DEPTH FUNCTIONS BASED ON A NUMBER OF OBSERVATIONS OF A RANDOM VECTOR

Ignacio Cascos

Abstract

Consider the sequence of random sets whose $n$-th element is the convex hull of $n$ independent observations of a random vector. We will study the coverage function (probability that a fixed point belongs to a random set) and the selection expectation of these random sets. The average number of independent observations of a random vector that are needed to cover a fixed point (include it in their convex hull) quantifies its centrality. Moreover, the sequence of selection expectations of an increasing number of independent observations of a random vector is a nested family of sets that are centered with respect to its distribution.

Keywords: convex hull; data depth; depth-trimmed region; halfspace depth; simplicial depth; zonoid trimming.

AMS 2000 subject classifications: 62H05; 60D05
Depth functions based on a number of observations of a random vector

Ignacio Cascos*
Department of Statistics, Universidad Carlos III de Madrid,
Av. Universidad 30, E-28911 Leganés (Madrid), Spain

Abstract

Consider the sequence of random sets whose \( n \)-th element is the convex hull of \( n \) independent observations of a random vector. We will study the coverage function (probability that a fixed point belongs to a random set) and the selection expectation of these random sets. The average number of independent observations of a random vector that are needed to cover a fixed point (include it in their convex hull) quantifies its centrality. Moreover, the sequence of selection expectations of an increasing number of independent observations of a random vector is a nested family of sets that are centered with respect to its distribution.

Key words: convex hull, data depth, depth-trimmed region, halfspace depth, simplicial depth, zonoid trimming

AMS 2000 subject classifications: 62H05, 60D05

1 Introduction

In a multivariate data analysis context, the notion of data depth refers to how central a fixed point is with respect to a data cloud. The mappings that assign to each point its centrality with respect to a multivariate probability distribution (or a data set) are called depth functions, see Zuo and Serfling (2000a). These depth functions introduce center-outward orderings in any multivariate data set. Such orderings make it possible to build multivariate analogues of several univariate data analysis techniques or constructions that are based on order statistics or ranks, from box-and-whiskers plots to \( L \)-statistics. See Liu et al. (1999) for a review of the impressive statistical toolbox built from depth functions.

Associated to each notion of depth, there is a family of depth-trimmed (or central) regions, which are sets of points whose centrality (depth) is, at least, some given value. The set of

*Partially supported by the Spanish Ministry of Science and Innovation under grants MTM2008-01519 and ECO2008-05080 and by the Region of Madrid and Universidad Carlos III de Madrid under grant CCG08-UC3M/HUM-4467.
points whose depth is exactly a given value constitute a kind of multivariate quantile and the depth-trimmed regions are a multivariate generalization of the quantile trimming, see Zuo and Serfling (2000b).

Throughout the manuscript, the concept of random set (set-valued random element) will be used profusely. However, we will always refer to random sets built as the convex hull of a given number of independent observations of a random vector. Clearly enough, the more central a fixed point $x$ is with respect to the distribution of a random vector, the less number of independent observations of the random vector that will be needed to include $x$ in their convex hull. As a consequence, the probability that $x$ is included in a random set built as the convex hull of a fixed number of independent observations of a random vector, as well as the mean number of independent observations of the random vector that are needed to include $x$, can be used to assess the depth of $x$ with respect to the distribution of the random vector, as we will show.

Furthermore, we will consider an increasing number of independent observations of the random vector. The selection (or set-valued) expectation of their convex hull is a set that is central with respect to the distribution of the random vector and grows larger, the more copies we are taking.

Section 2 is devoted to the introduction of depth functions and depth-trimmed regions together with some particular notions of data depth and depth-trimmed regions. In Section 3, we study the depth function given by the expected number of observations of a random vector that are needed so that a given point is contained in their convex hull. Section 4 treats about central regions defined as the (set-valued) expectation of the convex hull of a given number of independent copies of a random vector. We apply them to build a stochastic order and a volume statistic of scatter. Finally, some concluding remarks are briefly discussed in Section 5.

An Appendix describing how to compute the coverage probability of a random sample in terms of the probability of intersections of convex hulls in one fewer dimension is included at the end of the manuscript.

## 2 Classical depths

In the last years, several notions of data depth have been defined to assess the degree of centrality of a point with respect to a multivariate probability distribution. For any $x \in \mathbb{R}^p$ and any $p$-dimensional probability distribution $P$, the depth of $x$ with respect to $P$ is denoted by $D(x; P)$. If there is no possible confusion with the probability distribution, we write $D(x)$ shortly.

The *halfspace depth* of a point $x$ with respect to a probability distribution (see Tukey, 1975; Rousseeuw and Ruts, 1999) is given by

$$\text{HD}(x; P) := \inf \{ P(H) : H \text{ closed halfspace}, x \in H \}.$$  \hspace{1cm} (1)

Another classical depth function is the *simplicial depth*, see Liu (1990). It is defined as the probability that a point $x \in \mathbb{R}^p$ lies inside the convex hull of a simplex whose vertices
are \( p + 1 \) independent observations drawn from \( P \),
\[
\text{SD}(x; P) := \text{Pr}(x \in \text{co}\{X_1, \ldots, X_{p+1}\}),
\]
where \( X_1, X_2, \ldots, X_{p+1} \) is a random sample from \( P \) and \( \text{co} \) stands for the convex hull.

We will follow the definition of depth function given by Dyckerhoff (2004), see also Cascos (2009). If \( \mathbb{P} \) is the set of probabilities on the Borel sets of \( \mathbb{R}^p \), a statistical depth function is a bounded nonnegative mapping \( \text{D}(\cdot; \cdot) : \mathbb{R}^p \times \mathbb{P} \rightarrow \mathbb{R} \) satisfying

**D1** Affine invariance. \( \text{D}(Ax + b; P_{AX+b}) = \text{D}(x; PX) \) for all \( b \in \mathbb{R}^p \) and each \( p \times p \) nonsingular matrix \( A \);

**D2** Vanishing at infinity. \( \text{D}(x; P) \rightarrow 0 \) as \( \|x\| \rightarrow \infty \);

**D3** Upper semicontinuity. The level set of a depth function at any level \( \alpha \), \( \{x \in \mathbb{R}^p : \text{D}(x; P) \geq \alpha \} \) is closed;

**D4** Quasiconcavity. \( \text{D}(\alpha x + (1 - \alpha)y; P) \geq \min\{\text{D}(x; P), \text{D}(y; P)\} \).

In **D1**, notation \( PX \) refers to the probability distribution of the random vector \( X \).

A weaker requirement than **D4** is:

**D4’** Monotonicity with respect to deepest point. \( \text{D}(x; P) \leq \text{D}(\theta + \alpha(x - \theta); P) \) holds for \( \theta = \arg \max_x \text{D}(x; P) \) and any \( \alpha \in [0, 1] \).

The point (set of points) of maximal depth can be considered as a natural candidate for a location parameter, notion closely related to the one of multivariate symmetry (see Rousseeuw and Struyf, 2004; Zuo and Serfling, 2000). We will make use of the angular symmetry. A probability distribution \( P \) is angularly symmetric about \( \theta \) if for any Borel cone \( K \subseteq \mathbb{R}^p \) (i.e. \( rK = K \) for \( r > 0 \) and \( K \) Borel), it holds
\[
P(\theta + K) = P(\theta - K).
\]

The halfspace depth satisfies the properties **D1–D4** and is maximal at the point of angular symmetry, meanwhile the simplicial depth satisfies **D1–D3** and, on absolutely continuous angularly symmetric distributions, it is maximal at the point of angular symmetry and monotonic with respect to it (**D4’**).

From every depth function, it is possible to build a family of central regions. The depth-trimmed region of level \( \alpha \) associated with the depth function \( \text{D} \) is denoted by \( \text{D}^\alpha(P) \) and defined as
\[
\text{D}^\alpha(P) := \{x \in \mathbb{R}^p : \text{D}(x; P) \geq \alpha\}.
\]

By construction, a family of central regions is nested, that is, if \( \alpha \leq \beta \), then \( \text{D}^\beta(P) \subseteq \text{D}^\alpha(P) \).

Further, for each property of a depth function there is an equivalent one for depth-trimmed regions. In the sense that if each of **Di** holds, then the **Ri** below also holds, see Dyckerhoff (2004).
R1 Affine equivariance. $D^\alpha(P_{AX+b}) = AD^\alpha(P_X) + b$ for all $b \in \mathbb{R}^p$ and each $p \times p$ nonsingular matrix $A$;

R2 Boundedness. $D^\alpha(P)$ is bounded for every $\alpha > 0$;

R3 Closedness. $D^\alpha(P)$ is closed for every $\alpha > 0$;

R4 Convexity. $D^\alpha(P)$ is convex for every $\alpha > 0$.

R4’ Starshapedness. If $x$ is contained in all nonempty central regions, then any $D^\alpha(P)$ is starshaped with respect to $x$.

