Weighted-mean trimming of multivariate data

Rainer Dyckerhoff and Karl Mosler

November 1, 2010

Abstract

A general notion of trimmed regions for empirical distributions in d-space is introduced. The regions are called weighted-mean trimmed regions. They are continuous in the data as well as in the trimming parameter. Further, these trimmed regions have many other attractive properties. In particular they are subadditive and monotone which makes it possible to construct multivariate measures of risk based on these regions. Special cases include the zonoid trimming and the ECH (expected convex hull) trimming. These regions can be exactly calculated for any dimension. Finally, the notion of weighted-mean trimmed regions extends to probability distributions in d-space, and a law of large numbers applies.

Keywords: Central regions, continuous trimming, data depth, lift zonoid regions, expected convex hull, law of large numbers.

1 Introduction

The trimming of multivariate data or, more general, probability distributions in $\mathbb{R}^d$ has become an important tool in nonparametric multivariate analysis. Depending on a given distribution, a family of nested sets, called trimmed or central regions, is constructed each of which reflects the location, dispersion, and shape of the distribution.

The notion of trimmed regions is closely connected with that of data depth: Roughly speaking, each trimmed region can be considered as an upper level set of a function, the depth function, that measures sort of distance of a given point in $\mathbb{R}^d$ from a central point of the distribution, where the function takes its maximum.

Many special notions of data depth and trimmed regions have been proposed in the literature, among them the Mahalanobis depth, the halfspace depth, the simplicial depth, and the zonoid depth; for recent surveys, see Serfling (2006), Cascos (2009). Applications include multivariate data analysis (Liu et al., 1999), classification (Mosler and Hoberg, 2006), tests for multivariate location and scale (Dyckerhoff, 2002), and risk measurement (Cascos and Molchanov, 2007). For a general definition of data depth, see, e.g., Zuo and Serfling (2000a), Dyckerhoff (2004).

A general definition of trimmed regions is the following (see e.g. Zuo and Serfling (2000b), Mosler (2002), Dyckerhoff (2004)).

Definition 1 (Trimmed regions)**.** *Given an interval I in $\mathbb{R}$, a* family of trimmed regions *provides, for each set of data $\{x_1, \dots, x_n\} \in \mathbb{R}^d$ and $\alpha \in I$, a set $\mathrm{D}_\alpha(x_1, \dots, x_n) \subset \mathbb{R}^d$ such that:*

- **T1** (Convex body) $\mathrm{D}_\alpha(x_1, \dots, x_n)$ *is* convex*,* closed*, and* bounded*.*
- **T2** (Nested) *The mapping $\alpha \mapsto \mathrm{D}_\alpha(x_1, \dots, x_n)$ is* decreasing*, i.e. $\alpha < \beta$ implies $\mathrm{D}_\beta(x_1, \dots, x_n) \subset \mathrm{D}_\alpha(x_1, \dots, x_n)$.*
- **T3** (Affine equivariant) *The mapping $(x_1, \dots, x_n) \mapsto \mathrm{D}_\alpha(x_1, \dots, x_n)$ is* affine equivariant*.*

From here on, I will always denote an interval that constitutes the domain of the trimming parameter α. In most applications I will be equal to $(0, 1]$, however other choices are possible.

In some notions of trimmed regions convexity is weakened to starshapedness, and affine equivariance to translation-scale equivariance.

For practical use in data analysis, continuity is needed with respect to the data and as well with respect to the parameter:

- **T4** (Continuous in the data) The mapping $(x_1, \dots, x_n) \mapsto \mathrm{D}_\alpha(x_1, \dots, x_n)$ is continuous in terms of Hausdorff convergence.
- **T5** (Continuous in the parameter) The mapping $\alpha \mapsto \mathrm{D}_\alpha(x_1, \dots, x_n)$ is continuous in terms of Hausdorff convergence.

Continuity, of course, appears to be a natural postulate, as small deviations in the data should only slightly change the trimmed regions. The same applies for small changes of the depth parameter α. However, not all popular depth notions are continuous. While, e.g., trimmed regions based on Mahalanobis or zoonoid depth are continuous, those based on simplicial depth are not.

In applications to risk measurement two other properties of trimmed regions are important:

T6 (Subadditive in the data)

$$\mathrm{D}_\alpha(x_1 + y_1, \ldots, x_n + y_n) \subset \mathrm{D}_\alpha(x_1, \ldots, x_n) \oplus \mathrm{D}_\alpha(y_1, \ldots, y_n)\,.$$

T7 (Monotone in the data) *If $x_i \le y_i$ holds for all i (in the componentwise ordering of $\mathbb{R}^d$) then*

$$\begin{aligned} &\mathrm{D}_\alpha(y_1, \ldots, y_n) \subset \mathrm{D}_\alpha(x_1, \ldots, x_n) \oplus \mathbb{R}^d_+ \quad \textit{and} \\ &\mathrm{D}_\alpha(x_1, \ldots, x_n) \subset \mathrm{D}_\alpha(y_1, \ldots, y_n) \oplus \mathbb{R}^d_-\,, \end{aligned}$$

where $\oplus$ signifies the Minkowski sum of sets.

They allow for the construction of set-valued measures that are coherent multivariate risk measures (Cascos and Molchanov, 2007). While halfspace regions are neither subadditive nor monotone, e.g. zonoid regions are both. The univariate α-trimmed zonoid region is a closed interval whose lower extreme is the negative of a popular univariate risk measure, the α-expected shortfall. Multivariate zonoid regions can be transformed into multivariate risk measures, as well.

This paper provides a general notion of trimmed regions for empirical distributions in d-space which have all the above-mentioned properties **T1** to **T7**. Due to their construction *via* weighted means of data permutations we will call them *weighted-mean trimmed regions*. These regions are continuous in the data and in the trimming parameter, as well as subadditive and monotone. Special cases include the zonoid trimming and the ECH (expected convex hull) trimming and other meaningful notions of trimming. The regions can be exactly calculated for any dimension.

Overview of the paper: The definition of weighted-mean trimmed regions is developed in Section 2. Next, in Section 3, the continuity, monotonicity, and subadditivity of the regions are established as well as the intersection and projection properties. In Section 4 special cases of the new notion are investigated. They contain the zonoid regions, two continuous versions of the ECH regions, and the geometrically trimmed regions. Section 5 extends the notion of weighted-mean trimmings to trimmings of probability distributions on $\mathbb{R}^d$. A law of large numbers is proved: Under mild restrictions the trimmed regions of an independent sample converge, almost surely and in Hausdorff metric, to the trimmed regions of the underlying probability distribution. Section 6 concludes with remarks on robustness and computability of the new regions.

Some notation: By S^{d-1} we denote the $(d-1)$-dimensional unit sphere in $\mathbb{R}^d$, i.e., the set $\{x \in \mathbb{R}^d \mid \|x\| = 1\}$. Every element in S^{d-1} is interpreted as a direction in $\mathbb{R}^d$. For the set of the first n integers we use the notation $N = \{1, \dots, n\}$. As usual, the integer part of a real number x is denoted by $\lfloor x \rfloor$.

2 A general notion of trimmed regions

Assume that we are given n data points $x_1, \dots, x_n$ in $\mathbb{R}^d$. Our aim is to define a general notion of trimmed regions that satisfies the properties **T1** to **T7**. These regions will be constructed via their support functions (cf. e.g. Rockafellar, 1970). Recall that a closed convex set $K \subset \mathbb{R}^d$ is uniquely determined by its support function $h_K : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$,

$$h_K(p) = \sup \left\{ p'x \mid x \in K \right\}, \ p \in \mathbb{R}^d .$$

Further, the support function is finite for all $p \in \mathbb{R}^d$ if and only if K is a convex body, i.e., closed, convex and bounded. For any direction p in $\mathbb{R}^d$, let π_p denote a permutation that orders the values $p'x_i$ in ascending order,

$$p'x_{\pi_p(1)} \leq p'x_{\pi_p(2)} \leq \cdots \leq p'x_{\pi_p(n)} .$$

The permutation $\pi_{p,x}$ depends on the given data $x_1, \dots, x_n$. For ease of notation, we shall omit the subscript x whenever this is possible and write π_p instead. The permutation is unique if and only if the values $p'x_i$, $i = 1, \dots, n$, are pairwise distinct. The set of directions p for which π_p is not unique will be denoted by

$$H(x_1, \dots, x_n) = \left\{ p \in S^{d-1} \mid \text{there are } i \neq j \text{ such that } p'x_i = p'x_j \right\} .$$

In order to define a trimmed region *via* its support function we consider a function $h(p)$ that is a weighted average of values $p'x_i$,

$$h(p) = \sum_{j=1}^{n} w_{j,\alpha} p' x_{\pi_p(j)}\,, \quad p \in \mathbb{R}^d\,. \tag{1}$$

Here, $w_{j,\alpha}$, $j = 1, \ldots, n$, $\alpha \in I$, is a family of weights with

$$w_{j,\alpha} \geq 0 \text{ for all } j,\ \alpha, \qquad \sum_{j=1}^{n} w_{j,\alpha} = 1 \text{ for all } \alpha.$$

In the sequel, further restrictions will be imposed on the weights in order to make (1) the support function of a convex body that, in particular, satisfies the above properties **T1** to **T7**.

