

Chapter 2

Depth Statistics

Karl Mosler

2.1 Introduction

In 1975, John Tukey proposed a multivariate median which is the ‘deepest’ point in a given data cloud in \mathbb{R}^d (Tukey 1975). In measuring the depth of an arbitrary point z with respect to the data, Donoho and Gasko (1992) considered hyperplanes through z and determined its ‘depth’ by the smallest portion of data that are separated by such a hyperplane. Since then, this idea has proved extremely fruitful. A rich statistical methodology has developed that is based on data depth and, more general, nonparametric depth statistics. General notions of data depth have been introduced as well as many special ones. These notions vary regarding their computability and robustness and their sensitivity to reflect asymmetric shapes of the data. According to their different properties they fit to particular applications. The upper level sets of a depth statistic provide a family of set-valued statistics, named *depth-trimmed* or *central regions*. They describe the distribution regarding its location, scale and shape. The most central region serves as a *median*; see also the contribution by Oja, Chap. 1. The notion of depth has been extended from data clouds, that is empirical distributions, to general probability distributions on \mathbb{R}^d , thus allowing for laws of large numbers and consistency results. It has also been extended from d -variate data to data in functional spaces. The present chapter surveys the theory and methodology of depth statistics.

Recent reviews on data depth are given in Cascos (2009) and Serfling (2006). Liu et al. (2006) collects theoretical as well as applied work. More on the theory of depth functions and many details are found in Zuo and Serfling (2000) and the monograph by Mosler (2002).

The depth of a data point is reversely related to its *outlyingness*, and the depth-trimmed regions can be seen as multivariate set-valued *quantiles*. To illustrate the

K. Mosler (✉)

Universität zu Köln, Albertus-Magnus-Platz, 50923 Köln, Germany

e-mail: mosler@statistik.uni-koeln.de

Table 2.1 General government gross debt (% of GDP) and unemployment rate of the EU-27 countries in 2011 (Source: EUROSTAT)

Country	Debt %	Unempl. %	Country	Debt %	Unempl. %
Belgium	98.0	7.2	Luxembourg	18.2	4.9
Bulgaria	16.3	11.3	Hungary	80.6	10.9
Czech Republic	41.2	6.7	Malta	72.0	6.5
Denmark	46.5	7.6	Netherlands	65.2	4.4
Germany	81.2	5.9	Austria	72.2	4.2
Estonia	6.0	12.5	Poland	56.3	9.7
Ireland	108.2	14.4	Portugal	107.8	12.9
Greece	165.3	17.7	Romania	33.3	7.4
Spain	68.5	21.7	Slovenia	47.6	8.2
France	85.8	9.6	Slovakia	43.3	13.6
Italy	120.1	8.4	Finland	48.6	7.8
Cyprus	71.6	7.9	Sweden	38.4	7.5
Latvia	42.6	16.2	United Kingdom	85.7	8.0
Lithuania	38.5	15.4			

notions, we consider bivariate data from the EU-27 countries regarding unemployment rate and general government debt in percent of the GDP (Table 2.1). In what follows, we are interested which countries belong to a central, rather homogeneous group and which have to be regarded as, in some sense, outlying.

Section 2.2 introduces general depth statistics and the notions related to it. In Sect. 2.3, various depths for d -variate data are surveyed: multivariate depths based on distances, weighted means, halfspaces or simplices. Section 2.4 provides an approach to depth for functional data, while Sect. 2.5 treats computational issues. Section 2.6 concludes with remarks on applications.

2.2 Basic Concepts

In this section, the basic concepts of depth statistics are introduced, together with several related notions. First, we provide a general notion of depth functions, which relies on a set of desirable properties; then a few variants of the properties are discussed (Sect. 2.2.1). A depth function induces an outlyingness function and a family of central regions (Sect. 2.2.2). Further, a stochastic ordering and a probability metric are generated (Sect. 2.2.3).

2.2.1 Postulates on a Depth Statistic

Let E be a Banach space, \mathcal{B} its Borel sets in E , and \mathcal{P} a set of probability distributions on \mathcal{B} . To start with and in the spirit of Tukey's approach to data analysis,

we may regard \mathcal{P} as the class of empirical distributions giving equal probabilities $\frac{1}{n}$ to n , not necessarily different, data points in $E = \mathbb{R}^d$.

A *depth function* is a function $D : E \times \mathcal{P} \rightarrow [0, 1]$, $(z, P) \mapsto D(z | P)$, that satisfies the restrictions (or ‘postulates’) **D1** to **D5** given below. For easier notation, we write $D(z | X)$ in place of $D(z | P)$, where X is an arbitrary random variable distributed as P . For $z \in E$, $P \in \mathcal{P}$, and any random variable X having distribution P it holds:

- **D1 Translation invariant:** $D(z + b | X + b) = D(z | X)$ for all $b \in E$.
- **D2 Linear invariant:** $D(Az | AX) = D(z | X)$ for every bijective linear transformation $A : E \rightarrow E$.
- **D3 Null at infinity:** $\lim_{\|z\| \rightarrow \infty} D(z | X) = 0$.
- **D4 Monotone on rays:** If a point z^* has maximal depth, that is $D(z^* | X) = \max_{z \in E} D(z | X)$, then for any r in the unit sphere of E the function $\alpha \mapsto D(z^* + \alpha r | X)$ decreases, in the weak sense, with $\alpha > 0$.
- **D5 Upper semicontinuous:** The upper level sets $D_\alpha(X) = \{z \in E : D(z | X) \geq \alpha\}$ are closed for all α .

D1 and **D2** state that a depth function is *affine invariant*. **D3** and **D4** mean that the level sets D_α , $\alpha > 0$, are bounded and starshaped about z^* . If there is a point of maximum depth, this depth will w.l.o.g. be set to 1. **D5** is a useful technical restriction. An immediate consequence of restriction **D4** is the following proposition.

Proposition 2.1 *If X is centrally symmetric distributed about some $z^* \in E$, then any depth function $D(\cdot | X)$ is maximal at z^* .*

Recall that X is *centrally symmetric* distributed about z^* if the distributions of $X - z^*$ and $z^* - X$ coincide.

Our definition of a depth function differs slightly from that given in Liu (1990) and Zuo and Serfling (2000). The main difference between these postulates and ours is that they additionally postulate Proposition 2.1 to be true and that they do not require upper semicontinuity **D5**.

D4 states that the upper level set $D_\alpha(x^1, \dots, x^n)$ are starshaped with respect to z^* . If a depth function, in place of **D4**, meets the restriction

- **D4con:** $D(\cdot | X)$ is a *quasiconcave* function, that is, its upper level sets $D_\alpha(X)$ are convex for all $\alpha > 0$,

the depth is mentioned as a *convex depth*. Obviously, as a convex set is starshaped with respect to each of its points, **D4con** implies **D4**. In certain settings the restriction **D2** is weakened to

- **D2iso:** $D(Az | AX) = D(z | X)$ for every *isometric linear* transformation $A : E \rightarrow E$.