In (3), a family of depth-trimmed regions was built from a depth function. Alternatively, if a family of depth-trimmed regions satisfies the aforementioned properties, then it is possible to define a depth function satisfying the corresponding properties,

$$D(x; P) = \sup\{\alpha : x \in D^\alpha(P)\}.$$  \hspace{1cm} (4)

Since depth-trimmed regions are (set-valued) location parameters, it is also convenient that they are monotone. Notice however that in order to compare two random vectors, we need a multivariate less or equal relation. Given $x, y \in \mathbb{R}^p$, we write $x \leq y$ for the componentwise order, i.e., if $x = (x^{(1)}, \ldots, x^{(p)})$, then $x^{(i)} \leq y^{(i)}$ holds for each $1 \leq i \leq p$.

R5 Monotonicity. If $Y \leq X$ a.s., then for every $x \in D^\alpha(P_X)$, there is $y \in D^\alpha(P_Y)$ such that $y \leq x$.

One last property (subadditivity) is convenient if we aim to use depth-trimmed regions to assess financial risks, see Cascos and Molchanov (2007).

R6 Subadditivity. $D^\alpha(P_{X+Y}) \subseteq D^\alpha(P_X) \oplus D^\alpha(P_Y)$, where the summation of central regions is understood in the Minkowski (elementwise) sense. That is, given $A, B \subseteq \mathbb{R}^p$, we define $A \oplus B := \{a + b : a \in A, b \in B\}$.

For a given probability distribution $P$, Koshevoy and Mosler (1997a) built the zonoid trimmed region of level $\alpha \in (0, 1]$ as

$$ZD^\alpha(P) := \left\{ \int xg(x)dP(x) : g : \mathbb{R}^p \mapsto [0, \alpha^{-1}] \text{ measurable and } \int g(x)dP(x) = 1 \right\}.$$

The zonoid trimmed regions are a nested family of sets that satisfy properties R1–R6. From them, the zonoid depth is defined as in (4).
3 Degree-type depth functions

For a point \( x \in \mathbb{R}^p \), a \( p \)-dimensional distribution \( P \), and any \( k \geq 1 \), we denote by \( d_k(x; P) \) the probability that \( x \) is contained in the convex hull of a random sample from \( P \) of size \( k \),

\[
d_k(x; P) := \Pr(x \in \text{co}\{X_1, \ldots, X_k\}) \quad \text{if} \quad k \geq 1, \tag{5}
\]

where \( X_1, X_2, \ldots, X_k \) are \( k \) independent observations drawn from \( P \).

For technical reasons, we define \( d_0(x; P) = 0 \). Notice that when \( k = p + 1 \), we obtain the simplicial depth, \( \text{SD}(x; P) = d_{p+1}(x; P) \), which is, in fact, the first strictly positive element of the sequence \( \{d_k(x; P)\}_k \) if \( P \) is absolutely continuous.

The expected degree depth, denoted by \( \text{ED}(\cdot; P) \), is defined as the inverse of the expected number of independent observations drawn from \( P \) that are needed to contain a fixed point inside their convex hull. If \( x \in \mathbb{R}^p \), we define

\[
\text{ED}(x; P) := \left(\sum_{k=0}^{\infty} \Pr(x \notin \text{co}\{X_1, \ldots, X_k\})\right)^{-1} = \left(\sum_{k=0}^{\infty} (1 - d_k(x; P))\right)^{-1}. \tag{6}
\]

The term degree is borrowed from Chiu and Molchanov (2003), where the authors define the degree of a typical point of a point process as the number of neighboring points from the process (starting from the nearest one) that are needed to contain the given point inside their convex hull.

In the univariate case \((p = 1)\), if \( P \) is absolutely continuous and \( F \) is its associated cdf, i.e. \( F(x) = P((-\infty, x]) \), then

\[
d_k(x; P) = 1 - F(x)^k - (1 - F(x))^k, \tag{7}
\]

which for \( k = 2 \) amounts to the well-known expression \( \text{SD}(x) = 2F(x)(1 - F(x)) \). Further, the expected degree depth with respect to a continuous univariate distribution is

\[
\text{ED}(x; P) = \frac{F(x)(1 - F(x))}{1 - F(x)(1 - F(x))}. \tag{8}
\]

The expected degree depth has been defined in (6) in terms of an infinite sum, and thus we must study under what circumstances that sum is finite.

The point \( x \) does not belong to the convex hull of a set of points, if and only if all of them are in an open halfspace with \( x \) on its boundary. The supremum of the probabilities of all such halfspaces is related to the halfspace depth,

\[
\sup\{P(H) : H \text{ open halfspace, } x \in \partial H\} = 1 - \text{HD}(x; P),
\]

where \( \partial H \) stands for the boundary of \( H \).

If all the observations from \( P \) are on the same open halfspace, they will not contain \( x \) inside their convex hull, and thus

\[
(1 - \text{HD}(x))^k \leq 1 - d_k(x) \quad \text{for} \quad k \geq 1, \tag{9}
\]
which also holds for \( k = 0 \), since \( d_0(x) = 0 \). Adding on \( k \), we obtain

\[
\text{ED}(x) \leq \text{HD}(x) .
\] (10)

If \( P \) is absolutely continuous, we obtain a more accurate bound for \( \text{ED}(x) \). In that case, it holds that \( \text{HD}(x)^k + (1 - \text{HD}(x))^k \leq 1 - d_k(x) \) for \( k \geq 1 \) and thus

\[
\text{ED}(x) \leq \frac{\text{HD}(x)(1 - \text{HD}(x))}{1 - \text{HD}(x)(1 - \text{HD}(x))} ,
\]

which resembles the expression in (8).

A lower bound for \( d_k \) (and \( \text{ED} \)) at \( x \), whenever \( P \) assigns probability zero to each hyperplane containing \( x \), can be obtained the following way. Any set of \( p - 1 \) observations \( X_1, X_2, \ldots, X_{p-1} \) drawn from \( P \) will define, together with \( x \), a hyperplane. Let \( H(\{x, X_1, \ldots, X_{p-1}\}) \) be any of the two halfspaces that the hyperplane determines,

\[
1 - d_k(x) = \Pr(x \notin \text{co}\{X_1, \ldots, X_k\})
= \Pr(x \text{ and } p - 1 \text{ of the } Xs \text{ are in the boundary of an open halfspace containing the remaining } Xs)
\leq 2\binom{k}{p-1} \Pr(X_p, \ldots, X_k \in H(\{x, X_1, \ldots, X_{p-1}\}))
\leq 2\binom{k}{p-1} (1 - \text{HD}(x))^{k-p+1}, \quad k \geq p,
\]

where, obviously enough, \( \binom{k}{p-1} \) is the number of subsets of \( p - 1 \) random vectors out of the set of \( k \) and 2 is the number of halfspaces determined by a hyperplane.

Adding and inverting, we obtain a lower bound for the expected degree depth,

\[
\text{ED}(x) \geq \frac{1}{p - 2 + 2\text{HD}(x)^{-p}} .
\]

3.1 Properties

We will show that the expected degree depth is a statistical depth function. It satisfies the properties of

- **affine invariance**, vanishing at infinity, **upper semicontinuity**
- and, on angularly symmetric absolutely continuous distributions, **monotonicity with respect to deepest point**.

Further, for any \( k \geq p + 1 \), the function \( d_k(\cdot; P) \) satisfies the same properties.

**D1. Affine invariance** An affine transformation simultaneously applied to a point and a random vector does neither decrease \( d_k(\cdot) \) nor the expected degree depth.

**Proposition 1.** For any \( A \in \mathbb{R}^{q \times p} \) and \( b \in \mathbb{R}^q \), we have that

\[
d_k(Ax + b; P_{AX+b}) \geq d_k(x; P_X) \quad \text{for any } k \geq 1 .
\] (11)

If \( q = p \) and \( A \) is nonsingular, equality holds showing that \( d_k(\cdot) \) is affine invariant.

6
Proof. Let $A \in \mathbb{R}^{q \times p}$, $b \in \mathbb{R}^q$, $x \in \mathbb{R}^p$, and $k \geq 1$. If $x \in \text{co}\{X_1, \ldots, X_k\}$, then $Ax + b \in \text{co}\{AX_1 + b, \ldots, AX_k + b\}$, so $\Pr(x \in \text{co}\{X_1, \ldots, X_k\}) \leq \Pr(Ax + b \in \text{co}\{AX_1 + b, \ldots, AX_k + b\})$ and the first part of the result holds.

If $q = p$ and $A$ is nonsingular, $x \in \text{co}\{X_1, \ldots, X_k\}$ holds if and only if $Ax + b \in \text{co}\{AX_1 + b, \ldots, AX_k + b\}$ and the second part of the result is now straightforward.

Corollary 2. For any $A \in \mathbb{R}^{q \times p}$ and $b \in \mathbb{R}^q$, we have that
$$\text{ED}(Ax + b; P_{AX+b}) \geq \text{ED}(x; PX).$$

If $q = p$ and $A$ is nonsingular, equality holds showing that the expected degree depth is affine invariant.