Proposition 1 (Support function)**.** *The function h in (1) is the support function of a convex body if the weights $w_{j,\alpha}$ increase in j.*

Proof: It is well-known that a function $h : \mathbb{R}^d \to \mathbb{R}^d$ is the support function of a convex body if and only if it is sublinear, i.e., if the following two conditions hold:

(*i*) positive homogeneous: $h(\lambda u) = \lambda h(u)$ for all $\lambda > 0$, $u \in \mathbb{R}^d$,

(*ii*) subadditive: $h(u + v) \leq h(u) + h(v)$ for all $u, v \in \mathbb{R}^d$.

It is obvious that h is positive homogeneous. As by assumption $w_{j,\alpha}$ is increasing in j, it holds

$$\sum_{j=1}^{n} w_{j,\alpha} p' x_{\pi_p(j)} \geq \sum_{j=1}^{n} w_{j,\alpha} p' x_{\pi(j)} \quad \text{for every permutation } \pi. \tag{2}$$

Thus,

$$\begin{aligned}
h(u+v) &= \sum_{j=1}^{n} w_{j,\alpha} (u+v)' x_{\pi_{u+v}(j)} \\
&= \sum_{j=1}^{n} w_{j,\alpha} u' x_{\pi_{u+v}(j)} + \sum_{j=1}^{n} w_{j,\alpha} v' x_{\pi_{u+v}(j)} \\
&\leq \sum_{j=1}^{n} w_{j,\alpha} u' x_{\pi_u(j)} + \sum_{j=1}^{n} w_{j,\alpha} v' x_{\pi_v(j)} \\
&= h(u) + h(v)\,,
\end{aligned}$$

i.e., h is subadditive. $\square$

Remark: Note that also a converse of Proposition 1 can be shown: If the weights $w_{j,\alpha}$ are not increasing in j, then there exist $x_1, \ldots, x_n$ such that the function (1) fails to be subadditive.

Now, we are prepared to introduce the central notion of this article, the weighted-mean trimmed regions.

Definition 2 (Weighted mean trimmed regions)**.** *Assume that the weights $w_{j,\alpha}$ satisfy the following conditions (i) to (iii).*

(i) $\sum_{j=1}^{n} w_{j,\alpha} = 1$, $w_{j,\alpha} \geq 0$ *for* $j = 1, \ldots, n$, $\alpha \in I$.

(ii) $w_{j,\alpha}$ *increases in* j.

(iii) *If* $\alpha < \beta$ *then*

$$\sum_{j=1}^{k} w_{j,\alpha} \leq \sum_{j=1}^{k} w_{j,\beta}\,, \quad k = 1, \ldots, n\,. \tag{3}$$

The unique convex body, whose support function is given by

$$h_{\mathrm{D}_\alpha(x_1,\ldots,x_n)}(p) = \sum_{j=1}^{n} w_{j,\alpha} p' x_{\pi_p(j)}\,.$$

is denoted by $\mathrm{D}_\alpha(x_1, \ldots, x_n)$ *(or in short* D_α*). The sets* $\mathrm{D}_\alpha(x_1, \ldots, x_n)$, $\alpha \in I$, *are called the* weighted-mean trimmed regions, *in short the WMT regions, of* $x_1, \ldots, x_n$.

To illustrate these regions, consider the case $d = 1$. In this case the weighted mean trimmed regions are given by

$$\mathrm{D}_\alpha(x_1, \ldots, x_n) = \left[\sum_{j=1}^{n} w_{n+1-j,\alpha} x_{(j)}, \sum_{j=1}^{n} w_{j,\alpha} x_{(j)} \right],$$

where $x_{(j)}$ denotes the j-th smallest value of the data points $x_1, \ldots, x_n$. Thus, the regions are intervals whose endpoints are given by weighted averages of quantiles. This is closely related to the so-called halfspace trimming (see Tukey, 1975), where, in the univariate case, the trimmed regions are intervals whose endpoints are quantiles of the distribution. However, as the

WMT weights have to be increasing in j, the halfspace trimmed regions cannot be represented as WMT regions.

Condition (*ii*) is needed to ensure that the function h, defined by (1), is indeed the support function of a convex body.

Condition (*iii*) guarantees that the weighted-mean trimmed regions are nested. This will be shown in Proposition 3 below. Note that condition (*iii*) is equivalent to saying that the vector of weights is increasing with α in the sense of *majorization*, see, e.g., Marshall and Olkin (1979).

The next proposition provides representations of D_α and of its set of extreme points.

Proposition 2 (Extreme points of WMT regions)**.** *It holds that*

$$\mathrm{D}_\alpha(x_1, \ldots, x_n) = conv\left\{ \sum_{j=1}^{n} w_{j,\alpha} x_{\sigma(j)} \,\middle|\, \sigma \text{ permutation of } \{1, \ldots, n\} \right\}. \tag{4}$$

The set of extreme points of D_α *is given by*

$$Ext\big(\mathrm{D}_\alpha(x_1, \ldots, x_n)\big) = \left\{ \sum_{j=1}^{n} w_{j,\alpha} x_{\pi_p(j)} \,\middle|\, p \in S^{d-1} \setminus H(x_1, \ldots, x_n) \right\}. \tag{5}$$

Proof: Denote the right-hand side of (4) by C. The support function of C is

$$\begin{aligned} h_C(p) &= \max_\sigma p' \sum_{j=1}^{n} w_{j,\alpha} x_{\sigma(j)} = \max_\sigma \sum_{j=1}^{n} w_{j,\alpha} p' x_{\sigma(j)} \\ &= \sum_{j=1}^{n} w_{j,\alpha} p' x_{\pi_p(j)} = h_{\mathrm{D}_\alpha(x_1,\ldots,x_n)}(p) \end{aligned}$$

due to the inequality (2). From this follows equation (4).

To prove (5), observe that x_0 is an extreme point if and only if there exists a direction $p \in S^{d-1}$ such that the equation

$$p'x = h_{\mathrm{D}_\alpha}(p)\,, \quad x \in \mathrm{D}_\alpha\,, \tag{6}$$

has the unique solution x_0; see e.g., Dyckerhoff (2000). Obviously, for $p \in S^{d-1} \setminus H$ the unique solution of (6) is given by $\sum_{j=1}^{n} w_{j,\alpha} x_{\pi_p(j)}$. Hence, (5) follows. □

The first part of the preceding proposition shows that the trimmed regions can be represented as the convex hull of $n!$ points. Of course this is computationally intractable even for moderate values of n. However, the second part of the proposition shows that for computing the extreme points only those permutations have to be considered that are induced by a direction p. In fact, there are efficient algorithms to compute the trimmed regions, see Dyckerhoff (2000) and Cascos (2007) for the bivariate case and Mosler et al. (2009) and Bazovkin and Mosler (2010) for the general case.

For any $x_1, \ldots, x_n$ and any $\alpha \in I$, $\mathrm{D}_\alpha(x_1, \ldots, x_n)$ is bounded, closed and convex and, thus, satisfies **T1**.

Trimmed regions must decrease in the parameter. The next proposition shows that, due to condition (iii) in Definition 2, they are in fact nested.

Proposition 3 (Nestedness). *The WMT regions* D_α *satisfy* **T2**, *i.e.,* $\alpha < \beta$ *implies* $\mathrm{D}_\beta(x_1, \ldots, x_n) \subset \mathrm{D}_\alpha(x_1, \ldots, x_n)$.