Then, in case $E = \mathbb{R}^d$, D is called an *orthogonal invariant depth* in contrast to an *affine invariant* depth when **D2** holds. Alternatively, sometimes **D2** is attenuated to *scale invariance*,

- **D2sca:** $D(\lambda z | \lambda X) = D(z | X)$ for all $\lambda > 0$.

2.2.2 Central Regions and Outliers

For given P and $0 \leq \alpha \leq 1$, the level sets $D_\alpha(P)$ form a nested family of *depth-trimmed* or *central regions*. The innermost region arises at some $\alpha_{\max} \leq 1$, which in general depends on P . $D_{\alpha_{\max}}(P)$ is the set of *deepest points*. **D1** and **D2** say that the family of central regions is affine equivariant. Central regions describe a distribution X with respect to location, dispersion, and shape. This has many applications in multivariate data analysis. On the other hand, given a nested family $\{C_\alpha(P)\}_{\alpha \in [0,1]}$ of set-valued statistics, defined on \mathcal{P} , that are convex, bounded and closed, the function D ,

$$D(z \mid P) = \sup\{\alpha : z \in C_\alpha(P)\}, \quad z \in E, \quad P \in \mathcal{P}, \quad (2.1)$$

satisfies **D1** to **D5** and **D4con**, hence is a convex depth function.

A depth function D orders data by their degree of centrality. Given a sample, it provides a center-outward *order statistic*. The depth induces an *outlyingness function* $\mathbb{R}^d \rightarrow [0, \infty[$ by

$$\text{Out}(z \mid X) = \frac{1}{D(z \mid X)} - 1,$$

which is zero at the center and infinite at infinity. In turn, $D(z \mid X) = (1 + \text{Out}(z \mid X))^{-1}$. Points outside a central region D_α have outlyingness greater than $1/\alpha - 1$; they can be regarded as *outliers* of a specified level α .

2.2.3 Depth Lifts, Stochastic Orderings, and Metrics

Assume $\alpha_{\max} = 1$ for $P \in \mathcal{P}$. By adding a real dimension to the central regions $D_\alpha(P)$, $\alpha \in [0, 1]$, we construct a set, which will be mentioned as the *depth lift*,

$$\hat{D}(P) = \{(\alpha, y) \in [0, 1] \times E : y = \alpha x, x \in D_\alpha(P), \alpha \in [0, 1]\}. \quad (2.2)$$

The depth lift gives rise to an *ordering* of probability distributions in \mathcal{P} : $P \prec_D Q$ if

$$\hat{D}(P) \subset \hat{D}(Q). \quad (2.3)$$

The restriction $\hat{D}(P) \subset \hat{D}(Q)$ is equivalent to $D_\alpha(P) \subset D_\alpha(Q)$ for all α . Thus, $P \prec_D Q$ means that each central set of Q is larger than the respective central set of P . In this sense, Q is *more dispersed* than P . The depth ordering is antisymmetric, hence an *order*, if and only if the family of central regions completely characterizes the underlying probability. Otherwise it is a preorder only. Finally, the depth D introduces a *probability semi-metric* on \mathcal{P} by the Hausdorff distance of depth lifts,

$$\delta_D(P, Q) = \delta_H(\hat{D}(P), \hat{D}(Q)). \quad (2.4)$$

Recall that the *Hausdorff distance* $\delta_H(C_1, C_2)$ of two compact sets C_1 and C_2 is the smallest ε such that C_1 plus the ε -ball includes C_2 and vice versa. Again, the semi-metric is a metric iff the central regions characterize the probability.

2.3 Multivariate Depth Functions

Originally and in most existing applications depth statistics are used with data in Euclidean space. Multivariate depth statistics are particularly suited to analyze non-Gaussian or, more general, non-elliptical distributions in \mathbb{R}^d . Without loss of generality, we consider distributions of full dimension d , that is, whose convex hull of support, $\text{co}(P)$, has affine dimension d .

A random vector X in \mathbb{R}^d has a *spherical distribution* if AX is distributed as X for every orthogonal matrix A . It has an *elliptical distribution* if $X = a + BY$ for some $a \in \mathbb{R}^d$, $B \in \mathbb{R}^{d \times d}$, and spherically distributed Y ; then we write $X \sim \text{Ell}(a, BB', \varphi)$, where φ is the radial distribution of Y . Actually, on an elliptical distribution $P = \text{Ell}(a, BB', \varphi)$, any depth function $D(\cdot, P)$ satisfying **D1** and **D2** has parallel elliptical level sets $D_\alpha(P)$, that is, level sets of a quadratic form with *scatter matrix* BB' . Consequently, all affine invariant depth functions are essentially equivalent if the distribution is elliptical. Moreover, if P is elliptical and has a unimodal Lebesgue-density f_P , the density level sets have the same elliptical shape, and the density is a transformation of the depth, i.e., a function φ exists such that $f_P(z) = \varphi(D(z | P))$ for all $z \in \mathbb{R}^d$. Similarly, on a spherical distribution, any depth satisfying postulates **D1** and **D2iso** has analogous properties.

In the following, we consider three principal approaches to define a multivariate depth statistic. The first approach is based on distances from properly defined central points or on volumes (Sect. 2.3.1), the second on certain L-statistics (*viz.* decreasingly weighted means of order statistics; Sect. 2.3.2), the third on simplices and halfspaces in \mathbb{R}^d (Sect. 2.3.3). The three approaches have different consequences on the depths' ability to reflect asymmetries of the distribution, on their robustness to possible outliers, and on their computability with higher-dimensional data.

Figures 2.1, 2.2, 2.3 and 2.4 below exhibit bivariate central regions for several depths and equidistant α . The data consist of the unemployment rate (in %) and the GDP share of public debt for the countries of the European Union in 2011.

Most of the multivariate depths considered are convex and affine invariant, some exhibit spherical invariance only. Some are continuous in the point z or in the distribution P (regarding weak convergence), others are not. They differ in the shape of the depth lift and whether it uniquely determines the underlying distribution. A basic dispersion ordering of multivariate probability distributions serving as a benchmark is the *dilation order*, which says that Y spreads out more than X if $E[\varphi(X)] \leq E[\varphi(Y)]$ holds for every convex $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$; see, e.g., Mosler (2002). It is interesting whether or not a particular depth ordering is concordant with the dilation order.

2.3.1 Depths Based on Distances

The outlyingness of a point, and hence its depth, can be measured by a distance from a properly chosen center of the distribution. In the following notions, this is done with different distances and centers.