D2. Vanishing at infinity

Proposition 3. For any $k \geq 1$ and any probability distribution $P$, it holds that
$$\lim_{\|x\| \to \infty} d_k(x; P) = 0 \quad \text{and} \quad \lim_{\|x\| \to \infty} \text{ED}(x; P) = 0.$$

Proof. From (9) and (10), $d_k(x) \leq 1 - (1 - \text{HD}(x))^k$ and $\text{ED}(x) \leq \text{HD}(x)$. Since the halfspace depth satisfies D2, so do $d_k(\cdot)$ and the expected degree depth.

D3. Upper semicontinuity

Proposition 4. For any $k \geq 1$, the function $d_k(\cdot; P)$ is upper semicontinuous.

Proof. Let $x, x_m \in \mathbb{R}^p$ such that $\lim_m x_m = x$ and $d_k(x_m) \geq \alpha$ for all $m$. Let $A = \{(X_1, \ldots, X_k) : x \in \text{co}\{X_1, \ldots, X_k\}\}$ and $A_m = \{(X_1, \ldots, X_k) : x_m \in \text{co}\{X_1, \ldots, X_k\}\}$. Since $\limsup_m A_m = A$, we obtain
$$d_k(x) = \Pr(A) = \Pr(\limsup_m A_m) \geq \limsup_m \Pr(A_m) \geq \alpha.$$

Corollary 5. The expected degree depth is upper semicontinuous.

Proof. Since all $d_k(\cdot)$ are upper semicontinuous, each $(1 - d_k(\cdot))$ is lower semicontinuous and further, they are non-negative which implies that their infinite sum is also lower semicontinuous. Finally, the inverse of the sum is upper semicontinuous.
D4’. Monotonicity with respect to deepest point The argument of the proof of the next result follows the same lines of that of (Liu, 1990, Th. 3) where the simplicial depth is shown to be monotonic with respect to the point of angular symmetry.

Proposition 6. If $P$ is absolutely continuous and angularly symmetric about $\theta$, then for any $\alpha \in [0, 1]$ and $x \in \mathbb{R}^p$,

$$d_k(\theta + \alpha(x - \theta); P) \geq d_k(x; P) \quad \text{for every } k \geq 1.$$  \hfill (12)

\textit{Proof.} For convenience, we take the origin as the point of angular symmetry, $\theta = 0$, which supposes no restriction, since by Proposition 1, $d_k(\cdot)$ is affine invariant.

We will show that for any $\alpha \in [0, 1]$, the relation $d_k(x) \leq d_k(\alpha x)$ holds for any $k \geq 1$. Since $P$ is absolutely continuous, the probability of any hyperplane is equal to 0, and thus if $k \leq p$, it holds that $d_k(x) = d_k(\alpha x) = 0$. Let us take $k \geq p + 1$.

We consider the arrow from $\alpha x$ to $x$ and the events that it enters or leaves the random polytope $\text{co}\{X_1, \ldots, X_k\}$,

$$A_{\text{in}} := \{\text{the arrow from } \alpha x \text{ to } x \text{ enters } \text{co}\{X_1, \ldots, X_k\}\};$$

$$A_{\text{out}} := \{\text{the arrow from } \alpha x \text{ to } x \text{ leaves } \text{co}\{X_1, \ldots, X_k\}\}.$$

Since the probability of any hyperplane is 0, then with probability 1, the affine dimension of $\text{co}\{X_1, \ldots, X_k\}$ equals $p$ and the affine dimension of any subset of $p$ independent observations from $P$ equals $p - 1$.

Given $p$ points $x_1, \ldots, x_p \in \mathbb{R}^p$ that define a hyperplane, let $H(\{x_1, \ldots, x_p\})$ be the closed halfspace with those points in its boundary and containing the origin of coordinates. If the origin of coordinates is in the hyperplane defined by $x_1, \ldots, x_p$, let $H(\{x_1, \ldots, x_p\})$ be any of the two closed halfspaces with $x_1, \ldots, x_p$ in its boundary. We denote by $C$ the set of $p$-tuples of points of $\mathbb{R}^p$ whose convex hull intersects the segment $\text{co}\{x, \alpha x\}$, i.e.

$$C := \{(x_1, \ldots, x_p) : x_i \in \mathbb{R}^p, \text{co}\{x_1, \ldots, x_p\} \cap \text{co}\{x, \alpha x\} \neq \emptyset\}.$$

For any subset $S$ of $p$ elements from \{1,2,\ldots, k\}, we denote by $A_{\text{in}}^S$ the event that the convex hull of $p$ independent observations from $P$, given by $\{X_i : i \in S\}$, intersects with the arrow from $\alpha x$ to $x$ and none of the remaining observations belongs to $H(\{X_i : i \in S\})$, that is

$$A_{\text{in}}^S := \{\{X_i : i \in S\} \subseteq C \text{ and } X_j \notin H(\{X_i : i \in S\}) \text{ for every } j \notin S\},$$

and clearly

$$\Pr(A_{\text{in}}^S) = \int_C \left(1 - P(H(\{x_1, \ldots, x_p\}))\right)^{k-p} dP(x_1) \ldots dP(x_p).$$

Moreover, we denote by $A_{\text{out}}^S$ the event that the convex hull of the $p$ independent observations with index in $S$, $\{X_i : i \in S\}$, intersects with the arrow from $\alpha x$ to $x$ and all of the remaining observations belong to $H(\{X_i : i \in S\})$, that is

$$A_{\text{out}}^S := \{\{X_i : i \in S\} \subseteq C \text{ and } X_j \in H(\{X_i : i \in S\}) \text{ for every } j \notin S\},$$
and clearly
\[ \Pr(A_{\text{out}}^S) = \int_C P(H(\{x_1, \ldots, x_p\}))^{k-p} dP(x_1) \ldots dP(x_p). \]

Now, since \((\begin{pmatrix} k \\ p \end{pmatrix})\) is the number of different configurations for a set \(S \subset \{1, \ldots, k\}\) with \(p\) elements, we have \(\Pr(A_{\text{out}}) = \frac{1}{(\begin{pmatrix} k \\ p \end{pmatrix})} \Pr(A_{\text{out}}^S)\) and \(\Pr(A_{\text{in}}) = \frac{1}{(\begin{pmatrix} k \\ p \end{pmatrix})} \Pr(A_{\text{in}}^S)\).

If \(\beta \in \{\alpha, 1\}\), we denote by \(B_\beta\) the event \(\beta x\) belongs to the convex hull of the \(k\) independent observations from \(P\), that is, \(B_\beta := \{\beta x \in \text{co}\{X_1, X_2, \ldots, X_k\}\}\). We finally have that
\[ \Pr(\alpha x \in \text{co}\{X_1, \ldots, X_k\}) - \Pr(x \in \text{co}\{X_1, \ldots, X_k\}) = \Pr(B_\alpha \setminus B_1) - \Pr(B_1 \setminus B_\alpha). \]
The new events are easily obtained from the previous ones,
\[ B_\alpha \setminus B_1 = A_{\text{out}} \setminus A_{\text{in}}, \quad B_1 \setminus B_\alpha = A_{\text{in}} \setminus A_{\text{out}}, \]
which leads to
\[ d_k(\alpha x) - d_k(x) = \Pr(A_{\text{out}}) - \Pr(A_{\text{out}} \cap A_{\text{in}}) - (\Pr(A_{\text{in}}) - \Pr(A_{\text{in}} \cap A_{\text{out}})) \]
\[ = \Pr(A_{\text{out}}) - \Pr(A_{\text{in}}). \]

By the angular symmetry of \(P\) about the origin and the fact that the origin belongs to \(H(\{x_1, \ldots, x_p\})\), it holds that \(P(H(\{x_1, \ldots, x_p\})) \geq 1/2\), and thus
\[ \left(1 - P(H(\{x_1, \ldots, x_p\}))\right)^{k-p} \leq P(H(\{x_1, \ldots, x_p\}))^{k-p}, \]
or equivalently \(\Pr(A_{\text{out}}^S) \geq \Pr(A_{\text{in}}^S)\). As a consequence, we have \(\Pr(A_{\text{out}}) \geq \Pr(A_{\text{in}})\) and finally \(d_k(\alpha x) \geq d_k(x)\).

Proposition 6 can be immediately restated in terms of the expected degree depth.

**Corollary 7.** If \(P\) is absolutely continuous and angularly symmetric about \(\theta\), for any \(\alpha \in [0, 1]\) and \(x \in \mathbb{R}^p\), it holds that \(\text{ED}(x; P) \leq \text{ED}(\theta + \alpha(x - \theta); P)\).

The absolute continuity is a necessary assumption in Corollary 7.

**Example 8.** Let \(P\) be a univariate probability given by \(P(\{-1\}) = P(\{1\}) = 0.45\) and \(P(\{0\}) = 0.1\). Its median is 0, and therefore \(P\) is angularly symmetric about 0. However, \(\text{ED}(-1; P) = \text{ED}(1; P) = 0.45\) which is greater than the depth of the median, \(\text{ED}(0; P) = 11/29\).