Proof: Let $\alpha < \beta$. First, note that $\mathrm{D}_\beta(x_1, \ldots, x_n) \subset \mathrm{D}_\alpha(x_1, \ldots, x_n)$ holds for all $x_1, \ldots, x_n \in \mathbb{R}^d$ if and only if the corresponding support functions are ordered in the same way, i.e., if $h_{\mathrm{D}_\beta(x_1,\ldots,x_n)}(p) \le h_{\mathrm{D}_\alpha(x_1,\ldots,x_n)}(p)$ for every p and $x_1, \ldots, x_n \in \mathbb{R}^d$. The latter condition is equivalent to

$$\sum_{j=1}^{n} w_{j,\beta} p' x_{\pi_p(j)} \le \sum_{j=1}^{n} w_{j,\alpha} p' x_{\pi_p(j)} \quad \text{for every } p \text{ and } x_1, \ldots, x_n \in \mathbb{R}^d . \tag{7}$$

After (i) and (iii) in Definition 2 we have

$$\begin{aligned} &\sum_{j=1}^{k} w_{j,\alpha}(p' x_{\pi_p(k)} - p' x_{\pi_p(k+1)}) \\ &\qquad\qquad \ge \sum_{j=1}^{k} w_{j,\beta}(p' x_{\pi_p(k)} - p' x_{\pi_p(k+1)}) \quad \text{for } k = 1, \ldots, n-1 \end{aligned} \tag{8}$$

and

$$\sum_{j=1}^{n} w_{j,\alpha} p' x_{\pi_p(n)} = \sum_{j=1}^{n} w_{j,\beta} p' x_{\pi_p(n)} \, , \tag{9}$$

where both terms in the last equation equal $p' x_{\pi_p(n)}$. Now, adding the right- and left-hand sides of (8) and (9) we obtain (7). $\square$

Proposition 4 (Affine equivariance)**.** *The WMT regions* D_α *satisfy* **T3**, *i.e., for every matrix* $A \in \mathbb{R}^{m\times d}$ *and every* $b \in \mathbb{R}^d$ *it holds*

$$\mathrm{D}_\alpha(Ax_1 + b, \ldots, Ax_n + b) = A\mathrm{D}_\alpha(x_1, \ldots, x_n) + b\,.$$

Proof: Let $y_i = Ax_i + b$. Then, $p'y_i = (A'p)'x_i + p'b$. Note that $A\mathrm{D}_\alpha + b$ is a convex body that has support function

$$h_{A\mathrm{D}_\alpha+b}(p) = h_{\mathrm{D}_\alpha}(A'p) + b'p\,.$$

The permutation σ_p that orders the values $p'y_i$, $i = 1, \ldots, n$, in ascending order is identical with the permutation $\pi_{A'p}$ that orders the values $(A'p)'x_i$ in ascending order. Thus,

$$\begin{aligned}
h_{\mathrm{D}_\alpha(y_1,\ldots,y_n)}(p) &= \sum_{j=1}^n w_{j,\alpha}\left[(A'p)'x_{\sigma_p(j)} + p'b\right] \\
&= \sum_{j=1}^n w_{j,\alpha}\left[(A'p)'x_{\pi_{A'p}(j)} + p'b\right] \\
&= \sum_{j=1}^n w_{j,\alpha}(A'p)'x_{\pi_{A'p}(j)} + \left[\sum_{j=1}^n w_{j,\alpha}\right] p'b \\
&= h_{\mathrm{D}_\alpha(x_1,\ldots,x_n)}(A'p) + b'p \\
&= h_{A\mathrm{D}_\alpha(x_1,\ldots,x_n)+b}(p)\,,
\end{aligned}$$

from which the proposition follows. □

We summarize the results of the preceding propositions in a theorem.

Theorem 1. *The WMT regions are central regions in the sense of Definition 1, i.e., they satisfy the properties* ***T1***, ***T2*** *and* ***T3***.

The arithmetic mean of the data is contained in each WMT region and, hence, a deepest point. This important result is a consequence of Proposition 3:

Proposition 5 (Mean has maximal depth)**.** *For WMT regions it holds that*

$$\{\overline{x}\} \subset \mathrm{D}_\alpha(x_1, \ldots, x_n) \quad \textit{for each } \alpha \in I\,.$$

Proof: Since the weights $w_{j,\alpha}$ are increasing and sum up to unity, it follows that

$$\sum_{j=1}^{k} w_{j,\alpha} \leq \frac{k}{n} = \sum_{j=1}^{k} \frac{1}{n} \quad \text{for } k = 1, \ldots, n.$$

Thus, it follows from the proof of Proposition 3 that for every direction $p \in S^{d-1}$ we have

$$p'\overline{x} = p' \frac{1}{n} \sum_{j=1}^{n} x_{\pi_p(j)} = \sum_{j=1}^{n} \frac{1}{n} p' x_{\pi_p(j)} \leq \sum_{j=1}^{n} w_{j,\alpha}\, p' x_{\pi_p(j)} = h_{\mathrm{D}_\alpha(x_1,\ldots,x_n)}(p)\,.$$

Since this is equivalent to

$$\{\overline{x}\} \subset \mathrm{D}_\alpha(x_1, \ldots, x_n)\,,$$

the proposition follows. □

If the data are centrally symmetric[1], the center coincides with the mean and thus is included in all weighted-mean trimmed regions. In terms of data depth this is again tantamount saying that the center of symmetry is a deepest point.

Corollary 1 (Center has maximal depth)**.** *If the data are centrally symmetric about some $c \in \mathbb{R}^d$, it holds that*

$$c \in \mathrm{D}_\alpha(x_1, \ldots, x_n)\,, \quad \textit{for each } \alpha \in I.$$

3 Continuity, monotonicity, subadditivity

In this section we discuss three additional features that are common to the notion of trimmed regions defined in Definition 2. We start with continuity properties.

Proposition 6 (Continuity)**.**

(*i*) *The map $(x_1, \ldots, x_n) \mapsto \mathrm{D}_\alpha(x_1, \ldots, x_n)$ is continuous w.r.t. the Hausdorff metric, i.e., satisfies* **T4***.*

(*ii*) *If the map $\alpha \mapsto (w_{1,\alpha}, \ldots, w_{n,\alpha})$ is continuous, then the map $\alpha \mapsto \mathrm{D}_\alpha(x_1, \ldots, x_n)$ is continuous w.r.t. the Hausdorff metric, i.e., satisfies* **T5***.*

[1] A set $C \subset \mathbb{R}^d$ is *centrally symmetric* with center $c \in \mathbb{R}^d$, if for every point $c + d \in C$ the point $c - d$ is also in C

Proof: Recall that the convergence of convex bodies in the Hausdorff metric is equivalent to the pointwise convergence of their support functions. To prove (ii) we have to show that for every sequence $(\alpha_k)_{k\in\mathbb{N}}$ converging to α_0 the sequence of support functions $h_{\mathrm{D}_{\alpha_k}}$ converges pointwise to $h_{\mathrm{D}_{\alpha_0}}$. If the weights are continuous in α, obviously

$$\lim_{k\to\infty} h_{\mathrm{D}_{\alpha_k}}(p) \;=\; \sum_{j=1}^{n}\left(\lim_{k\to\infty} w_{j,\alpha_k}\right) x_{\pi_p(j)} \;=\; \sum_{j=1}^{n} w_{j,\alpha_0} x_{\pi_p(j)} \;=\; h_{\mathrm{D}_{\alpha_k}}(p)\,.$$

Regarding (i), observe that the map

$$(x_1,\ldots,x_n) \mapsto h_{\mathrm{D}_\alpha(x_1,\ldots,x_n)}(p) = \sum_{j=1}^{n} w_{j,\alpha} x_{\pi_p(j)}$$

is continuous. The rest follows immediately. □

One can consider the trimmed region of the data $x_1,\ldots,x_n$ as the trimmed region of a probability distribution that gives probability $1/n$ to each data point. If one adopts this point of view, it is interesting to ask, whether the weighted-mean trimmed regions are also continuous with respect to weak convergence of probability measures, that is qualitative robust in the sense of Hampel (1971). In fact, as it will be discussed in Section 5, the population version of WMT is continuous w.r.t. weak convergence of probability measures, provided the sequence is uniformly integrable. Unfortunately, without the assumption of uniform integrability, this result does not hold. The weighted mean trimmed regions have zero breakdown point since the mean of the data is always a deepest point. Since the mean is not robust, the weighted mean trimmed regions cannot be robust either. In those applications where robustness is an issue one has to preprocess the data with some outlier detection method.

The following two properties play an important role in constructing multivariate risk measures via central regions, see Cascos and Molchanov (2007).

Proposition 7 (Subadditivity)**.** *The trimmed regions* D_α *are subadditive in the data, i.e., satisfy* **T6**.

Proof: Let $z_i = x_i + y_i$. Recall that support functions are additive w.r.t. the Minkowski addition of sets, $h_K(p) + h_L(p) \;=\; h_{K\oplus L}(p)$. The support function of $\mathrm{D}_\alpha(x_1+y_1,\dots,x_n+y_n)$ is given by

$$\begin{aligned}
h_{\mathrm{D}_\alpha(x_i+y_i)}(p) &= \sum_{j=1}^n w_{j,\alpha} p' z_{\pi_{p,z}(j)} \\
&= \sum_{j=1}^n w_{j,\alpha} p' x_{\pi_{p,z}(j)} + \sum_{j=1}^n w_{j,\alpha} p' y_{\pi_{p,z}(j)} \\
&\le \sum_{j=1}^n w_{j,\alpha} p' x_{\pi_{p,x}(j)} + \sum_{j=1}^n w_{j,\alpha} p' y_{\pi_{p,y}(j)} \\
&= h_{\mathrm{D}_\alpha(x_i)}(p) + h_{\mathrm{D}_\alpha(y_i)}(p) \\
&= h_{\mathrm{D}_\alpha(x_i)\oplus \mathrm{D}_\alpha(y_i)}(p)\,.
\end{aligned}$$

Thus,

$$\mathrm{D}_\alpha(x_1+y_1,\dots,x_n+y_n) \subset \mathrm{D}_\alpha(x_1,\dots,x_n) \oplus \mathrm{D}_\alpha(y_1,\dots,y_n)\,.$$

□

Proposition 8 (Monotonicity)**.** *The trimmed regions* D_α *are monotone in the data, i.e., satisfy* **T7***.*

Proof: Assume $x_i \le y_i$ for $i=1,\dots,n$. First, we have to show that

(i) $\mathrm{D}_\alpha(y_1,\dots,y_n) \subset \mathrm{D}_\alpha(x_1,\dots,x_n) \oplus \mathbb{R}^d_+$.