L_2 -Depth The L_2 -depth, D^{L_2} , is based on the mean outlyingness of a point, as measured by the L_2 distance,

$$D^{L_2}(z | X) = (1 + E\|z - X\|)^{-1}. \quad (2.5)$$

It holds $\alpha_{\max} = 1$. The depth lift is $\hat{D}^{L_2}(X) = \{(\alpha, z) : E\|z - \alpha X\| \leq 1 - \alpha\}$ and convex. For an empirical distribution on points $x^i, i = 1, \dots, n$, we obtain

$$D^{L_2}(z | x^1, \dots, x^n) = \left(1 + \frac{1}{n} \sum_{i=1}^n \|z - x^i\|\right)^{-1}. \quad (2.6)$$

Obviously, the L_2 -depth vanishes at infinity (**D3**), and is maximum at the *spatial median* of X , i.e., at the point $z \in \mathbb{R}^d$ that minimizes $E\|z - X\|$. If the distribution is centrally symmetric, the center is the spatial median, hence the maximum is attained at the center. Monotonicity with respect to the deepest point (**D4**) as well as convexity and compactness of the central regions (**D4con**, **D5**) derive immediately from the triangle inequality. Further, the L_2 -depth depends continuously on z . The L_2 -depth converges also in the probability distribution: For a uniformly integrable and weakly convergent sequence $P_n \rightarrow P$ it holds $\lim_n D(z | P_n) = D(z | P)$.

However, the ordering induced by the L_2 -depth is no sensible ordering of dispersion, since the L_2 -depth contradicts the dilation order. As $\|z - x\|$ is convex in x , the expectation $E\|z - X\|$ increases with a dilation of P . Hence, (2.5) decreases (!) with a dilation.

The L_2 -depth is invariant against rigid Euclidean motions (**D1**, **D2iso**), but not affine invariant. An affine invariant version is constructed as follows: Given a positive definite $d \times d$ matrix M , consider the M -norm,

$$\|z\|_M = \sqrt{z' M^{-1} z}, \quad z \in \mathbb{R}^d. \quad (2.7)$$

Let S_X be a positive definite $d \times d$ matrix that depends continuously (in weak convergence) on the distribution and measures the dispersion of X in an affine equivariant way. The latter means that

$$S_{XA+b} = AS_X A' \text{ holds for any matrix } A \text{ of full rank and any } b. \quad (2.8)$$

Then an *affine invariant L_2 -depth* is given by

$$(1 + E\|z - X\|_{S_X})^{-1}. \quad (2.9)$$

Besides invariance, it has the same properties as the L_2 -depth. A simple choice for S_X is the covariance matrix Σ_X of X (Zuo and Serfling 2000). Note that the covariance matrix is positive definite, as the convex hull of the support, $\text{co}(P)$, is assumed to have full dimension. More robust choices for S_X are the *minimum volume ellipsoid* (MVE) or the *minimum covariance determinant* (MCD) estimators; see Rousseeuw and Leroy (1987), Lopuhaä and Rousseeuw (1991), and the contribution by Rousseeuw and Hubert, Chap. 4.

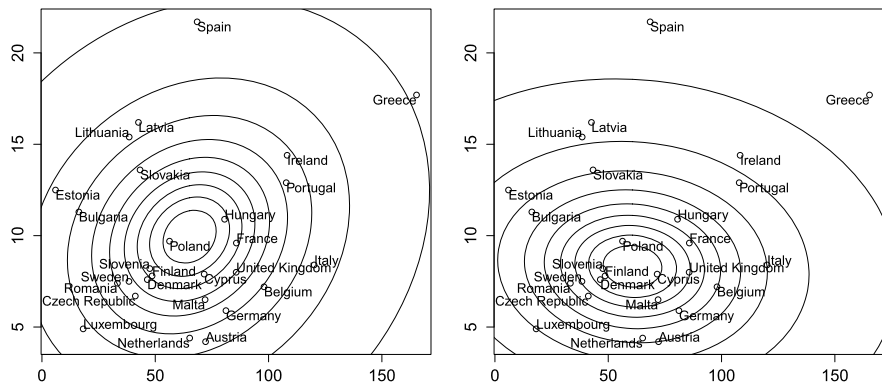


Fig. 2.1 Governmental debt (x -axis) and unemployment rate (y -axis); Mahalanobis regions (moment, *left*; MCD, *right*) with $\alpha = 0.1(0.1), \dots, 0.9$

Mahalanobis Depths Let c_X be a vector that measures the location of X in a continuous and affine equivariant way and, as before, S_X be a matrix that satisfies (2.8) and depends continuously on the distribution. Based on the estimates c_X and S_X a simple depth statistic is constructed, the *generalized Mahalanobis depth*, given by

$$D^{\text{Mah}}(z | X) = (1 + \|z - c_X\|_{S_X}^2)^{-1}. \quad (2.10)$$

Obviously, (2.10) satisfies **D1** to **D5** and **D4con**, taking its unique maximum at c_X . The depth lift is the convex set $\hat{D}^{\text{Mah}}(X) = \{(\alpha, z) : \|z - \alpha c_X\|_{S_X}^2 \leq \alpha^2(\alpha - 1)\}$, and the central regions are ellipsoids around c_X . The generalized Mahalanobis depth is continuous on z and P . In particular, with $c_X = E[X]$ and $S_X = \Sigma_X$ the (*moment*) *Mahalanobis depth* is obtained,

$$D^{\text{mMah}}(z | X) = (1 + (z - E[X])' \Sigma_X^{-1} (z - E[X]))^{-1}. \quad (2.11)$$

Its sample version is

$$D^{\text{mMah}}(z | x^1, \dots, x^n) = (1 + (z - \bar{x})' \hat{\Sigma}_X^{-1} (z - \bar{x}))^{-1}, \quad (2.12)$$

where \bar{x} is the mean vector and $\hat{\Sigma}_X$ is the empirical covariance matrix. It is easily seen that the α -central set of a sample from P converges almost surely to the α -central set of P , for any α . Figure 2.1 shows Mahalanobis regions for the debt-unemployment data, employing two choices of the matrix S_X , namely the usual moment estimate Σ_X and the robust MCD estimate. As it is seen from the Figure, these regions depend heavily on the choice of S_X . Hungary, e.g., is rather central (having depth greater than 0.8) with the moment Mahalanobis depth, while it is much more outlying (having depth below 0.5) with the MCD version.

Concerning uniqueness, the Mahalanobis depth fails in identifying the underlying distribution. As only the first two moments are used, any two distributions which have the same first two moments cannot be distinguished by their Mahalanobis depth

functions. Similarly, the generalized Mahalanobis depth does not determine the distribution. However, within the family of nondegenerate d -variate normal distributions or, more general, within any affine family of nondegenerate d -variate distributions having finite second moments, a single contour set of the Mahalanobis depth suffices to identify the distribution.