In conclusion, \(\text{ED}(\cdot)\) and \(d_k(\cdot)\) satisfy properties **D1–D3**, and under the extra assumptions of absolute continuity and angular symmetry, **D4’** also holds.

**Remark 9.** The depth-trimmed regions of the expected degree depth and the ones of \(d_k\) satisfy properties **R1–R3** and **R4’** if \(P\) is absolutely continuous and angularly symmetric. A simple coupling argument shows that they satisfy **R5**. The fact that, in the univariate case, these notions of depth produce the same center-outward ordering on the real line as the quantile function, see (7) and (8), which is not subadditive, shows that they are, in general, not subadditive.
3.1.1 Further properties

Wendel (1962) derived an expression from geometric probability that can be restated in terms of our degree-type depth functions.

Proposition 10. If a \( p \)-dimensional probability distribution \( P \) is angularly symmetric about \( \theta \) and assesses probability zero to every hyperplane through \( \theta \), then

\[
d_k(\theta; P) = 1 - \frac{1}{2^{k-1}} \sum_{i=0}^{p-1} \binom{k-1}{i} \quad \text{for any } k \geq p \quad \text{and} \quad \ED(\theta; P) = \frac{1}{2p+1}.
\]

Proof. The expression of \( d_k(\theta) \) is the main result in Wendel (1962) and \( \ED(\theta) \) is obtained performing the addition in (6).

From Proposition 10, it follows that \( \SD(\theta; P) = 2^{-p} \) which is well-known from (Liu, 1990, Th. 4) where it is shown under the extra assumption of absolute continuity of \( P \).

Wagner and Welzl (2001) showed that for absolutely continuous probability distributions, the value of \( d_k(\theta; P) \) in Proposition 10 is an upper bound for the probability that a point is contained in the convex hull of \( k \) i.i.d. random vectors in \( \mathbb{R}^p \).

Proposition 11. If \( P \) is an absolutely continuous \( p \)-dimensional probability distribution, for any \( x \in \mathbb{R}^p \), it holds that

\[
d_k(x; P) \leq 1 - \frac{1}{2^{k-1}} \sum_{i=0}^{p-1} \binom{k-1}{i} \quad \text{for any } k \geq p \quad \text{and} \quad \ED(x; P) \leq \frac{1}{2p+1}.
\]

From Proposition 11, it follows that \( \SD(x; P) \leq 2^{-p} \) if \( P \) is absolutely continuous.

3.2 Bivariate degree

Jewell and Romano (1982) described a method to compute the probability that \( k \) independent observations drawn from an absolutely continuous bivariate probability distribution \( P \) contain a fixed point \( x \in \mathbb{R}^2 \) in their convex hull. They showed that

\[
d_k(x; P) = 1 - k \int_{\mathbb{R}^2} P(H_{x,y})^{k-1} dP(y) \quad \text{for } k \geq 1,
\]

where \( H_{x,y} \) is the closed halfplane with \( x \) and \( y \) on its boundary which is on the clockwise side of the arrow from \( x \) to \( y \), that is,

\[
H_{x,y} = \{ x + (z^{(2)}, -z^{(1)}): \langle z, y-x \rangle \geq 0 \}.
\]

As a consequence, the expected degree depth of \( x \in \mathbb{R}^2 \) can be computed as

\[
\ED(x; P) = \left( 1 + \int_{\mathbb{R}^2} \frac{1}{P(H_{x,y})^2} dP(y) \right)^{-1}.
\]
For a random sample of size \( n \) drawn from \( P, X_1, X_2, \ldots, X_n \), let \( P_n \) be its empirical distribution. Let us assume that \( X_1, X_2, \ldots, X_n \) are in general position (i.e., no three of them lie on the same line) which happens with probability one if \( P \) is absolutely continuous. The probability that a fixed point \( x \in \mathbb{R}^2 \) does not belong to the convex hull of \( k \) observations drawn at random from \( P_n \) is the sum of the probabilities that all those \( k \) observations lie on the halfspace \( H_{x,X} \), and \( X_i \) is an extreme observation, that is, it does not belong to any \( H_{x,X_j} \) where \( j \neq i \) and \( X_j \) was drawn from the previous sample. We have

\[
d_k(x; P_n) = 1 - \sum_{i=1}^{n} \left( P_n(H_{x,X_i})^k - (P_n(H_{x,X_i}) - 1/n)^k \right).
\]

And, from the expression above, we obtain the empirical expected degree depth, denoted by \( \text{ED}_n(\cdot) := \text{ED}(\cdot; P_n) \), as

\[
\text{ED}_n(x) = \left( 1 + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{P_n(H_{x,X_i})} \left( 1 - 1/n \right) \right)^{-1}.
\]

Rousseeuw and Ruts (1996) built an algorithm, based on the angles of the projections of the data points over the circumference centered at \( x \in \mathbb{R}^2 \), to compute all the values \( P_n(H_{x,X_i}) \), with complexity \( O(n \log n) \) in time. That algorithm was originally designed to compute the bivariate halfspace and simplicial depths, however, we can take advantage of it and compute \( \text{ED}_n(\cdot) \) or any \( d_k(\cdot; P_n) \) for any point with complexity \( O(n \log n) \) in time.

To end the current section, we will show that \( \text{ED}_n \) is a uniformly consistent estimator of \( \text{ED} \), but in order to do so, we need a previous result.

**Lemma 12.** If \( P \) is absolutely continuous and \( \alpha \in (0, 1] \), on a set of probability one, it holds that

\[
\sup_{x \in \text{HD}^\alpha(P)} \left| \int \frac{1}{P(H_{x,y})^2} (dP(y) - dP_n(y)) \right| \to 0.
\]

**Proof.** Let \( x_0 \in \text{HD}^\alpha(P) \) and \( B_\delta(x_0) \) be the closed ball centered at \( x_0 \) of radius \( \delta \)

\[
\int \sup_{x_1 \in B_\delta(x_0)} \frac{1}{P(H_{x_1,y})^2} dP(y) - \int \inf_{x_2 \in B_\delta(x_0)} \frac{1}{P(H_{x_2,y})^2} dP(y)
= \int \sup_{x_1,x_2 \in B_\delta(x_0)} \frac{P(H_{x_1,y})^2 - P(H_{x_2,y})^2}{P(H_{x_1,y})^2P(H_{x_2,y})^2} dP(y)
\leq \frac{2}{\alpha^4} \int \sup_{x_1,x_2 \in B_\delta(x_0)} (P(H_{x_1,y}) - P(H_{x_2,y})) dP(y).
\]

Let \( \varepsilon > 0 \). If \( d = \text{diam}(\text{HD}^\alpha(P)) \) is the supremum of the distances between any two points from \( \text{HD}^\alpha(P) \), \( \gamma > 0 \) is such that \( P(B_\gamma(x)) < \varepsilon \alpha^4/4 \) for all \( x \in \mathbb{R}^2 \) and \( b > 0 \) satisfies that for any \( y, u \in \mathbb{R}^2 \) with \( \|u\| = 1 \), the probability of the band of width \( b \) with \( y \) on its boundary
and normal to \( u \), \( P(\{x + y : 0 \leq \langle x, u \rangle \leq b\}) < \varepsilon \alpha^4/4 \). Notice that \( \gamma \) and \( b \) exists since \( P \) is absolutely continuous, let \( \delta = r\gamma(2d)^{-1} \). We have

\[
\int \sup_{x_1 \in B_z(x_0)} \frac{1}{P(H_{x_1,y})^2} dP(y) - \int \inf_{x_2 \in B_z(x_0)} \frac{1}{P(H_{x_2,y})^2} dP(y) \\
\leq \frac{2}{\alpha^4} \left[ \int_{B_z(x_0)} dP(y) + \int_{\mathbb{R}^2 \setminus B_z(x_0)} \sup_{x_1, x_2 \in B_z(x_0)} (P(H_{x_1,y}) - P(H_{x_2,y})) dP(y) \right] \\
\leq \frac{2}{\alpha^4} \left[ \frac{\varepsilon \alpha^4}{4} + \int_{\mathbb{R}^2 \setminus B_z(x_0)} \frac{\varepsilon \alpha^4}{4} dP(y) \right] \leq \varepsilon .
\]

Therefore, condition 1.1.6 in Gaenssler and Stute (1979) is satisfied, and the result holds.

\[ \square \]

If \( P \) is absolutely continuous, then \( \text{ED}_n \) is uniformly consistent.

**Theorem 13.** If \( P \) is absolutely continuous, then

\[
\sup_x |\text{ED}_n(x) - \text{ED}(x)| \longrightarrow 0 \quad a.s. \text{ as } n \to \infty .
\]

**Proof.** Let \( \varepsilon > 0 \), the strong uniform consistency of the empirical halfspace depth (see Donoho and Gasko, 1992, pp. 1816–1817), guarantees that with probability 1, there exists \( n_0 \) such that \( \sup_{n} |\text{HD}_n(x) - \text{HD}(x)| < \varepsilon/6 \) if \( n \geq n_0 \).