Since $h_{\mathrm{D}_\alpha(x_1,\dots,x_n)\oplus\mathbb{R}^d_+}(p) = h_{\mathrm{D}_\alpha(x_1,\dots,x_n)}(p) + h_{\mathbb{R}^d_+}(p)$, condition (i) is equivalent to

$$h_{\mathrm{D}_\alpha(y_1,\dots,y_n)}(p) \le h_{\mathrm{D}_\alpha(x_1,\dots,x_n)}(p) + h_{\mathbb{R}^d_+}(p) \quad \text{for all } p \in \mathbb{R}^d.$$

Since

$$h_{\mathbb{R}^d_+}(p) = \begin{cases} 0 & \text{if } p \in \mathbb{R}^d_-, \\ \infty & \text{otherwise,} \end{cases}$$

condition (i) is equivalent to

$$h_{\mathrm{D}_\alpha(y_1,\dots,y_n)}(p) \le h_{\mathrm{D}_\alpha(x_1,\dots,x_n)}(p) \quad \text{for all } p \in \mathbb{R}^d_-\,,$$

which has to be checked. From $x_i \le y_i$ obtain $p'y_i \le p'x_i$ for all $p \in \mathbb{R}^d_-$ and all i. Thus, for all $p \in \mathbb{R}^d_-$,

$$\begin{aligned} h_{\mathrm{D}_\alpha(y_1,\dots,y_n)}(p) &= \sum_{j=1}^n w_{j,\alpha} p' y_{\pi_{p,y}(j)} \\ &\le \sum_{j=1}^n w_{j,\alpha} p' x_{\pi_{p,y}(j)} \\ &\le \sum_{j=1}^n w_{j,\alpha} p' x_{\pi_{p,x}(j)} \\ &= h_{\mathrm{D}_\alpha(x_1,\dots,x_n)}(p)\,, \end{aligned}$$

where the second inequality follows from (2). The proof of the second condition **T7**(*ii*) is similar. □

Proposition 4 shows that the WMT regions are *affine equivariant*. Moreover, as the proposition holds also for singular matrices, it implies that the regions D_α are equivariant w.r.t. projections,

$$p'\mathrm{D}_\alpha(x_1,\dots,x_n) = \mathrm{D}_\alpha(p'x_1,\dots,p'x_n)\,, \quad \text{for every } p \in \mathbb{R}^d\,.$$

This property has been named the *strong projection property* in Dyckerhoff (2004), where a number of important implications, concerning orderings between multivariate distributions, is demonstrated.

In order to coincide exactly with the level sets of a data depth, $\mathrm{D}_\alpha = \{y \in \mathbb{R}^d \,|\, \mathrm{depth}(y) \ge \alpha\}$, the regions have to satisfy the so called *intersection property* (Dyckerhoff, 2004),

$$\mathrm{D}_\alpha(x_1,\dots,x_n) = \bigcap_{\beta:\beta<\alpha} \mathrm{D}_\beta(x_1,\dots,x_n)\,.$$

Proposition 9 (Intersection property)**.** *If the map $\alpha \mapsto (w_{1,\alpha},\dots,w_{n,\alpha})$ is continuous from the left, then the WMT regions D_α satisfy the intersection property.*

Proof: From the left-continuity it follows that

$$\lim_{\beta \nearrow \alpha} h_{\mathrm{D}_\beta}(p) = h_{\mathrm{D}_\alpha}(p)\,,$$

which is equivalent to Hausdorff convergence of the sets D_β to D_α. It follows from Proposition 3 that the sets D_β are decreasing in β. Since the Hausdorff limit of a decreasing sequence of sets is its intersection, it follows

$$\mathrm{D}_\alpha(x_1,\dots,x_n) = \operatorname*{H\text{-}lim}_{\beta\nearrow\alpha} \mathrm{D}_\beta(x_1,\dots,x_n) = \bigcap_{\beta:\beta<\alpha} \mathrm{D}_\beta(x_1,\dots,x_n)\,.$$

□

4 Special families of regions

This section presents several special cases of WMT regions. Some of them are known from the literature. They are compared with each other and with the regions obtained by halfspace and simplicial trimmings. Let $x_1,\dots,x_n$ be given data in $\mathbb{R}^d$.

Zonoid regions

Koshevoy and Mosler (1997) introduced the *zonoid trimmed regions* $\mathrm{ZD}_\alpha(x_1, \dots, x_n)$ for $0<\alpha\le 1$ by

$$\mathrm{ZD}_\alpha(x_1,\dots,x_n) = \left\{ \sum_{i=1}^n \lambda_i x_i \,|\, 0\le\lambda_i\le\frac{1}{n\alpha}\,, \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Thus, the support function of the zonoid trimmed regions is given by

$$h_{\mathrm{ZD}_\alpha}(p) = \max \left\{ \sum_{i=1}^n \lambda_i p' x_i \,|\, 0\le\lambda_i\le\frac{1}{n\alpha}\,, \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Obviously, the sum is maximized by putting as much weight as possible on large values of $p'x_i$. Therefore, the support function is

$$h_{\mathrm{ZD}_\alpha}(p) = \sum_{j=1}^n w_{j,\alpha} p' x_{\pi_p(j)}\,,$$

where the weights $w_{j,\alpha}$ are given by

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

These weights obviously satisfy the conditions (i) to (iii) of Definition 2. Proposition 2 now gives a nice representation of the zonoid regions when $\alpha = k/n$ for some $k = 1, \ldots, n$:

$$\mathrm{ZD}_{k/n}(x_1, \ldots, x_n) = \mathrm{conv}\left\{ \frac{1}{k} \sum_{i \in J} x_i \,\middle|\, J \subset N, |J| = k \right\}$$

The above equation says that the zoonoid region of level $\alpha = k/n$ is simply the convex hull of all means of k data points. The support function in this case is given by

$$h_{\mathrm{ZD}_{k/n}}(p) = \sum_{j=n-k+1}^{n} \frac{1}{k}\, p' x_{\pi_p(j)} \,.$$

As the weights are continuous in α, the regions are continuous in α, too. Hence they satisfy all postulates **T1** to **T7**. Besides, they have many properties that are useful in applications, among them:

- Full information about the data; i.e., given the family of zonoid regions, the underlying distribution of data is uniquely determined.
- Multivariate expected shortfall; i.e., the zonoid region can be transformed into a notion of multivariate expected shortfall that is an extension of the univariate expected shortfall (Cascos and Molchanov, 2007).

For a comprehensive treatment of properties of zonoid regions, see Mosler (2002).

To calculate zonoid regions from given data, there exist efficient exact algorithms, in dimension $d = 2$ by Dyckerhoff (2000) and in dimension $d \geq 3$ by Mosler et al. (2009). For the latter algorithm an R-package is available.

Example 1. *In order to illustrate the various trimmed regions let us consider the daily returns on Intel Corp. and Adobe Systems Inc. shares in May and June 2008 ($n = 42$). The data are taken from the historical stock market database at the Center for Research in Security Prices (CRSP) at the University of Chicago. Figure 1 exhibits the data and their zonoid trimmed regions in $\mathbb{R}^2$ (for $\alpha = 0.01, 0.1, 0.2, \ldots 0.8, 0.9$).*

Continuous expected convex hull (CECH) regions

Cascos (2007) defined the *expected convex hull trimmed regions*, in short ECH *regions*, as follows:

$$\mathrm{ECH}_k(x_1,\dots,x_n) = \frac{1}{\binom{n}{k}} \sum_{\substack{J\subset N\\ |J|=k}} \mathrm{conv}\{x_i \,|\, i\in J\}\,. \tag{10}$$

The regions are parameterized by an integer parameter k that ranges from 1 to n. A comparison of the defining formula with that for zonoid regions shows that an ECH region is obtained by first calculating the *convex hull* of each subset of k data points, and second the *mean* of all these convex hulls, while a zonoid region is determined the other way round: First the *means* of all subsets of k data points are calculated, and second the *convex hull* of all these means. An algorithm to calculate the ECH regions in dimension $d = 2$, similar to that in Dyckerhoff (2000), has been given by Cascos (2007).

According to (10), the ECH regions are defined for integer values of k only. Further, the ECH-regions increase with k. Thus, although the regions are nested, they are nested in the reverse way, so that **T2** is not satisfied. However, validity of **T2** can be achieved by a simple reparameterization.