Projection Depth The *projection depth* has been proposed in Zuo and Serfling (2000):

$$D^{\text{proj}}(z | X) = \left(1 + \sup_{p \in S^{d-1}} \frac{|\langle p, z \rangle - \text{med}(\langle p, X \rangle)|}{\text{Dmed}(\langle p, X \rangle)} \right)^{-1}, \quad (2.13)$$

where S^{d-1} denotes the unit sphere in \mathbb{R}^d , $\langle p, z \rangle$ is the inner product (that is the projection of z on the line $\{\lambda p : \lambda \in \mathbb{R}\}$), $\text{med}(U)$ is the usual median of a univariate random variable U , and $\text{Dmed}(U) = \text{med}(|U - \text{med}(U)|)$ is the median absolute deviation from the median. The projection depth satisfies **D1** to **D5** and **D4con**. It has good properties, which are discussed in detail by Zuo and Serfling (2000). For breakdown properties of the employed location and scatter statistics, see Zuo (2000).

Oja Depth The Oja depth is not based on distances, but on average volumes of simplices that have vertices from the data (Zuo and Serfling 2000):

$$D^{\text{Oja}}(z | X) = \left(1 + \frac{E(\text{vol}_d(\text{co}\{z, X_1, \dots, X_d\}))}{\sqrt{\det \Sigma_X}} \right)^{-1},$$

where X_1, \dots, X_d are random vectors independently distributed as P , co denotes the convex hull, V_d the d -dimensional volume, and Σ_X is defined as above. In particular, we can choose $D_X = \Sigma_X$. The Oja depth satisfies **D1** to **D5**. It is continuous on z and maximum at the Oja median (Oja 1983), which is not unique; see also the contribution by Oja, Chap. 1. The Oja depth determines the distribution uniquely among those measures which have compact support of full dimension.

Figure 2.2 contrasts the projection depth regions with the Oja regions for our debt-unemployment data. The regions have different shapes, but agree in making Spain and Greece the most outlying countries.

2.3.2 Weighted Mean Depths

A large and flexible class of depth statistics corresponds to so called weighted-mean central regions, shortly WM regions (Dyckerhoff and Mosler 2011, 2012). These are convex compacts in \mathbb{R}^d , whose support function is a weighted mean of order statistics, that is, an L-statistic. Recall that a convex compact $K \subset \mathbb{R}^d$ is uniquely determined by its support function h_K ,

$$h_K(p) = \max\{p'x : x \in K\}, \quad p \in S^{d-1}.$$

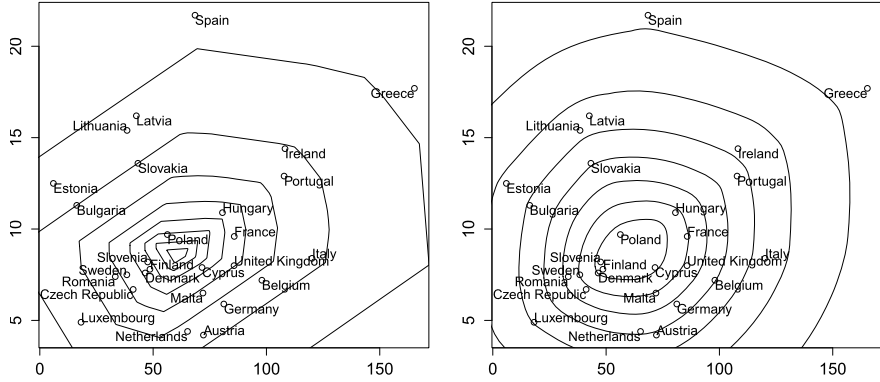


Fig. 2.2 Governmental debt and unemployment rate; projection depth regions (*left*), Oja regions (*right*); both with $\alpha = 0.1(0.1), \dots, 0.9$

To define the WM α -region of an empirical distribution on x^1, x^2, \dots, x^n , we construct its support function as follows: For $p \in S^{d-1}$, consider the line $\{\lambda p \in \mathbb{R}^d : \lambda \in \mathbb{R}\}$. By projecting the data on this line a linear ordering is obtained,

$$p'x^{\pi_p(1)} \leq p'x^{\pi_p(2)} \leq \dots \leq p'x^{\pi_p(n)}, \quad (2.14)$$

and, by this, a permutation π_p of the indices $1, 2, \dots, n$. Consider weights $w_{j,\alpha}$ for $j \in \{1, 2, \dots, n\}$ and $\alpha \in [0, 1]$ that satisfy the following restrictions (i) to (iii):

- (i) $\sum_{j=1}^n w_{j,\alpha} = 1$, $w_{j,\alpha} \geq 0$ for all j and α .
- (ii) $w_{j,\alpha}$ increases in j for all α .
- (iii) $\alpha < \beta$ implies $\sum_{j=1}^k w_{j,\alpha} \leq \sum_{j=1}^k w_{j,\beta}$, $k = 1, \dots, n$.

Then, as it has been shown in Dyckerhoff and Mosler (2011), the function $h_{D_\alpha(x^1, \dots, x^n)}$,

$$h_{D_\alpha(x^1, \dots, x^n)}(p) = \sum_{j=1}^n w_{j,\alpha} p'x^{\pi_p(j)}, \quad p \in S^{d-1}, \quad (2.15)$$

is the support function of a convex body $D_\alpha = D_\alpha(x^1, \dots, x^n)$, and $D_\alpha \subset D_\beta$ holds whenever $\alpha > \beta$. Now we are ready to see the general definition of a family of WM regions.

Definition 2.1 Given a weight vector $w_\alpha = w_{1,\alpha}, \dots, w_{n,\alpha}$ that satisfies the restrictions (i) to (iii), the convex compact $D_\alpha = D_\alpha(x^1, \dots, x^n)$ having support function (2.15) is named the *WM region* of x^1, \dots, x^n at level α , $\alpha \in [0, 1]$. The corresponding depth (2.1) is the *WM depth* with weights w_α , $\alpha \in [0, 1]$.

