 1973] Rao, C. (1973). *Linear statistical inference and its applications*. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley.

[Resnick, 1999] Resnick, S. I. (1999). *A Probability Path*. Birkhauser, Boston.

[Romanazzi, 1999] Romanazzi, M. (1999). Influence function of halfspace depth. *Journal of Multivariate Analysis*, 77, 138–161.

[Rousseeuw et al., 1999] Rousseeuw, P., Aelst, S., & Hubert, M. (1999). Regression depth: Rejoinder. *Journal of the American Statistical Association*, 94(446), 419–433.

[Rousseeuw & Hubert, 1999] Rousseeuw, P. & Hubert, M. (1999). Regression depth. *Journal of the American Statistical Association*, 94(446), 388–402.

[Rousseeuw & Ruts, 1996] Rousseeuw, P. & Ruts, I. (1996). Algorithm AS307: bivariate location depth. *Applied Statistics*, 45(4), 516–526.

[Rousseeuw & Struyf, 1998] Rousseeuw, P. & Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing*, 8, 193–203.

[Rousseeuw, 1985] Rousseeuw, P. J. (1985). Multivariate estimation with hight breakdown point. In W. Grossmann, G. Pflug, I. Vincze, & W. Wertz (Eds.), *Mathematical Statistics and Applications*, volume 8 of *Proceedings of the 4th Pannonian Symposium on Mathematical Statistics* (pp. 283–297).: Reidel, Dordrecht.

[Schimek, 2000] Schimek, M. (2000). *Smoothing and regression: approaches, computation, and application*. Willey.

[Serfling, 2006] Serfling, R. (2006). Depth functions in nonparametric multivariate inference. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72, 1.

[Silverman, 2000] Silverman, B. (2000). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC.

[Stahel, 1981] Stahel, W. A. (1981). *Robust estimation: Infinitesimal optimality and covariance matrix estimators*. PhD thesis (in German), ETH Zurich.

[Struyf & Rousseeuw, 2000] Struyf, A. & Rousseeuw, P. (2000). High-dimensional computation of the deepest location. *Computational Statistics and Data Analysis*, 34, 415–426.

[Struyf & Rousseeuw, 1999] Struyf, A. & Rousseeuw, P. J. (1999). Halfspace depth and regression depth characterize the empirical distribution. *J. Multivar. Anal.*, 69(1), 135–153.

[Tolstov & Silverman, 1976] Tolstov, G. & Silverman, R. (1976). *Fourier Series*. Dover Books on Mathematics. Dover Publications.

[Tukey, 1974] Tukey, J. (1974). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, volume 2 of *Canadian Mathematical Congress, Montreal* (pp. 523–531).

[Wei, 2008] Wei, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *Journal of the American Statistical Association*, 103(481), 397–409.

[Wickham, 2009] Wickham, H. (2009). *ggplot: Elegant Graphics for Data Analysis*. Springer.

[Zuo & Serfling, 2000a] Zuo, Y. & Serfling, R. (2000a). General notion of statistical depth function. *Annals of Statistics.*, 28(2), 461–482.

[Zuo & Serfling, 2000b] Zuo, Y. & Serfling, R. (2000b). Nonparametric notions of multivariate "scatter measure" and "more scattered" based on statistical depth functions. *Journal of Multivariate Analysis*, 75(1), 62 – 78.

[Zuo & Serfling, 2000c] Zuo, Y. & Serfling, R. (2000c). Structural properties and convergence results for contours of sample statistical depth functions. *Annals of Statistics.*, 28(2), 483–499.