We will extend the definition of ECH regions, so that they are defined not only for integer values of the trimming parameter but for a whole interval of real numbers, namely for all $\alpha\in(0,1]$. These regions will satisfy all properties **T1**,...,**T7**. The main difference to the ECH regions is that our regions satisfy **T5**, i.e., they are continuous in the trimming parameter.

The support function of an ECH region is given by (Cascos, 2007)

$$h_{\mathrm{ECH}_k}(p) = \sum_{j=k}^{n} \frac{\binom{j-1}{k-1}}{\binom{n}{k}} p' x_{\pi_p(j)} = \sum_{j=k}^{n} \frac{\binom{j}{k}-\binom{j-1}{k}}{\binom{n}{k}} p' x_{\pi_p(j)}\,, \quad k=1,\dots,n\,.$$

With

$$\binom{a}{b} = \begin{cases} \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)}\,, & \text{if } a\ge 0,\ 0\le b< a+1\,,\\ 0 & \text{otherwise}\,, \end{cases}$$

we define the weights

$$w_{j,\alpha} = \frac{\binom{j}{\beta}-\binom{j-1}{\beta}}{\binom{n}{\beta}}$$

where $\beta = \alpha^{-1}$. These weights satisfy the conditions of Definition 2. Inserting them into (1) yields a modification of the expected convex hull trimmed regions which we call *continuous expected convex hull trimmed regions*, shortly CECH *regions*. Note that these weights depend continuously on the parameter $\alpha = \beta^{-1}$, therefore the regions are continuous in α as well. For $\alpha = 1/k$ they coincide with the ECH regions in the sense that

$$\mathrm{CECH}_{1/k}(x_1, \dots, x_n) = \mathrm{ECH}_k(x_1, \dots, x_n)\,.$$

The data from Example 1 are again used to illustrate the CECH regions; see Figure 2.

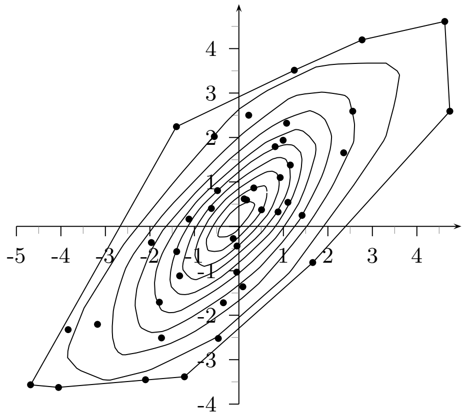

Figure 1: Zonoid regions;
$\alpha = 0.01, 0.1, 0.2, \dots 0.8, 0.9\,.$

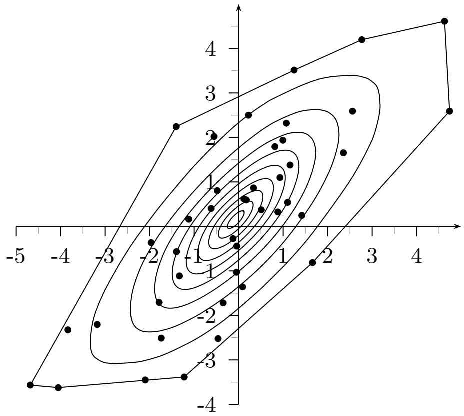

Figure 2: CECH regions;
$\alpha = 0.01, 0.1, 0.2, \dots 0.8, 0.9\,.$

Continuous ECH* **regions**

Cascos (2007) proposes a further type of trimmed regions which we shall call ECH* regions.

$$\mathrm{ECH}^*_k(x_1, \dots, x_n) = \frac{1}{n^k} \sum_{i_1, \dots i_k \in N} \mathrm{conv}\{x_{i_1}, \dots, x_{i_1}\}\,.$$

The support function of these regions is given by

$$h_{\mathrm{ECH}^*_k}(p) = \sum_{j=1}^{n} \frac{j^k - (j-1)^k}{n^k}\, p' x_{\pi_p(j)}\,, \quad k = 1, 2, \dots$$

Like the ECH regions these regions are defined only for integer values of k and do not satisfy **T2**. The same modification as for the ECH regions yields the *continuous* ECH* *regions*, in short CECH* *regions*, with support function

$$h_{\mathrm{CECH}^*_\alpha}(p) = \sum_{j=1}^{n} \frac{j^{1/\alpha} - (j-1)^{1/\alpha}}{n^{1/\alpha}}\, p' x_{\pi_p(j)}\,.$$

Obviously, the weights

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

satisfy all the above properties.

To illustrate the CECH* regions we use the same data; see Figure 3.

Geometrically trimmed regions

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 trimmed regions, which shall be named *geometrically trimmed regions.* They are illustrated in Figure 4.

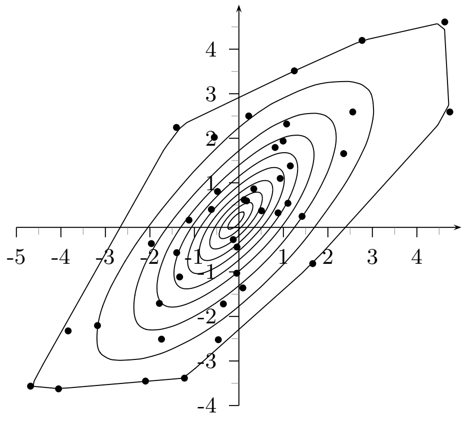

Figure 3: CECH* regions; $\alpha = 0.01, 0.1, 0.2, \ldots 0.8, 0.9\,.$

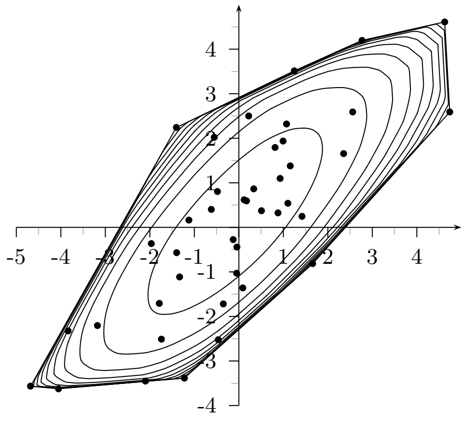

Figure 4: Geometrically trimmed regions; $\alpha = 0.01, 0.1, 0.2, \ldots 0.8, 0.9\,.$

Extreme points compared

If one compares the trimmed regions shown in the four foregoing examples, then one has the impression that the zonoid regions are less "smooth" than the other three regions. This results from the fact that the zonoid regions will in general have less extreme points than the other regions. According to Proposition 2 the set of extreme points is given by the set

$$\mathrm{Ext}\big(\mathrm{D}_\alpha(x_1,\ldots,x_n)\big) = \left\{ \sum_{j=1}^{n} w_{j,\alpha} x_{\pi_p(j)} \,\Big|\, p \in S^{d-1} \setminus H(x_1,\ldots,x_n) \right\}.$$

Consider, e.g., the zonoid regions for $\alpha = k/n$. Here the weights are

$$w_{j,k/n} = \begin{cases} \frac{1}{k}\,, & \text{if } n-k+1 \le j \le n, \\ 0\,, & \text{if } 1 \le j \le n-k. \end{cases}$$

Observe that only two different weights are used in this weighting scheme. Thus, two directions p and q that yield *different permutations* do in general *not generate different extreme points.* Only if there is an index l, such that $\pi_p(l) \le n-k$ and $\pi_q(l) > n-k$ we will get different extreme points.

On the other hand, for the geometrically trimmed regions or the CECH* regions we have n different weights for each α. Thus, two directions p and q that yield *different permutations* will in general also *generate different extreme points.*

Halfspace depth trimming and simplicial depth trimming

To contrast the weighted-mean trimmings with other well-known notions of trimmed regions, we consider the trimmed regions based on the halfspace depth (see Tukey, 1975) and on the simplicial depth (see Liu, 1990) for the same data as above. The trimmed regions for the halfspace depth are given in Figure 5 and for the simplicial depth in Figure 6. The most obvious difference is the lack of convexity of the simplicial depth trimmed regions.

5 Trimming of probability distributions and law of large numbers

So far we have investigated weighted-mean trimmings of data, in other words, of empirical distributions in $\mathbb{R}^d$. In this section we will introduce

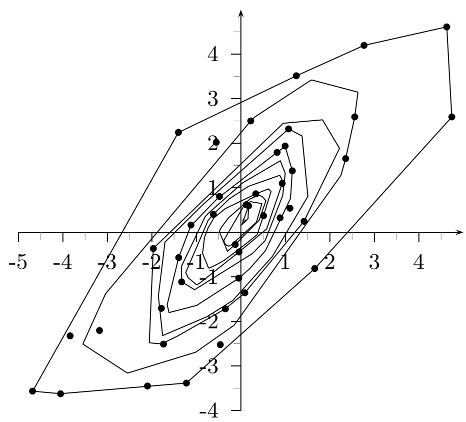

Figure 5: Halfspace depth trimmed regions; $\alpha = 0.02, 0.06, \ldots 0.38, 0.42$.