It follows that the WM depth satisfies the restrictions **D1** to **D5** and **D4con**. Moreover, it holds

$$D_\alpha(x^1, \dots, x^n) = \text{conv} \left\{ \sum_{j=1}^n w_{j,\alpha} x^{\pi(j)} : \pi \text{ permutation of } \{1, \dots, n\} \right\}. \quad (2.16)$$

This explains the name by stating that a WM region is the convex hull of weighted means of the data. Consequently, outside the convex hull of the data the WM depth vanishes. WM depths are useful statistical tools as their central regions have attractive analytical and computational properties. Sample WM regions are consistent estimators for the WM region of the underlying probability. Besides being *continuous* in the distribution and in α , WM regions are *subadditive*, that is,

$$D_\alpha(x^1 + y^1, \dots, x^n + y^n) \subset D_\alpha(x^1, \dots, x^n) \oplus D_\alpha(y^1, \dots, y^n),$$

and *monotone*: If $x^i \leq y^i$ holds for all i (in the componentwise ordering of \mathbb{R}^d), then

$$\begin{aligned} D_\alpha(y^1, \dots, y^n) &\subset D_\alpha(x^1, \dots, x^n) \oplus \mathbb{R}_+^d \quad \text{and} \\ D_\alpha(x^1, \dots, x^n) &\subset D_\alpha(y^1, \dots, y^n) \oplus \mathbb{R}_-^d, \end{aligned}$$

where \oplus signifies the Minkowski sum of sets.

Depending on the choice of the weights $w_{j,\alpha}$ different notions of data depths are obtained. For a detailed discussion of these and other special WM depths and central regions, the reader is referred to Dyckerhoff and Mosler (2011, 2012).

Zonoid Depth For an empirical distribution P on x^1, \dots, x^n and $0 < \alpha \leq 1$ define the zonoid region (Koshevoy and Mosler 1997)

$$D_\alpha^{\text{zon}}(P) = \left\{ \sum_{i=1}^n \lambda_i x^i : 0 \leq \lambda_i \leq \frac{1}{n\alpha}, \sum_{i=1}^n \lambda_i = 1 \right\}.$$

See Fig. 2.3. The corresponding support function (2.15) employs the weights

$$w_{j,\alpha} = \begin{cases} 0 & \text{if } j < n - \lfloor n\alpha \rfloor, \\ \frac{n\alpha - \lfloor n\alpha \rfloor}{n\alpha} & \text{if } j = n - \lfloor n\alpha \rfloor, \\ \frac{1}{n\alpha} & \text{if } j > n - \lfloor n\alpha \rfloor. \end{cases} \quad (2.17)$$

Many properties of zonoid regions and the zonoid depth $D^{\text{zon}}(z | X)$ are discussed in Mosler (2002). The zonoid depth lift equals the so called lift zonoid, which fully characterizes the distribution. Therefore the zonoid depth generates an antisymmetric depth order (2.3) and a probability metric (2.4). Zonoid regions are not only invariant to affine, but to general linear transformations; specifically any marginal projection of a zonoid region is the zonoid region of the marginal distribution. The zonoid depth is continuous on z as well as P .

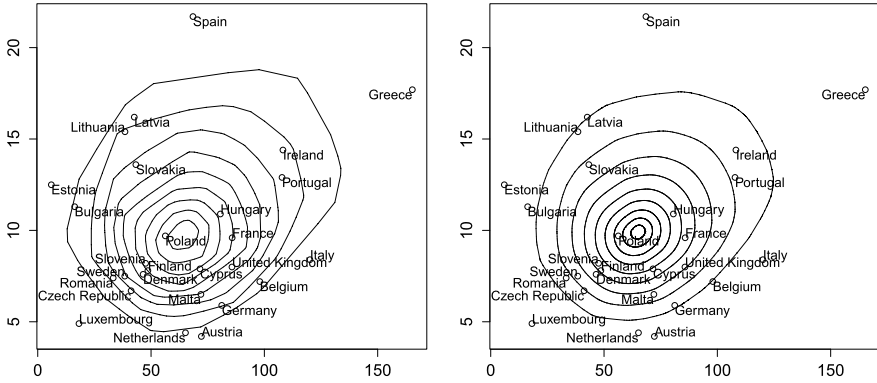


Fig. 2.3 Governmental debt and unemployment rate; zonoid regions (*left*), ECH* regions (*right*); both with $\alpha = 0.1(0.1), \dots, 0.9$

Expected Convex Hull Depth Another important notion of WMT depth is that of *expected convex hull* (ECH*) depth (Casco 2007). Its central region D_α (see Fig. 2.3) has a support function with weights

$$w_{j,\alpha} = \frac{j^{1/\alpha} - (j-1)^{1/\alpha}}{n^{1/\alpha}}. \quad (2.18)$$

Figure 2.3 depicts zonoid and ECH* regions for our data. We see that the zonoid regions are somewhat angular while the ECH* regions appear to be smoother; this corresponds, when calculating such regions in higher dimensions, to a considerably higher computation load of ECH*.

Geometrical Depth The weights

$$w_{j,\alpha} = \begin{cases} \frac{1-\alpha}{1-\alpha^n} \alpha^{n-j} & \text{if } 0 < \alpha < 1, \\ 0 & \text{if } \alpha = 1, \end{cases}$$

yield another class of WM regions. The respective depth is the *geometrically weighted mean depth* (Dyckerhoff and Mosler 2011).

2.3.3 Depths Based on Halfspaces and Simplices

The third approach concerns no distances or volumes, but the combinatorics of halfspaces and simplices only. In this it is independent of the metric structure of \mathbb{R}^d . While depths that are based on distances or weighted means may be addressed as *metric depths*, the following ones will be mentioned as *combinatorial depths*. They remain constant, as long as the compartment structure of the data does not change. By this, they are very robust against *location outliers*. Outside the convex support $\text{co}(X)$ of the distribution every combinatorial depth attains its minimal value, which is zero.

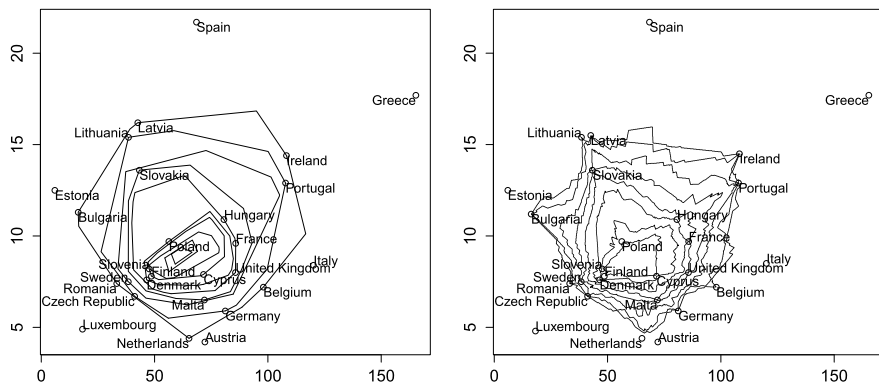


Fig. 2.4 Governmental debt and unemployment rate; Tukey regions (*left*) with $\alpha = \frac{2}{27}, \frac{1}{27}, \dots, \frac{1}{27}$, simplicial regions (*right*) with $\alpha = 0.25, 0.3(0.1), \dots, 0.9$

Location Depth Consider the population version of the *location depth*,

$$D^{\text{loc}}(z | X) = \inf\{P(H) : H \text{ is a closed halfspace, } z \in H\}. \quad (2.19)$$

The depth is also known as *halfspace* or *Tukey depth*, its central regions as *Tukey regions*. The location depth is affine invariant (**D1**, **D2**). Its central regions are convex (**D4con**) and closed (**D5**); see Fig. 2.4. The maximum value of the location depth is smaller or equal to 1 depending on the distribution. The set of all such points is mentioned as the *halfspace median set* and each of its elements as a *Tukey median* (Tukey 1975).