For all \( x \in \mathbb{R}^2 \) with \( \text{HD}(x) < \varepsilon/3 \), we have that with probability 1, it holds \( \text{HD}_n(x) \leq \text{HD}(x) + |\text{HD}_n(x) - \text{HD}(x)| < \varepsilon/2 \) for \( n \geq n_0 \). After (10), with probability 1,

\[
|\text{ED}_n(x) - \text{ED}(x)| \leq \text{ED}_n(x) + \text{ED}(x) \leq \text{HD}_n(x) + \text{HD}(x) < \varepsilon \quad \text{for } n \geq n_0 .
\]

If \( \text{HD}(x) \geq \varepsilon/3 \), i.e., \( x \in \text{HD}^{\varepsilon/3}(P) \), notice that for \( n \geq n_0 \), we have \( \text{HD}_n(x) > \varepsilon/6 \) a.s. The distance between the empirical and the population expected degree depth can be decomposed as:

\[
|\text{ED}_n(x) - \text{ED}(x)| \leq \left| \int \frac{1}{P(H_{x,y})^2} dP(y) - \int \frac{1}{P_n(H_{x,y})(P_n(H_{x,y}) - 1/n)} dP_n(y) \right| \\
\leq \left[ \left| \int \frac{1}{P(H_{x,y})^2} (dP(y) - dP_n(y)) \right| \\
+ \left| \int \frac{1}{P(H_{x,y})^2} - \frac{1}{P_n(H_{x,y})(P_n(H_{x,y}) - 1/n)} dP_n(y) \right| \right] .
\]

On the one hand, by Lemma 12, we have

\[
\sup_x \left| \int \frac{1}{P(H_{x,y})^2} (dP(y) - dP_n(y)) \right| \longrightarrow 0 \quad a.s.
\]
On the other hand

\[\left| \int \frac{1}{P(H_{x,y})^2 - 1/nP(H_{x,y})} \frac{1}{dP_n(y)} - \frac{1}{P_n(H_{x,y}) - 1/n} \frac{dP_n(y)}{dP_n(x)} \right| \leq \frac{\int (P_n(H_{x,y})^2 - 1/nP(H_{x,y}) - P(H_{x,y})^2) dP_n(x)}{\text{HD}(x)^2 \text{HD}_n(x)(\text{HD}_n(x) - 1/n)} \leq \frac{324n}{\varepsilon^3(n\varepsilon - 1)} \left[ 2 \int (P_n(H_{x,y}) - P(H_{x,y})) dP_n(x) + 1/n \right] \leq \frac{324n}{\varepsilon^3(n\varepsilon - 1)} \left[ 2 \sup_{x,y} |P_n(H_{x,y}) - P(H_{x,y})| + 1/n \right] \rightarrow 0 \text{ a.s.} \]

because the class of closed halfspaces is a Glivenko-Cantelli class.

Dümbgen (1992) showed the strong uniform consistency of the simplicial depth, \( \text{SD}(\cdot) = d_{p+1}(\cdot) \), for any dimension \( p \) when \( P \) is absolutely continuous. The same result can be obtained for \( d_k(\cdot) \) for any \( k \) using similar arguments.

In Figure 1 a), we present the results of the 37 athletes that competed in the 2008 Olympics Decathlon in long jump (in meters, axis X) and in the 100m race (in seconds, axis Y) finishing both events, source: http://en.beijing2008.cn/. The grey-scale image of the halfspace depth of its empirical probability is presented in b), the one of \( d_3(\cdot) \) in c) and the one of the expected degree depth in d).

From visual inspection of Figure 1, we can observe that, on the farthermost points of the data cloud, the expected degree depth and the halfspace depth have a similar behavior. Then again, on the central area of the data cloud, the expected degree depth and the simplicial depth are very much alike. What actually happens is that the exact location of an exterior point, very much influences the simplicial depth, which is neither the case of the halfspace depth, nor of the expected degree.

### 3.3 Empirical comparison

Liu (1995) proposes to use multivariate rank statistics built in terms of depth functions in order to monitor processes of multivariate quality measures. The rank of \( x \) with respect to \( P \) is

\[ r_P(x) = P(\{ y : \text{D}(y;P) \leq \text{D}(x;P) \}) . \]

In most practical applications, the theoretical distribution remains unknown and we are forced to use empirical ranks, obtained replacing \( P \) by an empirical probability.

In order to compare the empirical ranks obtained for the simplicial, expected degree and halfspace depth, we have taken a sample of \( n = 150 \) observations from a standard bivariate normal distribution. In Figure 2, we have plotted the empirical rank (w.r.t. the previous sample) versus the theoretical rank of 1000 observations drawn also from a standard bivariate normal distribution. Since the theoretical depth depends only on the distance to the origin, the theoretical rank is the same for the three depths. We observe that the adjustment of the
empirical ranks to the theoretical ones for the expected degree and halfspace depths is better than for the simplicial depth. At the same time, since the expected degree depth takes more possible values than the halfspace depth, there is a greater variety of empirical ranks.

4 Expected convex hull

The convex hull of $k \geq 1$ independent copies of a random vector $X$ constitutes a compact convex random set. That is, if $X_1, \ldots, X_k$ are the $k$ independent copies of $X$, then $\text{co}\{X_1, \ldots, X_k\}$ is a measurable mapping from a probability space, into the family of compact convex sets in $\mathbb{R}^p$. For measurability and other technical issues about random sets, the
interested reader is referred to Molchanov (2005).

Given a random set, we can define its selection (set-valued) expectation as the set of expectations of all the integrable random vectors which are (a.s.) in the random set.

For any \( k \geq 1 \), we define the expected convex hull region of level \( 1/k \) of a \( p \)-dimensional probability distribution with finite first moment \( P \) as

\[
CD^{1/k}(P) := \mathbb{E} \text{co}\{X_1, \ldots, X_k\} = \{\mathbb{E}Y : Y \in \text{co}\{X_1, \ldots, X_k\} \text{ a.s.}\},
\]

(13)

where \( X_1, \ldots, X_k \) are, as usual, independent observations drawn from \( P \).

In the univariate case \( (p = 1) \) we obtain intervals,

\[
CD^{1/k}(P) = [\mathbb{E}X_{1:k}, \mathbb{E}X_{k:k}],
\]

(14)

where \( X_{1:k} := \min\{X_1, \ldots, X_k\} \) and \( X_{k:k} := \max\{X_1, \ldots, X_k\} \).

The support function of \( K \subseteq \mathbb{R}^p \) is defined as \( h(K, u) := \sup\{\langle x, u \rangle : x \in K\} \) for all \( u \in \mathbb{R}^p \). The support function of the selection expectation of a random set coincides with the expectation of its support function, and thus, for any \( p \geq 1 \) and \( u \in \mathbb{R}^p \),

\[
h(CD^{1/k}(P), u) = \mathbb{E}h(\text{co}\{X_1, \ldots, X_k\}, u) = \mathbb{E}\langle X, u \rangle_{k:k}.
\]

(15)

The expected convex hull depth of a point with respect to a random vector with finite first moment is the inverse of the minimum number of independent copies of itself that are needed so that the point belongs to the selection expectation of their convex hull. Let \( x \in \mathbb{R}^p \), we define as in (4)

\[
CD(x; P) := \max\{1/k : x \in \mathbb{E} \text{co}\{X_1, \ldots, X_k\}\}.
\]

The expected convex hull regions constitute a family of depth-trimmed regions.

**Theorem 14.** The family of expected convex hull regions \( \{CD^{1/k}(P)\}_k \) of a probability distribution \( P \) with finite first moment is a nested family of sets, \( CD^{1/k}(P) \subseteq CD^{1/m}(P) \) for every \( k \leq m \), such that \( CD^1(P_X) = \{\mathbb{E}X\} \) and satisfies the properties \( R1-R6 \):
R1 $CD^{1/k}(P_{AX+b}) = ACD^{a}(P_X) + b$ for every $A \in \mathbb{R}^{q \times p}$ and $b \in \mathbb{R}^{q}$;

R2 $CD^{1/k}(P)$ is bounded;

R3 $CD^{1/k}(P)$ is closed;

R4 $CD^{1/k}(P)$ is convex;

R5 $CD^{1/k}(P)$ is monotonic;

R6 $CD^{1/k}(P)$ is subadditive.

Proof. The expected convex hull regions are affine equivariant, because for any $k$ it holds $\text{co}\{AX_1 + b, \ldots, AX_k + b\} = \text{Aco}\{X_1, \ldots, X_k\} + b$ and the selection expectation is affine equivariant. Further, since $P$ has finite first moment, $\mathbb{E}(X, u) < \infty$ for any $u \in \mathbb{R}^p$ and as a consequence $\mathbb{E}(X, u)_{k:k} < \infty$ for any $k \geq 1$ and $u \in \mathbb{R}^p$, therefore $CD^{1/k}(P)$ is bounded. It is also closed and convex since it is a selection expectation.