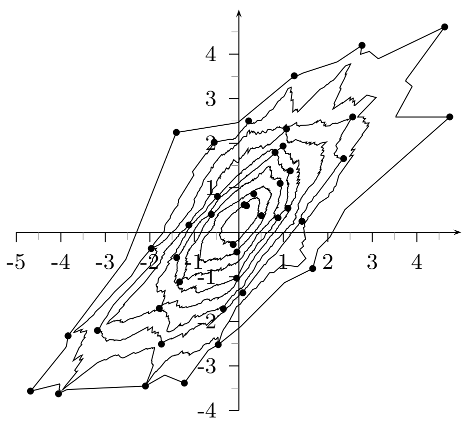

Figure 6: Simplicial depth trimmed regions; $\alpha = 0.01, 0.04, \ldots 0.21, 0.25$.

related trimmings of d-variate probability distributions, and show that under mild conditions a strong law of large numbers applies.

Consider a weighted-mean trimming having weights $w_{j,n,\alpha}$, and let $r_{n,\alpha}$ be an increasing function that generates these weights as follows,

$$r_{n,\alpha}(0) = 0\,, \quad w_{n,j,\alpha} = r_{n,\alpha}\left(\frac{j}{n}\right) - r_{n,\alpha}\left(\frac{j-1}{n}\right)\,, \qquad j = 1, \ldots, n.$$

Then, obviously, $r_{n,\alpha}(1) = 1$, and, in order to satisfy the restrictions of Definition 2, $r_{n,\alpha}(t)$ must be increasing in α for all t and convex for all α. Then, $r_{n,\alpha}$ is absolutely continuous, so that $r_{n,\alpha}$ has a derivative almost everywhere which we shall denote by $r'_{n,\alpha}$.

Now, let X be a d-variate random vector with a finite first moment. Assume that there is a function r_α having bounded derivative r'_α and that the $r_{n,\alpha}$ converge pointwise to r_α. Then, as is shown in Dyckerhoff and Mosler (2010), we may define a population version of the weighted mean trimming by defining $\mathrm{D}_\alpha(X)$ as the unique convex body that has the support function

$$h_{\mathrm{D}_\alpha(X)}(p) = \int_0^1 Q_{p'X}(t)\, dr_\alpha(t)\,.$$

Note that the assumptions on $r_{n,\alpha}$ carry over to r_α. Under these assumptions the map $p \mapsto \int_0^1 Q_{p'X}(t)\, dr_\alpha(t)$ is indeed a support function that defines central regions in the sense of Definition 1; see Dyckerhoff and Mosler (2010). For example, the subadditivity of $h_{\mathrm{D}_\alpha(X)}$ follows from the monotonicity of the weighting function r'_α, the finiteness of $h_{\mathrm{D}_\alpha(X)}$ from the boundedness of r'_α. Further, nestedness is a consequence of the fact that $\alpha \mapsto r_\alpha(t)$ is increasing in α for all t.

The population version of WMT trimming is continuous in α as well as in the distribution. More precisely, if the weight generating functions r_α are continuous in α (in the sense that r_{α_n} converges pointwise to r_α if α_n converges to α), then $\alpha \mapsto \mathrm{D}_\alpha(X)$ is continuous w.r.t. the Hausdorff metric. Further, $\mathrm{D}_\alpha(X_n)$ converges to $\mathrm{D}_\alpha(X)$ whenever X_n is a sequence of random vectors with finite first moments that is uniformly integrable and converges in distribution to X. Without the assumption of uniform integrability the result does in general not hold. Again, the proofs of these results are given in Dyckerhoff and Mosler (2010).

The following theorem states that, under mild conditions on the weight generating functions, for any α the weighted-mean trimmed regions $D_\alpha(X_1, \dots, X_n)$ of a sample $X_1, \dots, X_n$ from X converge with probability one to the weighted mean trimmed region $D_\alpha(X)$ of X.

Theorem 2 (Strong law of large numbers)**.** *Let X be a d-variate random vector with finite first moment and let $X_1, X_2, \dots$ be independent and identically distributed as X. If there is a function r_α such that*

$$\lim_{n\to\infty} r_{n,\alpha}(t) = r_\alpha(t) \qquad \text{for every } t \in (0,1)$$

and the derivatives $r'_{n,\alpha}$ are uniformly bounded in n, i.e., if

$$\sup_{n\in\mathbb{N}} \|r'_{n,\alpha}\|_\infty < \infty\,,$$

then the strong law of large numbers holds:

$$\mathrm{D}_\alpha(X_1, \dots, X_n) \xrightarrow{\text{Hausdorff}} \mathrm{D}_\alpha(X) \qquad P-a.s.$$

Proof: The support function

$$h_{\mathrm{D}_\alpha(X_1,\dots,X_n)}(p) = \sum_{i=1}^{n} w_{j,n,\alpha} p' X_{\pi_p(j)}$$

can be considered as an L-statistic, i.e., a linear function of the order statistics of $p'X_1, \dots, p'X_n$. Under the two conditions above it follows from the strong law for L-statistics (van Zwet, 1980) that

$$\lim_{n\to\infty} \sum_{i=1}^{n} w_{j,n,\alpha} p' X_{\pi_p(j)} = \int_0^1 Q_{p'X}(t)\, dr_\alpha(t) \quad P-a.s.$$

To prove that the central regions $\mathrm{D}_\alpha(X_1,\dots,X_n)$ converge in the Hausdorff metric to $\mathrm{D}_\alpha(X)$, we have to show that their support functions converge to $h_{\mathrm{D}_\alpha(X)}$ uniformly on the unit sphere.

From the above we know that for each single $p \in \mathbb{Q}^d$ we have

$$P\left(\lim_{n\to\infty} h_{\mathrm{D}_\alpha(X_1,\dots,X_n)}(p) = h_{\mathrm{D}_\alpha(X)}(p)\right) = 1\,.$$

Since the countable union of null sets is again a null set it also holds that

$$P\left(\lim_{n\to\infty} h_{\mathrm{D}_\alpha(X_1,\dots,X_n)}(p) = h_{\mathrm{D}_\alpha(X)}(p) \text{ for all } p \in \mathbb{Q}^d\right) = 1\,.$$

Thus, with probability one the support functions $h_{\mathrm{D}_\alpha(X_1,\dots,X_n)}$ converge pointwise on $\mathbb{Q}^d$ to the desired limit. Theorem 10.8 in Rockafellar (1970) says that pointwise convergence of convex functions on a dense subset implies uniform convergence on each closed bounded subset. Since support functions are convex it follows that

$$h_{\mathrm{D}_\alpha(X_1,\dots,X_n)}\Big|_{S^{d-1}} \xrightarrow{\text{uniform}} h_{\mathrm{D}_\alpha(X)}\Big|_{S^{d-1}} \qquad P-a.s.$$

We conclude

$$\mathrm{D}_\alpha(X_1,\dots,X_n) \xrightarrow{\text{Hausdorff}} \mathrm{D}_\alpha(X) \qquad P-a.s.$$

as it was to be shown. □

Example 2. *For the zonoid trimmed regions we have the weight generating functions*

$$r_{n,\alpha}(t) = \begin{cases} \frac{t-(1-\alpha)}{\alpha} & \text{if } 1-\alpha \le t \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $r_{n,\alpha}$ does not depend on n and that

$$r'_{n,\alpha}(t) = \begin{cases} 1/\alpha & \text{if } 1-\alpha \le t \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\sup_{n\in\mathbb{N}} \|r'_{n,\alpha}\|_\infty = 1/\alpha$*, and the assumptions of Theorem 2 are satisfied.*

Example 3. *The* CECH* *regions are generated by*

$$r_{n,\alpha}(t) = t^{1/\alpha}\,.$$

As in the case of the zonoid regions, these functions are independent of n. We obtain

$$r'_{n,\alpha}(t) = \frac{1}{\alpha} t^{\frac{1}{\alpha}-1},$$

and $\|r'_{n,\alpha}\|_\infty = 1/\alpha$. Thus, the assumptions of Theorem 2 are satisfied.

Example 4. *For the* CECH *regions we have*

$$r_{n,\alpha}(t) = \frac{\binom{tn}{1/\alpha}}{\binom{n}{1/\alpha}} = \frac{\Gamma(tn+1)\Gamma(n-\beta+1)}{\Gamma(n+1)\Gamma(tn-\beta+1)},$$

with $\beta = 1/\alpha$. It can be shown that $r_{n,\alpha}$ is indeed a weight generating function for each n, i.e., it is increasing and convex in t, increasing in α, and almost everywhere differentiable. The details of the proof can be found in Proposition 10 in the Appendix.