If X has an *angular symmetric* distribution, the location depth attains its maximum at the center and the center is a Tukey median; this strengthens Proposition 2.1. (A distribution is called *angular* ($=$ *halfspace*) *symmetric* about z^* if $P(X \in H) \geq 1/2$ for every closed halfspace H having z^* on the boundary; equivalently, if $(X - z^*)/\|X - z^*\|$ is centrally symmetric with the convention $0/0 = 0$.)

If X has a Lebesgue-density, the location depth depends continuously on z ; otherwise the dependence on z is noncontinuous and there can be more than one point where the maximum is attained. As a function of P the location depth is obviously noncontinuous. It determines the distribution in a unique way if the distribution is either discrete (Struyf and Rousseeuw 1999; Koshevoy 2002) or continuous with compact support. The location depth of a sample from P converges almost surely to the location depth of P (Donoho and Gasko 1992). The next depth notion involves simplices in \mathbb{R}^d .

Simplicial Depth Liu (1990) defines the *simplicial depth* as follows:

$$D^{\text{sim}}(z | X) = P(z \in \text{co}(\{X_1, \dots, X_{d+1}\})), \quad (2.20)$$

where X_1, \dots, X_{d+1} are i.i.d. by P . The sample version reads as

$$D^{\text{sim}}(z | x^1, \dots, x^n) = \frac{1}{\binom{n}{d+1}} \#\{i_1, \dots, i_{d+1} : z \in \text{co}(\{x^{i_1}, \dots, x^{i_{d+1}}\})\}. \quad (2.21)$$

The simplicial depth is affine invariant (**D1**, **D2**). Its maximum is less or equal to 1, depending on the distribution. In general, the point of maximum simplicial depth is not unique; the *simplicial median* is defined as the gravity center of these points. The sample simplicial depth converges almost surely uniformly in z to its population version (Liu 1990; Dümbgen 1992). The simplicial depth has positive breakdown (Chen 1995).

If the distribution is Lebesgue-continuous, the simplicial depth behaves well: It varies continuously on z (Liu 1990, Theorem 2), is maximum at a center of angular symmetry, and decreases monotonously from a deepest point (**D4**). The *simplicial central regions* of a Lebesgue-continuous distribution are connected and compact (Liu 1990).

However, if the distribution is discrete, each of these properties can fail; for counterexamples see, e.g., Zuo and Serfling (2000). The simplicial depth characterizes an empirical measure if the supporting points are in *general position*, that is, if no more than d of the points lie on the same hyperplane.

As Fig. 2.4 demonstrates, Tukey regions are convex while simplicial regions are only starshaped. The figure illustrates also that these notions are rather insensitive to outlying data: both do not reflect *how far* Greece and Spain are from the center. Whether, in an application, this kind of robustness is an advantage or not, depends on the problem and data at hand.

Other well known combinatorial data depths are the *majority depth* (Liu and Singh 1993) and the *convex-hull peeling depth* (Barnett 1976; Donoho and Gasko 1992). However, the latter possesses no population version.

2.4 Functional Data Depth

The analysis of functional data has become a practically important branch of statistics; see Ramsay and Silverman (2005). Consider a space E of functions $[0, 1] \rightarrow \mathbb{R}$ with the supremum norm. Like a multivariate data depth, a functional data depth is a real-valued functional that indicates how ‘deep’ a function $z \in E$ is located in a given finite cloud of functions $\in E$. Let E' denote the set of continuous linear functionals $E \rightarrow \mathbb{R}$, and E'^d the d -fold Cartesian product of E' . Here, following Mosler and Polyakova (2012), functional depths of a general form (2.22) are presented. Some alternative approaches will be addressed below.

Φ -Depth For $z \in E$ and an empirical distribution X on $x^1, \dots, x^n \in E$, define a *functional data depth* by

$$D(z | X) = \inf_{\varphi \in \Phi} D^d(\varphi(z) | \varphi(X)), \quad (2.22)$$

where D^d is a d -variate data depth satisfying **D1** to **D5**, $\Phi \subset E'^d$, and $\varphi(X)$ is the empirical distribution on $\varphi(x^1), \dots, \varphi(x^n)$. D is called a Φ -depth. A population version is similarly defined.

Each φ in this definition may be regarded as a particular ‘aspect’ we are interested in and which is represented in d -dimensional space. The depth of z is given as the smallest multivariate depth of z under all these aspects. It implies that all aspects are equally relevant so that the depth of z cannot be larger than its depth under any aspect.

As the d -variate depth D^d has maximum not greater than 1, the functional data depth D is bounded above by 1. At every point z^* of *maximal D -depth* it holds $D(z^* | X) \leq 1$. The bound is attained with equality, $D(z^* | X) = 1$, iff $D^d(\varphi(z^*)) | \varphi(X)) = 1$ holds for all $\varphi \in \Phi$, that is, iff

$$z^* \in \bigcap_{\varphi \in \Phi} \varphi^{-1}(D_1^d(\varphi(X))). \quad (2.23)$$

A Φ -depth (2.22) always satisfies **D1**, **D2sca**, **D4**, and **D5**.

It satisfies **D3** if for every sequence (z^i) with $\|z^i\| \rightarrow \infty$ exists a φ in Φ such that $\varphi(z^i) \rightarrow \infty$. (For some special notions of functional data depth this postulate has to be properly adapted.)

D4con is met if **D4con** holds for the underlying d -variate depth.

We now proceed with specifying the set Φ of functionals and the multivariate depth D^k in (2.22). While many features of the functional data depth (2.22) resemble those of a multivariate depth, an important difference must be pointed out: In a general Banach space the unit ball B is not compact, and properties **D3** and **D5** do not imply that the level sets of a functional data depth are compact. So, to obtain a meaningful notion of functional data depth of type (2.22) one has to carefully choose a set of functions Φ which is not too large. On the other hand, Φ should not be too small, in order to extract sufficient information from the data.