Let $k \geq 1$, if $Y \leq X$ (componentwisely) a.s., then $(X, u) \leq (Y, u)$ a.s. for any $u \in \mathbb{R}^p$ (the non-positive quadrant) and then $\mathbb{E}(X, u)_{k:k} \leq \mathbb{E}(Y, u)_{k:k}$ for any $u \in \mathbb{R}^p$.

Given $x \in CD^{1/k}(P_X)$, and $u \in \mathbb{R}^p$, from (15) we have $\langle x, u \rangle \leq h(CD^{1/k}(P_X), u) \leq h(CD^{1/k}(P_Y), u)$. As a consequence, by the convexity of $CD^{1/k}(P_Y)$, there is $y \in CD^{1/k}(P_Y)$ such that $y \leq x$.

The subadditivity follows from the subadditivity of the maximum $\max\{x_1 + y_1, x_2 + y_2\} \leq \max\{x_1, x_2\} + \max\{y_1, y_2\}$ and the linearity of the expectation. For any $u \in \mathbb{R}^p$,

$$h(CD^{1/k}(P_{X+Y}), u) = \mathbb{E}\langle X + Y, u \rangle_{k:k} \leq \mathbb{E}\left(\langle X, u \rangle_{k:k} + \langle Y, u \rangle_{k:k}\right)$$

$$= \mathbb{E}\langle X, u \rangle_{k:k} + \mathbb{E}\langle Y, u \rangle_{k:k} = h(CD^{1/k}(P_X), u) + h(CD^{1/k}(P_X), u).$$

After Theorem 14, the expected convex hull depth is affine invariant, vanishes at infinity, is upper semicontinuous, and is quasiconcave (D1–D4).

As an example of expected convex hull regions, we will consider the ones of a standard $p$-dimensional normal distribution $P$. From (15), it is simple to deduce that $CD^{1/k}(P)$ is a closed ball centered at the origin of radius $\mathbb{E}X_{k:k}$, where $X$ is a random variable with standard normal distribution. The expectations of those extreme order statistics for a wide selection of values of $k$ are provided, e.g., in Teichroew (1956). The contour plots of these regions are presented in Figure 3, a).

Given a $p$-dimensional random sample $X_1, \ldots, X_n$, Cascos (2007) defined the expected convex hull trimmed region of level $k$ of a sample as

$$CD^k_n := \left(\begin{array}{c} n \\ k \end{array}\right)^{-1} \bigoplus_{1 \leq i_1 < \ldots < i_k \leq n} \text{co}\{X_{i_1}, \ldots, X_{i_k}\}.$$
Using a circular sequence algorithm, it is possible to compute the extreme points of \(CD_n^k\) in \(O(n^2 \log n)\) time, see Cascos (2007).

Given \(K_1, K_2 \subseteq \mathbb{R}^p\) compact and convex, their Hausdorff distance is given by
\[
d_H(K_1, K_2) := \sup \{|h(K_1, u) - h(K_2, u)| : u \in \mathbb{R}^p, \|u\| = 1\}.
\]

**Theorem 15.** If \(P\) has finite first moment, then
\[
CD_n^k \longrightarrow CD_{1/k}(P) \quad \text{a.s. ,}
\]
where the convergence of sets is in the Hausdorff distance.

**Proof.** The family \(C^p\) of continuous functions from \(S^{p-1} := \{u \in \mathbb{R}^p : \|u\| = 1\}\), the unit sphere in \(\mathbb{R}^p\), with the supremum norm constitutes a Banach space. As it is argued by Artstein and Vitale (1975), the support functions restricted to the unit sphere belong to \(C^p\), which makes it possible to use results of Banach spaces. The support function of \(CD_{1/k}^n\) can be expressed as
\[
h(CD_{1/k}^n, \cdot) = \left(\frac{n}{k}\right)^{-1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} h(\text{co}\{X_{i_1}, \ldots, X_{i_k}\}, \cdot),
\]
i.e. \(h(CD_{1/k}^n, \cdot)\) is a \(U\)-statistic (of degree \(k\)) in a Banach space that estimates the population function \(h(\text{CD}_{1/k}(P), \cdot)\), see (15). After (Borovskikh, 1996, Th. 3.1.1) it holds that
\[
\lim \sup_{n} \sup_{u \in S^{p-1}} |h(CD_n^k, u) - h(CD_{1/k}(P), u)| = 0 \quad \text{a.s.}
\]

\(\square\)

Figure 3 b) represents the contour plots of the empirical expected convex hull trimmed regions of the Decathlon data.

### 4.1 Variability stochastic order based on the expected convex hull

Stochastic orders are partial order relations between probability distributions of random elements, for a comprehensive treatment see Müller and Stoyan (2002).

Although stochastic orders refer to probability distributions, they are usually written in terms of random elements with such distributions. The following univariate variability stochastic orders are of interest to us. Given \(X, Y\) two random variables:

- **increasing convex stochastic order.** \(X \leq_{icx} Y\) if \(E f(X) \leq E f(Y)\) for all increasing convex \(f\) such that both expectations exist;

- **convex stochastic order.** \(X \leq_{cx} Y\) if \(X \leq_{icx} Y\) and \(EX = EY\).

We will also define a classical multivariate stochastic order. Given \(X, Y\) two random vectors in \(\mathbb{R}^p\), they are ordered with respect to:
a) Standard bivariate standard distribution  b) Decathlon data

Figure 3: Contour pots of $CD^{1/k}(P)$ for $1 \leq k \leq 10$ and $CD^{1/k}_n$ for $1 \leq k \leq 37$.

- linear convex stochastic order. $X \leq_{\text{lex}} Y$ if $\langle X, u \rangle \leq_{\text{cx}} \langle Y, u \rangle$ for all $u \in \mathbb{R}^p$.

We will define a new stochastic order in terms of the inclusion of expected convex hull trimmed regions. Given $X$ and $Y$ two random vectors with finite first moment, we say that $X$ is smaller than $Y$ with respect to the convex hull order if

$$CD^{1/k}(P_X) \subseteq CD^{1/k}(P_Y) \text{ for every } k \geq 1.$$ 

Since the inclusion relation between compact convex subsets is characterized by the less or equal relation of their support functions, after (15), we have

$$X \leq_{\text{ch}} Y \text{ if } E\langle X, u \rangle_{k:k} \leq E\langle Y, u \rangle_{k:k} \text{ for every } u \in \mathbb{R}^p \text{ and } k \geq 1,$$

which for random variables becomes a relation between expectations of the extreme order statistics.

$$X \leq_{\text{ch}} Y \text{ if } EX_{k:k} \leq EY_{k:k} \text{ and } EX_{1:k} \geq EY_{1:k} \text{ for every } k \geq 1.$$ 

The convex hull order is a stochastic order in the sense that it satisfies the reflexivity, transitivity, and antisymmetry properties as a relation among probability distributions of random vectors.

**Proposition 16.** Given $X, Y$ two random vectors in $\mathbb{R}^p$ with finite first moment, the convex hull order satisfies the properties of:

1. reflexivity, $X \leq_{\text{ch}} X$;
2. **transitivity**, if \( X \leq_{\text{ch}} Y \) and \( Y \leq_{\text{ch}} Z \), then \( X \leq_{\text{ch}} Z \);

3. **antisymmetry**, if \( X \leq_{\text{ch}} Y \) and \( Y \leq_{\text{ch}} X \), then \( X \) and \( Y \) are identically distributed.

**Proof.** Since the convex hull order is defined by set inclusions and the set inclusion is reflexive and transitive, the first two statements clearly hold. The antisymmetry is satisfied by the fact that the sequence of expected convex hulls \( \{\text{CD}^{1/k}(P)\}_k \) characterizes \( P \), see Vitale (1987).

---

**General properties of the convex hull order**

**Proposition 17.** The convex hull order satisfies the following properties:

1. if \( X \leq_{\text{ch}} Y \), then \( \mathbb{E}X = \mathbb{E}Y \);

2. if \( \mathbb{E}X = 0 \), then \( X \leq_{\text{ch}} aX \) for every \( a \geq 1 \);

3. if \( X \leq_{\text{ch}} Y \), then \( AX + b \leq_{\text{ch}} AY + b \) for every \( A \in \mathbb{R}^{q \times p} \) and \( b \in \mathbb{R}^q \);

4. if \( X \leq_{\text{ch}} Y \), then \( \text{co(supp } X \text{) } \subseteq \text{co(supp } Y \text{)} \).

Where supp \( X \) stands for the support of \( X \).

We can show that it is a linear stochastic order.

**Lemma 18.** Given \( X, Y \) two random vectors in \( \mathbb{R}^p \)

\( X \leq_{\text{ch}} Y \) if and only if \( \langle X, u \rangle \leq_{\text{ch}} \langle Y, u \rangle \) for every \( u \in \mathbb{R}^p \).

**Proof.** Since \( X \leq_{\text{ch}} Y \), after (16) it holds that \( \mathbb{E}\langle X, u \rangle_{k,k} \leq \mathbb{E}\langle Y, u \rangle_{k,k} \) for every \( u \in \mathbb{R}^p \) and \( k \geq 1 \). If together with each \( u \in \mathbb{R}^p \) we consider \( -u \), the relation among maxima turns into a relation among minima and the result is then obvious.