It is well-known that

$$\lim_{x\to\infty} \frac{\Gamma(x+a)}{\Gamma(x+b)} x^{b-a} = 1,$$

see, e.g. Abramowitz and Stegun (1965). Therefore,

$$\lim_{n\to\infty} t^{-\beta} r_{n,\alpha(t)} = \lim_{n\to\infty} \left[\frac{\Gamma(n+1-\beta)}{\Gamma(n+1)} n^\beta \cdot \frac{\Gamma(tn+1)}{\Gamma(tn+1-\beta)} (tn)^{-\beta} \right] = 1.$$

Thus, the population versions of the CECH *and the* CECH* *regions coincide.*

Further, the first derivative of $r_{n,\alpha}$ is given by

$$r'_{n,\alpha}(t) = n r_{n,\alpha}(t) \left[\psi(tn+1) - \psi(tn-\beta+1)\right],$$

where ψ is the digamma function, see Abramowitz and Stegun (1965). Since $r_{n,\alpha}$ is convex

$$\|r'_{n,\alpha}\|_\infty = r'_{n,\alpha}(1) = n\left[\psi(n+1) - \psi(n-\beta+1)\right].$$

From the mean value theorem and the fact that ψ' is decreasing on $(0,\infty)$ we conclude

$$\|r'_{n,\alpha}\|_\infty = n\beta\psi'(\xi) \le n\beta\psi'(n+1-\beta),$$

where $\xi \in [n+1-\beta, n+1]$. It can be shown that for $x > 1$ the trigamma function $\psi'(x)$ is bounded above by $1/(x-1)$. Thus,

$$\|r'_{n,\alpha}\|_\infty \le n\beta\psi'(n+1-\beta) \le \beta\frac{n}{n-\beta} \xrightarrow{n\to\infty} \beta = \frac{1}{\alpha}.$$

From this now follows $\sup_{n\in\mathbb{N}} \|r'_{n,\alpha}\|_\infty < \infty$ and the assumptions of Theorem 2 are satisfied.

Cascos and Molchanov (2007) also considered a continuous version of the univariate population ECH regions. The same weighting function as in Examples 3 and 4, $r'_{n,\alpha}(t) = \alpha^{-1} t^{\alpha^{-1}-1}$, is proposed by them.

Example 5. *The geometrically trimmed regions are generated by*

$$r_{n,\alpha}(t) = \frac{\alpha^{n(1-t)} - \alpha^n}{1 - \alpha^n}\,,$$

the derivative being

$$r'_{n,\alpha}(t) = n \cdot \frac{\alpha^{n(1-t)}(-\ln\alpha)}{1 - \alpha^n}\,.$$

Therefore

$$\sup_{n\in\mathbb{N}} \left\| r'_{n,\alpha}(t) \right\|_\infty = \sup_{n\in\mathbb{N}} \left[n \cdot \frac{-\ln\alpha}{1-\alpha^n} \right] = \infty\,.$$

By this, the assumptions of Theorem 2 are violated. In fact,

$$\lim_{n\to\infty} r_{n,\alpha}(t) = \begin{cases} 0\,, & \text{if } 0 \le t < 1, \\ 1\,, & \text{if } t = 1. \end{cases}$$

Thus, all weight generating functions converge to the same limit, independent of the trimming parameter α. Of course, this suggests that there is no law of large numbers for the geometrical trimming.

6 Conclusions

A general notion of trimming multivariate data has been introduced, the weighted-mean trimming, which, in contrast to other existing trimmings like halfspace and simplicial trimming, yields central regions that are continuous, subadditive, and monotone in the data. Further, under mild restrictions the trimmed regions satisfy a strong law of large numbers. Also, the weighted-mean trimming satisfies the intersection property, by which the trimmed regions coincide with the upper level sets of a statistical depth function, and the strong projection property, which allows calculating the depth as the infimum of univariate depths regarding data projections in all directions.

Like the existing data depths, the new notion of data depth has many possible applications in describing multivariate data with respect to their location, dispersion and shape, and in testing hypotheses about this. It may be also used in risk analysis, clustering data, and similar tasks. Particularly its continuity and subadditivity properties make the new notion a good choice in many of these applications. However, like the Mahalanobis and the zonoid regions, the weighted-mean regions cannot be employed for the detection and elimination of outliers: As each weighted-mean region contains the mean of the data, these regions are *not robust*; their asymptotic breakdown point is zero.

A crucial issue in applying any notion of trimming and data depth to multivariate data is its computability. Simple is the calculation of *Mahalanobis trimmed regions*, which are ellipses around the mean and, by this, cannot reflect any asymmetry of the data. For most other notions of trimmed regions in the literature exact algorithms are available only in dimension $d = 2$, while at best approximate procedures have been proposed for higher dimensions. An exception are the zonoid trimmed regions; for them an exact algorithm has been constructed (Mosler et al. (2009), also as an R-package) by which they can be efficiently calculated in any dimension. For an extension of this algorithm to the case of general weighted-mean trimmed regions see Bazovkin and Mosler (2010).

Acknowledgment

We thank two anonymous referees for their careful reading and highly valuable suggestions concerning the presentation of the paper.

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York.

Bazovkin, P. and Mosler, K. (2010). An exact algorithm for weighted-mean trimmed regions in any dimension. Discussion Paper in Statistics and Econometrics 06/10, University of Cologne.

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 and I. Molchanov, eds., *New Perspectives in Stochastic Geometry.* Clarendon Press, Oxford University Press, Oxford.

Cascos, I. and Molchanov, I. (2007). Multivariate risks and depth-trimmed regions. *Finance and Stochastics* **11**, 373–397.

Dyckerhoff, R. (2000). Computing zonoid trimmed regions of bivariate data sets. In J. Bethlehem and P. van der Heijden, eds., *COMPSTAT 2000. Proceedings in Computational Statistics*, 295–300. Physica-Verlag, Heidelberg.

Dyckerhoff, R. (2002). Inference based on data depth. Chapter 5, in K. Mosler, *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*, Springer, New York.

Dyckerhoff, R. (2004). Data depths satisfying the projection property. *Allgemeines Statistisches Archiv* **88**, 163–190.

Dyckerhoff, R. and Mosler, K. (2010). Weighted-mean regions of a probability distribution. Working Paper, University of Cologne.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics* **42**, 1887–1896.

Koshevoy, G. and Mosler, K. (1997). Zonoid trimming for multivariate distributions. *Annals of Statistics* **25**, 1998–2017.

Liu, R. (1990). On a notion of data depth based on random simplices. *Annals of Statistics* **18**, 405–414.

Liu, R., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Annals of Statistics* **27**, 783–858. With discussion.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach.* Springer, New York.

Mosler, K. and Hoberg, R. (2006). Data analysis and classification with the zonoid depth. In R. Liu, R. Serfling and D. Souvaine, eds., *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, 49–59. American Mathematical Society.

Mosler, K., Lange, T. and Bazovkin, P. (2009). Computing zonoid trimmed regions in dimension $d > 2$. *Computational Statistics and Data Analysis* **53**, 2500–2510.

Rockafellar, R. T. (1970). *Convex Analysis.* John Wiley & Sons, New York.

Serfling, R. (2006). Depth functions in nonparametric multivariate inference. In R. Liu, R. Serfling and D. Souvaine, eds., *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, 1–16. American Mathematical Society.

Tukey, J. W. (1975). Mathematics and picturing data. In R. James, ed., *Proceedings of the 1974 International Congress of Mathematicians, Vancouver*, vol. 2, 523–531.

van Zwet, W. R. (1980). A strong law for linear functions of order statistics. *Annals of Probability* **8**, 986–990.

Zuo, Y. and Serfling, R. (2000a). General notions of statistical depth function. *Annals of Statistics* **28**, 461–482.

Zuo, Y. and Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *Annals of Statistics* **28**, 483–499.

Appendix

The following proposition shows that the generating function of the CECH regions is indeed a weight generating function.

Proposition 10. *For $n \in \mathbb{N}$ and $\alpha \in (\frac{1}{n+1}, 1]$ let*

$$r_{n,\alpha}(t) = \begin{cases} \frac{\binom{tn}{\beta}}{\binom{n}{\beta}} & \text{if } \frac{\beta-1}{n} < t \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\beta = 1/\alpha$.

Then, the following assertions hold:

(1) $r_{n,\alpha}(0) = 0$, $r_{n,\alpha}(1) = 1$.

(2) $r_{n,\alpha}$ *is continuous in t.*

(3) $r_{n,\alpha}$ *is differentiable for* $t \neq \frac{\beta-1}{n}$.

(4) $r_{n,\alpha}$ *is increasing in* t.

(5) $r_{n,\alpha}$ *is convex in* t.

(6) $r_{n,\alpha}$ *is increasing in* α.

(7) $\lim_{n\to\infty} r_{n,\alpha}(t) = t^{1/\alpha}$ *for all* $t \in [0,1]$.

(8) $\sup_{n\in\mathbb{N}} \|r'_{n,\alpha}\|_\infty < \infty$

Proof: If $\frac{1}{n+1} < \alpha \leq 1$ then $0 \leq \frac{\beta-1}{n} < 1$ and (1) follows.