Graph Depths For $x \in E$ denote $x(t) = (x_1(t), \dots, x_d(t))$ and consider

$$\Phi = \{\varphi^t : E \rightarrow \mathbb{R}^d : \varphi^t(x) = (x_1(t), \dots, x_d(t)), t \in T\} \quad (2.24)$$

for some $T \subset [0, 1]$, which may be a subinterval or a finite set. For D^d use any multivariate depth that satisfies **D1** to **D5**. This results in the *graph depth*

$$\text{GD}(z | x^1, \dots, x^n) = \inf_{t \in T} D^d(z(t) | x^1(t), \dots, x^n(t)). \quad (2.25)$$

In particular, with the univariate halfspace depth, $d = 1$ and $T = J$ we obtain the *halfgraph depth* (López-Pintado and Romo 2005). Also, with the univariate simplicial depth the *band depth* (López-Pintado and Romo 2009) is obtained, but this, in general, violates monotonicity **D4**.

Grid Depths We choose a finite number of points in J , t_1, \dots, t_k , and evaluate a function $z \in E$ at these points. Notate $\underline{t} = (t_1, \dots, t_k)$ and $z(\underline{t}) = (z_1(\underline{t}), \dots, z_d(\underline{t}))^T$. That is, in place of the function z the $k \times d$ matrix $z^{(k)}$ is considered. A *grid depth* RD is defined by (2.22) with the following Φ ,

$$\Phi = \{\varphi^r : \varphi^r(z) = (\langle r, z_1(\underline{t}) \rangle, \dots, \langle r, z_d(\underline{t}) \rangle), r \in S^{k-1}\}, \quad (2.26)$$

which yields

$$\text{RD}(z \mid x^1, \dots, x^n) = \inf_{r \in \mathcal{S}^{k-1}} D^d(\langle r, z(\underline{t}) \rangle \mid \langle r, x^1(\underline{t}) \rangle, \dots, \langle r, x^n(\underline{t}) \rangle). \quad (2.27)$$

A slight extension of the Φ -depth is the *principal components depth* (Mosler and Polyakova 2012). However, certain approaches from the literature are no Φ -depths. These are mainly of two types. The first type employs *random projections* of the data: Cuesta-Albertos and Nieto-Reyes (2008b) define the depth of a function as the univariate depth of the function values taken at a randomly chosen argument t . Cuevas et al. (2007) also employ a random projection method. The other type uses average univariate depths. Fraiman and Muniz (2001) calculate the univariate depths of the values of a function and integrate them over the whole interval; this results in kind of ‘average’ depth. Claeskens et al. (2012) introduce a multivariate ($d \geq 1$) functional data depth, where they similarly compute a weighted average depth. The weight at a point reflects the variability of the function values at this point (more precisely: is proportional to the volume of a central region at the point).

2.5 Computation of Depths and Central Regions

The moment Mahalanobis depth and its elliptical central regions are obtained in any dimension by calculating the mean and the sample covariance matrix, while robust Mahalanobis depths and regions are determined with the R-procedures “cov.mcd” and “cov.mve”. In dimension $d = 2$, the central regions of many depth notions can be exactly calculated by following a *circular sequence* (Edelsbrunner 1987). The R-package “depth” computes the exact location ($d = 2, 3$) and simplicial ($d = 2$) depths, as well as the Oja depth and an approximative location depth for any dimension. An exact algorithm for the location depth in any dimension is developed in Liu and Zuo (2012). Cuesta-Albertos and Nieto-Reyes (2008a) propose to calculate instead the *random Tukey depth*, which is the minimum univariate location depth of univariate projections in a number of randomly chosen directions. With the algorithm of Paindaveine and Šiman (2012), Tukey regions are obtained, $d \geq 2$. The bivariate projection depth is computed by the R-package “ExPD2D”; for the respective regions, see Liu et al. (2011). The zonoid depth can be efficiently determined in any dimension (Dyckerhoff et al. 1996). An R-package (“WMTregions”) exists for the exact calculation of zonoid and general WM regions; see Mosler et al. (2009), Bazovkin and Mosler (2012). The R-package “rainbow” calculates several functional data depths.

2.6 Conclusions

Depth statistics have been used in numerous and diverse tasks of which we can mention a few only. Liu et al. (1999) provide an introduction to some of them. In descriptive multivariate analysis, depth functions and central regions visualize the data regarding location, scale and shape. By bagplots and sunburst plots, outliers

can be identified and treated in an interactive way. In k -class supervised classification, each—possibly high-dimensional—data point is represented in $[0, 1]^k$ by its values of depth in the k given classes, and classification is done in $[0, 1]^k$. Functions of depth statistics include depth-weighted statistical functionals, such as $\int_{\mathbb{R}^d} x w(D(x | P)) dP / \int_{\mathbb{R}^d} w(D(x | P)) dP$ for location. In inference, tests for goodness of fit and homogeneity regarding location, scale and symmetry are based on depth statistics; see, e.g., Dyckerhoff (2002), Ley and Paindaveine (2011). Applications include such diverse fields as statistical control (Liu and Singh 1993), measurement of risk (Casco and Molchanov 2007), and robust linear programming (Bazovkin and Mosler 2011). Functional data depth is applied to similar tasks in description, classification and testing; see, e.g., López-Pintado and Romo (2009), Cuevas et al. (2007).

This survey has covered the fundamentals of depth statistics for d -variate and functional data. Several special depth functions in \mathbb{R}^d have been presented, metric and combinatorial ones, with a focus on the recent class of WM depths. For functional data, depths of infimum type have been discussed. Of course, such a survey is necessarily incomplete and biased by the preferences of the author. Of the many applications of depth in the literature only a few have been touched, and important theoretical extensions like regression depth (Rousseeuw and Hubert 1999), depth calculus (Mizera 2002), location-scale depth (Mizera and Müller 2004), and likelihood depth (Müller 2005) have been completely omitted.

Most important for the selection of a depth statistic in applications are the questions of computability and—depending on the data situation—robustness. Mahalanobis depth is solely based on estimates of the mean vector and the covariance matrix. In its classical form with moment estimates Mahalanobis depth is efficiently calculated but highly non-robust, while with estimates like the minimum volume ellipsoid it becomes more robust. However, since it is constant on ellipsoids around the center, Mahalanobis depth cannot reflect possible asymmetries of the data. Zonoid depth can be efficiently calculated, also in larger dimensions, but has the drawback that the deepest point is always the mean, which makes the depth non-robust. So, if robustness is an issue, the zonoid depth has to be combined with a proper preprocessing of the data to identify possible outliers. The location depth is, by construction, very robust but expensive when exactly computed in dimensions more than two. As an efficient approach the random Tukey depth yields an upper bound on the location depth, where the number of directions has to be somehow chosen.

A depth statistics measures the centrality of a point in the data. Besides ordering the data it provides numerical values that, with some depth notions, have an obvious meaning; so with the location depth and all WM depths. With other depths, in particular those based on distances, the outlyingness function has a direct interpretation.