---

**Relation with other stochastic orders**

The convex hull order is a variability stochastic order among random vectors strictly weaker than the linear convex order.

**Proposition 19.** If two random variables \( X, Y \) satisfy \( X \leq_{\text{cx}} Y \), then \( X \leq_{\text{ch}} Y \).

**Proof.** Let \( X \leq_{\text{cx}} Y \), then \( X \leq_{\text{cx}} Y \) and \( -X \leq_{\text{cx}} -Y \). After (Ross, 1996, p. 454), we have \( \mathbb{E}\langle X, u \rangle_{k,k} \leq \mathbb{E}\langle Y, u \rangle_{k,k} \) or equivalently \( \mathbb{E}X_{1:k} \geq \mathbb{E}Y_{1:k} \) for every \( k \geq 1 \) which after (17) leads to \( X \leq_{\text{ch}} Y \).

**Theorem 20.** If two random vectors \( X, Y \) satisfy \( X \leq_{\text{lex}} Y \), then \( X \leq_{\text{ch}} Y \).

**Proof.** Let \( X \leq_{\text{lex}} Y \), then \( \langle X, u \rangle \leq_{\text{cx}} \langle Y, u \rangle \) for every \( u \in \mathbb{R}^p \) and, by Proposition 19, we have \( \langle X, u \rangle \leq_{\text{ch}} \langle Y, u \rangle \) for every \( u \in \mathbb{R}^p \), which by Lemma 18 leads to \( X \leq_{\text{ch}} Y \).

In the framework of central regions, since the zonoid trimmed regions characterize the linear convex order by inclusion (see Koshevoy and Mosler, 1998), Theorem 20 means that given two probability distributions \( P, Q \) such that \( \text{ZD}^\alpha(P) \subseteq \text{ZD}^\alpha(Q) \) for all \( \alpha \in (0, 1] \), then \( \text{CD}^{1/k}(P) \subseteq \text{CD}^{1/k}(Q) \) for all \( k \geq 1 \). The reverse is not true.
4.2 Multivariate Gini mean difference

It is possible to obtain information about the scatter of a random vector studying the volumes of its depth-trimmed regions, see Zuo and Serfling (2000d).

We propose two scatter estimates based on the expected convex hull central regions. These scatter estimates generalize the Gini mean difference and the Gini index to the multivariate setting. There is a vast literature on multivariate extensions of such statistics (see e.g. Koshevoy and Mosler, 1997b; Oja, 1983).

Let $X$ be a random variable and $X_1, X_2$ be two independent copies of $X$. We will recall the definitions of the Gini mean difference and the Gini index of $X$.

The Gini mean difference of $X$ is defined as

$$M_1(X) := \frac{1}{2} \mathbb{E}|X_1 - X_2|,$$

and if $\mathbb{E}X \neq 0$, its Gini index is given by

$$G_1(X) := \frac{M_1(X)}{|\mathbb{E}X|}.$$

If $\text{vol}_p$ stands for the $p$-dimensional volume (the Lebesgue measure on $\mathbb{R}^p$), the volume of the expected convex hull central region of level $1/2$ of a random variable ($p = 1$) equals to

$$\text{vol}_1\text{CD}^{1/2}(P) = \text{vol}_1\mathbb{E}\text{co}\{X_1, X_2\} = \mathbb{E}X_{2:2} - \mathbb{E}X_{1:2} = \mathbb{E}|X_1 - X_2|.$$

As a consequence, given a random vector $X$ in $\mathbb{R}^p$, the value

$$M_p(X) := \frac{1}{2} \text{vol}_p\text{CD}^{1/2}(P)$$

is a natural candidate for a multivariate Gini mean difference.

In the same fashion, if all of the components of $\mathbb{E}X$ are nonzero, a natural candidate for a multivariate Gini index is the multivariate Gini mean difference, $M_p$, of the normalized random vector $(X^{(1)}/|\mathbb{E}X^{(1)}|, \ldots, X^{(p)}/|\mathbb{E}X^{(p)}|)$,

$$G_p(X) := M_p\left((X^{(1)}/|\mathbb{E}X^{(1)}|, \ldots, X^{(p)}/|\mathbb{E}X^{(p)}|)\right).$$

The next result together with $M_p(X) \leq M_p(Y)$ if $X \leq_{cx} Y$, which is true after Proposition 19 (the inclusion guarantees not having a greater volume), shows that $M_p$ fulfills the requirements of a multivariate measure of scatter proposed by Oja (1983).

**Proposition 21.** Given $A \in \mathbb{R}^{p \times p}$ and $b \in \mathbb{R}^p$, it holds that

1. $M_p(AX + b) = |\det(A)|M_p(X)$;

2. if $A$ is diagonal, $G_p(AX) = G_p(X)$.
Proof. Statement 1. follows from the R1 in Theorem 14 and (Zuo and Serfling, 2000d, Th. 2.1), meanwhile statement 2. is trivial after 1.

The expected convex hull central region of level 1/2 can be decomposed in terms of the expectation of the random vector and a convex body that is centrally symmetric about the origin,

\[ CD^{1/2}(P) = \mathbb{E}co\{X_1, X_2\} = \mathbb{E}X_1 + \mathbb{E}co\{0, X_2 - X_1\} = \{\mathbb{E}X + x : x \in \mathbb{E}co\{0, X_2 - X_1\}\}. \] (20)

The information about location contained in \( CD^{1/2}(P_X) \) is provided by \( \mathbb{E}X \) and the information about scatter by \( \mathbb{E}co\{0, X_2 - X_1\} \).

From (20), the multivariate Gini mean difference defined in (18) can be written as

\[ M_p(X) = \frac{1}{2} \text{vol}_p \mathbb{E}co\{0, X_2 - X_1\} = \frac{1}{2p!} \mathbb{E} | \det [Y_1, \ldots, Y_p] |, \] (21)

where \( Y_1, \ldots, Y_p \) are \( p \) independent random vectors distributed as \( X_1 - X_2 \) and \( [Y_1, \ldots, Y_p] \) stands for the matrix whose columns are \( Y_1, \ldots, Y_p \). The last equality in (21) follows from the fact that set \( \mathbb{E}co\{0, X_2 - X_1\} \) is a zonoid (see Schneider, 1993). The formula for the volume of the zonoid of a probability used in (21) can be found in (Mosler, 2002, Th. 2.10).

Finally, after Proposition 21, the multivariate Gini index built in (19) is related with the multivariate Gini mean difference from (18) through

\[ G_p(X) = \frac{M_p(X)}{\prod_{i=1}^p \mathbb{E}X^{(i)}}. \]

5 Concluding remarks

In this paper, we have introduced a depth function, the expected degree depth, and a family of central regions, the expected convex hull regions.

The expected degree depth satisfies properties similar to the ones of the halfspace and simplicial depths, the two most popular depth functions. Its bivariate sample version can be computed in \( O(n \log n) \) time, complexity achieved by the most efficient algorithms for the halfspace and simplicial depths, whose optimality was shown by Aloupis et al. (2002). In comparison with such depths, it has smooth behavior on regions far from the center of a data cloud and the fact that it takes many possible different values facilitates the discrimination between distinct points in terms of their centrality. The estimation of the expected degree on higher dimensions remains an open problem. Parallel to obtaining the properties of the expected degree depth, it was shown that, for any \( k \), the mapping \( d_k(\cdot) \) is also a depth function. These \( d_k \) depths, can be used to detect features from a data cloud not captured by other depths.

From their construction and properties, the expected convex hull regions are similar to the zonoid ones. Their most remarkable property is the (Minkowski) subadditivity that
makes them particularly suitable to assess financial risks to vector portfolios, see Cascos and Molchanov (2007). In the framework of financial mathematics, the subadditivity has an interpretation in terms of risk diversification and the only subadditive depth-trimmed regions described in the existing literature are the zonoid ones.

Acknowledgements

Thanks are due to Ilya Molchanov from Universität Bern, for posing the problem that originated this research and many enlightening discussions, and to Miguel López-Díaz from Universidad de Oviedo, for carefully reading an early version of this manuscript. Useful hints from Javier Cárcamo from Universidad Autónoma de Madrid and Rosa Lillo and Ángel López from Universidad Carlos III de Madrid are also acknowledged.

References


Appendix

Coverage and convex hull intersection  We will present a technique to compute the probability that a point is contained in the convex hull of a given number of independent copies of a random vector.

Lemma 22. Given $x_1, \ldots, x_k \in \mathbb{R}^p$, $i < k$ negative scalars $\alpha_1, \ldots, \alpha_i < 0$, and $k - i$ positive scalars $\alpha_{i+1}, \ldots, \alpha_k > 0$, then $0 \in \text{co}\{(\alpha_1, x_1), \ldots, (\alpha_k, x_k)\}$ if and only if $\text{co}\{x_1/\alpha_1, \ldots, x_i/\alpha_i\} \cap \text{co}\{x_{i+1}/\alpha_{i+1}, \ldots, x_k/\alpha_k\} \neq \emptyset$.