For $t > (\beta-1)/n$ the function $r_{n,\alpha}$ can be written as

$$r_{n,\alpha}(t) = \frac{\Gamma(tn+1)\Gamma(n-\beta+1)}{\Gamma(n+1)\Gamma(tn-\beta+1)} .$$

From $t > (\beta-1)/n$ it follows that $tn-\beta+1 > 0$. Thus, all arguments of the gamma function are positive. Since $\Gamma(x)$ is positive and continuous for $x > 0$ it follows that $r_{n,\alpha}$ is continuous for $t \neq (\beta-1)/n$. Since $\lim_{x\searrow 0}\Gamma(x) = \infty$ we have $\lim_{t\searrow(\beta-1)/n} r_{n,\alpha}(t) = 0$. Therefore, $r_{n,\alpha}$ is also continuous at $t = (\beta-1)/n$ and (2) is proved.

The derivatives of the gamma functions are given by

$$\Gamma'(z) = \Gamma(z)\psi(z) \quad \text{and} \quad \Gamma''(z) = \Gamma(z)\left[\psi^2(z) + \psi'(z)\right] ,$$

where ψ and ψ' are the digamma and trigamma functions, respectively, see Abramowitz and Stegun (1965). Thus, the first two derivatives of $r_{n,\alpha}$ are given by

$$\begin{aligned}
\frac{dr_{n,\alpha}(t)}{dt} &= nr_{n,\alpha}(t)\left[\psi(tn+1) - \psi(tn-\beta+1)\right] , \\
\frac{d^2r_{n,\alpha}(t)}{dt^2} &= n^2 r_{n,\alpha}(t)\big[(\psi(tn+1) - \psi(tn-\beta+1))^2 \\
&\qquad + \big(\psi'(tn+1) - \psi'(tn-\beta+1)\big)\big] .
\end{aligned}$$

Therefore, (3) holds. Since ψ is increasing for $x > 0$, we see that $r'_{n,\alpha}(t) \geq 0$ for $t > (\beta-1)/n$, which proves (4).

Differentiating w.r.t. α yields

$$\frac{dr_{n,\alpha}(t)}{d\alpha} = \frac{1}{\alpha^2} r_{n,\alpha}(t)\left[\psi(tn+1) - \psi(tn-\beta+1)\right] ,$$

which is also positive for $\alpha > (tn+1)^{-1}$. This proves (6).

It is well-known that
$$\lim_{x\to\infty} \frac{\Gamma(x+a)}{\Gamma(x+b)} x^{b-a} = 1\,,$$
see, e.g. Abramowitz and Stegun (1965). Therefore,
$$\lim_{n\to\infty} t^{-\beta} r_{n,\alpha(t)} = \lim_{n\to\infty} \left[\frac{\Gamma(n+1-\beta+1)}{\Gamma(n+1)} n^{\beta} \cdot \frac{\Gamma(tn+1)}{\Gamma(tn+1-\beta)} (tn)^{-\beta} \right] = 1\,,$$
which implies (7).

To prove (5) we need the following two lemmas.

Lemma 1. *For $a \in (0,\infty)$ let $f_a : (0,\infty) \to \mathbb{R}$ be defined by*
$$f_a(x) = \frac{1-e^{-ax}}{1-e^{-x}}\,.$$
Then, the following assertions hold:

1. *f_a is decreasing iff $a \geq 1$. f_a is increasing iff $a \leq 1$.*
2. *$\lim_{x\to 0} f_a(x) = a$ and $\lim_{x\to\infty} f_a(x) = 1$*
3. *$1 \leq f_a(x) \leq a$ if $a > 1$, and $a \leq f_a(x) \leq 1$ if $a < 1$.*

Proof: The assertions follow by routine calculation. □

Lemma 2. *For $a \in (0,\infty)$ let the function $g_a : (0,\infty) \to \mathbb{R}$ be defined by $g_a(x) = \Gamma(x+a)/\Gamma(x)$. Then, g_a is convex iff $a \geq 1$ and concave iff $a \leq 1$.*

Proof: The second derivative of g_a is given by
$$g_a''(x) = \frac{\Gamma(x+a)}{\Gamma(x)} \left[(\psi(x+a) - \psi(x))^2 + \big(\psi'(x+a) - \psi'(x)\big) \right].$$
Since $\Gamma(x+a)/\Gamma(x) > 0$ we just have to consider the factor in brackets.

For ψ and ψ' the following integral representations hold (see Abramowitz and Stegun, 1965)
$$\psi(x) = \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-xt}}{1-e^{-t}} \right) dt \quad \text{and} \quad \psi'(x) = \int_0^\infty \frac{te^{-xt}}{1-e^{-t}}\, dt\,.$$

Therefore

$$\psi(x+a)-\psi(x)=\int_0^\infty \frac{e^{-xt}-e^{-(x+a)t}}{1-e^{-t}}\,dt=\int_0^\infty \frac{1-e^{-at}}{1-e^{-t}}\,e^{-xt}\,dt$$

and

$$\psi'(x+a)-\psi'(x)=\int_0^\infty \frac{te^{-(x+a)t}-te^{-xt}}{1-e^{-t}}\,dt=-\int_0^\infty \frac{1-e^{-at}}{1-e^{-t}}\,te^{-xt}\,dt\,.$$

Now,

$$\begin{aligned}&[\psi(x+a)-\psi(x)]^2+\big[\psi'(x+a)-\psi'(x)\big]\\ &\qquad=\left(\int_0^\infty \frac{1-e^{-at}}{1-e^{-t}}\,e^{-xt}\,dt\right)^2-\int_0^\infty \frac{1-e^{-at}}{1-e^{-t}}\,te^{-xt}\,dt\end{aligned}$$

The term in parentheses is the Laplace transform of the function f_a defined above. Since the product of Laplace transforms is the Laplace transform of the convolution we see that

$$\begin{aligned}&[\psi(x+a)-\psi(x)]^2+\big[\psi'(x+a)-\psi'(x)\big]\\ &\qquad=\int_0^\infty\int_0^t f_a(u)f_a(t-u)\,du\,e^{-xt}\,dt-\int_0^\infty f_a(t)\,te^{-xt}\,dt\\ &\qquad=\int_0^\infty\left[\frac{1}{t}\int_0^t f_a(u)f_a(t-u)\,du-f_a(t)\right]\,te^{-xt}\,dt\,.\end{aligned}$$

From Lemma 1 it follows that, for $a>1$,

$$f_a(u)f_a(t-u)\geq f_a(t)\geq 1\,.$$

Therefore,

$$\frac{1}{t}\int_0^t f_a(u)f_a(t-u)\,du-f_a(t)\geq\frac{1}{t}\int_0^t f_a(t)\,du-f_a(t)=f_a(t)-f_a(t)=0\,,$$

and thus

$$[\psi(x+a)-\psi(x)]^2+\big[\psi'(x+a)-\psi'(x)\big]\geq 0$$

Conversely, if $a<1$ we get $f_a(u)f_a(t-u)\leq f_a(t)\leq 1$ and thus

$$[\psi(x+a)-\psi(x)]^2+\big[\psi'(x+a)-\psi'(x)\big]\leq 0\,,$$

which finishes the proof of the lemma. $\square$

Now we are able to prove (5). Since $\alpha \leq 1$ we have $\beta \geq 1$. Further,

$$\frac{d^2 r_{n,\alpha}(t)}{dt^2} = n^2 \frac{\Gamma(n-\beta+1)}{\Gamma(n+1)} \, g''_\beta(tn+1-\beta)\,.$$

Now, (5) follows from Lemma 2.

Since $r_{n,\alpha}$ is convex

$$\begin{aligned}\|r'_{n,\alpha}\|_\infty = r'_{n,\alpha}(1) = n r_{n,\alpha}(1)\,[\psi(n+1) - \psi(n-\beta+1)] \\ = n\,[\psi(n+1) - \psi(n-\beta+1)]\ .\end{aligned}$$

From the mean value theorem and the fact that ψ' is decreasing on $(0,\infty)$ now follows

$$\|r'_{n,\alpha}\|_\infty = n\beta\psi'(\xi) \leq n\beta\psi'(n+1-\beta)\,,$$

where $\xi \in [n+1-\beta, n+1]$. The series representation of the trigamma function (see Abramowitz and Stegun, 1965) gives the following upper bound for $x > 1$

$$\psi'(x) = \sum_{k=0}^{\infty} \frac{1}{(x+k)^2} \leq \sum_{k=0}^{\infty} \left[\frac{1}{x+k-1} - \frac{1}{x+k}\right] = \frac{1}{x-1}\,.$$

Thus,

$$\|r'_{n,\alpha}\|_\infty \leq n\beta\psi'(n+1-\beta) \leq \beta\frac{n}{n-\beta}\,.$$

From

$$\lim_{n\to\infty} \beta\frac{n}{n-\beta} = \beta = \frac{1}{\alpha}$$

now follows (8), which completes the proof. $\square$