References

- Barnett, V. (1976). The ordering of multivariate data (with discussion). *Journal of the Royal Statistical Society. Series A. General*, 139, 318–352.

- Bazovkin, P., & Mosler, K. (2011). *Stochastic linear programming with a distortion risk constraint*. [arXiv:1208.2113v1](#).
- Bazovkin, P., & Mosler, K. (2012). An exact algorithm for weighted-mean trimmed regions in any dimension. *Journal of Statistical Software*, 47(13).
- Cascos, I. (2007). The expected convex hull trimmed regions of a sample. *Computational Statistics*, 22, 557–569.
- Cascos, I. (2009). Data depth: multivariate statistics and geometry. In W. Kendall & I. Molchanov (Eds.), *New perspectives in stochastic geometry*, Oxford: Clarendon/Oxford University Press.
- Cascos, I., & Molchanov, I. (2007). Multivariate risks and depth-trimmed regions. *Finance and Stochastics*, 11, 373–397.
- Chen, Z. (1995). Bounds for the breakdown point of the simplicial median. *Journal of Multivariate Analysis*, 55, 1–13.
- Claeskens, G., Hubert, M., & Slaets, L. (2012). Multivariate functional halfspace depth. In *Workshop robust methods for dependent data*, Witten.
- Cuesta-Albertos, J., & Nieto-Reyes, A. (2008a). A random functional depth. In S. Dabo-Niang & F. Ferraty (Eds.), *Functional and operatorial statistics* (pp. 121–126). Heidelberg: Physica-Verlag.
- Cuesta-Albertos, J., & Nieto-Reyes, A. (2008b). The random Tukey depth. *Computational Statistics & Data Analysis*, 52, 4979–4988.
- Cuevas, A., Febrero, M., & Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22, 481–496.
- Donoho, D. L., & Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20, 1803–1827.
- Dümbgen, L. (1992). Limit theorems for the simplicial depth. *Statistics & Probability Letters*, 14, 119–128.
- Dyckerhoff, R. (2002). *Datentiefe: Begriff, Berechnung, Tests*. Mimeo, Fakultät für Wirtschafts- und Sozialwissenschaften, Universität zu Köln.
- Dyckerhoff, R., Koshevoy, G., & Mosler, K. (1996). Zonoid data depth: theory and computation. In A. Pratt (Ed.), *Proceedings in computational statistics COMPSTAT* (pp. 235–240). Heidelberg: Physica-Verlag.
- Dyckerhoff, R., & Mosler, K. (2011). Weighted-mean trimming of multivariate data. *Journal of Multivariate Analysis*, 102, 405–421.
- Dyckerhoff, R., & Mosler, K. (2012). Weighted-mean regions of a probability distribution. *Statistics & Probability Letters*, 82, 318–325.
- Edelsbrunner, H. (1987). *Algorithms in combinatorial geometry*. Heidelberg: Springer.
- Fraiman, R., & Muniz, G. (2001). Trimmed means for functional data. *Test*, 10, 419–440.
- Koshevoy, G. (2002). The Tukey depth characterizes the atomic measure. *Journal of Multivariate Analysis*, 83, 360–364.
- Koshevoy, G., & Mosler, K. (1997). Zonoid trimming for multivariate distributions. *The Annals of Statistics*, 25, 1998–2017.
- Ley, C., & Paindaveine, D. (2011). *Depth-based runs tests for multivariate central symmetry*. ECARES discussion papers 2011/06, ULB, Bruxelles.
- Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, 18, 405–414.
- Liu, R. Y., Parelius, J. M., & Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion). *The Annals of Statistics*, 27, 783–858.
- Liu, R. Y., Serfling, R., & Souvaine, D. L. (2006). *Data depth: robust multivariate analysis, computational geometry and applications*. Providence: Am. Math. Soc.
- Liu, R. Y., & Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88, 252–260.
- Liu, X., & Zuo, Y. (2012). *Computing halfspace depth and regression depth*. Mimeo.
- Liu, X., Zuo, Y., & Wang, Z. (2011). *Exactly computing bivariate projection depth contours and median*. [arXiv:1112.6162v1](#).

- López-Pintado, S., & Romo, J. (2005). *A half-graph depth for functional data*. Working papers 01/2005, Universidad Carlos III, Statistics and Econometrics.
- López-Pintado, S., & Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104, 718–734.
- Lopuhaä, H. P., & Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, 19, 229–248.
- Mizera, I. (2002). On depth and deep points: a calculus. *The Annals of Statistics*, 30, 1681–1736.
- Mizera, I., & Müller, C. H. (2004). Location-scale depth. *Journal of the American Statistical Association*, 99, 949–989.
- Mosler, K. (2002). *Multivariate dispersion, central regions and depth: the lift zonoid approach*. New York: Springer.
- Mosler, K., Lange, T., & Bazovkin, P. (2009). Computing zonoid trimmed regions in dimension $d > 2$. *Computational Statistics & Data Analysis*, 53, 2500–2510.
- Mosler, K., & Polyakova, Y. (2012). *General notions of depth for functional data*. [arXiv:1208.1981v1](https://arxiv.org/abs/1208.1981v1).
- Müller, C. H. (2005). Depth estimators and tests based on the likelihood principle with application to regression. *Journal of Multivariate Analysis*, 95, 153–181.
- Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1, 327–332.
- Paindaveine, D., & Šíman, M. (2012). Computing multiple-output regression quantile regions. *Computational Statistics & Data Analysis*, 56, 840–853.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York: Springer.
- Rousseeuw, P. J., & Hubert, M. (1999). Regression depth (with discussion). *Journal of the American Statistical Association*, 94, 388–433.
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York: Wiley.
- Serfling, R. (2006). Depth functions in nonparametric multivariate inference. In R. Liu, R. Serfling, & D. Souvaine (Eds.), *Data depth: robust multivariate analysis, computational geometry and applications* (pp. 1–16). Providence: Am. Math. Soc.
- Struyf, A., & Rousseeuw, P. J. (1999). Halfspace depth and regression depth characterize the empirical distribution. *Journal of Multivariate Analysis*, 69, 135–153.
- Tukey, J. W. (1975). Mathematics and picturing data. In R. James (Ed.), *Proceedings of the 1974 international congress of mathematicians*, Vancouver (Vol. 2, pp. 523–531).
- Zuo, Y. (2000). A note on finite sample breakdown points of projection based multivariate location and scatter statistics. *Metrika*, 51, 259–265.
- Zuo, Y., & Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics*, 28, 461–482.

Robustness and Complex Data Structures

Festschrift in Honour of Ursula Gather

Becker, C.; Fried, R.; Kuhnt, S. (Eds.)

2013, X, 379 p., Hardcover

ISBN: 978-3-642-35493-9