Proof. Given $x$’s and $\alpha$’s as in the statement, let $0 \in \text{co}\{(\alpha_1, x_1), \ldots, (\alpha_k, x_k)\}$. If we multiply each of the $(\alpha_j, x_j)$ by a positive scalar, 0 remains in the convex hull, $0 \in \text{co}\{(-1, -x_1/\alpha_1), \ldots, (-1, -x_i/\alpha_i), (1, x_{i+1}/\alpha_{i+1}), \ldots, (1, x_k/\alpha_k)\}$. This is equivalent to $\text{co}\{x_1/\alpha_1, \ldots, x_i/\alpha_i\} \cap \text{co}\{x_{i+1}/\alpha_{i+1}, \ldots, x_k/\alpha_k\} \neq \emptyset$. \hfill \qed

Figure 4: Four points in $\mathbb{R}^2$ that do not contain another fixed one, marked by ‘◦’, in their convex hull.

Figure 5: Five points in $\mathbb{R}^3$ that contain another fixed one, marked by ‘◦’, in their convex hull.

Figures 4 and 5 explain Lemma 22 graphically in $\mathbb{R}^2$ and $\mathbb{R}^3$ respectively. In both figures, the origin is marked by ‘◦’, the points whose first coordinate is negative by ‘+’ and the points whose first coordinate is positive by ‘×’. The projections of the points in Figure 5 on the hyperplane $\{(1, x^{(2)}, x^{(3)}) : x^{(2)}, x^{(3)} \in \mathbb{R}\}$ are marked by the same symbol that was used for the corresponding point. After Lemma 22, the origin belongs to the convex hull of
the other points if and only if the convex hulls of certain projections of them have nonempty intersection.

Given a $p$-dimensional probability distribution $P$, we will explain how to take advantage of Lemma 22 in order to compute $d_k(x)$, whenever $P$ does not have an atom at $x$. By the affine invariance of $d_k(\cdot)$, it is enough to show how to compute $d_k(0)$.

Let $H$ be any closed halfspace such that the origin belongs to its boundary and $P(\partial H) = 0$. By the affine invariance of $d_k(\cdot)$, it supposes no restriction to consider the halfspace $H = \{(x^{(1)}, x^{(2)}, \ldots, x^{(p)}) : x^{(1)} \geq 0, x^{(2)}, \ldots, x^{(p)} \in \mathbb{R}\}$ which we will assume hereafter.

Take a sequence of $k \geq 2$ independent observations from $P$ and consider the events

$$A_k := \{0 \text{ belongs to the convex hull of the } k \text{ observations}\};$$

$$B_k^i := \{\text{out of the } k \text{ observations, exactly } i \text{ lie in } H\}.$$ 

In order for $A_k$ to hold, there must be, at least, one observation in $H$ and, at least, one out of $H$, so $\Pr(A_k|B_k^0) = \Pr(A_k|B_k^k) = 0$ and, by the law of total probability, we have

$$d_k(0) = \Pr(A_k) = \sum_{i=1}^{k-1} \Pr(A_k|B_i^i)\Pr(B_i^i). \quad (22)$$

Clearly $\Pr(B_i^k)$ is the probability of $i$ successes out of $k$ independent trials, with $q = P(H)$ the probability of a single success. Further, let $X_1, \ldots, X_k$ be the $k$ independent observations from $P$ ordered in such a way that $X_j^{(1)} > 0$ if $j \leq i$ and $X_j^{(1)} < 0$ if $j \geq i + 1$, where $X_j^{(1)}$ is the first coordinate of $X_j$. If $X_i^* := (X_i^{(2)}, \ldots, X_i^{(p)})$ is the projection of $X_i$ on its last $p - 1$ coordinates, then after Lemma 22 and (22)

$$d_k(0) = \sum_{i=1}^{k-1} \binom{k}{i} q^i (1 - q)^{k-i} \Pr(\co\{Y_1, \ldots, Y_i\} \cap \co\{Z_1, \ldots, Z_{k-i}\} = \emptyset), \quad (23)$$

where the $Y$’s are independent random vectors distributed as $X^*/X^{(1)}|_{X^{(1)}>0}$ and the $Z$’s are independent random vectors distributed as $X^*/X^{(1)}|_{X^{(1)}<0}$.

**Bivariate degree** Let $X = (X^{(1)}, X^{(2)})$ in $\mathbb{R}^2$ such that $\Pr(X^{(1)} = 0) = 0$. Let $Y$ be distributed as $X^{(2)}/X^{(1)}|_{X^{(1)}>0}$ and $Z$ as $X^{(2)}/X^{(1)}|_{X^{(1)}<0}$, then

$$\Pr(A_k|B_i^k) = 1 - \Pr(\{Y_{i;i} < Z_{1:n-i}\} \cup \{Y_{1;i} > Z_{n-i:n-i}\})$$

$$= 1 - \Pr(Y_{i;i} < Z_{1:n-i}) - \Pr(Z_{n-i:n-i} < Y_{1;i}). \quad (24)$$

**Example 23.** Let $P$ be a bivariate probability distribution, angularly symmetric about the origin and such that $P(L) = 0$ for any line $L$ through the origin. If $X$ is an observation from $P$, the random variables $Y$ and $Z$ from (24) would be absolutely continuous and identically distributed, therefore $\Pr(Y_{i;i} < Z_{1:n-i}) = \binom{n}{i}^{-1}.$
Given $0 < q < 1$, let $Q(A) = qP(A \cap H) + (1 - q)P(A \setminus H)$ for any $A \subseteq \mathbb{R}^2$ Borel, then we have the following depths for the origin with respect to $Q$,

$$d_k(0; Q) = 1 - q^k - (1 - q)k - 2 \sum_{i=1}^{k-1} q^i (1 - q)^{k-i};$$  \hspace{1cm} (25)

$$\text{ED}(0; Q) = \frac{q(1-q)}{1+q-q^2};$$

$$\text{SD}(0; Q) = q(1-q).$$  

An equivalent expression for probability (25) was obtained in (Jewell and Romano, 1982, p. 549).

We can further obtain the simple recursive formula

$$d_k(0; Q) = \frac{1}{2} \left( d_{k-1}(0; Q) + 1 - q^{k-1} - (1 - q)^{k-1} \right), \text{ if } k \geq 2.$$

**Degree in $\mathbb{R}^3$**  
Given two probability distributions on $\mathbb{R}^2$, Rogers (1978) described a method to compute the probability of nonempty intersection for the convex hulls of $n$ independent observations from the first distribution and $m$ independent observations from the second. We will use it in order to compute the coverage probability of $k$ i.i.d. random vectors in $\mathbb{R}^3$.

**Example 24.** Let $X$ be a random vector in $\mathbb{R}^3$ distributed as $P$ and whose projection on the unit sphere $S^2$, $X/\|X\|$, is uniformly distributed (this would be the case, e.g., of any spherically symmetric distribution about the origin).

Using the same notation as in (22) and (23), we have

$$\text{Pr}(A_k|B_i^k) = \text{Pr}(\text{co}\{Y_1, \ldots, Y_i\} \cap \text{co}\{Z_1, \ldots, Z_{k-i}\} \neq \emptyset),$$  \hspace{1cm} (26)

where the $Y$'s are $i$ independent random vectors distributed as $X^*/X^{(1)}|_{X^{(1)}>0}$ and the $Z$'s are $k-i$ independent random vectors distributed as $X^*/X^{(1)}|_{X^{(1)}<0}$. Since $X/\|X\|$ is uniformly distributed on $S^2$, the random vectors $Y$ and $Z$ are identically distributed, with density

$$f(s,t) = \frac{1}{2\pi} (1 + s^2 + t^2)^{-3/2}, \quad s,t \in \mathbb{R}.$$

Using the techniques described by Rogers (1978), it is possible to compute the probabilities in (26). We will do it for $k = 4$ and $i = 1, 2$,

$$\text{Pr}(A_4|B_1^4) = \text{Pr}(\text{co}\{Y_1\} \cap \text{co}\{Z_1, Z_2, Z_3\} \neq \emptyset) = (12 - \pi^2)/(2\pi^2),$$

$$\text{Pr}(A_4|B_2^4) = \text{Pr}(\text{co}\{Y_1, Y_2\} \cap \text{co}\{Z_1, Z_2\} \neq \emptyset) = (\pi^2 - 8)/(\pi^2).$$

Notice that the fact that $Y$ and $Z$ are identically distributed, guarantees $\text{Pr}(A_4|B_1^4) = \text{Pr}(A_4|B_2^4)$.

Finally, given $0 < q < 1$, let $Q(A) = qP(A \cap H) + (1 - q)P(A \setminus H)$ for any $A \subseteq \mathbb{R}^3$ Borel. The simplicial depth of the origin with respect to $Q$ is now

$$d_4(0; Q) = \text{SD}(0; Q) = \frac{2q(1-q)}{\pi^2} ((1-q)^2 + q^2)(12 - \pi^2) + 3q(1-q)(\pi^2 - 8).$$

27